М. Эддоус, Р. Стэнсфилд

МЕТОДЫ ПРИНЯТИЯ РЕШЕНИЙ

Перевод с английского под редакцией
член-корр. РАН ***И.И.Елисеевой***

Рекомендовано к изданию
Министерством общего и профессионального
образования Российской Федерации

Москва
"Аудит"
Издательское объединение "ЮНИТИ"
1997

;К 65.050.9(4Вл)я73
Э18

Перевод с английского
С.А. Лукина (гл. 1–3), *И.И. Елисеевой* (гл. 4–6, 8),
Т.В. Костеевой (гл. 7, 9–14, приложения)

Рецензенты:
кафедра статистики Санкт-Петербургского торгово-экономического института и д-р экон. наук проф. М.М. Юзбашев

Главный редактор издательства *Н.Д.Эриашвили*

Эддоус М., Стэнсфилд Р.
Э18 Методы принятия решений/ Пер. с англ. под ред. член-корр. РАН И.И. Елисеевой. — М.: Аудит, ЮНИТИ, 1997. — 590 с.
ISBN 0-85121-832-6 (англ.)
ISBN 5-85177-027-9 (русск.)

Предлагаемая книга представляет собой учебное пособие, которое подготовлено и используется Ассоциацией присяжных дипломированных бухгалтеров Великобритании (The Chartered Association of Certified Accountants — ACCA) в подготовке дипломированных бухгалтеров.

Излагаются теоретические основы принятия управленческих решений в условиях неопределенности и оценки риска: элементы теории вероятностей, минимаксные правила принятия решений, статистический анализ для принятия решений, включающий процедуры статистического вывода, контроля качества, анализа временных рядов и прогнозирования, методы оптимизации решений.

ISBN 0-85121-832-6 (англ.)
ISBN 5-85177-027-9 (русск.)

ББК 65.050.9(4Вл)я73

От имени Ассоциации дипломированных корпоративных бухгалтеров (АССА) я с большим удовольствием представляю всем российским профессионалам, работающим в области бухгалтерского учета, финансов, управления и бизнеса, а также студентам учебник "Методы принятия решений" авторов М. Эддоуса, Р. Стенсфилда.

Учебник охватывает широкий круг вопросов, связанных с принятием решений, относящихся ко всем уровням управления. Изложены ключевые методы принятия решений: методы теории вероятностей, испытание гипотез, статистический контроль качества, прогнозирование, линейное программирование, транспортная задача, имитационное моделирование.

АССА очень рада продолжению сотрудничества с Издательским объединением ЮНИТИ по обучению финансовым профессиям в России. Первая книга, выпущенная Издательским объединением ЮНИТИ из серии АССА,— Р. Адамс "Основы аудита" (1995).

Антея Л. Роуз

Главный Исполнительный Директор

Ассоциации Дипломированных

Корпоративных Бухгалтеров (АССА)

ОТ НАУЧНОГО РЕДАКТОРА

Каждый день в нашей жизни мы принимаем решения — большие и малые, связанные с бизнесом, с личными и общественными делами. В древности люди принимали решения, основываясь на интуиции, заключениях астрологов, прорицателей и т. д.

Развитие науки, усложнение экономических и социальных связей и отношений привели к разработке специальной области научного знания — теории принятия решений, основанной прежде всего на теории вероятностей и математической статистике. Формирование ее основ относится к концу XVII в. — началу XVIII в. Задолго до этого времени люди стали задумываться о характере вероятностных процессов и пытаться изучать их. Мощный толчок практическому использованию подсчета шансов дало открытие Колумбом Америки. Оживились внешнеторговые морские перевозки, а значит и страховое дело, расширились банковские операции, стала развиваться кредитная система. Все это требовало оценки рисков, и подсчет шансов из мира азартных игр перешел в мир экономики.

Предлагаемая книга представляет собой учебное пособие, которое подготовлено и используется Ассоциацией присяжных дипломированных бухгалтеров Великобритании (The Chartered Association of Certified Accountants — ACCA) в подготовке дипломированных бухгалтеров.

В пособии излагается широкий круг проблем и методов, относимых в нашей учебной литературе к разным дисциплинам: теории вероятностей, математической статистике, линейному программированию, исследованию операций, сетевому анализу, теории очередей, определению предпринимательского риска и т.д.

Все названные разделы излагаются на высоком уровне и в то же время без усложнения, без доказательств множества теорем, которыми злоупотребляют отечественные учебники по экономико-математическим методам и статистике. Математический аппарат здесь полностью подчинен решению конкретных экономических задач. Это сказывается и в "облегченном" написании формул: нередко отсутствуют подстрочные значки, не указаны пределы суммирования и т. д.

Достоинством книги является то, что в ней объединены в системном изложении знания, разбросанные в нашей литературе по многим разнородным учебным пособиям. Еще одной отличительной чертой является продуманность изложения, высокий методический уровень. Изложение строится следующим образом: сначала кратко дается теоретическая основа, затем иллюстрации на примерах, в конце каждой главы имеется краткое резюме, после чего следуют вопросы и упражнения по данной теме. Ценно то, что ответы на все вопросы и упражнения приведены в конце книги. Не ограничившись этим, авторы включили дополнительные упражнения и привели полный набор экзаменационного задания для получения диплома

"Certified Accountant". Эти экзаменационные требования могут быть чрезвычайно полезны для подготовки в России новых высококлассных специалистов.

Хотя в учебных планах вузовской подгтовки отечественных специалистов по экономическим специальностям обычно нет дисциплины "Методы принятия решений", но, во-первых, такой курс может появиться, а во-вторых, разделы этого курса входят в такие дисциплины, как теория вероятностей, статистика, исследование операций, управление финансами, общий менеджмент, маркетинг.

Весь материал разделен авторами книги на 4 части и 14 глав. Первая часть посвящена общим проблемам принятия решений в условиях неопределенности. Она включает основные понятия и правила теории вероятностей, в том числе теорему Байеса, знакомит читателя с основными законами распределений (которые в книге называются "вероятностными распределениями"). В отличие от наших пособий, где основные законы распределений — распределение Пуассона, биномиальное, нормальное распределения — часто рассматриваются вне связи друг с другом, в пособии подчеркивается возможность замены одного распределения другим (при определенных условиях). Особое значение в этом разделе имеет изложение правил принятия решений, где дается понятие дерева решений и рассматриваются методы принятия решений без использования и с использованием численных значений вероятностей отдельных исходов. Специальное внимание уделяется оценке чувствительности решений: их зависимости от изменений вероятностей исходов. Разнообразный подбор примеров включает и принятие инвестиционных решений.

Во второй части рассматриваются методы анализа данных как составной части принятия решений. В этом разделе вводится понятие статистической выборки и выборочного распределения; излагается процедура статистического вывода — методы статистического оценивания и проверки гипотез. В этой же части излагаются методы статистического контроля качества. Рассматривается регрессионный анализ, включая множественную регрессию, правда, в весьма ограниченном объеме —, только при двух независимых переменных, затрагиваются проблемы описания нелинейных связей, а также применение ранговой корреляции. В этой части авторы повторяют формальное утверждение о том, что свободный член линейного уравнения регрессии a характеризует значение зависимой переменной y при нулевом значении независимой переменной x, тогда как давно доказано, что свободный член выполняет функцию "пилота" — доводки до функционального соотношения между средними величинами x и y[1]. Эта функция очевидна из выражения: $a = \bar{y} - \sqrt{b}\,\bar{x}$. Известно, что a>0, если x варьирует сильнее, чем y, т.е. Vx>Vy; в противном случае получаем a<0.

Отмечая, что формула коэффициента корреляции К. Пирсона основана на моменте произведения, авторы приводят лишь преобразованное выражение *r*:

$$r = \frac{n\sum xy - \sum x \cdot \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)\,(n\sum y^2 - (\sum y)^2)}}$$

---

[1] См., например, Энгвер Н.Н. Проблема выбора формы связи при математико-статистической обработке экономической информаии. Автореф. дисс... канд. экон. наук.– М.: 1970.– 19 с.

тогда как сущность коэффициента корреляции яснее, если в числителе указать именно момент произведения $n \sum (x - \bar{x}) \cdot (y - \bar{y})$. Авторы умолчали о сложностях оценки генерального коэффициента корреляции ввиду нарушения нормальности распределения выборочных коэффициентов корреляции и необходимости использования z-преобразования Р. Фишера. При испытании гипотез делается вывод не только от испытуемой, но и от альтернативной гипотезы, тогда как, строго говоря, вывод может быть сделан только в отношении нулевой гипотезы.

Специальная глава содержит приемы изучения динамики и прогнозирования. В этой части интерес представляет методика анализа, ориентированная на разные модели данных — аддитивную и мультипликативную. Дается методика полного разложения ряда динамики на компоненты, включая выделение тренда, периодической и случайной компоненты.

Третья часть содержит методы сетевого анализа и планирования и их приложения; планирования и управления запасами и т.д. Заключительная часть посвящена моделированию. Рассматриваются линейное программирование, транспортная задача, имитационные модели и их применение. И в этом разделе акцент делается на прикладные аспекты без глубокой теории.

Подробно рассмотрены графический и симплексный методы, представлены интересные, содержательные примеры. На наш взгляд, было бы целесообразно дополнить этот раздел более подробным изложением двойственного симплекс-метода, наиболее эффективного при решении задач линейного программирования типа "задачи о диете", а также включить специальные методы решения задач целочисленного линейного программирования.

Принятие решения всегда включает определенную последовательность действий столь многообразных и взаимосвязанных, что в какой-то момент нужно остановиться и решить, что делать дальше. Статистические тесты, "деревья решений" и другие методы позволяют уменьшить частоту неправильных действий. В любом случае авторам удалось показать, что принятие решения носит вероятностный характер. Во всех разделах уделяется внимание исследованию чувствительности того или иного метода к изменению исходных данных.

Таким образом, предлагаемое пособие содержит и необходимые теоретические сведения и практикум с разбором решений. Повсеместно показаны возможности графических методов. Пособие оснащено статистико-математическими таблицами, основными формулами. Содержание и методические достоинства предлагаемого учебного пособия позволяют надеяться, что все, кто будет им пользоваться, испытают такое же удовлетворение, которое испытали мы при подготовке перевода книги британских коллег.

И. И. Елисеева

Часть 1

ПРИНЯТИЕ РЕШЕНИЙ В УСЛОВИЯХ НЕДОСТАТКА ИНФОРМАЦИИ

В первых двух частях этой книги рассматриваютс торые статистические проблемы. Часть 1 посвящена б понятиям теории вероятностей, биномиальному и норма распределениям, распределению Пуассона, правилам и принятия решений.

То, что статистика рассматривает непостоянс неопределенность, делает ее необходимой для бизнеса и терского учета. Например, затраты на производство измен порой самым непредсказуемым образом из-за того, что дс сложно их точно измерить. На затраты влияют кач сырья, объемы производства, неполадки машин и механ Эти же причины могут изменять спрос на данный това надеемся, что данное пособие поможет вам понять важ статистики в управлении вариацией и неопределенност бизнесе и бухгалтерском учете, быстрее ориентироват практических ситуациях и заложить основы для более глу го изучения данного предмета.

Глава 1. ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. ВВЕДЕНИЕ

Одна из задач этой книги состоит в том, чтобы помочь вам в практической деятельности. И чем более непредсказуем результат, тем интереснее и сложнее принимать решение. Как правило, возможные исходы известны, но вопрос в том, какому из них суждено сбыться. Теория вероятностей предлагает нам пути уменьшения неопределенности, именно поэтому важно ею овладеть. Студентам бывает трудно ее освоить из-за множества новых концепций, правил, понятий. Поэтому старайтесь представить себе реальные ситуации, в которых может встретиться предложенная проблема, что логично в такой ситуации предпринять. Просчитайте, какие последствия повлечет то или иное решение. Иногда это сделать сложно, но именно в этом заключается первый и самый важный этап принятия решения.

Итак, рассмотрев теоретически возможные исходы (с практической точки зрения), вы делаете вычисление вероятности каждого из них простым делом.

В следующем разделе кратко рассказывается об основных понятиях теории вероятностей.

1.2. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Теория вероятностей родилась из математической теории игр.

В середине XVII века французские математики Паскаль и Ферма разработали математическую модель, описывающую вероятность исходов в играх, зависящих от случая, по заказу известных игроков в азартные игры. При игре в «кости», рулетку, как и при опросах, исследованиях (физических, экономических, социологических и т.д.), результаты меняются от раза к разу даже при сохранении неизменных условий.

Деловые люди принимают решения в таких же условиях. Например, специалист по маркетингу никогда не сможет точно предсказать объемы реализации нового товара. Так же, как и заключая пари, невозможно предвидеть, выиграешь или проиграешь. И в том, и в другом случае присутствует *неопределенность*. Теория вероятностей как раз и оперирует этим понятием. Изучение теории вероятностей, основанной на игре случая, обеспечит надежный инструмент измерения и контроля различных форм неопределенности, с которыми имеют дело бизнесмены.

Для начала несколько слов о терминологии.

Опыт — действие, результат которого заранее неизвестен. Например, результат бросания монеты или игральной кости.

Эксперимент — один или несколько опытов. Например, бросание монеты 6 раз.

Исход — возможный результат эксперимента. Например, монета брошена 6 раз, результат: «решка», «решка», «решка», «орел», «орел», «решка».

Событие — один или несколько исходов эксперимента. Например, монета брошена 6 раз, событие: 2 «орла», 4 «решки».

Хотя может существовать неопределенность в отношении результата единственного эксперимента, ряд идентичных экспериментов определяет наиболее вероятный результат, который и используется в анализе ситуации в бизнесе.

1.2.1. Вероятность — что это такое?

В качестве иллюстрации можно взять бросание монеты. Существуют два возможных исхода — «орел» и «решка». С какой вероятностью будет выпадать «решка»? Бросим монету 10 раз, а результаты запишем. А потом увеличим число экспериментов до 100, 1000 и так далее. Возможные результаты могут быть таковы.

Таблица 1.1. **Возможное распределение результатов**

Число бросков	*Число*		*Соотношение*	
	«орлов»	*«решек»*	*«орлов»*	*«решек»*
10	6	4	0,6	0,4
100	64	36	0,64	0,36
1000	643	357	0,643	0,357
10000	6431	3569	0,6431	0,3569

По мере увеличения числа бросков выявляется стремление частоты появления «орлов» к определенной величине. В данном примере их доля — 0,643 при точности трех знаков после запятой.

На основе данных табл. 1.1 можно предсказать, что при 10001-м броске вероятность выпадения «орла» больше, нежели решки, и составит примерно 0,643.

Если считать «успехом» тот результат, в котором мы заинтересованы, то определение вероятности будет следующим:

Вероятность — это число «успехов», полученное в результате большого числа экспериментов, так что вероятность определяет возможность «успеха» в следующем эксперименте.

1.2.2. Свойства вероятности

Из того, что вероятность является соотношением, следуют два важных вывода. Если мы обозначим вероятность исхода эксперимента p, то можно сказать следующее:

1. Числовое значение вероятности находится в интервале от 0 до 1 включительно:

$$0 < p < 1,$$

т.е. p не может быть отрицательным или быть больше 1.

2. Сумма вероятностей результатов (вероятность полной группы событий) равна 1, т.е. вероятность того, что что-то произойдет, равна 1: $\sum P = 1$.

Следовательно, вероятность события E есть P(E), тогда:

$$0 \le P(E) \le 1.$$

1.2.3. Как найти значение вероятности

Зная, что вероятность можно измерить, попробуем выразить ее в цифрах. Существуют три возможных пути.

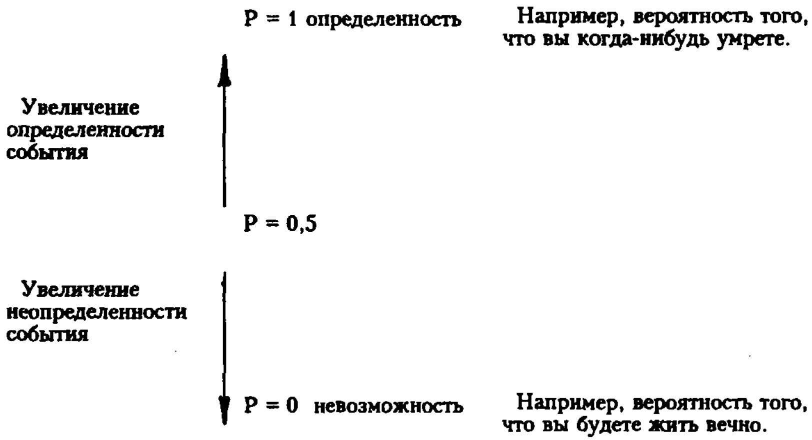


Рис. 1.1. Измерение вероятности

ВЕРОЯТНОСТЬ, ОПРЕДЕЛЯЕМАЯ СИММЕТРИЕЙ

Существуют ситуации, в которых возможные исходы равновероятны. Например, при бросании монеты один раз, если монета стандартная, вероятность появления «орла» или «решки» одинакова, т.е. **P(«орел») = P(«решка»)**. Так как возможны лишь два исхода, то **P(«орел») + P(«решка») = 1**, следовательно, **P(«орел») = P(«решка») = 0,5**.

В экспериментах, где исходы имеют равные шансы появления, вероятность события E, P (E) равна:

$$P(E) = \frac{\text{Количество равновероятных исходов, составляющих E}}{\text{Общее количество исходов}}.$$

❑ **Пример 1.1.** Монета брошена три раза. Какова вероятность двух «орлов» и одной «решки»?

Решение.

Для начала найдем все возможные исходы: (ррр, рро, рор, орр, роо, оро, оор,
Чтобы убедиться, все ли возможные варианты мы нашли, воспольз
диаграммой в виде дерева (см. гл.1 раздел 1.3.1).

Итак, имеются 8 равновозможных исходов, следовательно, вероятность ка
из них равна 1/8. Событие Е — два «орла» и «решка» — произошло три
(роо, оро, оор). Поэтому:

$$P(E) = \frac{\text{Количество исходов, дающих E}}{\text{Общее количество исходов}} = \frac{3}{8}.$$

❒ **Пример 1.2.** Стандартная игральная кость брошена два раза. Какова вероят
того, что сумма очков равна 9 или больше?

Решение.

Найдем все возможные исходы.

Таблица 1.2. **Общее количество очков, получаемое при двукратном бросани
игральной кости**

Очки при втором броске	*Очки при первом броске*					
	1	*2*	*3*	*4*	*5*	*6*
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	**9**	**1(**
5	6	7	8	**9**	**10**	**1:**
6	7	8	**9**	**10**	**11**	**1:**

Итак, в 10 из 36 возможных исходов сумма очков равна 9 или бол
следовательно:

$$P(\text{9 или больше очков в результате двух бросаний игральной кости}) = \frac{10}{36}$$

ВЕРОЯТНОСТЬ, ОПРЕДЕЛЯЕМАЯ ЭМПИРИЧЕСКИ

Пример с монетой из табл. 1.1 наглядно иллюстрирует механизм определ
вероятности.

При общем числе экспериментов n, из которых m удачных, вероят
требуемого результата подсчитывается так:

$$P(\text{успешных исходов}) = \frac{m}{n} \text{ при большом } n.$$

Отношение m/n есть относительная частота появления определен
результата при достаточно продолжительном эксперименте. Вероятность об
подсчитывается либо на основе данных проведенного эксперимента, либ
основе прошлых данных.

Пример 1.3. Из пятисот протестированных электроламп 415 проработали более 1000 часов. На основе данных этого эксперимента можно заключить, что вероятность нормального функционирования лампы данного типа более 1000 часов составляет:

$$\frac{415}{500} = 0{,}83 .$$

Примечание. Контроль имеет разрушающий характер, поэтому не все лампы могут быть проверены. Если бы была протестирована только одна лампа, то вероятность составила бы 1 или 0 (т.е. сможет проработать 1000 часов или нет). Отсюда следует необходимость повторения эксперимента.

Пример 1.4. В табл. 1.3 приведены данные о стаже мужчин, работающих в фирме:

Таблица 1.3. **Стаж работы мужчины**

Стаж, лет	*Число работников*	
Менее 1	26	
От 1 до 2	36	
От 2 до 3	16	= 38
От 3 до 4	20	
От 4 до 5	2	
5 и более	0	
Всего: 100		

Какова вероятность того, что следующий принятый на работу в фирму человек проработает не меньше двух лет:

Решение.

Из таблицы видно, что 38 из 100 работников работают в компании больше двух лет. Эмпирическая вероятность того, что следующий работник останется в компании на срок более двух лет равна:

$$\frac{38}{100} = 0{,}38 .$$

При этом мы предполагаем, что новый работник «типичен», а условия работы неизменны.

СУБЪЕКТИВНАЯ ОЦЕНКА ВЕРОЯТНОСТИ

В бизнесе часто возникают ситуации, в которых отсутствует симметрия, и экспериментальных данных тоже нет. Поэтому определение вероятности благоприятного исхода под влиянием взглядов и опыта исследователя носит субъективный характер.

Пример 1.5.

1. Эксперт по инвестициям считает, что вероятность получения прибыли в течение первых двух лет равна 0,6.
2. Прогноз менеджера по маркетингу: вероятность продажи 1000 единиц товара в первый месяц после его появления на рынке равна 0,4.

1.3. ВЕРОЯТНОСТЬ СЛОЖНЫХ СОБЫТИЙ

Сложные события представляют собой серию опытов и комбинацию всех возможных исходов. Очень важно учесть все возможные исходы, для чего составляется «дерево вероятностей», на котором отражаются последовательность экспериментов и их результаты.

1.3.1. Дерево вероятностей

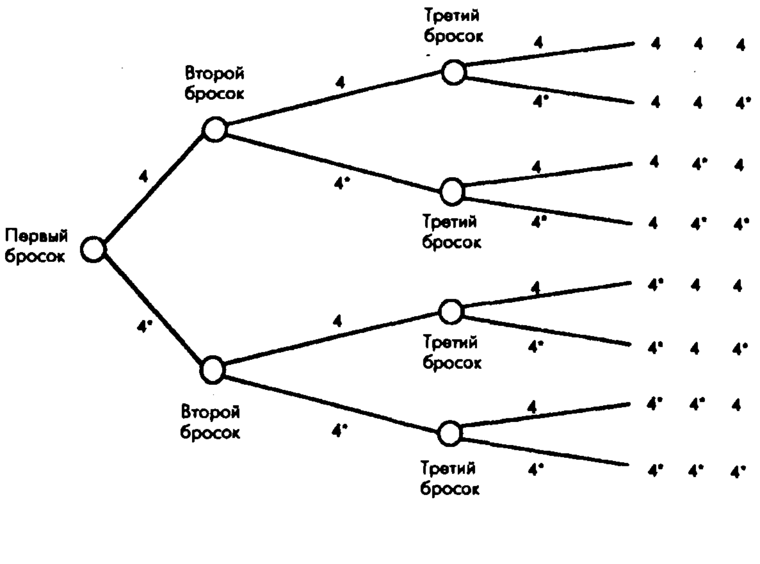
Опыты представлены здесь последовательностью кружков, а каждый исход — «ветвью» (линией) от соответствующего кружка. Вероятность соответствующего исхода указана около «ветви», а вероятность всего сложного события в ее конце.

❐ **Пример 1.6.** Игральная кость брошена 3 раза. Какова вероятность того, что «четверка» выпадет 1 раз, 2 или 3 раза, или не выпадет ни разу?

Решение

Эксперимент состоит из трех опытов. Существуют два исхода для каждого опыта: или «четверка» выпадает (4), или нет (4*).

Для того, чтобы подсчитать вероятность выпадения «четверки» 1, 2, 3 раза или ни разу, нужно знать, как вычисляется вероятность сложных событий, что мы и рассмотрим в следующем разделе и тогда вернемся к этому примеру.



(4 — выпала "четверка"; 4* — выпала "не четверка")

Рис. 1.2. Исходы трех бросаний игральной кости

1.4. ДЕЙСТВИЯ С ВЕРОЯТНОСТЯМИ

Для начала определимся с терминологией:

Независимыми событиями А и В называются такие, если появление одного из них не изменяет вероятности появления другого. Например: по одному разу брошены монета и кость, выпали — «решка» и «6». Результаты обоих событий друг на друга не влияют, поэтому называются независимыми.

Несовместимыми событиями А и В называются такие, если может произойти только одно из них. Например, брошена игральная кость: А — выпало четное число, В — нечетное. Если кость брошена только один раз, А и В произойти одновременно не могут, поэтому они — несовместные события.

Вероятность сложных событий определяется двумя правилами — правилом сложения вероятностей и правилом умножения вероятностей.

1.4.1. Правило сложения вероятностей

Для простоты рассмотрим лишь два события — А и В. Правило сложения вероятностей применяется для подсчета вероятности осуществления событий А или В, или их обоих сразу:

$$P(A + B) = P(A) + A(B) - P(AB).$$

Если события А и В несовместимы, то:

$$P(A+B) = P(A) + P(B).$$

Так как события А и В — несовместимые, то они не могут произойти одновременно, значит:

$$P(AB) = 0.$$

❑ **Пример 1.7.** Игральная кость брошена один раз. Какова вероятность выпадения «двойки» или нечетного числа?

Решение.

Возможны 6 исходов — 1, 2, 3, 4, 5, и 6. Назовем событием А выпадение «двойки», а событием В — выпадение «единицы», «тройки» или «пятерки».

Решить задачу можно либо по правилу симметрии, либо используя правило сложения вероятностей.

1. По правилу симметрии:

$$P\text{ (2 или нечетное число)} = \frac{\text{Число благоприятных исходов (т.е. 2 или 1, 3, 5)}}{\text{Общее число исходов}} = \frac{4}{6}.$$

2. По формуле сложения вероятностей:
два события несовместимы, значит:

$$P(AB) = 0, \text{ поэтому } P(A+B) = P(A) + P(B).$$

Отсюда

$$P(A+B)=P(A)+P(B)=\frac{1}{6}+\frac{3}{6}=\frac{4}{6}.$$

❏ **Пример 1.8.** Игральная кость брошена один раз. Какова вероятность выпадения «двойки» или четного числа?

Решение.

События совместимы, когда они могут произойти одновременно. Поэтому:

1. По правилу симметрии: существуют три исхода – 2, 4, 6, следовательно, вероятность появления «двойки» или любого другого четного числа равна 3/6.
2. По правилу сложения вероятностей:

$$P(A+B)=P(A)+P(B)-P(AB),$$

а так как A и B не являются взаимоисключающими, то

$$P(A+B)=\frac{1}{6}+\frac{3}{6}-\frac{1}{6}=\frac{3}{6}.$$

❏ **Пример 1.9.** Число дефектов в изделии может быть любым — 0, 1, 2, 3, 4, и т.д. По оценке компании вероятность отсутствия дефекта составляет 0,9, а вероятность наличия одного дефекта – 0,05. Какова вероятность, что в изделии не больше, чем один дефект?

Решение.

Р(не больше, чем один дефект) = Р(отсутствие дефектов или один).

Опять события несовместимы, поэтому:

Р (отсутствие дефектов или один) =
= Р(отсутствие дефектов) + Р(один дефект) + 0,9 + 0,05 = 0,95.

❏ **Пример 1.10.** Прогноз метеорологов:

Р(дождь) = 0,4; Р(ветер) = 0,7; Р(дождь и ветер) = 0,2.

Какова вероятность того, что будет дождь или ветер?

Решение.

По формуле сложения вероятностей:

Р(дождь или ветер или то, и другое) =
= Р (дождь) + Р(ветер) – Р(дождь и ветер) =
=0,4 + 0,7 – 0,2 = 0,9.

Мы использовали таблицы, схемы, логические «деревья» для иллюстрации возможных исходов эксперимента. Диаграмма Венна – еще один вид представления результата в графической форме. Здесь события изображены в виде окружностей, помещенных внутри прямоугольника. В данном случае одна окружность обозначает дождь, другая – ветер. Их общее пространство – дождь и ветер. Свободная площадь – отсутствие дождя и ветра. Значения вероятностей должны быть проставлены в соответствующих местах диаграммы (см. рис. 1.3):

Если A и B **независимы**, то P (B/A) = P (B), и правило выглядит так:

$$P(AB) = P(A) \times P(B).$$

Пример 1.12. Игральная кость брошена дважды. Событие A — выпадение "двойки" при первом бросании, событие B — выпадение нечетного числа при втором бросании. Какова вероятность того, что события A и B произойдут в одном эксперименте?

Решение.

Так как результат второго опыта не зависит от результата первого, то события A и B — независимы, тогда по формуле умножения вероятностей:

$$P(AB) = P(A) \times P(B) = \frac{1}{6} \times \frac{3}{6} = \frac{3}{36}.$$

Покажем решение на основе «дерева вероятностей». Во-первых, изобразим последовательность исходов: (рис. 1.4).

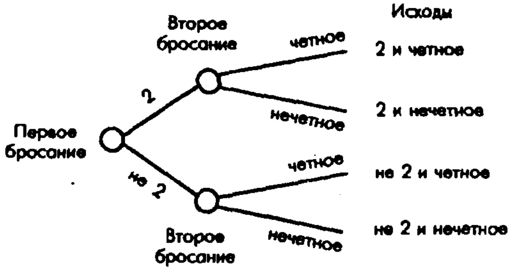


Рис.1.4. Последовательность исходов

Во-вторых, проставим вероятности на каждой «ветви» (рис.1.5.). В-третьих, последовательно перемножим вероятности по «стволу» каждой «ветви» и получим значение вероятности каждого конкретного исхода.

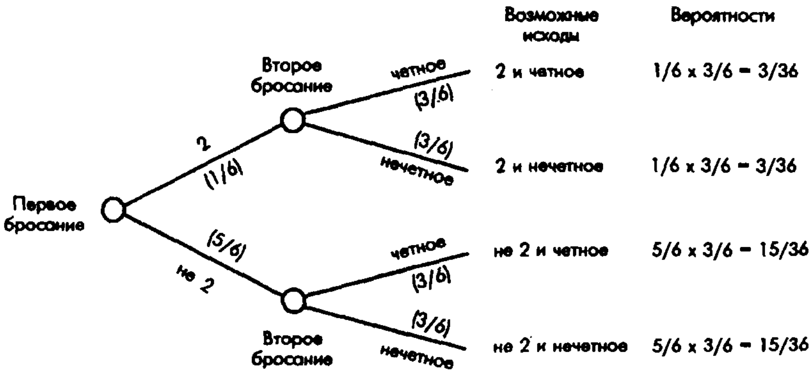


Рис.1.5. Вероятность возможных исходов

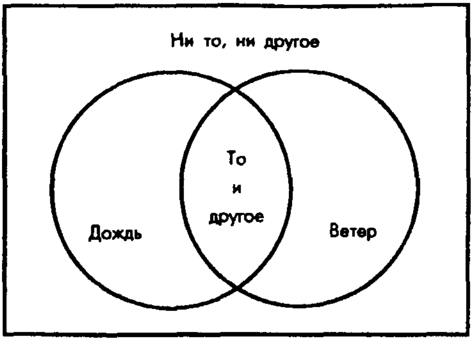


Рис. 1.3. Прогноз погоды в виде диаграммы Венна

1.4.2. Условная вероятность

Рассмотрим два события — E и F, которые происходят друг за другом. P (E) — вероятность события E. Отсюда возникают две альтернативные ситуации:

1. F от E **не зависит**, и на вероятность события F не влияет то, произошло ли уже событие E или нет.
2. E и F — **зависимы**, т.е. вероятность события F зависит от того, произошло ли уже событие E или нет. В этом случае вероятность события F называется **условной**.

Вероятность F при условии, что E произошло, обозначается так:

P(F при условии E) или P(F/E).

Если E и F независимы, тогда:

P (F при условии E) = P (F).

❑ **Пример 1.11.** В коробке 8 красных шаров и 6 голубых. Какова вероятность того, что из двух вытащенных наугад шаров последний будет красным?

Решение

Возможны два варианта:

1. Первым вытащен красный шар, в коробке осталось 7 красных и 6 голубых шаров;
2. Первым вытащен голубой шар, осталось 8 красных и 5 голубых шаров.

В случае 1 вероятность того, что второй вытащенный шар будет красным, равна 7/13.

В случае 2 вероятность равна 8/13.

1.4.3. Правило умножения вероятностей

Это правило применяется, когда требуется найти вероятность того, что события A и B произойдут одновременно. Правило умножения вероятностей состоит в следующем:

$$P(AB) = P(A) \times P(B/A).$$

Результат верхней «ветви» — это решение нашей задачи — (3/36), как и в первом варианте решения.

❑ **Пример 1.13.** На станции отправления имеется 8 заказов на отправку товара: пять — внутри страны, а три — на экспорт. Какова вероятность того, что два выбранных наугад заказа окажутся предназначенными для потребления внутри страны?

Решение.

Событие А — первый взятый наугад заказ — внутри страны. Событие В — второй, тоже взятый наугад заказ. Нам необходимо найти вероятность Р (АВ), поэтому по формуле:

$$P(AB) = P(A) \times P(B/A) = \frac{5}{8} \times \frac{4}{7} = \frac{20}{56}.$$

1.4.4. Правило вычисления вероятностей для более чем двух событий

Рассмотренные правила применимы также, если событий более, чем два. Для несовместимых событий правило сложения вероятностей приобретает следующий вид:

$$P(A + B + C + \ldots) = P(A) + P(B) + P(C) + \ldots$$

Для совместимых событий формула приобретает очень сложный вид.

Для независимых событий правило умножения вероятностей имеет следующий вид:

$$P(A \text{ и } B \text{ и } C \text{ и } \ldots) = P(A) \times P(B) \times P(C) \times \ldots$$

Если события не являются независимыми, то правило умножения вероятностей запишется как:

$$P(A \times B \times C \times \ldots) = P(A) \times P(B/A) \times P(C/AB) \times \ldots$$

❑ **Пример 1.14.** Совет директоров состоит из трех бухгалтеров, трех менеджеров и двух инженеров. Планируется создать подкомитет из его членов. Какова вероятность того, что все трое в этом подкомитете будут бухгалтеры?

Решение.

Используем схему «дерево», где А — выбран бухгалтер, А* — выбран не бухгалтер (см. рис. 1.6).

После составления «дерева» проставляем на каждой «ветви» вероятность исхода, которая изменяется по мере уменьшения числа возможных кандидатов: 8 — 7 — 6. Аналогично при избрании одного бухгалтера число оставшихся неизбранными уменьшается на единицу: от 3 до 2. Вероятности зависят от произошедших ранее событий.

Из диаграммы следует, что вероятность избрания всех трех бухгалтеров в члены подкомитета равна: (3/8) × (2/7) × (1/6) = 1/56. Однако для решения задачи можно обойтись и без диаграммы:

Р(все члены — бухгалтеры) = Р(1-й член — бухгалтер и
2-й член — бухгалтер и
3-й член — бухгалтер)
= Р(1-й член — бухгалтер) ×

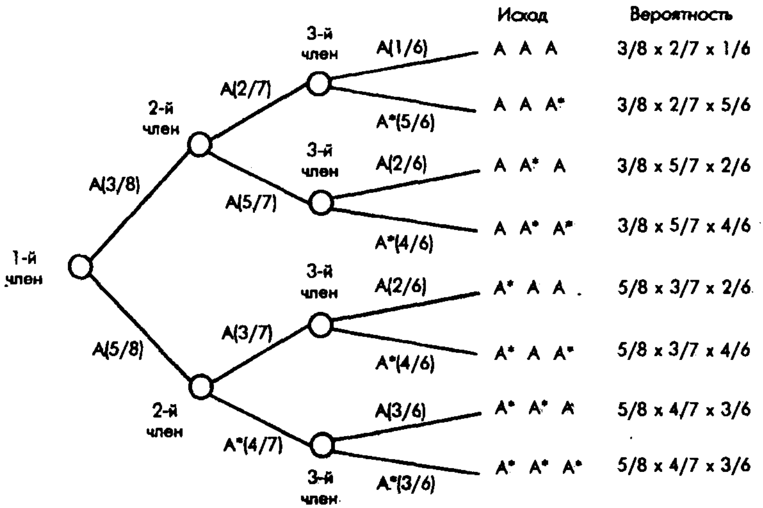


Рис. 1.6. Выборы подкомитета из трех человек (А — выбран бухгалтер, А* — выбран не бухгалтер)

P(2-й член — бухгалтер при условии, что 1-й член — бухгалтер) ×
P(3-й член — бухгалтер при условии, что 1-й и 2-й члены — бухгалтеры) =
$= (3/8) \times (2/7) \times (1/6) = 1/56.$

❐ **Пример 1.15.** Станок работает при условии одновременного функционирования узлов А, В и С, которые работают независимо друг от друга. Вероятность поломки этих узлов равна 0,2; 0,3; 0,1 соответственно. Какова вероятность, что станок выйдет из строя?

Решение

Станок функционирует только в случае бесперебойной работы каждого узла, в противном случае происходит остановка оборудования.

Для каждого узла вероятности таковы:

P (поломка узла А) = 0,2, следовательно, P (узел А работает) = 0,8;

P (поломка узла В) = 0,3, следовательно, P (узел В работает) = 0,7;

P (поломка узла С) = 0,1, следовательно, P (узел С работает) = 0,9.

По диаграмме рис. 1.7 определим вероятность бесперебойной работы узлов в течение года:

P (работает А и работает В и работает С) =
= P (работает А) × P (работает В) x P (работает С) =
$= 0{,}8 \times 0{,}7 \times 0{,}9 = 0{,}504.$

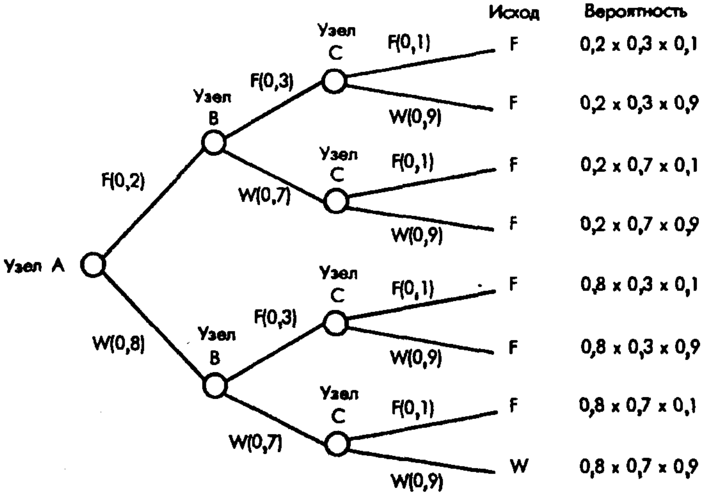


Рис. 1.7. Вероятность поломки оборудования:
F — поломка узла; W — узел работает

Однако нужно вычислить вероятность поломки оборудования в течение года. Эта вероятность равна сумме семи остальных «ветвей» или, так как вероятность полной группы событий (т.е. бесперебойная работа и поломка) равна 1, то:

$$P\ (\text{поломка}) = 1 - P\ (\text{бесперебойная работа}) = 1 - 0{,}504 = 0{,}496.$$

1.5. ФОРМУЛА БАЙЕСА

На основании правила умножения вероятностей мы можем вывести формулу Байеса:

$$PP\ (AB) = P\ (A) \times P\ (B/A) \text{ или } P\ (AB) = P\ (B) \times P\ (A/B),$$

откуда

$$P\ (A/B) = \frac{P\ (AB)}{P\ (B)};$$

Это и есть формула Байеса.

Вероятность P(A) подсчитывается до проведения опыта, поэтому носит теоретический, предварительный характер. Вероятность P(A/B) основывается на данных уже проведенного эксперимента, поэтому более точна с практической точки зрения.

❑ **Пример 1.16.** Фирма имеет три источника поставки комплектующих — фирмы А, В, С. На долю фирмы А приходится 50% общего объема поставок, В — 30% и С — 20%. Из практики известно, что 10% поставляемых фирмой А деталей бракованных, фирмой В — 5% и фирмой С — 6%.

1. Какова вероятность, что взятая наугад деталь была получена от фирмы А?
2. Какова вероятность, что взятая наугад и оказавшаяся бракованной деталь получена от фирмы А?

Решение.

1. Так как 50% комплектующих деталей поставляются фирмой А, то

$$P\ (\text{деталь поставлена А}) = 50\%,\ \text{или}\ P\ (A) = 0{,}5.$$

Следовательно, вероятность того, что взятая наугад деталь была поставлена фирмой А, **равна 0,5**.

2. Рассмотрим рис. 1.8.

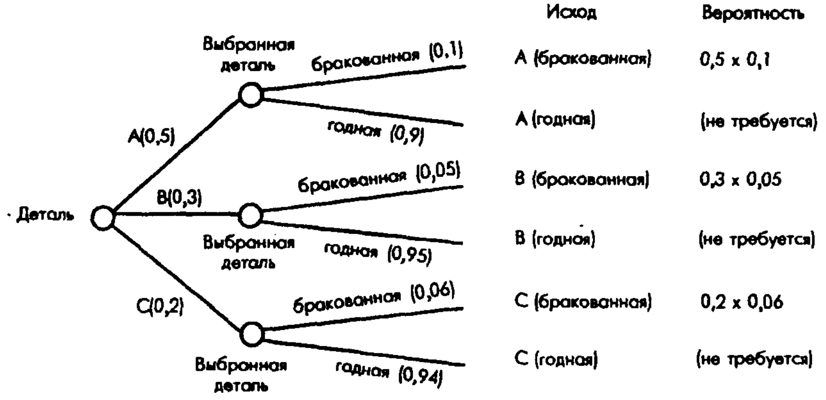


Рис. 1.8. Дерево вероятностей

На основе рис. 1.8 вероятность брака в поставках равна:

$$(0{,}5 \times 0{,}1) + (0{,}3 \times 0{,}5) + (0{,}2 \times 0{,}06) = 0{,}077,$$

т.е. число бракованных деталей в общем количестве составляет 77 из 1000 штук.

Вычислим долю фирмы А в общем количестве бракованной продукции:

$$P\ (\text{фирма А, деталь бракованная}) = \frac{(0{,}5 \times 0{,}1)}{0{,}077} = 0{,}65\,.$$

Иначе говоря, вероятность того, что выбранная деталь была получена от фирмы А при условии, что она оказалась бракованной, изменилась с 0,5 до 0,65. Используя формулу Байеса, это можно выразить так:

$$P\,(A\ /\ \text{с браком}) = \frac{P\,(A \text{ и с браком})}{P\,(\text{с браком})} =$$

$$= \frac{0{,}5 \times 0{,}1}{(0{,}5 \times 0{,}1) + (0{,}3 \times 0{,}05) + (0{,}2 \times 0{,}06)} = \frac{0{,}5 \times 0{,}1}{0{,}077} = 0{,}65.$$

Итак, вероятность того, что поступившая бракованная деталь была поставлена фирмой А, равна 0,65.

❑ **Пример 1.17.** В центральную бухгалтерию корпорации поступили пачки накладных для проверки и обработки. 90% пачек были признаны удовлетворительными: они содержали только 1% неправильно оформленных накладных. Остальные 10% пачек накладных были признаны неудовлетворительными, так как содержали 5% неправильно оформленных накладных.

1. Какова вероятность того, что следующая партия поступивших накладных будет признана неудовлетворительной?
2. Взятая наугад из пачки накладная оказалась оформленной неправильно. Учитывая это, какова вероятность того, что вся пачка накладных будет признана не соответствующей стандартам?
3. Вторая взятая наугад накладная тоже была неправильно формленной. Приняв во внимание оба факта, определите вероятность, что вся пачка накладных является неудовлетворительной.

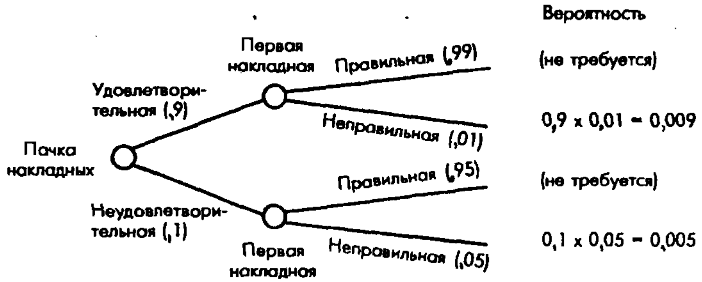


Рис. 1.9. Вероятность появления неправильной накладной

Решение.

Ввиду объемности каждой пачки накладных будем считать, что вероятность появления неправильной накладной существенно не изменится.

1. Без дополнительной информации, основываясь на данных прошлых экспериментов, вероятность того, что пачка накладных будет признана недействительной, равна 0,1.
2. Для последующего решения может быть применена диаграмма в виде «дерева» (см. рис. 1.9).

Полная вероятность обнаружения неправильной накладной равна:

$$0{,}009 + 0{,}005 = 0{,}014.$$

Р (пачка накладных признана неудовлетворительной после того, как первая отобранная накладная признана неправильной

$$\frac{\text{Р (пачка неудовлетворительна и накладная неправильна)}}{\text{Р (неправильная накладная)}} = \frac{0{,}005}{0{,}014} = 0{,}357 .$$

3. Для ответа на последний вопрос используем дополнительную схему.

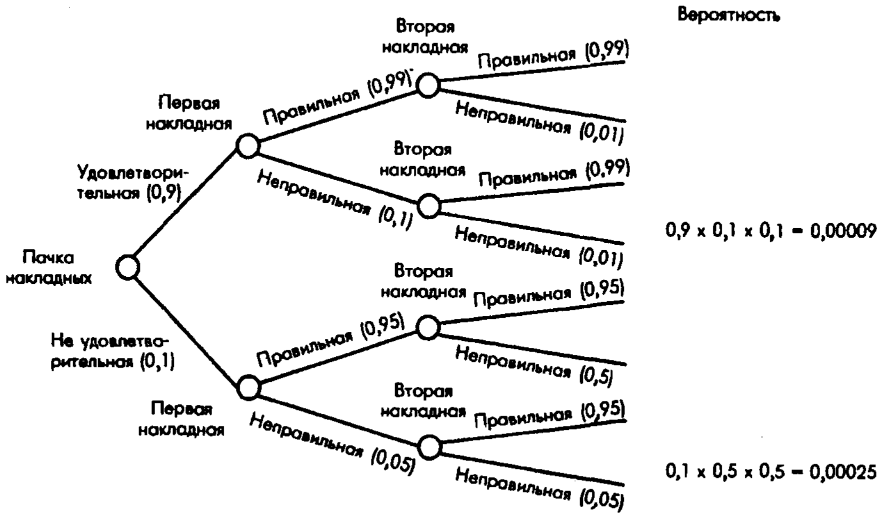


Рис. 1.10. Вероятность обнаружения двух неправильных накладных

Вероятность того, что первые две накладные неправильны, равна:

$$0{,}00009 + 0{,}00025 = 0{,}00034,$$

Р (пачка накладных признана неудовлетворительной из-за двух накладных, признанных неправильными) =

$$= \frac{\text{Р (пачка накладных признана неудовлетворительной и две накладные неправильные)}}{\text{Р (две накладные неправильны)}} =$$

$$= \frac{0{,}00025}{0{,}00034} = 0{,}735 .$$

Итак, если, по предварительным данным, вероятность того, что пачка накладных будет неудовлетворительна, составляла 0,1, то после первого эксперимента она составила 0,36, после второго — 0,74.

1.6. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

Рассмотрим теперь только те эксперименты, результаты которых имеют численное значение. Например, бросив монету 10 раз, зафиксируем выпавшее число «решек».

При многократном повторении эксперимента можно вычислить среднее значение величины. Среднее значение величины (полученное при неограниченно большом числе опытов) иначе называется **математическим ожиданием** E(x). Метод определения математического ожидания имеет много общего с нахождением среднего значения частотного распределения, которое выглядит так:

$$\bar{x} = \frac{\sum fx}{\sum f}.$$

Если значения частот f заменить на относительные частоты (или вероятности) p, то мы получим среднюю, или математическое ожидание:

$$E(x) = \frac{\sum px}{\sum p}.$$

Так как $\sum p = 1$, то $E(x) = \sum px$.

☐ **Пример 1.18**. Стандартная монета брошена 4 раза. Каково ожидаемое число «решек»?

Решение.

Возможны 16 исходов:

(pppp, pppo, ppop, popp, oppp, ppoo, poop, oopp, opop, popo, oppo, pooo, opoo, oopo, ooop, oooo).

Вероятность каждого из исходов равна 1/16. Поэтому

P (0 «решек» за 4 броска) = 1/16;

P (1 «решка» за 4 броска) = 4/16;

P (2 «решки» за 4 броска) = 6/16;

P (3 «решки» за 4 броска) = 4/16;

P (4 «решки» за 4 броска) = 1/16;

Общая вероятность = 16/16 = 1.

Следовательно, ожидаемое количество «решек» за 4 броска равно:

$$E(x) = \sum px = \frac{1}{16} \times 0 + \frac{4}{16} \times 1 + \frac{6}{16} \times 2 + \frac{4}{16} \times 3 + \frac{1}{16} \times 4 = 2,$$

т.е. ожидаемое количество «решек» равно 2.

☐ **Пример 1.19.** Вероятность того, что игрок выиграет 1000 ф. ст. составляет 0,1. Вероятность выигрыша 500 ф. ст. равна 0,2. В случае проигрыша ему нужно будет уплатить 300 ф. ст. Какова ожидаемая прибыль от игры?

Решение

Вероятность проигрыша = 1 − P (выигрыш) =
= 1 − (0,1 + 0,2) =
= 1 − 0, 3 = 0,7.

Ожидаемая прибыль такова:

$$E(x) = \sum px = 0{,}1 \times 1000 + 0{,}2 \times 500 + 0{,}7 \times (-300) = -10 \text{ ф. ст. за игру.}$$

Таков средний размер убытка за одну игру, если играется множество игр при идентичных условиях. В каждой отдельной игре игрок может выиграть 1000, 500 ф. ст. или проиграть 300 ф. ст., но при большом количестве игр убыток составит 10 ф. ст. в расчете на одну игру.

Математическое ожидание помогает определить необходимость проведения эксперимента, так как при большом значении математического ожидания эксперимент целесообразен (см. гл. 3).

1.7. ПЕРЕСТАНОВКИ И СОЧЕТАНИЯ

Производя выборки из совокупности, важно знать, сколько выборок может быть сделано. Для этого мы должны, во-первых, решить влияет ли порядок отбора элементов на результат.

Перестановки: элементы совокупности идентифицированы и порядок, в котором они отобраны имеет большое значение.

☐ **Пример 1.20.** Из пяти букв — A, B, C, Д, E — нужно составить как можно больше трехбуквенных сочетаний.

Решение

На первое место можно поставить любую из пяти букв. Для второй позиции остаются вакантными четыре буквы, для третьей — три. Итак, для первых пяти вариантов существуют четыре вторых и три третьих буквы, т.е. общее число перестановок из трех букв равно:

$$5 \times 4 \times 3 = 60\ .$$

Иными словами, количество перестановок пять по три: $5 \times 4 \times 3 = 60$.

Возможны следующие сочетания:

ABC	ACB	BAC	BCA	CAB	CBA
ABD	ADB	BAD	BDA	DAB	DBA
ABE	AEB	BAE	BEA	EAB	EBA
ADC	ACD	DAC	DCA	CAD	CDA
AEC	ACE	EAC	ECA	CAE	CEA
ADE	AED	DAE	DEA	EAD	EDA
DBC	DCB	BDC	BCD	CDB	CBD

EBC	ECB	BEC	BCE	CEB	CBE
DEC	DCE	EDC	ECD	CDE	CED
DBE	DEB	BDE	BED	EDB	EBD

Теперь необходимо вывести общую формулу для перестановок, чтобы иметь возможность подсчитывать их количество, зная общее количество элементов и сколько элементов нужно выбрать.

Допустим, общее количество элементов n, нужно выбрать три. Тогда на первое место претендуют n элементов, на второе — $(n-1)$, на третье — $(n-2)$ элемента. Число перестановок составит $n \times (n-1) \times (n-2)$. Однако этот метод получения формулы приемлем при выборе группы небольшого объема, в противном случае формула становится громоздкой.

Для краткости перестановка из n элементов по r обозначается как ${}^{n}P_{r}$. Символ !, будучи помещенным после числа, указывает на факториал. Например:

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

или

$$10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 .$$

В общем виде: $n! = n \times (n-1) \times (n-2) \times (n-3) \times \dots \times 1$.

Особо отметим, что $1! = 1$ и по определению $0! = 1$. С небольшими манипуляциями факториал может использоваться для вычисления перестановок. Вернемся к примеру **1.20**. Нужно подсчитать число перестановок из пяти по три. Мы нашли, что ${}^{5}P_{3} = 5 \times 4 \times 3$.

Можем переписать это следующим образом:

$${}^{5}P_{3} = 5 \times 4 \times 3 \times \frac{2 \times 1}{2 \times 1} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = \frac{5!}{2!} .$$

Аналогично

$${}^{n}P_{3} = n \times (n-1) \times (n-2) \times \frac{(n-3) \times (n-4) \times \dots \times 1}{(n-3) \times (n-4) \times \dots \times 1} = \frac{n!}{(n-3)!} .$$

Так как мы не будем всегда отбирать по три элемента, то в общем виде эта формула выглядит так:

$${}^{n}P_{r} = \frac{n!}{(n-r)!} .$$

Сочетания:

число сочетаний означает число выборок, которые могут быть сделаны из n элементов по r элементов — это все r-элементные подмножества n-элементного множества, где различными подмножествами считаются те, которые имеют различный состав элементов, при этом порядок отбора не важен.

❑ **Пример 1.21.** Сколько существует 3-буквенных сочетаний из литер A, B, C, D, E? Порядок букв в сочетании не важен.

Решение.

Вернувшись к решению примера **1.20**, видим, что в каждом ряду, в строке перестановок использованы одни и те же буквы. Каждый ряд представляет возможные комбинации из трех букв. Количество комбинаций из 5 букв, взятых по 3, равно 10. Отметим, что число перестановок трех букв равно:

$$^3P_3 = \frac{3!}{(3-3)!} = \frac{3!}{0!} = \frac{3!}{1} = 6 .$$

Обозначив число комбинаций из 5 элементов, взятых по 3 $^5C_3$, получим:

$$^5C_3 = \frac{^5P_3}{^3P_3} = \frac{5!\,/2!}{3!} = \frac{5!}{3! \times 2!} = 10 .$$

Итак, что же получится если обобщить решение этой задачи, каково число комбинаций n элементов, взятых по 3 за раз?

$$^nC_3 = \frac{^nP_3}{^3P_3} = \frac{n!\,/(n-3)!}{3!} = \frac{n!}{3! \times (n-3)!} .$$

Теперь мы можем записать формулу числа сочетаний из n элементов, взятых по r. В общем виде это выглядит так:

$$^nC_r = \frac{^nP_r}{^rP_r} = \frac{n!\,/(n-r)!}{r!} = \frac{n!}{r! \times (n-r)!} .$$

1.7.1. Использование перестановок и сочетаний для вычисления вероятности

❑ **Пример 1.22**. Вернемся к примеру **1.14.** Совет директоров компании состоит из трех бухгалтеров, трех менеджеров и двух инженеров. Во вновь создаваемый подкомитет должны войти три человека. Необходимо найти вероятность того, что подкомитет будет состоять полностью из бухгалтеров.

Решение.

Число сочетаний трех бухгалтеров из трех бухгалтеров равно:

$$^3C_3 = \frac{3!}{3!\,(3-3)!} = 1.$$

Число сочетаний из 8 чел. по 3 равно:

$$^8C_3 = \frac{8!}{3!\,(8-3)!} = 56 .$$

Следовательно, вероятность того, что все члены подкомитета будут бухгалтерами, равна:

$$\frac{{}^{3}C_{3}}{{}^{8}C_{3}}=\frac{1}{56},$$

т.е. тот же результат, что и в примере **1.14**.

□ **Пример 1.23**. Из трех бухгалтеров, восьми менеджеров и шести научных сотрудников необходимо случайным отбором сформировать комитет из десяти человек. Какова вероятность того, что в комитете окажутся: один бухгалтер, пять менеджеров и четверо научных сотрудников?

Решение.

Число сочетаний для одного бухгалтера $\quad {}^{3}C_{1}=\frac{3!}{1!\,(3-1)!}=3.$

Число сочетаний для пяти менеджеров $\quad {}^{8}C_{5}=\frac{8!}{5!\,(8-5)!}=56.$

Число сочетаний для четырех ученых $\quad {}^{6}C_{4}=\frac{6!}{4!\,(6-4)!}=15.$

Число сочетаний для десяти человек $\quad {}^{17}C_{10}=\frac{17!}{10!\,(17-10)!}=19448.$

Следовательно, итоговая вероятность формирования комитета заданного состава равна:

$$\frac{{}^{3}C_{1}\times{}^{8}C_{5}\times{}^{6}C_{4}}{{}^{17}C_{10}}=\frac{3\times 56\times 15}{19448}=0{,}130.$$

РЕЗЮМЕ

Теория вероятностей имеет дело с понятием неопределенности. При проведении эксперимента мы знаем все возможные исходы, однако не все они произойдут. Ответ на вопрос, каковы же шансы реализации того или иного исхода, дает вероятность.

Событие — это один или несколько исходов, которые нас интересуют. Численные границы вероятности — от 0 до 1 включительно. Сумма вероятностей всех возможных исходов (вероятность полной группы событий) всегда равна 1. Вероятность определяется или на основе свойства симметрии эксперимента, или путем повторения измерений, или на основе субъективной оценки.

Существуют два основных правила подсчета вероятности: правило сложения и правило умножения вероятностей. Правило сложения вероятностей применяется для вычисления вероятности появления события A или B, или обоих вместе:

$$P(A+B)=P(A)+P(B)-P(AB).$$

Для несовместимых событий формула преобразуется в следующую:

$$P(A + B) = P(A) + P(B).$$

Правило умножения вероятностей применяется, когда события происходят вместе:

$$P(AB) = P(A) \times P(B/A).$$

Для независимых событий эта формула преобразуется в следующую:

$$P(AB) = P(A) \times P(B).$$

Перед вычислением вероятности необходимо составить список всех возможных исходов эксперимента. Лучше всего изобразить их графически, для чего используются таблицы, диаграммы в виде «деревьев», диаграммы Венна. Только после этого приступают к расчетам.

Расчет по формуле Байеса применяется с целью модификации вероятности в случае, когда появилась новая дополнительная информация. Этот расчет основан на правиле умножения вероятностей:

$$P(A/B) = \frac{P(AB)}{P(B)}.$$

Математическое ожидание определяется следующим образом:

$$E(x) = \sum px.$$

Мы можем решить — проводить эксперимент или нет, вычислив среднюю величину.

Для определения числа возможных вариантов выборки r элементов из группы в n элементов используются перестановки и сочетания.

Перестановки:

$${}^nP_r = \frac{n!}{(n-r)!} \text{ — порядок отбора элементов важен.}$$

Сочетания:

$${}^nC_r = \frac{n!}{r!\,(n-r)!} \text{ — порядок отбора элементов не важен.}$$

УПРАЖНЕНИЯ

Упражнение 1.1

Монета брошена 4 раза. Какова вероятность выпадения двух «решек» и двух «орлов»

(i) именно в таком порядке;

(ii) в любом порядке?

Упражнение 1.2
Игральная кость брошена дважды. Результаты двух исходов обобщаются. Какова вероятность получения

(i) 23;
(ii) 12?

Упражнение 1.3
Служащие компании «Tutial plc» распределены в таблице по отделам и полу:

Подразделение	*Женщины*	*Мужчины*
Производственный отдел	6	20
Ремонтная мастерская	3	10
Склады	5	5
Автобаза	2	8
Отдел реализации	5	10

Наудачу выбран один служащий. Найти вероятность того, что это:

(i) женщина;
(ii) работник ремонтной мастерской;
(iii) мужчина, работающий на складе;
(iv) женщина, работающая на складе или автобазе;
(v) работник производственного отдела или отдела реализации?

В той же компании решено организовать консультационный комитет из двух человек. Какова вероятность того, что они будут:

(vi) женщины;
(vii) оба из производственного отдела;
(viii) один из магазина, а другой из автобазы;
(ix) женщина из ремонтной мастерской и мужчина, работающий на складе;
(x) обе женщины, работающие на складе, или один человек из производст венного отдела, а другой — мужчина, работающий в отдела реализации?

Упражнение 1.4
На курсах повышения квалификации бухгалтеров учат определять правильность накладной. В качестве проверки преподаватель предлагает обучающимся проверить 10 накладных, 4 из которых содержат ошибки. Он берет наугад из этих 10 две накладные и просит проверить. Какова вероятность того, что они окажутся:

(i) обе ошибочные;
(ii) одна ошибочная, а другая нет?

При условии, что обучающийся идентифицирует неправильную накладную с вероятностью 0,8, а правильную — с вероятностью 0,9, какова вероятность правильной идентификации двух предложенных ему накладных, если:

(iii) обе ошибочные;

(vi) одна ошибочная, а другая нет?

Упражнение 1.5

Магазин получает товар партиями по 100 штук. Если пять, взятых наугад, образцов соответствуют стандартам, партия товара поступает на реализацию. В очередной партии 8 единиц товара с дефектом. Какова вероятность того, что товар поступит на реализацию?

Упражнение 1.6

К каждому из десяти этапов сборки станка прилагается документация. Для каждого этапа вероятность ошибки в документации составляет 0,002. Какова вероятность того, что документы:

(i) в полном порядке;

(ii) имеется одна ошибка?

(iii) реорганизация концелярской работы могла бы свести количество этапов до пяти, однако, тогда вероятность ошибки возрастет до 0,003.

Как изменились бы ответы на вопросы (i) и (ii) при новых условиях?

Упражнение 1.7

R, S, T — компоненты электронной системы. Вероятность бесперебойной работы каждого из компонентов в течение года 0,95; 0,9; 0,93 соответственно.

(i) Какова вероятность безотказной работы всей системы на протяжении этого срока, если необходимо, чтобы работали все три компонента?

(ii) Допустим, достаточно, чтобы работали два из трех компонентов. Какова вероятность безотказной работы системы в этом случае?

(iii) Внесенные усовершенствования сделали эксплуатацию системы возможной, если работает хотя бы один из компонентов. Какова вероятность функционирования системы в течение всего года.

Упражнение 1.8

В электронной системе имеются два компонента G и H, работы одного из которых достаточно для ее функционирования. Вероятности работы компонентов приведены ниже по годам.

Год эксплуатации	*I*	*II*	*III*	*IV*	*V*
Вероятность для компонента G	0,1	0,2	0,2	0,3	0,2
компонента H	0,2	0,3	0,3	0,2	0,0

(i) Какова вероятность функционирования системы на каждом году эксплуатации?

(ii) Какова вероятная продолжительность работы каждого компонента и системы в целом?

Упражнение 1.9

По вероятности попадания в дорожную аварию водители подразделяются страховой компанией на три группы — низкая, средняя, высокая. Среди застрахованных водители этих групп составили 25%, 60%, 15% соответственно. Ниже приведена таблица вероятностей попадания водителей каждой группы в аварии:

Риск водителя	*Вероятность попадания в аварию*		
	низкая	*средняя*	*высокая*
1 авария в год	0,01	0,03	0,10
2 аварии в год	0	0,01	0,05
3 аварии в год	0	0	0,01
4 аварии в год	0	0	0

Требуется найти:

(i) Если м-р А не попадал в аварию в течение года, какова вероятность того, что он принадлежит к группе водителей с высокой вероятностью попадания в дорожную аварию?

(ii) Если м-р В не попадал в аварию в течение четырех лет, какова вероятность того, что он принадлежит к группе водителей с низкой вероятностью попадания в дорожную аварию?

Упражнение 1.10

Фирма собирается выпускать новый товар на рынок. Подсчитано, что вероятность хорошего сбыта продукции равна 0,6; плохого — 0,4. Компания собирается провести маркетинговое исследование, вероятность правильности которого 0,8. Как изменятся первоначальные вероятности уровня реализации, если это исследование предскажет плохой сбыт?

Упражнение 1.11

Какова должна быть сумма страхового взноса за год за дом, оцененный в 60000 ф. ст., чтобы компания могла полностью возместить убытки, если установлено, что в течение года подвергаются разрушению два из каждых ста подобных домов? Из них 5% восстановлению не подлежат, для 25% — убытки составляют 8000 ф. ст.; для остальных — 4000 ф. ст.

Упражнение 1.12
25 участников годового собрания акционеров претендуют на посты председателя, секретаря, казначея и четырех других поста в правлении.

Определите:
(i) Сколько существует способов замещения вакантных мест претендентами?
(ii) Сколько существует способов замещения четырех остальных вакансий после избрания председателя, секретаря и казначея?
(iii) Сколько комитетов могут быть сформированы из 25 человек?

Упражнение 1.13
В транспортной компании работают 10 водителей.

(i) На каждый из пяти обслуживаемых через день заводов требуется послать одного из водителей. Сколькими способами это может быть осуществлено?
(ii) Каждый второй день требуются два водителя для этих пяти заводов. Сколькими способами это может быть осуществлено?

Упражнение 1.14
Компания "Trophit plc" выпускает гайки и болты в стандартной британской и метрической системах мер. Однажды коробка с пятнадцатью 5-мм болтами опрокинулась в ящик с 30-дюймовыми болтами, а коробка с пятнадцатью 5-мм гайками – в ящик с 30-дюймовыми гайками.

Найти:
(i) Какова вероятность, что взятые наугад болт и гайка подойдут друг к другу?
(ii) Какова вероятность, что два болта и две гайки, взятые наугад, образуют две подходящие друг к другу пары?

Упражнение 1.15
Инвестор имеет 10000 ф. ст. для покупки акций либо химической компании, либо пивоваренной. Его брокер привел ему следующие данные о вероятной отдаче денег в последующие 12 месяцев:

Вероятная годовая прибыль	*– 2000*	*– 1000*	*0*	*1000*	*2000*	*3000*	*4000*
Химическая компания	0,05	0,1	0,2	0,2	0,2	0,2	0,05
Пивоваренная компания	0,05	0,2	0,3	0,2	0,1	0,1	0,05

Каков ожидаемый годовой результат вложения денег:
(i) в химическую компанию?
(ii) в пивоварню?

Упражнение 1.16

Международная компания страхования жизни пользуется в своей работе статистическими данными смертности. Ниже приведены цифры о количестве умерших по достижении определенного возраста.

Возраст	0	10	20	30	40	50	60	70	80	90	100
Число доживающих до данного возраста	1000	981	966	944	912	880	748	525	261	45	0

1. Используя таблицу, определите вероятность того, что:
   а) выбранный наугад новорожденный умрет до достижения 60-летнего возраста;
   б) выбранный наугад 30-летний человек умрет до достижения 60-летнего возраста;
   в) выбранный наугад 50-летний человек умрет до достижения 60-летнего возраста.

   Объясните, почему вероятности в одном случае больше, чем в другом.
2. Корпорация собирается начать страхование 50-летних. Предполагается, что страхуемый делает единственный взнос в возрасте 50 лет и, если в течение 10 лет он умирает, то наследник[1] получает 10000 ф. ст., если же нет, тогда компания ничего не платит. Каков должен быть размер взноса, чтобы компания не понесла убытки?
3. Найдите вероятность того, что компания потерпит убыток, если компания назначит сумму взноса в 2000 ф. ст., а не в 10000 ф. ст. и на условиях вопроса 2 страховой полис подпишут 12 человек?
4. В вышеприведенной таблице использовались данные по смертности 1986 г. Оцените надежность, целесообразность использования этих данных при расчетах страховых тарифов на перспективу.
5. Используя следующие данные, ответьте на вопросы:

Возраст	50	52	54	56	58	60
Число доживающих до этого возраста	880	866	846	822	788	748

   а) Известно, что 50-летний человек умрет до достижения 60-летия. Вычислите вероятный год смерти.
   б) Подсчитайте размер страхового взноса по данным вопроса 2, если:

   (i) ожидаемый год смерти до 60 лет оценивается, как в пункте а);
   (ii) постоянная ставка — 8% на годовую премию;
   канцелярские расходы по каждому полису — 100 ф. ст.;
   компания выплачивает страховым агентам за каждый полис 200 ф. ст. комиссионных.

Упражнение 1.17

Каждую пятницу бронированный автомобиль доставляет заработную плату из местного отделения банка в пять фирм. В качестве меры предосторожности стараются использовать различные маршруты. Водитель выбирает один из предложенных диспетчером вариантов.

Нужно найти:

1. а) Какова вероятность того, что нынешний маршрут не повторит предыдущий?
   б) Какова вероятность того, что маршрут не повторится ни разу в течение месяца?
   в) Какова вероятность того, чтс фирмы А и В будут обслужены одна за другой в течение одной пятницы?
2. Для сокращения времени доставки денег диспетчер решил оптимизировать маршрут. Сначала бронеавтомобиль завозит деньги в расположенные поблизости друг от друга фирмы А, В и С, а затем — в остальные D и E. Порядок фирм выбирается произвольно. Как эти данные изменяют ответы на вопросы 1а; 1б; 1в?
3. В дополнение к условию вопроса 2: водителю было приказано не повторять маршрут предыдущей недели. Какова вероятность, что маршрут не повторится в течение месяца?

Глава 2. ВЕРОЯТНОСТНЫЕ РАСПРЕДЕЛЕНИЯ

2.1. ВВЕДЕНИЕ

В предыдущей главе мы рассмотрели эксперименты со случайным исходом. Для численного выражения появления того или иного исхода используется понятие вероятности. Вероятность сложных событий была получена путем обобщения вероятностей отдельных исходов.

В этой главе мы рассмотрим вероятностные распределения сначала дискретных, а затем и непрерывных случайных величин. **Дискретные** случайные величины представляют собой целочисленные значения исходов, **непрерывные** — любые возможные значения. Основные виды вероятностных распределений дискретных случайных величин — биномиальное и распределение Пуассона. Они особенно часто используются в аудиторском деле. Например, при проверке может строиться распределение счетов по доле ошибок.

Для непрерывных случайных величин также существует несколько видов вероятностных распределений, среди которых наиболее часто используется нормальное распределение (см. § 2.7.). Особенно важную роль нормальное распределение играет при рассмотрении средних значений случайных величин. (см гл. 4).

Часто расчеты биномиального и пуассоновского распределения отнимают много времени, поэтому используется приближение, что позволяет упростить расчеты, почти не снижая точности.

В § 2.5 рассматривается использование пуассоновского распределения в качестве приближения к биномиальному распределению. В § 2.8 и 2.9 показывается, как использовать нормальное распределение для аппроксимации биномиального и пуассоновского.

2.2. ВЕРОЯТНОСТНЫЕ РАСПРЕДЕЛЕНИЯ ДИСКРЕТНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

2.2.1. Дискретная случайная величина

Если значения исходов эксперимента целочисленны, то они представляют собой дискретные величины (см. табл. 2.1.)

Обозначим случайную величину R, а ее значение — r, отсюда из табл. 2.1. (e) количество проданных за день тортов:

$$r = 0 \text{ или } 1 \text{ или } 2 \text{ или } 3 \text{ или } 4 \text{ или } 5.$$

Таблица 2.1. **Дискретные случайные величины**

	Эксперимент со случайным исходом	*Случайная величина*	*Значение случайной величины*
(а)	Бросание монеты 2 раза и регистрация числа выпавших «решек»	Число «решек»	0, 1, 2
(б)	Бросание игральной кости один раз и регистрация результата со случайным исходом	Выпавшее число	0, 1, 2, 3, 4, 5, 6
(в)	Отбор 5 образцов из партии товаров и регистрация числа бракованных	Число бракованных образцов	0, 1, 2, 3, 4, 5
(г)	Регистрация числа дорожных происшествий на определенном участке дороги в течение недели	Число дорожных происшествий за неделю	0, 1, 2, 3, ...
(д)	Регистрация спроса на машины	Число заказов на машины в течение дня	0, 1, 2, 3, ...
(е)	Количество проданных за день тортов, (если их всего изготавливается пять)	Количество проданных тортов	0, 1, 2, 3, 4, 5

2.2.2. Распределение дискретной случайной величины

Рассмотрим пример с проверкой 10 накладных. Существует 11 исходов эксперимента (табл. 2.2):

Таблица 2.2. **Число правильных накладных**

Номер исхода	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Число правильных накладных	0	1	2	3	4	5	6	7	8	9	10
Число неправильных накладных	10	9	8	7	6	5	4	3	2	1	0

Число правильных накладных представляет собой дискретную случайную величину. Когда мы собираемся произвести эксперимент, заранее неизвестно, какой исход из одиннадцати возможных будет иметь место, однако можно просчитать вероятность каждого из них.

Если принять за R эксперимент, состоящий из дискретных случайных величин, то набор вероятностей, соответствующий каждому из исходов эксперимента, будет называться вероятностным распределением этого R. Вероятность того, что дискретная случайная величина примет какое-то значение r, обозначается так: $P(R = r)$.

❑ **Пример 2.1.**

а) Эксперимент: дважды бросаем монету и регистрируем число "решек".
Возможные исходы: {ОО, РО, ОР, РР}.
Значения дискретной случайной величины: $r = 0,1,2$.
Вероятности: $P\ (R = r) = \{1/4,\ 2/4,\ 1/4\}$.

б) Эксперимент: игральная кость брошена один раз, регистрируем выпавшие очки.

Возможные исходы: {1, 2, 3, 4, 5, 6}.
Значения дискретной случайной величины: r = 1, 2, 3, 4, 5, 6.
Вероятности: P (R = r) = {1/6, 1/6, 1/6, 1/6, 1/6, 1/6}.

в) Эксперимент: эксперт по инвестициям следующим образом оценивает вероятность получения прибыли на инвестиции:
Возможные исходы (тыс. ф. ст.): 1, 2, 3, 4, 5.
Значения дискретной случайной величины: r = 1, 2, 3, 4, 5.
Вероятности: P = (R = r) = {0,4; 0,3; 0,2; 0,1; 0,0}.

В вышеприведенных примерах вероятностные распределения были представлены списком вероятностей, что неудобно при больших экспериментах.

Так как существует связь между составляющими вероятностного распределения, то ее можно выразить математической функцией:

В примере 2.1 (а): $P(R= r) = {}^2C_r \times (1/4)$ $\qquad r = 0, 1, 2;$

В примере 2.1 (б): $P(R = r) = 1/6$ $\qquad r = 1, 2, ..., 6;$

В примере 2.1 (в): $P(R = r) = 0{,}1\ (5-r)$ $\qquad r = 1, 2, ..., 5.$

В случае необходимости каждый элемент распределения может быть вычислен по этим формулам. Каждый раз, когда это возможно, распределение вероятностей дискретной случайной величины выражается математической функцией f (r), где P (R = r) = f (r).

2.2.3. Графическое представление распределения дискретной случайной величины

□ **Пример 2.2.** Распределение дискретной случайной величины может быть представлено в виде линейного графика.

а) Для примера **2.1 а**), в котором два раза бросается монета, вероятностное распределение может быть проиллюстрировано рис. 2.1.

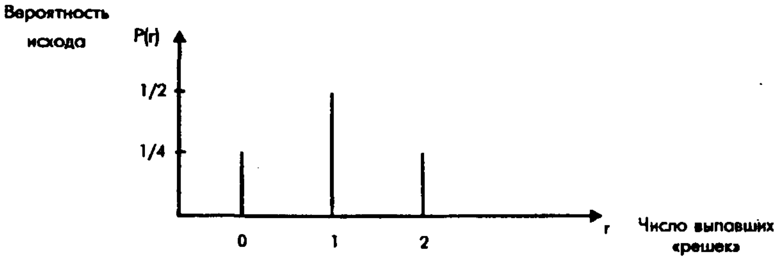


Рис. 2.1. Распределение количества "решек" при двукратном бросании монеты

б) Для примера **2.1 б**), в котором регистрируется количество очков на брошенной один раз игральной кости иллюстрация дана на рис. 2.2.

В каждом примере действуют правила, касающиеся вероятностей:

$$0 \le p \le 1 .$$

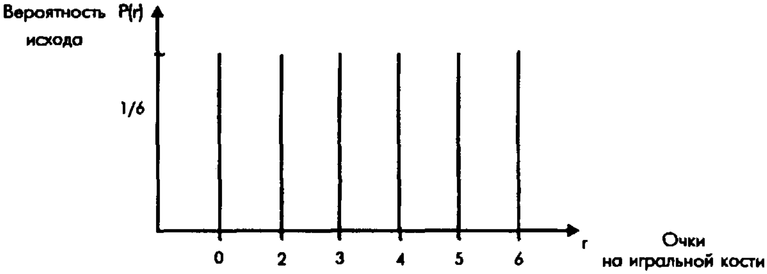


Рис.2.2. Распределение исходов — игральная кость брошена один раз

Сумма вероятностей всех исходов (полной группы событий) равна 1, в примере **2.1 а):**

$$1/4 + 2/4 + 1/4 = 1;$$

в примере **2.1 б)**:

$$6 \times (1/6) = 1\ .$$

2.2.4. Математическое ожидание и стандартное отклонение вероятностного распределения

Таблица 2.3. **Количество автомобилей, проданных в течение одного дня (данные за месяц)**

Количество проданных за день автомобилей	*Количество дней*	*Расчет*	
r	*f*	*fr*	*fr* ^{2}
0	5	0	0
1	4	4	4
2	8	16	32
3	4	12	36
4	6	24	96
5	3	15	75
Всего	30	71	243

Среднее количество машин, проданных в день:

$$\bar{r} = \frac{\sum f\,r}{\sum f} = \frac{(5 \times 0) + (4 \times 1) + (8 \times 2) + (4 \times 3) + (6 \times 4) + (3 \times 5)}{5 + 4 + 8 + 4 + 6 + 3},$$

$$\bar{r} = \frac{71}{30} = 2{,}37$$ — машин в день.

Дисперсия количества машин $\bar{r}^2$ проданных в день:

$$\sigma^2 = \frac{\sum f\, r^2}{\sum f} - \bar{r}^2 = \frac{243}{30} - 2{,}367^2 = 2{,}50\ .$$

Среднее квадратическое отклонение = $\sqrt{\text{дисперсия}}$.
Отсюда:

$$\sigma = \sqrt{\frac{243}{30} - 2{,}367^2} = 1{,}58 \text{ машин в день.}$$

Частотное распределение используется для оценки вероятности продажи машин в день:

Таблица 2.4. **Вероятность продажи машин**

Количество машин в день	*Число дней*	*Вероятность продажи r машин в день*	*Расчет*	
r	f	$P(r)$	$r\,P(r)$	$r^2\,P(r)$
0	5	5/30	0	0
1	4	4/30	4/30	4/30
2	8	8/30	16/30	32/30
3	4	4/30	12/30	36/30
4	6	6/30	24/30	96/30
5	3	3/30	15/30	75/30
Всего	30	1	71/30	243/30

Вероятность — это относительная частота появления каждого значения дискретной случайной величины. Среднее значение и стандартное отклонение можно найти с помощью вероятностного распределения и с помощью частотного распределения. В этом случае используется относительная частота (вероятность), которая заменяет частоту. Для вероятностного распределения:

$$\textit{Среднее значение} = \frac{\sum r\, P(r)}{\sum P(r)} =$$

$$= \frac{(0 \times 5/30) + (1 \times 4/30) + (2 \times 8/30) + (3 \times 4/30) + (4 \times 6/30) + (5 \times 3/30)}{1} = 2{,}37$$

машин в день.

$$\textit{Дисперсия} = \frac{\sum r^2\, P(r)}{\sum r\, P(r)} - \left(\frac{\sum r^2\, P(r)}{\sum P(r)}\right)^2 = \frac{243/30}{1} - 2{,}37^2 = 2{,}50\ .$$

Среднее значение, основанное на вероятностном распределении, называется обычно **математическим ожиданием** (при условии многократного проведения эксперимента). Математическое ожидание дискретной случайной величины обозначается E(r) или μ.

Так как $\sum P(r) = 1$, то математическое ожидание имеет вид:

$$E(r) = \sum r\,P(r)\,.$$

Вариация вероятностного распределения может быть измерена при помощи среднего квадратического отклонения или дисперсии дискретной случайной величины:

$$\textit{Дисперсия} = (\textit{Стандартное отклонение})^2.$$

А так как $\sum P(r) = 1$:

$$\sigma^2 = \sum r^2\,P(r) - (E(r))^2.$$

Пример 2.3. Бросаем монету три раза и регистрируем число «решек».

Возможные исходы:

{ppp, ppo, pop, opp, poo, opo, oop, ooo}

Дискретная случайная величина — количество «решек», ее возможные значения:

$$r = 0 \text{ или } 1 \text{ или } 2 \text{ или } 3\,.$$

Вероятностное распределение, основанное на значениях возможных исходов, следующее (табл. 2.5):

Таблица 2.5. **Распределение вероятностей «решек»**

Число «решек», r	*1*	*2*	*3*	*4*
P(r)	1/8	3/8	3/8	1/8

Математическое ожидание «решек»:

$$E(r) = \sum r\,P(r) = (0 \times 1/8) + (1 \times 3/8) + (2 \times 3/8) + (3 \times 1/8) = 12/8 = 1{,}5.$$

Дисперсия «решек»:

$$\sigma^2 = \sum r^2\,P(r) - (E(r))^2 = (0 \times 1/8 + 1 \times 3/8 + 4 \times 3/8 + 9 \times 1/8) - 1{,}5^2 = 3{,}0 - 2{,}25 = 0{,}75.$$

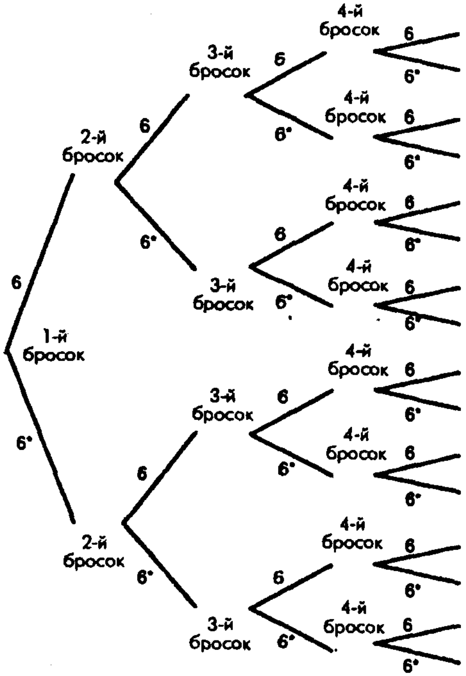
Стандартное отклонение числа «решек»

$$\sigma = \sqrt{0{,}75} = 0{,}87.$$

2.3. БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

2.3.1. Что такое биномиальное распределение

Биномиальное распределение можно проиллюстрировать следующим примером. Игральная кость брошена четыре раза. Нас интересуют исходы с "шестеркой", следовательно, "успех", если выпало "шесть", в противном случае — "неудача". Вероятность первого — $1/6$, второго — $5/6$.



Исходы четырех опытов	Количество "успехов", т.е. "шестерок"	Вероятность каждого исхода
6,6,6,6	4	1/6 X 1/6 X 1/6 X 1/6
6,6,6,6*	3	1/6 X 1/6 X 1/6 X 5/6
6,6,6*,6	3	1/6 X 1/6 X 5/6 X 1/6
6,6,6*,6*	2	1/6 X 1/6 X 5/6 X 5/6
6,6*,6,6	3	1/6 X 5/6 X 1/6 X 1/6
6,6*,6,6*	2	1/6 X 5/6 X 1/6 X 5/6
6,6*,6*,6	2	1/6 X 5/6 X 5/6 X 1/6
6,6*,6*,6*	1	1/6 X 5/6 X 5/6 X 5/6
6*,6,6,6	3	5/6 X 1/6 X 1/6 X 1/6
6*,6,6,6*	2	5/6 X 1/6 X 1/6 X 5/6
6*,6,6*,6	2	5/6 X 1/6 X 5/6 X 1/6
6*,6,6*,6*	1	5/6 X 1/6 X 5/6 X 5/6
6*,6*,6,6	2	5/6 X 5/6 X 1/6 X 1/6
6*,6*,6,6*	1	5/6 X 5/6 X 1/6 X 5/6
6*,6*,6*,6	1	5/6 X 5/6 X 5/6 X 1/6
6*,6*,6*,6*	0	5/6 X 5/6 X 5/6 X 5/6

Рис. 2.3. Возможные исходы эксперимента: 6 — выпало "шесть", 6* — выпало не "шесть"

Для того, чтобы вывести формулу распределения вероятностей, рассмотрим пример появления двух "шестерок", т.е. два " успеха" и две " неудачи". Это может произойти шестью способами.

6 6 6* 6*; 6 6* 6 6*; 6 6* 6* 6; 6* 6* 6 6; 6* 6 6* 6; 6* 6 6 6*.

Вероятности этих исходов равны:

$$\frac{1}{6}\times\frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}=\left(\frac{1}{6}\right)^2\times\left(\frac{5}{6}\right)^2.$$

Отсюда вероятность появления двух "шестерок":

$$\left(\frac{1}{6}\right)^2\times\left(\frac{5}{6}\right)^2\times\text{(количество возможных способов получения двух "шестерок" при четырех бросках)}$$

$$P(\text{две "шестерки" при четырех бросках})=\left(\frac{1}{6}\right)^2\times\left(\frac{5}{6}\right)^2\times 6=\frac{150}{1296}.$$

Для иллюстрации вероятностей остальных исходов воспользуемся схемой:

$$P\text{ (0 "шестерок" при четырех бросках)}=\left(\frac{1}{6}\right)^0\times\left(\frac{5}{6}\right)^4\times 1=\frac{625}{1296};$$

$$P\text{ (1 "шестерка" при четырех бросках)}=\left(\frac{1}{6}\right)^1\times\left(\frac{5}{6}\right)^3\times 4=\frac{500}{1296};$$

$$P\text{ (2 "шестерки" при четырех бросках)}=\left(\frac{1}{6}\right)^2\times\left(\frac{5}{6}\right)^2\times 6=\frac{150}{1296};$$

$$P\text{ (3 "шестерки" при четырех бросках)}=\left(\frac{1}{6}\right)^3\times\left(\frac{5}{6}\right)^1\times 4=\frac{20}{1296};$$

$$P\text{ (4 "шестерки" при четырех бросках)}=\left(\frac{1}{6}\right)^4\times\left(\frac{5}{6}\right)^0=\frac{1}{1296}.$$

$$\text{Полная вероятность}=\frac{1296}{1296}=1.$$

Отсюда можно вывести формулу, описывающую вероятность появления "шестерки" при четырех бросках:

$$P\text{ (r "шестерок" при четырех бросках)}=\left(\frac{1}{6}\right)^r\times\left(\frac{5}{6}\right)^{4-r}\times\text{(число опытов с r "шестерками")}.$$

Из главы 1 известно, что число "успехов" r при четырех попытках представляет собой число комбинаций r элементов из четырех элементов, т.е.:

$$^4C_r = \frac{4!}{r!\,(4-r)!}\,.$$

Отсюда:

$$P\ (r\ \text{"шестерок" при четырех бросках}) = \left(\frac{1}{6}\right)^r \times \left(\frac{5}{6}\right)^{4-r} \times {}^4C_r\,.$$

Важные характеристики этого биномиального эксперимента:

1. Все четыре опыта абсолютно идентичны.
2. Результаты опыта друг от друга не зависят.
3. Для каждого опыта возможны два исхода: 6 и не 6 — "успех" и "неудача".
4. Для каждого опыта вероятность "успеха" одинакова — 1/6.

Эти характеристики являются необходимым условием биномиального распределения. Теперь от конкретного эксперимента перейти к общей формуле.

2.3.2. Биномиальное распределение

В любом эксперименте, где для n идентичных, независимых опытов с двумя возможными исходами ("успех" и "неудача") вероятность "успеха" одна и та же, вероятность r "успехов" в n опытах равна:

P (r) = (вероятность r "успехов") × (вероятность (n-r) "неудач") ×
× (число способов, которыми это может быть сделано)

$$P\,(r) = \frac{n!}{r!\ (n-r)!}\,p^r q^{n-r} \qquad r = 0,\ 1,\ 2\ldots,\ n,$$

где p — вероятность "успеха" в очередном опыте;
q — вероятность "неудачи" $p + q = 1$.

❐ **Пример 2.4.** Вероятность поломки одного из пяти работающих независимо друг от друга станков равна 0,2. Если происходит поломка, станок до конца дня не работает. Какова вероятность, что 0, 1, 2, 3, 4, 5 станков сломаются в течение дня?

Решение.

В данном случае присутствуют все условия биномиального распределения:

1. Пять станков представляют собой пять идентичных опытов.
2. Станки работают независимо друг от друга.
3. Возможны два исхода для каждого из станков — или он ломается или нет.
4. Вероятность поломки одинакова — 0,2.

Следовательно, вероятность бесперебойной работы равна 0,8, отсюда вероятность r поломок в течение дня:

$$P\,(r) = {}^5C_r\,(0{,}2)^r\,(0{,}8)^{5-r}\,, \qquad r = 0,\ 1,\ \ldots,\ 5.$$

Таблица 2.6. **Вероятность поломки станков в течение дня**

Число станков		*Число вариантов появления*	*Вероятность*
сломавшихся r	*несломавшихся* $5-r$	${}^5C_r$	$P(r)$
P (r сломавшихся станков в день)			
0	5	$\frac{5!}{0!\,(5-0)!}=1$	$1 \times (0,2)^0 \times (0,8)^5 = 0,3277$
1	4	$\frac{5!}{1!\,(5-1)!}=5$	$5 \times (0,2)^1 \times (0,8)^4 = 0,4096$
2	3	$\frac{5!}{2!\,(5-2)!}=10$	$10 \times (0,2)^2 \times (0,8)^3 = 0,2048$
3	2	$\frac{5!}{3!\,(5-3)!}=10$	$10 \times (0,2)^3 \times (0,8)^2 = 0,0512$
4	1	$\frac{5!}{4!\,(5-4)!}=5$	$5 \times (0,2)^4 \times (0,8)^1 = 0,0064$
5	0	$\frac{5!}{5!\,(5-5)!}=1$	$1 \times (0,2)^5 \times (0,8)^0 = 0,0003$
		Всего:	1,0000

Дискретное вероятностное распределение $P(r)$ для всех r можно проиллюстрировать графиком:

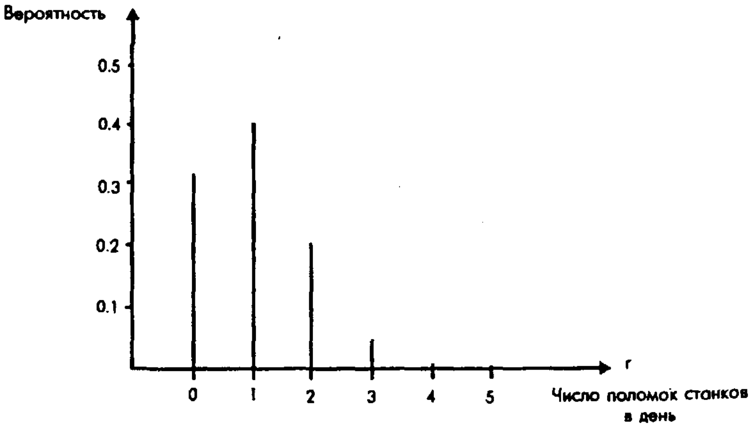


Рис. 2.4. Вероятностное распределение поломок одного из пяти станков в течение дня

2.3.3. Математическое ожидание и стандартное отклонение для биномиального распределения

Для любой дискретной случайной величины математическое ожидание составляет:

$$E(r) = \sum r\,P(r),$$

Может быть показано, что для случайной величины с биномиальным распределением вероятностей:

$$E(r) = \sum r\,P(r) = np,$$

где n — число опытов;
p — вероятность успеха в каждом из них;
P (r) — биномиальная вероятность.

Аналогично стандартное отклонение для дискретной случайной величины равно:

$$\sigma = \sqrt{\sum r^2 P(r) - \left(\sum r\,P(r)\right)^2}\ .$$

Для случайной величины с биномиальным распределением вероятностей:

$$\sigma = \sqrt{\sum r^2 P(r) - \left(\sum r\,P(r)\right)^2} = \sqrt{n\,p\,(1-p)} = \sqrt{n\,p\,q}\ .$$

Следовательно, дисперсия равна $\sigma^2 = \sqrt{n\,p\,q}$, где q — вероятность "неудачи" в любом из опытов.

Эти характеристики биномиального распределения расчитаны в примере 2.5.

❑ **Пример 2.5.** По данным примера 2.4 найдем математическое ожидание сломавшихся за день станков и стандартное отклонение. Для расчетов используем сначала общую формулу для E(r) и σ, а потом формулу для биномиального распределения.

Таблица 2.7. **Вероятность поломки r станков в день**

Число сломавшихся за день станков r	*Вероятность поломки P(r)*	*r P(r)*	*r^2 P(r)*
0	0,3277	0	0
1	0,4096	0,4096	0,4096
2	0,2048	0,4096	0,8192
3	0,0512	0,1536	0,4608
4	0,0064	0,0256	0,1024
5	0,0003	0,0015	0,0075
Всего	1,0000	0,9999	1,7995

Примечание. Значения вероятностей округлены до четырех знаков после запятой, что приводит к небольшим погрешностям, а именно: итоговые значения должны быть равны 1,0000 и 1,8000 соответственно.

По общей формуле ожидаемое количество поломок в день:

$$E(r) = \sum r\,P(r)\,,\quad E(r) = 1,0,$$

по формуле биномиального распределения:

$$E(r) = n\,p = 5 \times 0{,}2 = 1{,}0\,,$$

т.е. ожидаемое количество поломок — один станок в день. Стандартное отклонение по общей формуле:

$$\sigma = \sqrt{\sum r^2\,P(r) - \left(\sum r\,P(r)\right)^2}\,,$$

отсюда

$$\sigma = \sqrt{1{,}8 - 1{,}0^2} = 0{,}8944$$

(до четырех знаков после запятой).

По формуле биномиального распределения:

$$\sigma = \sqrt{n\,p\,q} = \sqrt{5 \times 0{,}2 \times 0{,}8} = 0{,}8944$$

(до четырех знаков после запятой),
дисперсия:

$$\sigma^2 = n\,p\,q = 5 \times 0{,}2 \times 0{,}8 = 0{,}8\,.$$

Иногда нужно знать долю "успехов" в общем количестве опытов. Исходя из формулы математического ожидания, получим:

$$E\,(\text{доля " успехов" в } n \text{ опытах}) = \frac{n\,p}{n} = p\,.$$

Формула стандартного отклонения доли "успехов" имеет следующий вид:

$$\sigma\,(\text{доля "успехов"}) = \frac{\sqrt{n\,p\,q}}{n} = \sqrt{\frac{p\,q}{n}}\,.$$

Отсюда в примере **2.5**, предполагаемая доля поломок станков в день:

$$\frac{n\,p}{n} = \frac{1}{5} = 0{,}2 = p,$$

а стандартное отклонение:

$$\sigma = \sqrt{\frac{0{,}2 \times 0{,}8}{5}} = 0{,}179$$

(до трех знаков после запятой).

Дисперсия доли поломок в день:

$$\sigma^2 = \frac{0{,}2 \times 0{,}8}{5} = 0{,}032.$$

☐ **Пример 2.6.** Компания производит пружины, 10% из которых оказываются бракованными. Сто пружин отобраны для контроля качества. Требуется найти ожидаемое количество бракованных пружин и стандартное отклонение бракованных в отобранных образцах, а также вероятность того, что в выборке по меньшей мере 15 бракованных пружин.

Решение.

Используем биномиальное распределение, так как:

1. Имеются 100 идентичных опытов.
2. Опыты независимы, так как пружины отбираются наугад.
3. Для каждого опыта возможны два исхода: пружина может быть с дефектом и без оного.
4. Вероятность, что любая из пружин имеет дефект, равна 0,1. Поскольку выборка делается из массовой партии, доля бракованных пружин сильно измениться не может.

Ожидаемое количество пружин с дефектом:

$$E(r) = np = 100 \times 0{,}1 = 10 \text{ пружин в выборке.}$$

Стандартное отклонение брака: $\sigma = \sqrt{npq}$:

$$\sigma = \sqrt{100 \times 0{,}1 \times 0{,}9} = 3 \text{ пружины в выборке.}$$

Вероятность того, что имеется r бракованных образцов в выборке:

$$P(r \text{ бракованных образцов из 100 едениц}) = {}^{100}C_r\,(0{,}1)^r\,(0{,}9)^{100-r};\quad r = 0, 1, 2, \ldots, 100.$$

$$P(r \geq 15) = P(15) + P(16) + \ldots + P(100) = 1 - P(0) + P(1) + \ldots + P(14)\,.$$

Расчеты в данном случае займут много времени и места, поэтому методы приблизительного вычисления вероятности будут рассмотрены в § 2.5 и 2.8.

2.4. РАСПРЕДЕЛЕНИЕ ПУАССОНА

2.4.1. Что такое распределение Пуассона

Например, регистрируется количество дорожных происшествий за неделю на определенном участке дороги. Это число представляет собой случайную величину, которая может принимать значения: 0, 1, 2, 3, ... (верхнего предела нет). Число дорожных происшествий может быть каким угодно большим. Если рассмотреть какой-либо короткий временной промежуток в течение недели, скажем минуту, то происшествие либо произойдет на его протяжении, либо нет. Вероятность дорожного происшествия в течение отдельно взятой минуты очень мала, и примерно такая же она для всех минут.

Распределение вероятностей числа происшествий описывается формулой:

$$P(r \text{ происшествий в неделю}) = \frac{m^r e^{-m}}{r!}, \quad r = 0, 1, 2, 3, \ldots,$$

где m — среднее количество происшествий за неделю на определенном участке дороги;
e — константа, равная 2,718...

Характерные особенности данных, для которых наилучшим образом подходит распределение Пуассона, следующие:

1. Каждый малый интервал времени может рассматриваться как опыт, результатом которого является одно из двух: либо происшествие ("успех"), либо его отсутствие ("неудача"). Интервалы столь малы, что может быть только один "успех" в одном интервале, вероятность которого мала и неизменна.
2. Число "успехов" в одном большом интервале не зависит от их числа в другом, т.е. "успехи" беспорядочно разбросаны по временным промежуткам.
3. Среднее число "успехов" постоянно на протяжении всего времени.

Распределение вероятностей Пуассона может быть использовано не только при работе со случайными величинами на временных интервалах, но и при учете дефектов дорожного покрытия на километр пути или опечаток на страницу текста.

Общая формула распределения вероятностей Пуассона:

$$P(r \text{ "успехов" на заданном интервале}) = \frac{m^r e^{-m}}{r!}, \quad r = 0, 1, 2, 3, \ldots,$$

где m — среднее число "успехов" на единицу.

В таблицах распределения вероятностей Пуассона значения $P(r)$ табулированы для определенных значений m и r.

❑ **Пример 2.7.** В среднем на телефонной станции заказывают три телефонных разговора в течение пяти минут. Какова вероятность, что будет заказано 0, 1, 2, 3, 4 или больше четырех разговоров в течение пяти минут?

Решение.

Применим распределение вероятностей Пуассона, так как:

1. Существует неограниченное количество опытов, т.е. маленьких отрезков времени, когда может появиться заказ на телефонный разговор, вероятность чего мала и постоянна.
2. Считается, что спрос на телефонные разговоры беспорядочно распределен во времени.
3. Считается, что среднее число телефонных разговоров в любом 5-минутном отрезке времени одинаково.

В этом примере среднее число заказов равно 3 за 5 минут. Отсюда, распределение Пуассона:

$$P(r \text{ заказов за 5 минут}) = \frac{3^r e^{-3}}{r!}, \quad r = 0, 1, 2, 3, 4, \ldots,$$

$$P\,(0\ \text{заказов за 5 минут}) = P\,(0) = \frac{3^0\,e^{-3}}{0!} = \frac{1}{1} \times 0{,}0498 = 0{,}0498\ ;$$

$$P\,(1\ \text{заказа за 5 минут}) = P\,(1) = \frac{3^1\,e^{-3}}{1!} = \frac{3}{1} \times P\,(0) = 3 \times 0{,}0498 = 0{,}1494\ ;$$

$$P\,(2\ \text{заказа за 5 минут}) = P\,(2) = \frac{3^2\,e^{-3}}{2!} = \frac{3}{2} \times P\,(1) = \frac{3}{2} \times 0{,}1494 = 0{,}2240\ ;$$

$$P\,(3\ \text{заказа за 5 минут}) = P\,(3) = \frac{3^3\,e^{-3}}{3!} = \frac{3}{3} \times P\,(2) = \frac{3}{3} \times 0{,}2240 = 0{,}2240\ ;$$

$$P\,(4\ \text{заказа за 5 минут}) = P\,(4) = \frac{3^4\,e^{-3}}{4!} = \frac{3}{4} \times P\,(3) = \frac{3}{4} \times 0{,}2240 = 0{,}1680\ ;$$

$$\begin{aligned} P\,(\text{более 4 заказов за 5 минут}) &= \{P(5) + P(6) + P(7) + P(8) + \ldots\} = \\ &= 1 - \{P(0) + P(1) + P(2) + P(3) + P(4)\} \\ &= 1 - \{0{,}0498 + 0{,}1494 + 0{,}2240 + 0{,}2240 + 0{,}1680\} \\ &= 1 - 0{,}8152 = 0{,}1848. \end{aligned}$$

При распределении вероятностей Пуассона, зная среднее число "успехов" на 5-минутном промежутке (например m, как в примере **2.7**), для того чтобы узнать r среднее число "успехов" за один час, нужно просто умножить m на 12. В примере **2.7** среднее число заказов в час составит: $3 \times 12 = 36$. Аналогично, если требуется определить среднее число заказов в минуту: $m = 3/5 = 0{,}6$.

❑ **Пример 2.8.** В среднем за пять дней рабочей недели на автоматической линии происходят 3,4 неполадок. Какова вероятность двух неполадок в каждый день работы?

Решение.

Можно применить распределение Пуассона:

1. Существует неограниченное количество опытов, т.е. малых промежутков времени, в течение каждого из них может произойти или не произойти неполадка на автоматической линии. Вероятность этого для каждого промежутка времени мала и постоянна.
2. Предполагается, что неполадки беспорядочно расположены во времени.
3. Предполагается, что среднее число неполадок в течение любых пяти дней постоянно.

Среднее число неполадок равно 3, 4 за пять дней. Отсюда число неполадок в день:

$$\frac{3{,}4}{5} = 0{,}68 = m\ .$$

Следовательно,

$$P(r\ \text{неполадок в день}) = \frac{0{,}68^r\,e^{-0{,}68}}{r!}\ , \qquad r = 0,\ 1,\ 2,\ 3,\ \ldots$$

Поэтому

$$P(2\ \text{неполадки в день}) = \frac{0{,}68^2\, e^{-0{,}68}}{2!} = 0{,}1171.$$

☐ **Пример 2.9.** В компании, сдающей на прокат две машины, каждодневный спрос на автомобили подчиняется распределению Пуассона и в среднем составляет 1,3 машины в день. Предположим, машины используются в равной степени. Какова вероятность, что в любой из дней:

1) ни на одну машину не будет заказов;
2) одна из машин совершенно точно будет арендована, а другая — то ли будет, то ли нет;
3) на обе поступят заказы.

Решение.

Число заказов на машину в день — это дискретная величина. Вероятность r заказов такова:

$$P(r) = \frac{1{,}3^r\, e^{-1{,}3}}{r!}, \qquad r = 0, 1, 2, 3, \ldots$$

1. P(нет заказов на машины) = P(спрос за день равен 0):

$$P(r = 0) = \frac{1{,}3^0\, e^{-1{,}3}}{0!} = 0{,}2725\,.$$

2. P(ни один автомобиль не заказан) = P(спрос на одну из машин 0 или 1, а на другую — 1)

$$P(r = 0) + P(r = 1) \times P(\text{другая машина арендована}) =$$

$$= 0{,}2725 + \frac{1{,}3\, e^{-1{,}3}}{1!} \times \frac{1}{2} = 0{,}4497.$$

Примечание: P(другая машина арендована)=1/2, так как машины используются в равной степени.

3. P(обе машины арендованы) = P(спрос в день ≥ 2) =

$$= 1 - P(r \le 2) =$$

$$= 1 - \{P(r = 0) + P(r = 1)\} =$$

$$= 1 - 0{,}6268 = 0{,}3732.$$

2.4.2. Математическое ожидание и дисперсия распределения Пуассона

Математическое ожидание (или среднее число "успехов" на каком-то интервале) может быть определено по данным конкретной ситуации. Если найдено математическое ожидание, то и дисперсия известна, так как одно из свойств распределения вероятностей Пуассона состоит в следующем:

Математическое ожидание = Дисперсия.

Отсюда стандартное отклонение числа "успехов" на интервале равно:

$$\text{Стандартное отклонение} = \sqrt{\text{Дисперсия}} = \sqrt{\text{Математическое ожидание}}.$$

Это свойство полезно в тех случаях, когда имеются данные о случайной величине и требуется узнать, применимо или нет распределение Пуассона.

2.5. РАСПРЕДЕЛЕНИЕ ПУАССОНА КАК АППРОКСИМАЦИЯ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ

При определенных обстоятельствах распределение Пуассона может быть использовано как замена биномиального распределения. Как, например, в примере **2.6**, где расчеты могут отнять много времени при использовании биномиального распределения. Если же условия позволяют, то расчеты можно произвести, воспользовавшись распределением Пуассона. Тогда гораздо легче мы получим почти тот же результат. Аппроксимация биномиального распределения Пуассона дает хорошие результаты, если имеются:

1. Большое количество опытов; т.е. n — большое, предпочтительно $n \geq 30$;
2. Малая вероятность "успеха" в каждом опыте; т.е. p — маленькое, предпочтительно $p \leq 0,10$;
3. Предполагаемое количество "успехов" меньше пяти; т.е. $np \leq 5$.

В этих обстоятельствах распределение Пуассона со средним $m=pn$ вполне может быть использовано вместо биномиального распределения. Чем больше n и меньше p, тем точнее результат. В § 2.8 мы рассмотрим пример, как поступать с биномиальными расчетами при большом p. В этом случае для замены может быть использовано нормальное распределение.

☐ **Пример 2.10.** Производители карманных калькуляторов знают из опыта работы, что 1% произведенных и проданных калькуляторов имеют дефекты и их должны заменить по гарантии. Большая аудиторская фирма купила 500 калькуляторов. Какова вероятность, что пять или больше калькуляторов нужно будет заменить?

Решение.

Ситуация прекрасно укладывается в рамки биномиального распределения:

$$n = 500,\ p = 0,01,\ q = 0,99,\ np = 5.$$

Так как n — большое, p — маленькое, а np — меньше или равно пяти, то в качестве замены биномиального распределения может быть использовано распределение Пуассона.

а) Расчеты с использованием биномиального распределения, вероятности r дефектов в выборке дали следующие результаты:

$$P(r) = {}^{500}C_r \times 0,01^r \times 0,99^{500-r}; \quad r = 0, 1, 2, \ldots, 500;$$

$$P(r \geq 5) = 1 - P(r \leq 5) = 1 - \{P(r = 0) + P(r = 1) + P(r = 2) + P(r = 3) + P(r = 4)\};$$

$$P(0) = 0{,}01^0 \times 0{,}99^{500} \times \frac{500!}{500!\,0!} = 0{,}99^{500} \qquad = 0{,}00657;$$

$$P(1) = 0{,}01^1 \times 0{,}99^{499} \times \frac{500!}{499!\,1!} = 0{,}01 \times 0{,}99^{499} \times 500 \qquad = 0{,}03318;$$

$$P(2) = 0{,}01^2 \times 0{,}99^{498} \times \frac{500!}{498!\,2!} = 0{,}01^2 \times 0{,}99^{498} \times \frac{500 \times 499}{2 \times 1} \qquad = 0{,}08363;$$

$$P(3) = 0{,}01^3 \times 0{,}99^{497} \times \frac{500!}{497!\,3!} = 0{,}01^3 \times 0{,}99^{497} \times \frac{500 \times 499 \times 498}{3 \times 2 \times 1} \qquad = 0{,}14023;$$

$$P(4) = 0{,}01^4 \times 0{,}99^{496} \times \frac{500!}{496!\,4!} = 0{,}01^4 \times 0{,}99^{496} \times \frac{500 \times 499 \times 498 \times 497}{4 \times 3 \times 2 \times 1} \qquad = 0{,}17600;$$

Всего 0,43961.

Отсюда

$$P(r \geq 5) = 1 - \{0{,}43961\} = 0,\ 56039$$

(до пяти знаков после запятой),

Вероятность, что придется заменять 5 или больше калькуляторов, равна 0,560 (до трех знаков после запятой).

б) Расчеты с использованием распределения Пуассона:

$$m = np = 500 \times 0{,}01 = 5.$$

Вероятность r бракованных изделий в выборке приблизительно равна:

$$P(r) = \frac{5^r e^{-5}}{r!}, \qquad r = 0, 1, 2, 3, \ldots .$$

Отсюда

$$P(r \geq 5) = 1 - \{P(r = 0) + P(r = 1) + P(r = 2) + P(r = 3) + P(r = 4)\};$$

$$P(0) = e^{-5} \qquad = 0{,}00674;$$

$$P(1) = e^{-5} \times 5^1 = P(0) \times 5 \qquad = 0{,}03369;$$

$$P(2) = e^{-5} \times \frac{5^2}{2!} = P(1) \times \frac{5}{2} \qquad = 0{,}08422;$$

$$P(3) = e^{-5} \times \frac{5^3}{3!} = P(2) \times \frac{5}{3} \qquad = 0{,}14037;$$

$$P(4) = e^{-5} \times \frac{5^4}{4!} = P(3) \times \frac{5}{4} = 0{,}17547;$$

Всего: 0,44049.

Отсюда

$$P(r \geq 5) = 1 - 0{,}44049 = 0{,}55951$$

(до пяти знаков после запятой).

Вероятность, что 5 или больше калькуляторов придется заменить, равна 0,560 (до трех знаков после запятой). С такой точностью результат совпадает с тем, что мы получили в пункте (а), используя биномиальное распределение.

2.6. РАСПРЕДЕЛЕНИЕ НЕПРЕРЫВНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

2.6.1. Непрерывная случайная величина и плотность ее вероятности

В предыдущих параграфах этой главы мы рассмотрели вероятностное распределение дискретной случайной величины. Остальные разделы посвящены равномерному и нормальному распределению непрерывных случайных величин. Если в ходе эксперимента все значения случайной величины оказываются на определенном замкнутом участке и могут принимать в нем любые значения, то значит мы имеем дело с непрерывной случайной величиной. Непрерывная случайная величина имеет специфику в распределении вероятностей.

Например, если измерить объемы производимых заводом пластмассовых бутылок для сока, которые должны быть равны 200 мл, то полученные цифры окажутся на каком-то определенном интервале, допустим, от 190 до 210 мл. В данном случае непрерывная случайная величина будет иметь неограниченное множество значений в этих пределах. Получим функцию распределения вероятностей для непрерывной случайной величины. Предположим, мы имеем непрерывную случайную величину X, которая принимает значение x в интервале x_1 и x_2, для которого функция вероятности является непрерывной, то функция распределения непрерывной случайной величины равна f(x), $x_1 < x < x_2$.

График функции распределения непрерывной случайной величины представляет собой кривую, в отличие от линейной диаграммы для распределения дискретной случайной величины.

Пример 2.11.

а) Функция распределения:

$$f(x) = 0{,}125\,(5-x) \text{ для } 1<x<5.$$

График распределения вероятностей для заданной функции будет иметь вид (рис. 2.5.):

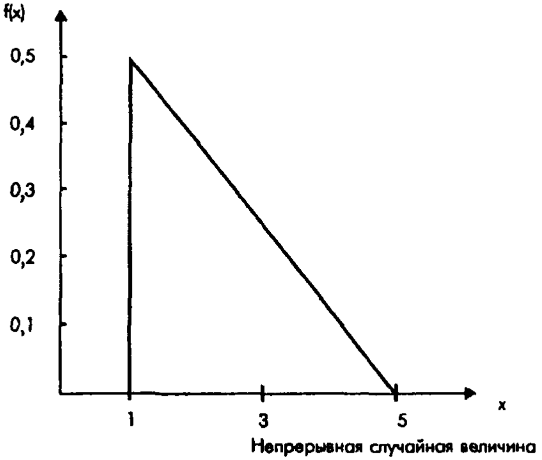


Рис. 2.5. Функция плотности вероятности f (x) = 0,125 (5–x)

б) Предположим, функция плотности вероятности: f (x)=6x (1–x); 0 ≤ x ≤ 1

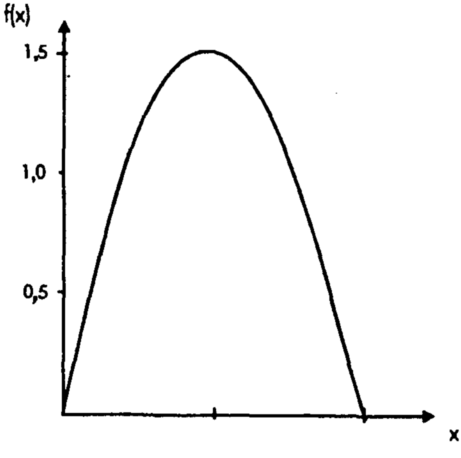


Рис. 2.6. Функция плотности вероятности f (x) = 6x (1-x)

Любому значению дискретной случайной величины соответствует определенная вероятность. Очевидно, что это невозможно для непрерывной переменной. Здесь вероятность соответствует некой области значений непрерывной случайной величины. Например, какова вероятность того, что объем пластмассовой бутылки находится в пределах от 195 до 197 мм?

Графически вероятность изображается как площадь под кривой, ограниченная некими пределами значений переменной. Общая площадь прямоугольников на гистограмме равна общей частоте. Общая площадь под кривой распределения соответствует общей вероятности 1.

В примере 2.11 а) вероятность, что X примет значение от 2 до 2,5 представлена закрашенной площадью графика функции распределения.

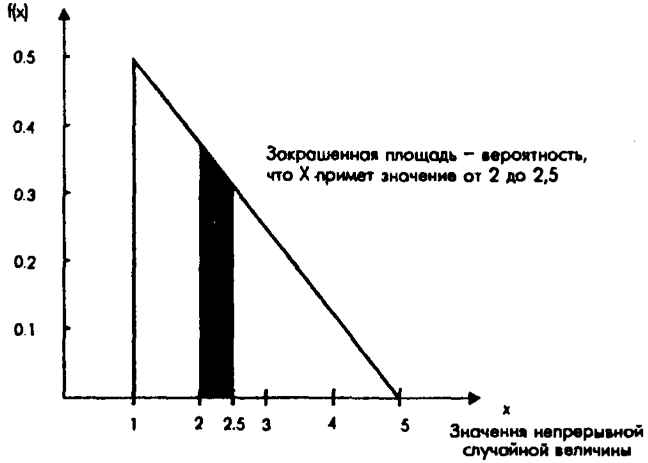


Рис. 2.7. Вероятность, что x примет значения от 2 до 2,5

Функция вероятности f (x) равна плотности вероятности:

$$\text{Плотность вероятности } P(x) = f(x)\ ;\ x_1 \le x \le x_2 .$$

Вероятность того, что непрерывная случайная величина примет значение из какого-то промежутка, равна площади под функцией плотности вероятности на этом промежутке (рис. 2.7). Однако непрерывная случайная величина никогда не принимает какое-то значение, допустим 5 мин., оно может быть меньше или больше пяти минут, или от 21,0 до 21,5 м, или более чем 3,6 кг.

2.6.2. Равномерное распределение

Равномерное распределение — простейший пример распределения непрерывной случайной величины. Мы используем его для иллюстрации некоторых положений, рассмотренных в предыдущем параграфе.

❐ **Пример 2.12**. Два бухгалтера ездят на работу, у первого дорога отнимает 20–25 мин., у второго — 20–30. Любое время на дорогу в этих пределах равновероятно. Ниже приведены графики равномерных распределений для обоих случаев (рис. 2.8 и 2.9).

Так как вероятность полной группы событий строго равна 1, то и площадь под кривой тоже должна быть равна 1. Это правило облегчает подсчет плотности вероятности для равномерного распределения.

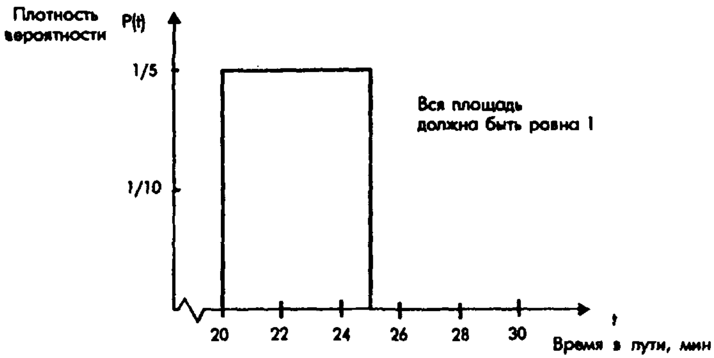


Рис. 2.8. Распределение времени в пути на работу первого бухгалтера

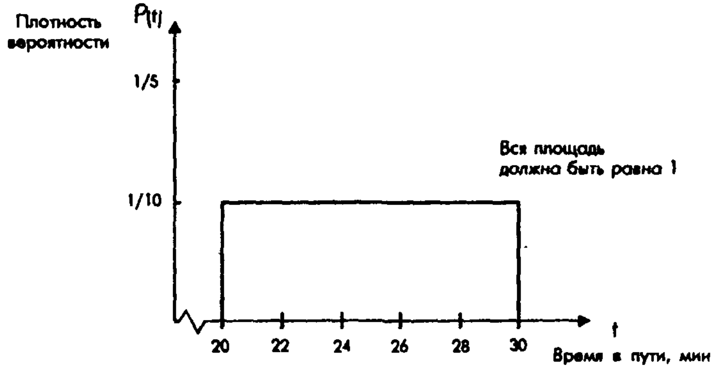


Рис. 2.9. Распределение времени в пути на работу второго бухгалтера

Для первого бухгалтера: интервал времени — 5 мин., значит ширина прямоугольника равна 5, следовательно, его высота (плотность вероятности) должна быть 1/5, тогда площадь под кривой (1/5) × 5 =1. Аналогично для второго бухгалтера: интервал времени — 10 мин., следовательно, плотность вероятности — 1/10, тогда 1/10 × 10 =1–полная вероятность. Теперь мы в состоянии подсчитать вероятность того, что дорога на работу занимает у каждого бухгалтера от 20,5 до 22,8 мин. Эта вероятность представляет собой площадь, ограниченную сверху кривой, а по бокам этими двумя пределами.

Для первого бухгалтера вероятность составит:

P(дорога занимает от 20,5 до 22,8 мин.) = (1/5) × (22,8 – 20,5) = (1/5) × 2,3 = 0,46.

Для второго бухгалтера вероятность составит:

P(дорога занимает от 20,5 до 22,8 мин.) = (1/10) × (22,8 – 20,5) = (1/10) × 2,3 = 0,23.

2.7. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ НЕПРЕРЫВНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

2.7.1. Природа нормального распределения

Нормальное распределение используется в ситуациях, связанных с измерениями веса или объема товаров, роста мужчин, проходящих медкомиссию, срока работы электроламп и т.д. Характерные свойства равномерного распределения, рассмотренные нами в предыдущем параграфе, относятся также и к нормальному распределению:

1. Площадь, образуемая кривой нормального распределения, представляет собой вероятность, что непрерывная случайная величина примет значения из заданного интервала.
2. Общая площадь под кривой нормального распределения равна полной вероятности, т.е. 1.
3. Невозможно точно определить вероятность того, что непрерывная случайная величина принимает какое-то конкретное значение.

Нормальное вероятностное распределение — это симметричное относительно среднего случайного значения величины распределение. Теоретически значения случайной величины находятся в интервале от минус до плюс бесконечности, т.е. непрерывная случайная величина может принимать любые значения, как положительные, так и отрицательные. Однако на практике нормальное распределение обычно используется для случайной величины, значения которой расположены в ограниченном интервале. Ниже представлен график типичного нормального распределения (рис. 2.10).

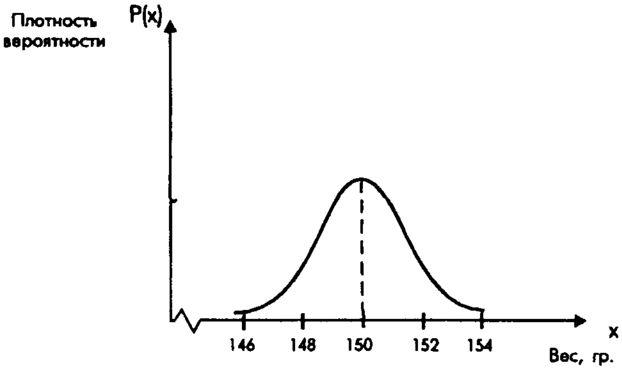


Рис. 2.10. Распределение пачек чая по весу

Функция плотности вероятности для нормального распределения представляет собой сложную математическую функцию, которая зависит от среднего значения случайной величины m и ее дисперсии σ^2. К счастью, все нормальные распределения можно свести к единому стандартному вероятностному распределению. Для этого распределения были составлены таблицы, в которых указываются площади под кривой для различных значений непрерывной случайной величины и по которым мы можем определить ту вероятность, которая нам требуется.

В стандартном нормальном распределении среднее значение случайной величины равно 0, дисперсия равна 1 и, следовательно, стандартное отклонение равно 1. Замена нормального распределения стандартным распределением означает то, что значения случайной величины (минуты, граммы, сантиметры и т.д.) выражены стандартным отклонением среднего значения случайной величины. После замены реальных цифр единицами стандартного отклонения (z) требуемые вероятности могут быть найдены по таблице.

Существует несколько вариантов таблиц нормального распределения. Все они дают один и тот же результат, только несколько различными путями. В Приложении 2 представлены таблицы вероятностей случайной величины, значения которой на z стандартных отклонений выше среднего.

2.7.2. Стандартное нормальное распределение

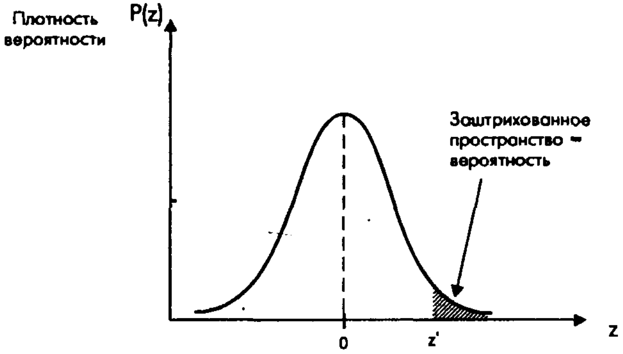


Рис. 2.11. Стандартное нормальное распределение

В таблицах нормального распределения приведены значения вероятностей для z от 0 до 3,5. Так как распределение симметрично, то эти же значения могут быть использованы для отрицательных z (см. рис. 2.12).

Значение непрерывной случайной величины можно выразить числом стандартных отклонений от среднего значения следующим образом:

$$z = \frac{x - \mu}{\sigma},$$

где x — значение случайной величины;

μ — среднее значение;

σ — стандартное отклонение;

z — на сколько стандартных отклонений отличается интересующее нас значение случайной величины от среднего.

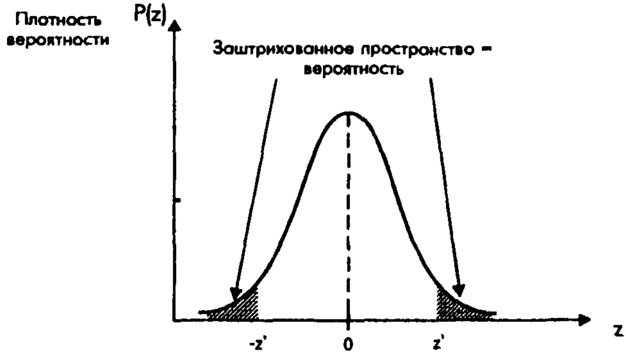


Рис. 2.12. Стандартное нормальное распределение

❒ **Пример 2.13.** Производителю электроламп известно, что средний срок работы лампы составляет 600 ч, а стандартное отклонение срока работы – 40 ч. Какова вероятность, что срок работы:

1) менее 700 ч;
2) менее 550 ч;
3) от 550 до 700 ч;
4) 2% ламп имеют минимальнай срок работы. Какова его величина?

Решение.

На рис. 2.13 приведена функция плотности распределения электроламп по сроку работы.

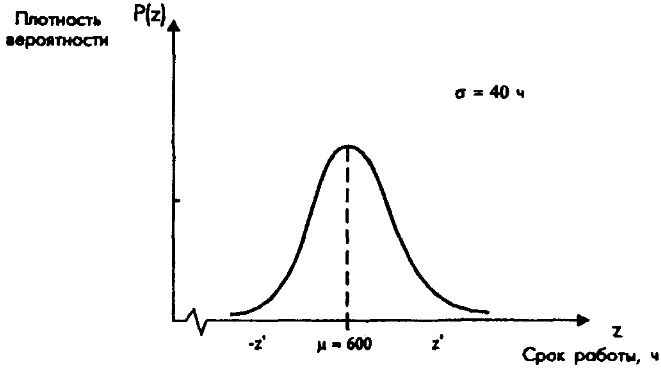


Рис. 2.13. Распределение электроламп по сроку работы

1. На рис. 2.14 вероятность того, что электролампа проработает менее 700 ч, представлена заштрихованным пространством.

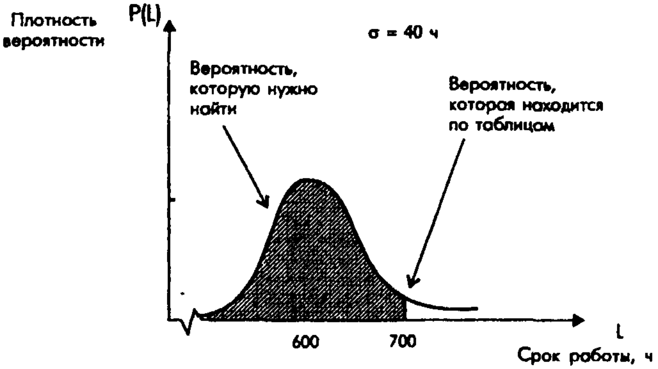


Рис. 2.14. Вероятность того, что электролампа проработает менее 700 ч

Теперь перейдем к стандартному нормальному распределению со средним, равным 0, и стандартным отклонением, равным 1. Подсчитаем, сколько стандартных отклонений находится между средним μ и 700 ч, т. е. найдем значение z:

$$z = \frac{L - \mu}{\sigma};$$

$$z = \frac{700 - 600}{40} = 2{,}5$$

(700 ч отличается от среднего на 2,5 стандартных отклонений).

По таблице нормального распределения находим:

$$P\ (z > 2{,}5) = 0{,}0062.$$

Так как общая вероятность равна 1, то:

$$P\ (z < 2{,}5) = 1 - 0{,}0062 = 0{,}9938\ ,$$

т.е. вероятность того, что лампа проработает меньше 700 ч, равна 99,38%. Иными словами, 99,38% ламп проработают 700 ч и меньше.

2. На рис. 2.15 вероятность того, что лампа проработает меньше 550 ч, представлена заштрихованным пространством.

Теперь сведем нормальное распределение к единому виду со средним, равным 0, и стандартным отклонением, равным 1. Подсчитаем, сколько стандартных отклонений находится между средним μ и 550 ч:

$$z = \frac{L - \mu}{\sigma};$$

$$z = \frac{550 - 600}{40} = -1{,}25$$

(550 ч на 1,25 стандартных отклонений меньше среднего).

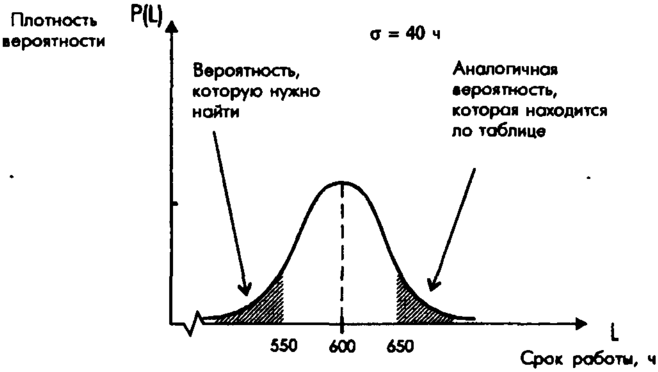


Рис. 2.15. Вероятность того, что срок работы лампы будет меньше 550 ч

Так как нормальное распределение симметрично, то:

$$P(z \le -1{,}25) = P(z \ge 1{,}25)\ .$$

По таблице нормального распределения находим:

$$P(z \ge -1{,}25) = 0{,}1056\ ,$$

или

$$P(z \le -1{,}25) = 0{,}1056\ ,$$

т.е. вероятность того, что срок работы лампы будет меньше 550 ч, равна 0,1056. Иными словами, 10,56% ламп проработают меньше 550 ч.

3. На рис. 2.16 площадь заштрихованного пространства равна вероятности того, что лампа проработает от 550 до 700 ч.

$$P(550 < L < 700\text{ ч}) = P(L < 700\text{ ч}) - P(L < 550\text{ ч}).$$

Требуемые вероятности мы рассчитали в пунктах **1** и **2**:

$$P(550 < L < 700\text{ ч}) = 0{,}9938 - 0{,}1056 = 0{,}8882,$$

т.е. вероятность того, что электролампа проработает от 550 до 700 ч, равна 0,8882. Иными словами, 88,82% ламп будут работать 550–700 ч.

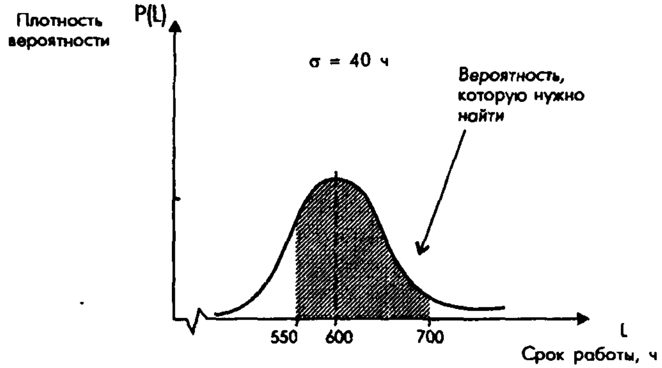


Рис. 2.16. Вероятность того, что лампа проработает от 550 до 700 ч

4. Этот вопрос несколько отличается от остальных. Отталкиваясь от процента лампочек, найдем соответствующий срок работы L (см. рис. 2.17).

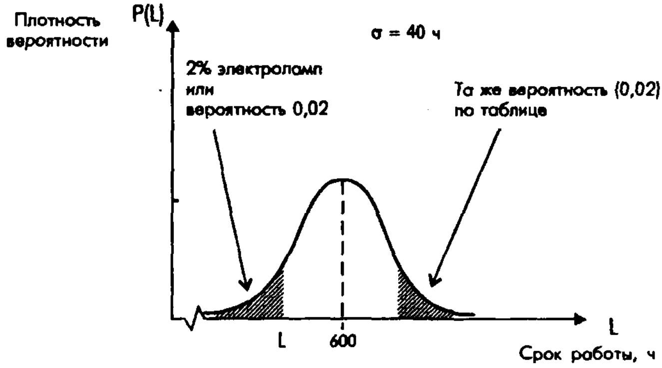


Рис. 2.17. Продолжительность работы 2% ламп

Во-первых, находим в таблице значение z, соответствующее вероятности 0,02; z равно 2,055. Это означает, что случайная величина, имеющая вероятность 0,02, на 2,055 стандартных отклонений больше среднего. Ввиду симметричности нормального распределения ту же самую вероятность имеет случайная величина, на 2,055 стандартных отклонений меньше среднего.

Тогда искомый срок работы равен: L = 600 − 82,2 = 517,8 ч. Т.е. для 2% ламп с минимальной продолжительностью работы срок работы составляет 517,8 ч.

2.8. ИСПОЛЬЗОВАНИЕ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ В КАЧЕСТВЕ АППРОКСИМАЦИИ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Возникает небольшая проблема в связи с тем, что нормальное распределение оперирует с непрерывными случайными величинами, в то время как биномиальное и пуассоновское — с дискретными. Но ее можно легко решить при помощи поправочного коэффициента, называемого "поправка на непрерывность".

Например, при помощи биномиального распределения можно вычислить вероятность существования двух бракованных образцов в выборке, состоящей из n штук. Используя нормальное распределение как замену биномиального, мы делаем допущение, что значение дискретной случайной величины 2 является значением непрерывной случайной величины на промежутке от 1,5 до 2,5. Это и называется поправкой на непрерывность. Нормальное распределение, которое мы использовали, имеет ту же среднюю, и стандартное отклонение, что и обычное биномиальное распределение. Площадь, покрываемая кривой нормального распределения на промежутке от 1,5 до 2,5, представляет собой приблизительное значение дискретной вероятности появления двух бракованных образцов.

Замена распределений производится только, если обычное биномиальное распределение очень трудоемко и к тому же сеществуют определенные предпосылки. В **2.5** мы использовали распределение Пуассона как замену биномиального. Это было возможно, если n — велико, p — мало и $np \le 5$. Биномиальное распределение можно заменить нормальным, если np, как и nq, больше 5, т.е. n должно быть большим, p больше 0,1, лучше всего около 0,5. Безусловно, указанные значения носят приблизительный характер. Тем не менее, чем больше n, np и nq, тем точнее замена.

Среднее и стандартное отклонение биномиального распределения имеют вид:

$$E(r) = np\ [= \mu]\ \text{ и }\ \sigma = \sqrt{npq}\ .$$

Эти величины используются для вычисления z при применении нормального распределения (как показано в **2.7**).

☐ **Пример 2.14.** Каждый день завод производит огромное количество чипсов, 40% из которых бракованные. Для проверки качества отбираются 20 образцов из произведенных за день чипсов. Какова вероятность, что 14 или больше из 20 бракованные?

Решение.

Произведем расчеты, используя обычное нормальное распределение:

$$P\ (r \text{ дефектов в 20 образцах}) = (0{,}4)^r \times (0{,}6)^{20-r} \times {}^{20}C_r; \qquad r = 0,\ 1,\ \ldots,\ 20;$$

$$P\ (14 \text{ или больше дефектов}) = P(14) + P(15) + P(16) + P(17) + P(18) + P(19) + P(20);$$

$$P\ (14) = 0{,}4^{14} \times 0{,}6^{6} \times \frac{20!}{14!\ 6!} = 0{,}004854;$$

$$P\ (15) = 0{,}4^{15} \times 0{,}6^{5} \times \frac{20!}{15!\ 5!} = 0{,}001294;$$

$$P(16) = 0{,}4^{16} \times 0{,}6^{4} \times \frac{20!}{16!\,4!} = 0{,}000270;$$

$$P(17) = 0{,}4^{17} \times 0{,}6^{3} \times \frac{20!}{17!\,3!} = 0{,}000042;$$

$$P(18) = 0{,}4^{18} \times 0{,}6^{2} \times \frac{20!}{18!\,2!} = 0{,}000005;$$

$$P(19) = 0{,}4^{19} \times 0{,}6^{1} \times \frac{20!}{19!\,1!} = 0{,}000000;$$

$$P(20) = 0{,}4^{20} \times 0{,}6^{0} \times \frac{20!}{20!\,0!} = 0{,}000000.$$

0,006465

P (14 или больше чипсов в выборке из 20 — бракованные) = 0,006465.

Расчеты с заменой биномиального распределения нормальным чрезвычайно просты. Сначала проверим, можно ли произвести замену: $np = 20 \times 0{,}4 = 8$; $nq = 20 \times 0{,}6 = 12$. Полученные результаты показывают, что применение нормального распределения в качестве приближения биномиального распределения возможно.

Теперь произведем проверку на непрерывность. Дискретное значение, равное 14, заменяем непрерывной случайной величиной на промежутке от 13,5 до 14,5. Вместо того, чтобы находить вероятность дискретной величины 14 и более дефектов, мы найдем вероятность значения непрерывной случайной величины более чем 13,5 дефектов. Среднее нормального распределения: $np = 8$, отсюда стандартное отклонение равно:

$$\sqrt{npq} = \sqrt{20 \times 0{,}4 \times 0{,}6} = 2{,}19.$$

Рассчитаем значение z для 13,5:

$$z = \frac{13{,}5 - 8}{2{,}19} = 2{,}51$$

(значение случайной величины на 2,51 стандартных отклонения больше среднего).

По таблице стандартного нормального распределения находим:

$$P(z \geq 2{,}51) = 0{,}0060.$$

Следовательно,

$$P(\text{число дефектов} \geq 13{,}5) = 0{,}0060.$$

Очевидно, что полученный с малыми затратами труда результат P(число дефектов ≥ 13,5) = 0,0060, почти ничем не отличается от результата расчетов с биномиальным распределением P(14 и более дефектов в выборке из 20 образцов) = 0,0065.

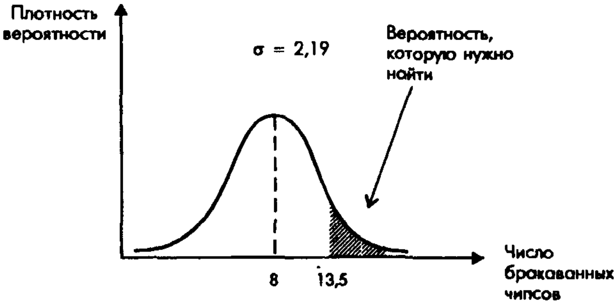


Рис. 2.18. Вероятность того, что бракованных чипсов более чем 13,5

☐ **Пример 2.15.** Рассмотрим пример замены биномиального распределения нормальным для определения вероятности пропорции (доли). Из прошлого опыта аудиторы "ABC и Co Ltd" знают, что в среднем из 1000 бухгалтерских проводок 35 бывают с ошибками. Какова вероятность, что при ближайшей проверке ошибок будет больше 5%?

Решение.

$$\text{Средняя доля} = p = \frac{35}{1000} = 0{,}035;$$

$$\text{Стандартное отклонение} = \sqrt{\frac{p\,q}{n}} = \sqrt{\frac{0{,}035 \times 0{,}965}{1000}} = 0{,}0058.$$

В качестве иллюстрации результата см. рис. 2.19.

В данном случае доля ошибок в общем числе проводок является непрерывной случайной величиной и поэтому можно использовать нормальное распределение без поправки на непрерывность. Последствия применения поправки на непрерывность в такого рода задаче мы рассмотрим в примере **6.13** гл. 6.

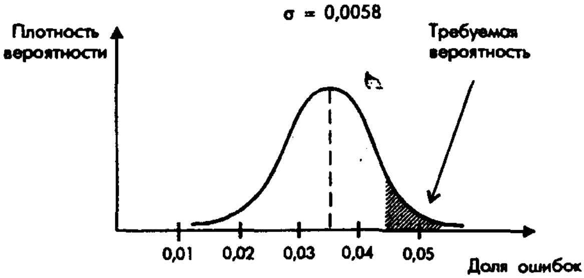


Рис. 2.19. Вероятность доли ошибок

Вычислим значение z для p = 0,05:

$$z = \frac{0,05 - 0,035}{0,0058} = 2,586$$

(на 2,586 стандартных отклонения выше средней доли).

По таблицам стандартного нормального распределения вероятность равна:

$$P(z > 2,586) = 0,0048.$$

Следовательно,

$$P(\text{доля ошибок} > 0,05) = 0,0048.$$

2.9. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ КАК ЗАМЕНА РАСПРЕДЕЛЕНИЯ ПУАССОНА

Нормальное распределение может использоваться и как замена распределения Пуассона. Для замещения дискретной случайной величины непрерывной применяется та же поправка на непрерывность с учетом среднего m и стандартного отклонения $\sqrt{m}$. Чем больше m, тем точнее результат, обычно $m \geq 10$, но в приведенном ниже примере при меньшем значении среднего получен вполне приемлемый итог.

❐ **Пример 2.16.** В среднем при установке компьютеров происходят две неполадки в неделю. Найдите вероятность, что в течение 4-х недель будет не больше 10 неполадок.

Решение.

Среднее m равно восьми неполадкам в месяц. Вероятность неполадок в месяц вычисляется по распределению Пуассона:

$$P(r\ \text{неполадок в месяц}) = \frac{8^r e^{-8}}{r!};\quad r=0,1,2,3,\ldots$$

Требуется найти:

$$P(r \geq 10) = 1 - \{P(0) + P(1) + P(2) + \ldots + P(10)\}.$$

По распределению Пуассона:

$$P(r \geq 10) = 1 - 0,816 = 0,184.$$

Вероятность того, что произойдет более чем 10 неполадок в течение месяца равна 0,184. Несмотря на то, что среднее меньше 10, мы будем использовать нормальное распределение.

В соответствии с поправкой на непрерывность находим вероятность более, чем 10,5 неполадок в месяц. Для использования нормального распределения при среднем, равном 8, стандартном отклонении $\sqrt{8}$, для более чем 10,5 неполадок в месяц получаем значение z:

$$z = \frac{10{,}5 - 8}{\sqrt{8}} = 0{,}884$$

(значение случайной величины на 0,884 стандартных отклонения больше среднего).

По таблице стандартного нормального распределения получаем вероятность:

$$P\ (Z \geq\ 0{,}884) = 0{,}1894.$$

Таким образом, вероятность более, чем 10,5 неполадок в месяц с использованием нормального распределения, равна 0,1894.

По распределению Пуассона эта же вероятность равна 0,184. Как видим, использование нормального распределения дало близкий результат.

2.10. КОМБИНАЦИИ СЛУЧАЙНЫХ ВЕЛИЧИН

2.10.1. Независимые случайные величины

Иногда возникает необходимость объединить случайные величины в новую случайную величину. Например, сборка продукции состоит из двух этапов. Время, которое уходит на каждый из них, независимо друг от друга, поэтому общее время равно:

$$t_E = t_1 + t_2,$$

где t_1 и t_2 — время сборки на первом и втором этапах соответственно.

На другой фабрике специальные коробочки заполняются апельсиновым соком. Затем, взятые наугад четыре коробочки, объединяются в упаковку. Если обозначить вес каждой коробочки — w, а вес упаковки — T, то:

$$T = w_1 + w_2 + w_3 + w_4.$$

Пока нас не интересует тип распределения исходных случайных величин. При условии, что случайные величины не зависимы друг от друга, мы можем вывести формулы средней и дисперсии объединенной случайной величины. Итак, если значения двух не зависимых друг от друга случайных величин обозначить x и y, то значение объединенной случайной величины z будет:

$$z = x + y.$$

Вне зависимости от типа распределения x и y среднее z есть сумма средних x и y:

$$\mu_z = \mu_x + \mu_y\ .$$

То же соотношение для дисперсий:

$$Дис\ (z) = Дис\ (x) + Дис\ (y).$$

Следует отметить, что стандартные отклонения не суммируются, суммируются только дисперсии.

❑ **Пример 2.17.** Ежедневный спрос на товар — 100 единиц, стандартное отклонение — 12 единиц в день. Используя данные средних продаж за день, вычислите среднее количество товара, продаваемое за неделю.

Решение

Так как проданное за день количество товара не зависит от предыдущих дней, то общее количество товара, проданное за семь дней рассчитывается как:

$$\text{Средний недельный спрос} = 7 \text{ дней} \times 100 = 700 \text{ ед. в неделю.}$$

Для того, чтобы рассчитать стандартное отклонение для недельных данных, сначала вычислим дисперсию. Для дневного спроса дисперсия равна $12^2 = 144$. Для недельного спроса: $7 \times 12^2 = 7 \times 144$; следовательно, стандартное отклонение равно: $\sqrt{7 \times 144} = 31{,}75$ ед./в неделю.

Может потребоваться формирование новой случайной величины z как разности исходных случайных величин x и y. Порядок действий остается тем же самым, что в случае суммирования. Если мы имеем две независимые случайные величины x и y, тогда:

$$z = x - y.$$

Вне зависимости от типа распределения x и y среднее z есть разность средних x и y:

$$\mu_z = \mu_x - \mu_y.$$

Однако дисперсия z равна сумме дисперсий x и y:

$$\text{Дис}(z) = \text{Дис}(x) + \text{Дис}(y).$$

❑ **Пример 2.18.** На лесопилке из готовых досок (L1), средняя длина которых 200 см, а стандартное отклонение — 1,7 см, делают доски (L2) со средней длиной 70 см и стандартным отклонением 0,3 см. Каково среднее и стандартное отклонение обрезков (L3)?

Решение

Исходная длина и длина обрезка не зависят друг от друга, т.е.:

$$\text{Средняя длина} = \text{Среднее}(L1) - \text{Среднее}(L2) = 200 - 70 = 130 \text{ см.}$$

$$\text{Дисперсия} = \text{Дис}(L1) + \text{Дис}(L2) = 1{,}7^2 + 0{,}3^2 = 2{,}98.$$

Следовательно, стандартное отклонение (L3) = $\sqrt{2{,}98} = 1{,}73$ см.

Независимо от того, складываем мы случайные величины или вычитаем, их дисперсии складываются.

Рассмотренные выше формулы могут быть применены к любому количеству независимых случайных величин. Мы не будем рассматривать ситуацию, когда случайные величины зависимы, кроме одного случая, приведенного ниже.

2.10.2. Особый случай зависимых случайных величин

Если новая случайная величина (z) получается при умножении x на какое-то число a, т.е. $z = ax$, то мы имеем дело уже не с комбинацией независимых случайных величин. Среднее величины z равно:

$$\mu_z = a\,\mu_x,$$

а дисперсия:

$$\text{Дис}\,(z) = a^2\ \text{Дис}\,(x).$$

❑ **Пример 2.19.** Товар, упомянутый в примере **2.17**, состоит из пяти идентичных компонентов. Каково среднее и стандартное отклонение дневного спроса на компоненты?

Решение.

На первый взгляд может показаться, что эта ситуация подобна ситуации с недельным спросом, но на самом же деле это не так. Мы имеем не пять независимых случайных величин, а одну величину за день, которую умножаем на 5, чтобы получить число необходимых компонентов.

Дневной спрос на компоненты:

$$\text{Среднее} = 5 \times 100 = 500\ \text{компонентов в день.}$$

$$\text{Дисперсия} = 5^2 \times 12^2.$$

Следовательно, стандартное отклонение = 5 × 12 = 60 компонентов в день.

2.10.3. Природа распределения объединенных случайных величин

Если объединяемые независимые случайные величины являются нормально распределенными, тогда и объединенная случайная величина будет подчиняться нормальному распределению. Например, если $z = x + y$, x и y — независимые нормально распределенные переменные, то z тоже будет распределено нормально. То же для $z = x - y$. Аналогично, если $z = ax$, где a — константа, x распределено нормально, то и z имеет нормальное распределение. Но это правило верно не для всех распределений. Например, если x и y — независимы и имеют биномиальное распределение, то z не будет иметь биномиального распределения. Это правило верно для любого числа независимых случайных величин. Вернемся к задаче с коробочками, которые наполняются апельсиновым соком и затем объединяются в упаковки по 4 штуки. Если вес полной коробочки распределяется нормально со средним весом $\overline{w}$ (гр) и стандартным отклонением σ_w, то вес упаковки подчинен нормальному распределению со средним весом, равным:

$$\overline{T} = \overline{w} + \overline{w} + \overline{w} + \overline{w} = 4 \times \overline{w},$$

а дисперсия:

$$\sigma_T^2 = \sigma_w^2 + \sigma_w^2 + \sigma_w^2 + \sigma_w^2 = 4 \times \sigma_w^2 .$$

РЕЗЮМЕ

Случайная величина представляет собой численный исход эксперимента. Случайная величина может быть или дискретной (только целые числа) или непрерывной (любые). Для дискретной случайной величины вероятность значения r находится по формуле дискретной вероятности:

$$P(r) = f(r) \text{ на определенном промежутке } r.$$

Математическое ожидание случайной величины находится по формуле:

$$E(r) = \sum r\,P(r).$$

Стандартное отклонение:

$$\sigma = \sqrt{\sum r^2\,P(r) - (E(r))^2}\,.$$

При биномиальном распределении мы рассматриваем n независимых, идентичных опытов. Каждый опыт имеет два возможных исхода: "успех" и "неудачу". Вероятность "успеха" одинакова для всех опытов. Формула вероятности биномиального распределения:

$$P(r) = \frac{n!}{r!\,(n-r)!}\,p^r\,q^{n-r};\quad r = 0, 1, 2, 3, \ldots, n,$$

$$E(r) = n\,p \quad \text{и} \quad \sigma = \sqrt{n\,p\,q}\,.$$

Распределение Пуассона используется для большого числа идентичных независимых опытов, каждый из которых имеет малую постоянную вероятность "успеха". Формула вероятности биномиального распределения:

$$P(r) = \frac{m^r\,e^m}{r!};\quad r = 0, 1, 2, 3, \ldots,$$

$$E(r) = m \quad \text{и} \quad \sigma = \sqrt{m}.$$

Распределение Пуассона используется для упрощения расчета биномиальных вероятностей при $n = 30$, $p \le 0{,}1$ и $np \le 5$.

Вероятность непрерывной случайной величины представляет собой площадь под кривой плотности вероятности. Вероятность существует только для интервала значений случайной величины, но не для отдельно взятого значения.

Все нормальные распределения зависят от средней и стандартного отклонения и могут быть сведены к единому распределению. Это стандартное нормальное распределение. Значения случайной величины выражаются в количестве стандартных отклонений от средней z. Существуют таблицы вероятностей для каждого значения z.

Нормальное распределение может использоваться в качестве замены биномиального, если np и nq больше 5. Нормальное распределение также может заменять пуассоновское, если $m \geq 10$.

Если складывать или вычитать независимые нормальные случайные величины получается также нормальная случайная величина. Среднее значение объединенной случайной величины есть сумма или разность индивидуалных средних. Дисперсия объединеной случайной величины есть сумма индивидуальных дисперсий.

Если x — нормальная случайная величина и $z = ax$, где a — константа, тогда средняя z является нормально распределенной величиной со средней $\mu = a\,x$ и дисперсией $\sigma^2_{(z)} = a^2\,\sigma^2_{(x)}$.

УПРАЖНЕНИЯ

Упражнение 2.1

Какая из следующих ситуаций описывается биномиальным распределением? Обоснуйте ответы (без расчетов).

1. "Sparks Garage Ltd" продает 10 машин одной марки в месяц. Какова вероятность, что владельцы двух купленных машин обратятся за гарантийным обслуживанием, если известно, что 5% покупателей машин данной марки им пользуются?
2. Оцените вероятность того, что у недавно купленной машины расход бензина будет превышать 30 л на 100 км, если известно по статистическим данным, что среднее потребление бензина у машин этого типа составляет 33 л на 100 км.

Упражнение 2.2

Практика показывает, что 7% накладных, проходящих проверку в бухгалтерии, оказываются неправильно оформленными. Наугад отобраны 20 накладных. Какова вероятность, что:

1. Три из них оформлены правильно.
2. Как минимум три оформлены неправильно?

Упражнение 2.3

Старинная семейная фирма решила начать продажу своих акций на бирже. Известно, что 80% брокеров посоветовали своим клиентам купить эти акции. Предположим, что совет был правильным.

Наугад отобраны 6 брокеров. Найти вероятность, что по крайней мере четверо из них предложили своим клиентам купить акции фирмы.

Упражнение 2.4

Торговец фруктами и овощами закупает бананы у заготовителей большими партиями. Но учитывая, что это товар скоропортящийся, он предполагает, что до 10% бананов будут подпорчены. Не имея возможности проверить всю закупаемую партию, он разработал следующую процедуру выборочной проверки качества.

Из поступившей партии наугад отбираются 30 гроздьев бананов, если подпорченные бананы имееются не более, чем в двух гроздях, то он покупает всю партию товара. Если подпорченные бананы имеются более чем в двух гроздях, сделка не состоится.

Какова вероятность того, что сделка не состоится, если в партии имеется 5% недоброкачественных бананов?

Упражнение 2.5

Производитель транзисторов хочет добиться того, чтобы брак составлял не более 1%. Для проверки качества с поточной линии берется 10 образцов.

Назовите тип распределения вероятностей, который описывает число бракованных образцов в выборке.

Упражнение 2.6

Среднее число грузовиков, прибывающих на склад под разгрузку в течение часа равно 3.

1. Назовите тип распределения вероятностей, описывающий распределение числа грузовиков, прибывающих в течение часа.
2. Какова вероятность, что в течение часа под разгрузку придет более четырех машин?

Упражнение 2.7

Компания “Pearsons Engineering Ltd” имеет оборудование для производства болтов и других метизов. В среднем два станка в течение часа выходят из строя. Для устранения неполадок на заводе работает специальный инженер, однако ему приходится вызывать ассистента, если происходит более двух поломок в час. Как часто в среднем потребуется помощь ассистента в течение 120 ч рабочей недели?

Упражнение 2.8

В компьютере компании "Star Holdings plc" хранится вся использованная информация. Специалист, ответственный за компьютеры в этой фирме, предлагает сохранить все файлы пятилетней давности. Для изучения спроса на архивную информацию он запросил данные за последние 100 дней:

Число использовавшихся старых файлов	*Число дней*
0	10
1	30
2	27
3	17
4	6
5	5
6	4
7	1

1. Вычислите среднее и число старых файлов, запрашиваемых в день, и стандартное отклонение.
2. Объясните, почему число старых файлов, запрашиваемых в день, подчиняется распределению Пуассона?
3. Предположим, число использовавшихся старых файлов подчиняется распределению Пуассона со средним значением, вычисленным в п. 1. Определите ожидаемые частоты этого распределения.

Упражнение 2.9

В среднем за месяц бухгалтерия компании "Star Holdings plc" использует 4 коробки флоппи-дискет. Предполагая, что потребность во флоппи-дискетах подчиняется распределению Пуассона, найдите число коробок дискет, которые бухгалтерия должна заказывать к началу очередного месяца такое, чтобы вероятность перерасхода была меньше, чем 4%.

Упражнение 2.10

Для транспортировки апельсины упаковываются в специальные ящики по 250 шт. в каждом. При вскрытии обнаруживается, что в среднем 0,6% апельсинов испорчены.

Какова вероятность, что во взятом для проверки ящике окажется не более двух испорченных плодов?

Упражнение 2.11

Упаковочный аппарат расфасовывает стиральный порошок в пакеты, средний вес которых 930 гр., а стандартное отклонение — 20 гр.

Какая доля пакетов будет иметь вес до 900 гр.? Если требуется, чтобы не более чем 2,5% пакетов содержали меньше, чем 900 гр., то как должна быть переналажена машина, чтобы соответствовать этому требованию?

Упражнение 2.12

Срок работы электрических компонент подчиняется нормальному распределению со средней продолжительностью работы 80 ч и стандартным отклонением — 30 ч. Допустим, поизводитель решил заменить все компоненты, которые вышли из строя до гарантийного срока работы, составляющего 45 ч. Какую долю общего выпуска составит эта часть продукции?

Допустим, производитель решил заменить только 10% общего выпуска, т.е. компоненты с самым коротким сроком работы. Какой гарантийный срок работы он должен назначить, чтобы выполнить это условие?

Упражнение 2.13

Фирма "HJ Woolley & Sons Ltd" производит вязальные спицы. Наиболее популярны размеры:

	Диаметр (мм)	*Точность (мм)*
Номер 10	3,25	± 0,125
Номер 11	3,00	± 0,125
Номер 12	2,75	± 0,125

"Wolley" производит нарезку игл из проволоки и их дальнейшую обработку. В результате чего средний диаметр заготовок становится 3,10 мм, а его стандартное отклонение — 0,10 мм. Допустим, значение диаметра подчиняется закону нормального распределения.

Определите число заготовок, пригодных для производства спиц № 11, учитывая, что дальнейшая обработка не изменяет диаметр заготовок.

Упражнение 2.14

Работа машины, расфасовывающей сахар, подчиняется правилам нормального распределения, со стандартным отклонением 20 гр. Машина может быть настроена на любой средний вес упаковки с точностью до грамма. В данном случае требуемый вес упаковки составляет 1000 гр. Чтобы вес товара соответствовал требованиям "Закона о мерах и весах" 1979 г., упаковка сахара должна удовлетворять трем следующим условиям:

1. Средний вес упаковки — не менее 1000 гр.
2. Вес не более, чем 2,5% упаковок, может быть меньше 975 гр.
3. Вес не более, чем одной из 10000 упаковок, может быть меньшее 950 гр.

На данный момент машина налажена на средний вес упаковки — 1010 гр.

Какова доля упаковок, содержащих менее 975 гр?

Какова доля упаковок, содержащих менее 950 гр?

Какой новый минимум среднего веса должен быть задан машине при том же стандартном отклонении, чтобы все упаковки продукции соответствовали Закону 1979 г.?

Упражнение 2.15

Компания "EG Mersoe & Co" — небольшая фирма, занимающаяся помолом муки. Фасовочный аппарат нужно заменить и м-ру Mersoe было предложено два варианта. По первому можно обеспечить работу фасовочного аппарата в соответствии с правилом нормального распределения, при стандартном отклонении 15 гр. Если заплатить на 50000 ф. ст. больше, то можно купить аппарат, который обеспечивает расфасовку со стандартным отклонением до 5 гр.

Расфасовывая муку в упаковки по 1 кг, аппарат должен соответствовать трем условиям:

1. Средний вес упаковки как минимум 1000 гр.
2. Вес не более 2,5% упаковок может быть меньше 985 гр.
3. Вес не более 0,01% упаковок может быть меньше 970 гр.

Компания продает 4 млн. килограммовых упаковок муки в год. Цена килограмма муки примерно 15 центов.

Сколько денег компания сэкономит в год, если купит более дорогой аппарат.

Упражнение 2.16

По данным опроса общественного мнения, из 10800 зарегистрированных избирателей 45% собираются проголосовать за либеральную партию. Если все, кто имеет право голосовать, придут на избирательные участки и данные опроса объективны, то какова вероятность, что либеральная партия получит менее 5000 голосов?

Упражнение 2.17

Для получения работы оператора компьютера кандидаты должны пройти письменное тестирование, состоящее из 100 вопросов с тремя вариантами ответа на каждый из них, причем лишь один из них правильный. Чтобы успешно пройти тестирование, нужно правильно ответить на 40 и более вопросов.

Какова вероятность того, что кандидат, выбирающий правильный ответ наугад, сдаст экзамен?

Упражнение 2.18

Компания "Achilles" продает специальные фломастеры для оформления плакатов в упаковке по 5 шт. Ввиду различных причин бракованные фломастеры можно обнаружить только после продажи. Менеджер, ответственный за качество, считает, что число бракованных фломастеров примерно одинаково. Тем не менее в последнее время стало поступать все больше нареканий от покупателей. Поэтому было решено отобрать наугад 500 упаковок для проверки качества:

Число бракованных фломастеров в упаковке	*Число упаковок*
0	392
1	73
2	30
3	3
4	2
5	0

а) По данным выборки найдите долю бракованных фломастеров.

б) Подразумевается, что число бракованных образцов подчиняется биномиальному распределению. Учитывая это, найдите количество упаковок (из этих пятисот), в которых 0, 1, 2 бракованных фломастеров соответственно.

Упражнение 2.19

В одной из трех предложенных ситуаций используйте биномиальное, пуассоновское или нормальное распределение. Объясните, почему вы выбрали тот или иной тип распределения, обоснуйте свою точку зрения.

Ситуация 1. Средний срок работы электросхемы — 800 ч, стандартное отклонение — 160 ч.

а) Минимальный гарантированный срок работы — 600 ч, сколько электросхем не проработает этого времени, т.е. сколько электросхем производителю придется заменить?

б) Если производитель согласен заменять только 1% электросхем с наиболее коротким сроком работы, какой срок гарантии ему следует установить?

в) Какова вероятность, что средний срок работы 25 выбранных наугад электросхем превысит 850 ч?

Ситуация 2. Зеленщик покупает персики большими партиями. Учитывая скоропортящийся характер товара, он допускает, что 15% фруктов будут подпорчены. Для проверки качества зеленщик выбирает 10 персиков, и если не более двух плодов оказались подпорчены, он покупает всю партию. Найдите вероятность того, что при нормальных условиях поставки партия товара будет куплена.

Ситуация 3. Участок дороги с односторонним движением два автомобиля проходят за 10-секундный интервал. Найдите вероятность того, что более трех автомобилей пройдут узкий участок дороги за 20-секундный интервал.

Упражнение 2.20

По данным Лондонского банка, недельная потребность в купюрах 1-, 5-, 10-фунтового достоинства подчиняется закону нормального распределения:

Номинал (ф. ст.)	*Число банкнот*	
	среднее	*стандартное отклонение*
1	1,200	250
5	600	100
10	50	5

Информация о других купюрах существенного значения не имеет. Данные по этим трем видам купюр независимы друг от друга.

1. Какова вероятность, что спрос на наличность в какой-либо из дней превысит 5000 ф. ст.?
2. На начало одного из дней в банке имелось 1600 ф. ст. в 1-фунтовых, 3500 ф. ст. в 5-фунтовых и 600 ф. ст. — 10-фунтовых банкнотах. Учитывая, что поступления наличности в течение дня не будет, ответьте на вопрос, какова вероятность, что банк сможет удовлетворить дневной спрос на наличность? (Замена купюры большего номинала на купюры меньшего не допускается; т.е. нельзя выдать две 5-фунтовых банкноты вместо 10-фунтовой.) По какому из трех видов банкнот положение банка более трудное?
3. Если допустить возможность замены купюр, объясните (не подсчитывая), каким образом изменится ситуация.

Упражнение 2.21

Компания "Classic Cars Ltd " сдает внаем автомобили. Из 15 имеющихся машин 10 берут внаем постоянные клиенты, а спрос на остальные подчиняется закону распределения Пуассона со средним 4.

1. Определите вероятность, что 10, 11, 12, 13, 14 или 15 машин будут сданы внаем в течение дня. Вычислите математическое ожидание числа машин, сданных внаем в течение дня.
2. Расходы на содержание машины вне зависимости от того, используется она или нет, составляют 6 ф. ст. Если же машина используется, то расходы увеличиваются на 2 ф. ст.. Подсчитайте математическое ожидание дневной прибыли компании, если плата за аренду составляет 20 ф. ст. в день.
3. Определите, целесообразно ли компании изменить количество машин с 15 до 14 или 16, учитывая, что распределение дневного спроса не изменяется. В каких пределах расходы на содержание 15 машин будут оптимальны?

Упражнение 2.22

"Electra Ltd" — маленькая электронная компания, специализирующаяся на сборке электроники из покупных компонентов. Из-за недавних изменений в номенклатуре выпускаемых товаров на складе скопилось 2000 соленоидов, основной характеристикой которых является электрическое сопротивление. Скопившиеся на складе соленоиды могут быть использованы только в приборе с сопротивлением 210–220 Ом. Так как первоначально соленоиды закупались для товаров с другими характеристиками, то имеются некоторые проблемы с их дальнейшим использованием. Известно, что сопротивление соленоидов, имеющихся на складе, колеблется от 205 до 225 Ом и распределение подчинено нормальному закону со стандартным отклонением, равным 5 Ом.

Существуют два пути использования запасов. Первый — продать все 2000 соленоидов на металлолом по 0,50 ф. ст. за штуку и закупить новые с требуемыми характеристиками. Второй — измерить сопротивление всех соленоидов и разделить их на две группы: те, у которых сопротивление не попадает в требуемые 210–220 Ом, продать на металлолом по 0,50 ф. ст. за штуку; остальные использовать для сборки приборов. Стоимость измерения и сортировки соленоидов составляет 1,40 ф. ст. за штуку, тогда как покупка нового соленоида потребует 4 ф. ст. за штуку.

1. Если среднее сопротивление соленоидов на складе составляет 210 Ом, сколько из них попадут в требуемый интервал 210–220 Ом? Что экономичнее — пустить все соленоиды на металлолом или измерить сопротивление каждого и продать в металлалом только ненужные?
2. Рассчитайте результаты обоих путей использования соленоидов, если среднее сопротивление равно: 205 Ом, 210, 215, 220, 225 Ом. Представьте результаты графически. Используя полученные графики, определите, каково должно быть среднее сопротивление, чтобы было экономически выгодно протестировать каждый соленоид?
3. Если вероятности средних сопротивлений соленоидов, приведенные в п. 2, равны, какой из вариантов их использования вы рекомендуете? Изменятся ли ваши рекомендации, если станет известно, что среднее сопротивление — 215 Ом в два раза более вероятно, чем остальных?

Глава 3. ПРАВИЛА И СХЕМЫ ПРИНЯТИЯ РЕШЕНИЙ

3.1. ВВЕДЕНИЕ

Принятие правильного решения вовремя — главная задача управленческого персонала любой компаниии. Неправильное или просто глупое решение может дорого стоить компании, иметь фатальные, непоправимые последствия. Поэтому важно, чтобы те, кто вовлечен в процесс принятия решений, использовали все имеющиеся у них средства и приняли "наилучшее" решение.

Прежде всего определимся с приоритетами: "наилучшее" — для кого или для чего? Перед тем, как принимать решение, следует тщательно продумать его цель. Трудность состоит в том, что задачи различных подразделений предприятия очень противоречивы. Например, такой простой вопрос, как размер запасов на складе. Интересы производства требуют иметь большой запас сырья и материалов, что приводит к возникновению незавершенного производства и несоразмерно меньшему выпуску готовой продукции. Однако процесс производства не прекращается. В то же самое время отдел маркетинга требует больших запасов готовой продукции на складе и гибких производственных линий, которые можно переналадить для небольших заказов. А бухгалтерия настаивает на увеличении оборачиваемости капитала и соответственно минимизации запасов, чтобы освободившиеся деньги можно было использовать для других целей.

Как определить оптимальные для компании запасы? Рассчитывать ли эти цифры для компании в целом или выбрать приоритетные направления в ущерб остальным и осуществлять контроль запасов так, чтобы обеспечить оптимальные затраты для выполнения отдельных функций без учета других? Этот процесс называется **субоптимизацией**. Принятие решений — достаточно сложный и интересный процесс, который носит исключительно субъективный характер. Как вы, например, выбираете любимый фильм, книгу, спектакль? То же самое и в бизнесе — как организовать фирму: по принципу жесткой централизации или наоборот; какого стиля придерживаться в менеджменте: авторитарного или демократического. Но в этой главе мы рассмотрим другие вопросы.

Бухгалтеры часто сталкиваются с проблемами, которые приводят их к количественному подходу принятия решений. Поэтому в этой главе мы займемся количественным анализом. Однако вы должны помнить, что обычно решение является результатом применения как количественного, так и субъективного подходов.

Итак, чтобы найти хорошее решение, следует:

1. Определить, цель решения.
2. Определить возможные варианты решения проблемы.

3. Определить возможные исходы каждого решения.
4. Оценить каждый исход.
5. Выбрать оптимальное решение на основе поставленной цели.

Как видите, поиск решения начинается с перечисления возможных вариантов и их исходов, затем производится оценка каждого исхода. Такова схема рассуждений при проведении количественного анализа. Вышеперечисленные этапы важны как в очень сложных случаях, так и в очень простых. Несмотря на то, что эта глава охватывает лишь малую часть серьезной и обширной темы, мы рассмотрим лишь некоторые из возможных целей принятия решений, но в любом случае выбор "лучшего варианта" зависит от обстоятельств и точки зрения того, кто принимает решение.

❑ **Пример 3.1.** Отдел маркетинга компании "Singles plc" представил своему руководству данные об ожидаемом объеме сбыта программных продуктов при трех вариантах цены.

Таблица 3.1. **Предполагаемые объемы продаж программных продуктов по разным ценам, ф. ст.**

Возможная цена за единицу	*8,00*	*8,60*	*8,80*
Предполагаемый объем продаж при данной цене (единиц в год):			
лучший из возможного	16000	14000	12500
наиболее вероятный	14000	12500	12000
худший из возможного	10000	8000	6000

Постоянные затраты составляют 40000 ф. ст. в год, переменные — 4,00 ф. ст. на единицу.

Решение состоит в том, чтобы назначить оптимальную цену. Заметим, у нас имеется всего лишь три варианта цены, т.е. только три возможных решения, и, чтобы облегчить расчеты, для каждого из вариантов по три исхода — различные объемы продаж.

Решение.

Для каждого исхода рассчитаем доход. В данном случае доход — это годовая прибыль.

Таблица 3.2. **Расчет прибыли за год, ф. ст.**

Цена за единицу	*8,00*	*8,60*	*8,80*
Переменные затраты на единицу продукции	4,00	4,00	4,00
Прибыль на единицу продукции	4,00	4,60	4,80
Общая прибыль за год:			
лучшая из возможного	64000	64000	60000
наиболее вероятная	56000	57500	57600
худшая из возможного	40000	36800	28800

Для того чтобы объяснить, какие трудности возникают в результате неопределенности, мы будем использовать данные из этой таблицы. Можно представить убедительные аргументы, которые приведут нас к одному из трех возможных решений. Наибольшая прибыль для наиболее вероятного объема продаж равна 57600 ф. ст. Эта цифра будет получена, если назначить цену в 8,80 ф. ст. Однако цена 8,60 ф. ст. предпочтительнее для компании, так как наиболее вероятная прибыль составляет примерно ту же величину, в то время как прибыль двух остальных исходов выше, чем для цены 8,80 ф. ст. Однако если мы примем во внимание постоянные расходы, то цена 8,00 ф. ст. — единственная, при которой "Singles" не терпит убытков, так как низкая прибыль здесь не меньше, чем постоянные расходы — 40000 ф. ст.

Таким образом, для любого из трех решений существуют свои аргументы. Какое решение будет принято, зависит от целей, которые оно преследует, и от отношения к риску того, кто принимает решение. Осторожный менеджер предпочтет цену 8,00 ф. ст. двум другим: возможные прибыли меньше, но и потери сведены к минимуму. Поэтому в числе прочих должен решаться вопрос об отношении к риску. Сейчас мы рассмотрим, как правила принятия решений могут применяться в каждом конкретном случае, а к вопросу о риске вернемся позже.

3.2. ПРАВИЛА ПРИНЯТИЯ РЕШЕНИЙ

Как уже отмечалось, принимая решения, следует руководствоваться соответствующими правилами. На первом этапе — определение цели. Принимающий решение сам выбирает, каким правилом ему воспользоваться, потому что для каждого случая применимо какое-то определенное правило. Итак, они делятся на две группы:

— правила принятия решений без использования численных значений вероятностей исходов;

— правила принятия решений с использованием численных значений вероятностей исходов.

3.2.1. Правила принятия решений без использования численных значений вероятностей исходов

1. **Максимаксное решение** — максимизация максимума доходов.
2. **Максиминное решение** — максимизация минимума доходов.
3. **Минимаксное решение** — минимизация максимума возможных потерь.

❐ **Пример 3.2.** Предположим, что вы владелец кондитерской "Cake Box". В начале каждого дня вам нужно решить вопрос, сколько пирожных следует иметь в запасе, чтобы удовлетворить спрос. Каждое пирожное обходится вам в 0,70 ф. ст., а вы его продаете по 1,30 ф. ст. Продать невостребованные пирожные на следующий день невозможно, поэтому остаток распродается в конце дня по 0,30 ф. ст. за штуку. В табл. 3.3 приведены данные по продажам в предыдущие периоды.

Таблица 3.3. **Спрос на пирожные**

Спрос на пирожные в день, шт.	*1*	*2*	*3*	*4*	*5*
Частота	5	10	15	15	5
Относительная частота (вероятность)	0,1	0,2	0,3	0,3	0,1

Нужно определить, сколько пирожных должно быть закуплено в начале каждого дня.

Решение.

Итак, в начале дня можно закупить для последующей продажи 1, 2, 3, 4 или 5 пирожных в день. В общем решение и его исходы примерно равны, но имея возможность принимать решения, нельзя контролировать исходы. Покупатели определяют их сами, поэтому исходы представляют также "фактор неопределенности". Чтобы определить вероятность каждого исхода, составим список возможных решений и соответствующих им исходов. В табл.3.4 рассчитаны доходы, иначе говоря, отдача в денежном выражении для любой комбинации решений и исходов.

Таблица 3.4. **Доход (прибыль) в день, ф. ст.**

Возможные исходы: спрос пирожных в день	*Число закупленных для продажи пирожных (возможные решения)*				
	1	*2*	*3*	*4*	*5*
1	0,60	0,20	0,20	0,60	1,00
2	0,60	1,20	0,80	0,40	0,00
3	0,60	1,20	1,80	1,40	1,00
4	0,60	1,20	1,80	2,40	2,00
5	0,60	1,20	1,80	3,40	3,00

Используя каждое из правил принятия решений, упомянутых в начале раздела, нужно ответить на вопрос: "Сколько пирожных должна закупить фирма "Cake Box" в начале каждого дня?"

1. **Правило максимакса** — максимизация максимума доходов. Каждому возможному решению в приведенной таблице соответствуют следующие максимальные доходы.

По этому правилу вы закупите в начале дня пять пирожных. Это подход карточного игрока — игнорируя возможные потери, рассчитывать на максимально возможный доход.

Таблица 3.5. **Максимальные доходы**

Количество закупаемых в день пирожных	*Максимальный доход (прибыль) в день, ф. ст.*
1	0,60
2	1,20
3	1,80
4	2,40
5	3,00 ← **максимум**

2. **Правило максимина** — максимизация минимального дохода. Каждому возможному решению в табл. 3.4 соответствуют минимальные доходы (табл. 3.6.).

По этому правилу вы закупите в начале дня одно пирожное, чтобы максимизировать минимальный доход. Это очень осторожный подход к принятию решений.

Таблица 3.6. **Минимальные доходы**

Количество закупаемых в день пирожных	*Минимальный доход (прибыль) в день, ф. ст.*
1	0,60 ← **максимум**
2	0,20
3	– 0,20
4	– 0,60
5	– 1,00

3. Правило минимакса – минимизация максимально возможных потерь. В данном случае больше внимания уделяется возможным потерям, чем доходам. Таблица возможных потерь дает представление о прибылях каждого исхода, потерянных в результате принятия неправильного решения. Например, если спрос составляет два пирожных и было закуплено два, то доход составит 1,20 ф. ст., если же вы приобрели три, то доход – 0,80 ф. ст. и вы недополучили 0,40 ф. ст. Эти 0,40 ф. ст. – то, что называется **возможными потерями** или **упущенным доходом**. Таблицу возможных потерь можно получить из таблицы доходов, находя наибольший доход для каждого исхода и сопоставляя его с другими доходами этого же исхода (см. табл. 3.7).

Как уже отмечалось, правило, которое используется для работы с таблицей упущенных доходов, – это правило минимакса. Оно также называется **минимаксное правило возможных потерь**. Состоит оно в том, чтобы для каждого решения выбрать максимально возможные потери. Затем выбирается то решение, которое ведет к минимальному значению максимальных потерь (табл. 3.8).

Таблица 3.7. **Возможные потери в день, ф. ст.**

Возможные исходы: спрос пирожных в день	*Число закупленных для продажи пирожных (возможные решения)*				
	1	*2*	*3*	*4*	*5*
1	0,0	0,40	0,80	1,20	1,60
2	0,60	0,0	0,40	0,80	1,20
3	1,20	0,60	0,0	0,40	0,80
4	1,80	1,20	0,60	0,0	0,40
5	2,40	1,80	1,20	0,60	0,0

Таблица 3.8. **Максимальные возможные потери**

Количество закупаемых в день пирожных	*Максимальные возможные потери в день, ф. ст. (из таблицы выше)*
1	2,40
2	1,80
3	1,20 ← **минимум**
4	1,20 ← **минимум**
5	1,60

Минимальная величина максимальных потерь возникает в результате закупки трех или четырех пирожных в день. Следовательно, по правилу минимакса вы выберете одно из этих решений.

Все рассмотренные критерии принятия решений приводят к различным результатам. Поэтому сначала выбирается тот критерий, который считается "лучшим", и тогда вы получаете "наилучшее" для вас решение.

3.2.2. Критерий Гурвича — компромиссный способ принятия решений

Этот способ принятия решений представляет собой компромисс между осторожным правилом максимина и оптимистичным правилом максимакса. В нем некоторым образом объединяются правила, не рассматривающие индивидуальные вероятности отдельных исходов, и те, в которых учитываются вероятности исходов.

При использовании критерия Гурвича (Hurwicz criterion) таблица доходов составляется как обычно. Для каждого решения рассматриваются лучший и худший результаты, т.е. то, о чем раньше говорилось в правилах максимина и максимакса. Принимающий решение придает вес обоим результатам, и, умножив результаты на соответствующие веса и суммируя, получает общий результат. Выбирается решение с наибольшим результатом. Такое решение задачи предполагает, что имеется достаточно информации для определения весов.

Пример с закупкой пирожных (пример 3.2) не очень приемлем для иллюстрации критерия Гурвича, так как высокие доходы встречаются более, чем в одном исходе. Например, если мы решили закупать три пирожных в день, наивысший доход в 1,80 ф. ст. существует для спроса 3, 4 и 5 пирожных.

Упростим таблицу доходов (табл. 3.4), чтобы проиллюстрировать вышесказанное, и рассмотрим низкие доходы для каждого решения и исходы с высокими доходами. Напоминаем, что принимающий решение не располагает данными о спросе из табл. 3.3, поэтому ему нужно самому вычислить веса для исходов с низкими и высокими доходами. В данном случае самый низкий доход из возможных — при одном пирожном в день, самый высокий — при пяти.

Допустим, принимающий решение определил вес для спроса одного пирожного в день, равным 0,4, а для спроса пяти пирожных — 0,6. Используя эти веса, составим таблицу.

Таблица 3.9. **Критерий Гурвича**

Количество пирожных, закупаемых в день	*Доход в день, ф. ст.*		*Вес*		*Всего в день, ф. ст.*
	низкий	*высокий*	*× 0,4*	*× 0,6*	
1	0,6	0,6	0,24	+ 0,36	= 0,6
2	0,2	1,2	0,08	+ 0,72	= 0,8
3	– 0,2	1,8	– 0,08	+ 1,08	= 1,0
4	– 0,6	2,4	– 0,24	+ 1,44	= 1,2
5	– 1,0	3,0	– 0,40	+ 1,80	= 1,4 ← **максимум**

Если принимающий решение использует указанные веса, то его решение по правилу Гурвича, будет состоять в том, чтобы закупать пять пирожных в день.

3.2.3. Правила принятия решений с использованием численных значений вероятностей исходов

В предыдущем разделе мы не использовали данные о вероятностях исходов, теперь попробуем при принятии решений использовать эти данные.

1. Правило максимальной вероятности – максимизация наиболее вероятных доходов. Рассмотрим относительные частоты (вероятности) дневного спроса на пирожные.

Таблица 3.10. **Относительные частоты (вероятности) дневного спроса на пирожные**

Количество пирожных, закупаемых в день	*1*	*2*	*3*	*4*	*5*
Частота	5	10	15	15	5
Относительная частота (вероятность)	0,1	0,2	0,3	0,3	0,1

Наибольшая вероятность 0,3 соответствует спросу в три и четыре пирожных в день. Теперь рассмотрим доходы каждого из исходов и выберем наибольший.

Таблица 3.11. **Максимальный доход для каждого из решений**

Количество пирожных, закупаемых в день	*Максимальный доход в день, ф. ст.*
3	1,80, когда исход равен 3 или больше
4	2,40, когда исход равен 4 или больше ← **максимум**

По этому правилу фирма "Cake Box" должна закупать четыре пирожных в день.

2. Оптимизация математического ожидания. Наиболее распространенный способ использования вероятностей при принятии решений – это вычисление математического ожидания. Оно рассчитывается для каждого решения либо для доходов, либо для возможных потерь. Выбирается решение либо с наибольшим ожидаемым доходом, либо с наименьшими возможными потерями.

а) Максимизируем ожидаемый доход для решений:

Е(доход от какого-либо решения) = $\sum$ (вероятность × доход) (суммируем для всех исходов рассматриваемого решения).

В примере с "Cake Box" ожидаемый доход в случае, если решено закупать пять пирожных в начале каждого дня, равен:

Е(доход, если закупается пять пирожных) = (0,1 × –1,0) + (0,2 × 0,0) +

+ (0,3 × 1,0) + (0,3 × 2,0) + (0,1 × 3,0) = 1,1 ф. ст. (в день).

При большем временном промежутке это означает, что при закупке пяти пирожных в день **средняя** прибыль составит 1,1 ф. ст. в день.

Ниже приведена таблица доходов фирмы "Cake Box", дополненная вероятностями. Следом за ней — таблица ожидаемых доходов для каждого решения.

Таблица 3.12. **Таблица доходов**

Возможные исходы: дневной спрос на пирожные	*Доход (прибыль) в день, ф. ст., Количество пирожных, закупаемых в день (возможные решения)*					*Вероятность*
	1	*2*	*3*	*4*	*5*	
1	0,60	0,20	– 0,20	0,60	1,00	0,1
2	0,60	1,20	0,80	0,40	0,0	0,2
3	0,60	1,20	1,80	1,40	1,00	0,3
4	0,60	1,20	1,80	2,40	2,00	0,3
5	0,60	1,20	1,80	2,40	3,00	0,1

Таблица 3.13. **Расчет возможных доходов (вероятность × доход из табл. 3.10)**

Возможные исходы: дневной спрос на пирожные	*Количество пирожных, закупаемых в день (возможные решения)*				
	1	*2*	*3*	*4*	*5*
1	0,06	0,02	–0,02	–0,06	–0,10
2	0,12	0,24	0,16	0,08	0,0
3	0,18	0,36	0,54	0,42	0,30
4	0,18	0,36	0,54	0,72	0,60
5	0,06	0,12	0,18	0,24	0,30
Ожидаемый доход в день всего, ф. ст.	0,60	1,10	1,40	1,40	1,10

Итак, максимальное значение ожидаемого дохода 1,40 ф. ст. в день, следовательно, используя критерий максимизации ожидаемого дохода, фирма "Cake Box" должна закупать три или четыре пирожных в день. В примерах этого типа, где решение повторяется множество раз, использование критерия математического ожидания наиболее приемлемо.

б) Минимизация ожидаемых возможных потерь. В данном случае производится та же последовательность действий, только с использованием таблицы возможных потерь и вероятности каждого из исходов. Выбирается решение, ведущее к наименьшим ожидаемым возможным потерям, вместо максимума ожидаемых доходов.

Таблица 3.14. **Возможные потери**

Возможные исходы: дневной спрос на пирожные	*Возможные потери: количество пирожных, закупаемых в день (возможные решения)*					*Вероятность*
	1	*2*	*3*	*4*	*5*	
1	0,0	0,40	0,80	1,20	1,60	0,1
2	0,60	0,0	0,40	0,80	1,20	0,2
3	1,20	0,60	0,0	0,40	0,80	0,3
4	1,80	1,20	0,60	0,0	0,40	0,3
5	2,40	1,80	1,20	0,60	0,0	0,1

Как мы видим, минимальные ожидаемые возможные потери равны 0,46 ф. ст. в день, т.е. наилучшее решение — закупать три или четыре пирожных в день. То же решение следует принять при использовании критерия максимизации ожидаемых доходов.

Таблица 3.15. **Расчет ожидаемых возможных потерь (вероятность × возможные потери)**

Возможные исходы: дневной спрос на пирожные	*Количество пирожных, закупаемых в день (возможные решения)*				
	1	*2*	*3*	*4*	*5*
1	0,0	0,04	0,08	0,12	0,16
2	0,12	0,0	0,08	0,16	0,24
3	0,36	0,18	0,0	0,12	0,24
4	0,54	0,36	0,18	0,0	0,12
5	0,24	0,18	0,12	0,06	0,0
Ожидаемые возможные потери в день — всего, ф. ст.	1,26	0,76	0,46	0,46	0,76

3.2.4. Зависимость решения от изменений значений вероятностей

Значения вероятностей, которые мы используем, основаны либо на уже имеющейся информации, либо на расчетах. Однако эти значения непостоянны, и поэтому полезно знать, насколько велика зависимость выбора решения от изменения величины вероятности, т.е. какова чувствительность решений.

Анализ чувствительности является важной темой, к которой мы будем обращаться на протяжении всей книги. Суть анализа заключается в числовой оценке изменения вероятности, определяющей выбор решения. Для иллюстрации возьмем пример с максимизацией ожидаемых доходов. Ниже рассмотрена ситуация с одним основным и одним альтернативным вариантом решения, хотя, как правило, на практике альтернативных вариантов больше.

Таблица 3.16. **Зависимость выбора решения от изменений значений вероятностей**

	Количество пирожных, закупаемых в день (возможные решения)				
	1	*2*	*3*	*4*	*5*
Базовые вероятности	0,1	0,2	0,3	0,3	0,1
Ожидаемый доход в день, ф. ст.	0,6	1,1	1,4	1,4	1,1
Альтернативные вероятности	0,2	0,2	0,2	0,2	0,2
Ожидаемый доход* в день, ф. ст.	0,6	1,0	1,2	1,2	1,0

* Вычисляется, как в 3.2.2.

Решение, дающее максимальный доход, — закупать три или четыре пирожных, не претерпело изменений, однако средняя прибыль в альтернативном варианте снизилась с 1,40 до 1,20 ф. ст. в день. В данном случае выбор решения нечувствителен к незначительным изменениям вероятности, т.е. не происходит замены выбранного варианта решения на новый.

3.2.5. Стоимость достоверной информации

Неопределенность при принятии решений может быть уменьшена путем сбора дополнительной информации, однако за нее нужно платить. Максимальная сумма денег, которую стоит заплатить, и является **стоимостью достоверной информации**. Если заранее известно, какой из исходов осуществится, то можно принять решение, ведущее к максимальному доходу, тем не менее это не означает, что мы можем контролировать исходы.

Например, фирма "Cake Box" принимает заказы на следующий день. Контролировать их количество нельзя, однако можно, корректируя количество закупаемых пирожных, максимизировать доход. На число закупаемых пирожных теперь влияет число поступающих заказов.

Ожидаемый доход равен:

$$E = \sum (\text{доход на поступивший объем заказов} \times \text{вероятность данного объема заказов})$$

$$E = (0{,}60 \times 0{,}1) + (1{,}20 \times 0{,}2) + (1{,}80 \times 0{,}3) + (2{,}40 \times 0{,}3) + (3{,}00 \times 0{,}1) = 1{,}86 \text{ ф. ст.}$$

Стоимость достоверной информации есть разница полученной цифры и максимального ожидаемого дохода без достоверной информации. Для "Cake Box" стоимость достоверной информации (ф. ст.): 1,86 – 1,40 = 0,46 (в день). Эта цифра равна минимальным ожидаемым возможным потерям.

Если известна стоимость достоверной информации, то известен максимум, который можно заплатить за дополнительную информацию о вероятностях исходов. Таким образом, фирма "Cake·Box" может заплатить 0,46 ф. ст. в день, чтобы получать информацию о спросе, т.е. это плата за своего рода "маркетинговые данные".

3.3. ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ И СТАНДАРТНОГО ОТКЛОНЕНИЯ ДЛЯ ОЦЕНКИ РИСКА

В результате использования правила максимизации ожидаемых доходов (или минимизации ожидаемых возможных потерь) мы получаем оценку для каждого исхода в виде таблицы доходов, чтобы выбрать "наилучшее" решение. В ней приводится разброс доходов для каждого исхода, анализ которого дает возможность оценить риск каждого решения. Альтернативный подход к оценке риска заключается в вычислении стандартного отклонения доходов, как это делается для любого другого вида распределений. Именно таким образом в нижеприведенном примере сравниваются два варианта инвестиций. Несмотря на то, что в этом случае и в примере с закупкой пирожных арифметически два варианта решаются совершенно одинаково, между ними существует значительная разница. Решение, принятое для закупки пирожных, остается неизменным изо дня в день, и идея ожидаемых (средних) доходов проста для понимания, тогда как решение об инвестициях принимается лишь однажды, что затрудняет понимание значения ожидаемых доходов на практике.

❐ **Пример 3.3.** Ниже приведены возможные чистые доходы и их вероятности для двух вариантов вложений.

Таблица 3.17. **Вероятности возможной чистой прибыли**

	Сравнение вариантов решений							
Чистая прибыль, тыс. ф. ст.	−3	−2	−1	0	1	2	3	4
Вероятности:								
Инвестиция 1	0	0	0,1	0,2	0,3	0,2	0,2	0
Инвестиция 2	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2

Ожидаемая прибыль:

$$E\ (\text{инвестиция } 1) = \sum\ (\text{доход} \times \text{вероятность}).$$

Отсюда

$$E\ (\text{ивестиция } 1) = (-\ 3 \times 0) + (-\ 2 \times 0) + (-\ 1 \times 0{,}1) + (0 \times 0{,}2) + (1 \times 0{,}3) +$$
$$+ (2 \times 0{,}3) + (3 \times 0{,}2) + (4 \times 0).$$

Следовательно,

$$E\ (\text{инвестиция } 1) = 1200 \text{ ф. ст.}$$

Аналогично для инвестиции 2:

$$E\ (\text{инвестиция } 2) = (-\ 3 \times 0{,}1) + (-\ 2 \times 0{,}1) + (-\ 1 \times 0{,}1) + (0 \times 0{,}1) +$$
$$+ (1 \times 0{,}1) + (2 \times 0{,}1) + (3 \times 0{,}2) + (4 \times 0{,}2).$$

Следовательно,

$$E\ (\text{инвестиция } 2) = 1100 \text{ ф. ст.}$$

Если принимать во внимание только ожидаемую прибыль, то инвестиция 1 безусловно лучше. Если бы решение об инвестициях принималось много раз при одних и тех же условиях, то тогда прибыль в среднем составляла бы 1200 ф. ст. Однако правило принятия решений не учитывает риск, связанный с инвестициями, т.е. "разброс" возможных исходов. Этот риск может быть определен с помощью дисперсии и стандартного отклонения прибыли.

Из гл. 2 мы знаем, что дисперсия вероятностного распределения представляет собой:

$$\text{Дисперсия} = \sum px^2 - (E\ (x))^2;$$

$$E\ (x) = \sum p\ x,$$

где x — прибыль на инвестиции;
p — вероятность получения данной прибыли.

Таблица 3.18. **Расчет средней прибыли и дисперсии для инвестиций**

Прибыль, тыс. ф. ст.	*Инвестиция 1*			*Инвестиция 2*		
x	p	px	px^2	p	px	px^2
−3	0	0	0	0,1	− 0,3	0,9
−2	0	0	0	0,1	− 0,2	0,4
−1	0,1	−0,1	0,1	0,1	− 0,1	0,1
0	0,2	0	0	0,1	0	0
1	0,3	0,3	0,3	0,1	0,1	0,1
2	0,2	0,4	0,8	0,1	0,2	0,4
3	0,2	0,6	1,8	0,2	0,6	1,8
4	0	0	0	0,2	0,8	3,2
Всего	1,0	1,2	3,0	1,0	1,1	6,9

Инвестиция 1:

$$\text{Дисперсия} = 3{,}0 - 1{,}2^2 = 1{,}56^2 \text{ (тыс. ф. ст.)}.$$

Следовательно,

$$\text{Стандартное отклонение прибыли} = \sqrt{1{,}56} = 1250 \text{ ф. ст.}$$

Инвестиция 2:

$$\text{Дисперсия} = 6{,}9 - 1{,}1^2 = 5{,}69^2 \text{ (тыс. ф. ст.)}.$$

Следовательно,

$$\text{Стандартное отклонение прибыли} = \sqrt{5{,}69} = 2385 \text{ ф. ст.}$$

Риск по варианту для инвестиции 1 меньше, так как дисперсия прибыли намного меньше, чем для инвестиции 2.

Таблица 3.19. **Математическое ожидание и стандартное отклонение для двух вариантов инвестиций, ф. ст.**

Инвестиция	*Ожидаемая прибыль*	*Стандартное отклонение*
1	1200	1250
2	1100	2385

Анализируя данные таблицы, можно прийти к выводу, что как большая ожидаемая прибыль, так и меньший "разброс" говорят в пользу инвестиции 1.

3.4. ИСПОЛЬЗОВАНИЕ ПОНЯТИЯ ПОЛЕЗНОСТИ ПРИ ОПРЕДЕЛЕНИИ РАЗМЕРОВ РИСКА

До сих пор мы рассматривали только правила принятия решений: кто-то выбирает правило, которое он предпочитает, и получает "лучшее" решение. Во внимание не принималось, кто же делает выбор — миллионер или студент, предпочитает ли он риск или стабильность, хотя его предпочтения уже частично определены тем выбором, который он сделал. Теория полезности позволяет принимающему решение влиять на денежный результат исходов согласно своим оценкам их полезности. Одно и то же правило в данном случае приводит к разным решениям у разных людей, каждый может приспосабливать процесс принятия решений к своим запросам.

Для примера рассмотрим два варианта вложений 1000 ф. ст.. По первому варианту без какого-либо риска можно получать 10% прибыли на вложенный капитал, т.е. через год сумма возрастет до 1100 ф. ст. По второму варианту можно, либо потерять весь капитал, либо его удвоить через год.

Итак, таблица доходов такова:

Таблица 3.20. **Доход за один год**

Возможные исходы варианта 2	*Возможные варианты вложений 1000 ф. ст.*	
	Доход, ф. ст.	
	вариант 1	*вариант 2*
Успех	1100	2000
Неудача	1100	0

Доход в 1100 ф. ст. по 100-балльной шкале — около 55, если точка отсчета — 0, а верхний предел — 2000 ф. ст. Теперь рассмотрим, какова полезность 1100 ф. ст. для двух разных людей, принимающих решение.

Например, для студентки, у которой исходная сумма 1000 ф. ст. — последние деньги, их потеря невосполнима, поэтому полезность 1000 ф. ст. высока. Чтобы выразить ее в цифрах, попросим ее оценить вероятность (P) максимального дохода (2000 ф. ст.) до того, как она примет решение и воспользуется вариантом 2. Если вероятность успеха варианта 2 мала, предположим, 0,1, то рисковать не стоит, лучше воспользоваться безрисковым вариантом 1. Однако если вероятность успеха равна 1, тогда она воспользуется вариантом 2, как наиболее предпочтительным. Где-то в интервале значения вероятности $P = \{0{,}1;\ 1\}$ находится точка замены безрискового варианта 1 более прибыльным, но опасным вариантом 2. Значение вероятности, где происходит смена варианта решения, представляет собой оценку полезности 1100 ф. ст. Допустим, $P = 0{,}95$; тогда полезность: $0{,}95 \times 100 = 95$. Таким образом, денежная шкала: 0 ф. ст. – 1100 ф. ст. – 2000 ф. ст. была заменена на шкалу полезности: 0 – 95 – 100.

Второй инвестор обладает капиталом в 500000 ф. ст. Потеря 1000 ф. ст. вряд ли серьезно ударит по его карману, и риск большой роли не играет. Поэтому в данном случае тоже устанавливается значение вероятности P, когда один вариант решения может быть заменен на другой. Допустим, $P = 0{,}2$; т.е. если вероятность успеха меньше 0,2, то стоит остановиться на варианте 1 и ограничиться доходом в 1100 ф. ст. Если же $P > 0{,}2$, тогда выбирается вариант 2, т.е. полезность 1100 ф. ст.

будет 0,2 × 100 = 20, и денежная шкала: 0 ф. ст. – 1100 ф. ст. – 2000 ф. ст. заменяется на шкалу полезности: 0 – 20 – 100 ф. ст..

Таким образом, одна и та же денежная шкала может быть заменена разными шкалами полезности в зависимости от возможностей и критериев инвесторов.

Для того, чтобы проиллюстрировать преимущества шкалы полезности по сравнению с денежной шкалой, в следующем разделе приводится пример, в котором используется правило максимизации математических ожиданий.

3.4.1. Преимущества шкалы полезности

В примере 3.4, используя правило максимизации математических ожиданий, мы продемонстрируем плюсы оценок полезности по сравнению с денежными доходами. Сначала воспользуемся критерием максимизации дохода. Переоценим доходы с помощью оценок полезности, а затем применим правило максимизации ожиданий полезности.

❑ **Пример 3.4.** Допустим, вы накопили 5000 ф. ст., чтобы купить дом в следующем году. И вдруг знакомый предлагает вам вложить деньги в его бизнес. В случае неудачи вы теряете 5000 ф. ст. и возможность купить дом. В случае успеха через год вы получаете 30000 ф. ст. Специалист по маркетингу оценивает вероятность успеха в 0,3. Альтернативный вариант – положить деньги в банк под 9% годовых, и никакого риска.

Таблица 3.21. **Доходы**

Возможные исходы	*Возможные решения: вложить 5000 ф. ст. в:*		*Вероятность*
	бизнес	*банк*	
Успех в бизнесе	30000	5450	0,3
Неудача в бизнесе	0	5450	0,7
Ожидаемый доход, ф. ст.	9000	5450	
	↑ *выбираем на основе максимизации ожидаемого дохода*		

По денежной шкале вложение денег в бизнес дает наибольший ожидаемый доход. Поэтому использование этого правила влечет за собой риск в расчете на большую прибыль. Однако этот выбор несколько опрометчив, так как в случае потери денег покупка дома останется лишь мечтой.

Шкала полезности для данного примера выглядит следующим образом:

0 – наименьший доход – 0 ф. ст.,

100 – наибольший доход – 30000 ф. ст., т. е.

$$U(0) = 0 \text{ и } U(30000) = 100.$$

На практике неважно, как будет градуирована шкала полезности – от 0 до 100 или от 0 до 1, имеет значение лишь соразмерность.

Для дохода 5450 ф. ст. не требуется оценка полезности, нужно только определить, какова должна быть вероятность P дохода 5450 ф. ст., если вы посчитаете его настолько же привлекательным, насколько и доход 30000 ф. ст. с вероятностью P, и 0 с вероятностью (1 – P).

Предположим, для вас достаточна вероятность по меньшей мере 60% успеха, т.е. P = 0,6, тогда полезность 5450 ф. ст.:

$$U(5450) = p \times 100 = 0{,}6 \times 100 = 60.$$

Таблица оценок полезности для этого примера следующая:

Таблица 3.22. **Таблица полезности**

Возможные исходы	*Возможные решения: инвестирование 5000ф. ст. в:*		*Вероятность*
	бизнес	*банк*	
Бизнес преуспевает	100	60	0,3
Крах бизнеса	0	60	0,7
Ожидаемая полезность	30	60	
	↑ *выбираем из-за наибольшей ожидаемой полезности*		

Вложение денег в банк – решение с наибольшей ожидаемой полезностью, однако это – прямо противоположно выбору, сделанному на основе критерия ожидаемого дохода, из-за учета риска, связанного с возможным исходом бизнеса. Для того чтобы оценить этот риск, начертим график, учитывающий оценки полезности и доходы. Сделать это можно, проставив значения U(0) и U(100) и соединив их прямой линией. Если оценка полезности 5450 ф. ст. находится выше этой линии, то принимающий решение принадлежит к числу тех, кто избегает риска, если ниже, то наоборот.

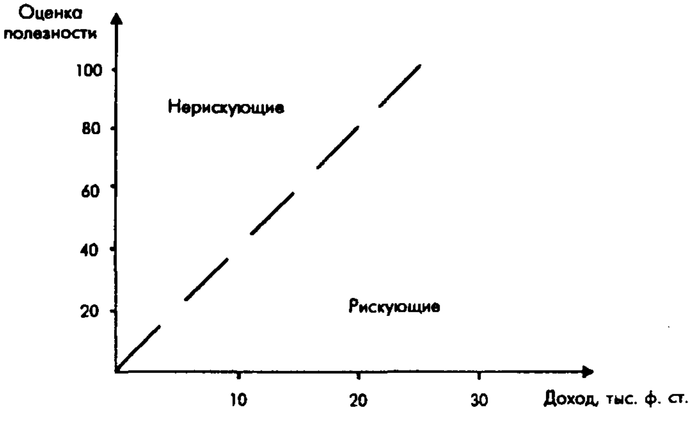


Рис. 3.1. График полезности

Как видно из графика, принимающий такого рода решение относится к **нерискующим**. Идею полезности можно использовать для решения задач с несколькими возможными решениями.

3.5. "ДЕРЕВО" РЕШЕНИЙ

Примеры, которые мы рассматривали до сих пор в этой главе, включали в себя единственное решение. Однако на практике результат одного решения заставляет нас принимать следующее и т.д. Эту последовательность нельзя выразить таблицей доходов, поэтому нужно использовать какой-то другой процесс принятия решений.

Схема **"дерево" решений** очень похожа на схему "дерево" вероятностей. Ее используют, когда нужно принять несколько решений в условиях неопределенности, когда каждое решение зависит от исхода предыдущего или исходов испытаний. Составляя "дерево" решений, нужно нарисовать "ствол" и "ветви", отображающие структуру проблемы. Располагаются "деревья" слева направо. "Ветви" обозначают возможные альтернативные решения, которые могут быть приняты, и возможные исходы, возникающие в результате этих решений. На схеме мы используем два вида "ветвей": первый — пунктирные линии, соединяющие квадраты возможных решений, второй — сплошные линии, соединяющие кружки возможных исходов.

Квадратные "узлы" обозначают места, где принимается решение, круглые "узлы" — появление исходов. Так как принимающий решение не может влиять на появление исходов, ему остается лишь вычислять вероятность их появления.

Когда все решения и их исходы указаны на "дереве", просчитывается каждый из вариантов, и в конце проставляется его денежный доход. Все расходы, вызванные решением, проставляются на соответствующей "ветви".

❑ **Пример 3.5.** Для финансирования проекта бизнесмену нужно занять сроком на один год 15000 ф. ст. Банк может одолжить ему эти деньги под 15% годовых или вложить в дело со 100%-ным возвратом суммы, но под 9% годовых. Из прошлого опыта банкиру известно, что 4% таких клиентов ссуду не возвращают. Что делать? Давать ему заем или нет? Перед вами пример задачи с одним решением, поэтому можно воспользоваться как таблицей доходов, так и "деревом". Рассмотрим оба варианта.

Решение 1 (по таблице доходов).

Максимизируем ожидаемый в конце года чистый доход, который представляет собой разность суммы, полученной в конце года, и инвестированной в его начале. Таким образом, если заем был выдан и возвращен, то чистый доход составит:

Чистый доход = ((15000 + 15% от 15000) − 15000) = 2250 ф. ст.

Таблица 3.23. **Чистый доход в конце года, ф. ст.**

Возможные исходы	*Возможные решения*		*Вероятность*
	выдавать заем	*не выдавать заем*	
Клиент заем возвращает	2250	1350	0,96
Клиент заем не возвращает	−15000	1350	0,04
Ожидаемый чистый доход	1560	1350	

Если банк решает выдать заем, то максимальный ожидаемый чистый доход равен 1560 ф. ст.

Решение 2 (по "дереву" решений).

В данном случае также используем критерий максимизации ожидаемого в конце года чистого дохода.

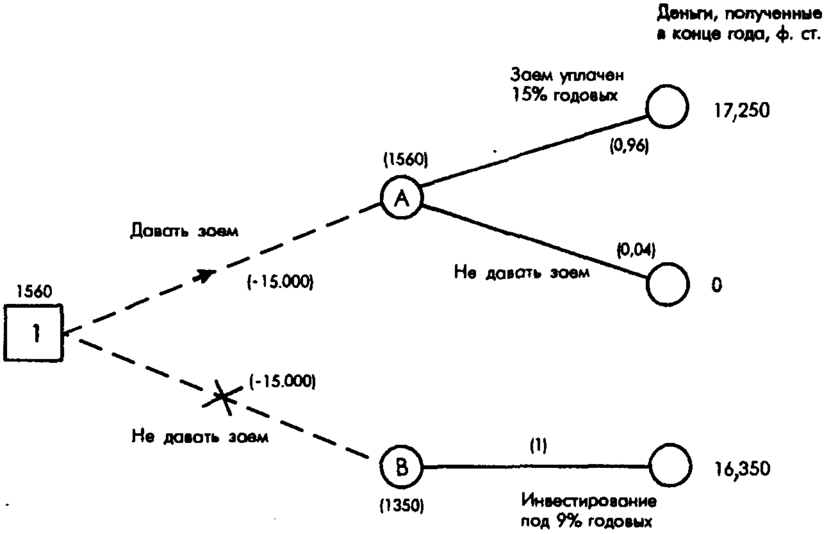


Рис. 3.2. "Дерево" решений для примера 3.5

Далее расчет ведется аналогично расчетам по таблице доходов. Ожидаемый чистый доход в кружках A и B вычисляется следующим образом:

В кружке А:

$$E \text{ (давать заем)} = \{17250 \times 0,96 + 0 \times 0,04\} - 15000 =$$

$$= 16500 - 15000 = 1560 \text{ ф. ст.}$$

В кружке Б:

$$E \text{ (не давать заем)} = \{16350 \times 1,0 - 15000\} = 1350 \text{ ф. ст.}$$

Поскольку ожидаемый чистый доход больше в кружке А, то принимается решение выдать заем.

3.5.1. Расчет двухуровневого "дерева" решений

❑ **Пример 3.6.** Рассмотрим ситуацию более сложную, чем в предыдущем примере, а именно: банк решает вопрос, проверять ли конкурентоспособность клиента перед тем, как выдавать заем. Аудиторская фирма берет с банка 80 ф. ст. за каждую проверку. В результате этого перед банком встают две проблемы: первая — проводить или нет проверку, вторая — выдавать после этого заем или нет.

Решая первую проблему, банк проверяет правильность выдаваемых аудиторской фирмой сведений. Для этого выбираются 1000 человек, которые были проверены и которым впоследствии выдавались ссуды:

Таблица 3.24. **Рекомендации аудиторской фирмы и возврат ссуды**

Рекомендации после проверки кредитоспособности	*Фактический результат*		
	Клиент ссуду вернул	*Клиент ссуду не вернул*	*Всего*
Давать ссуду	735	15	750
Не давать ссуду	225	25	250
Всего	960	40	1000

Какое решение должен принять банк?

Решение.

Этап 1. Построим "дерево", как показано ниже. Вероятности проставляются по данным этапа 2.

Этап 2. Используя данные табл. 3.24, вычислим вероятность каждого исхода:

$$P\ (\text{клиент ссуду вернет; фирма рекомендовала}) = 7{,}35/750 = 0{,}98;$$

$$P\ (\text{клиент ссуду не вернет; фирма рекомендовала}) = 15/750 = 0{,}02;$$

$$P\ (\text{клиент ссуду вернет; фирма не рекомендовала}) = 225/250 = 0{,}9;$$

$$P\ (\text{клиент ссуду не вернет; фирма не рекомендовала}) = 25/250 = 0{,}1.$$

Этап 3. На этом этапе слева направо проставим денежные исходы каждого из "узлов", используя конечные результаты, вычисленные ранее. Любые встречающиеся расходы вычитаем из ожидаемых доходов. Таким образом подсчитываем все "дерево", опираясь на ранее полученные результаты. После того, как пройдены квадраты "решений", выбирается "ветвь", ведущая к наибольшему из возможных при данном решении ожидаемому доходу. Другая "ветвь" зачеркивается, а ожидаемый доход проставляется над квадратом решения.

Сначала посмотрим на кружки исходов B и C, являющиеся следствием квадрата 2 (выдавать ли заем клиенту?)

Доход, ожидаемый от исхода B:

$$E\ (B) = 17250\ \text{ф. ст.} \times 0{,}98 + 0 \times 0{,}02 = 16905\ \text{ф. ст.},$$

чистый ожидаемый доход:

$$NE\ (B) = 16905 - 15000 = 1905\ \text{ф. ст.}$$

Доход, ожидаемый от исхода С:

$$E(C) = 16350 \text{ ф. ст.} \times 1{,}0 = 16350 \text{ ф. ст.},$$

чистый ожидаемый доход:

$$NE(C) = 16350 - 15000 = 1350 \text{ ф. ст.}$$

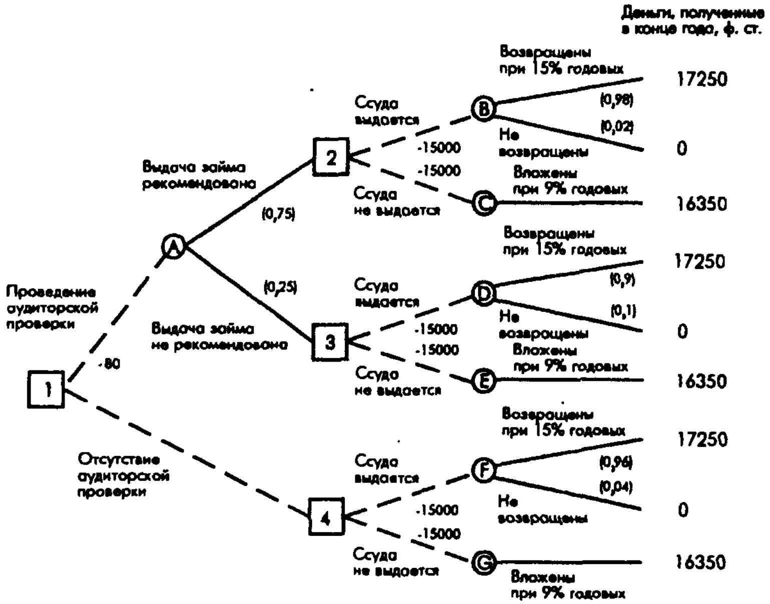


Рис.3.3. "Дерево" решений для банка с учетом аудиторской проверки

Предположим, что мы сейчас в квадрате 2. Максимальный ожидаемый здесь доход 1905 ф. ст. в кружке В, поэтому принимаем решение выдать заем.

Приняв решение, корректируем "дерево", проставив чистый ожидаемый доход 1905 ф. ст. над квадратом 2. "Ветвь" — не давать заем — зачеркивается, как это показано на рис. 3.4.

То же самое с кружками исходов D и Е — результатами решения 3.

Доход, ожидаемый от исхода D:

$$E(D) = (17250 \text{ ф. ст.} \times 0{,}9) + (0 \times 0{,}1) = 15525 \text{ ф. ст.},$$

чистый ожидаемый доход:

$$NE(D) = 15525 - 15000 = 525 \text{ ф. ст.}$$

Аналогично для исхода E:

$$E(E) = 16350 \text{ ф. ст.} \times 1{,}0 = 16350 \text{ ф. ст.},$$

чистый ожидаемый доход:

$$NE(E) = 16350 - 15000 = 1350 \text{ ф. ст.}$$

Если бы мы были в квадрате 3, то максимальный ожидаемый доход был бы равен 1350 ф. ст. и можно было бы принять решение не выдавать заем. Теперь скорректируем эту часть схемы: над квадратом 3 пишем чистый ожидаемый доход и принимаем решение выдать заем.

Наконец приступаем к расчету кружков исходов F и G, которые являются результатами решения 4.

$$E(F) = 17250 \text{ ф. ст.} \times 0{,}96 + 0 \times 0{,}04 = 16560 \text{ ф. ст.};$$

$$NE(F) = 16560 - 15000 = 1560 \text{ ф. ст.};$$

$$E(G) = 16350 \times 1{,}0 = 16350 \text{ ф. ст.};$$

$$NE(G) = 16350 - 15000 = 1350 \text{ ф. ст.}$$

В квадрате 4 максимальный ожидаемый чистый доход составляет 1560 ф. ст., и поэтому принимаем решение выдать клиенту ссуду. Сумма 1560 ф. ст. надписывается над квадратом 4, а альтернативная "ветвь" перечеркивается.

Теперь вернемся к "узлам" A и 1. Используя ожидаемые чистые доходы над квадратами 2 и 3, рассчитаем математическое ожидание для кружка A:

$$E(A) = (1905 \text{ ф. ст.} \times 0{,}75) + (1350 \text{ ф. ст.} \times 0{,}25) = 1766 \text{ ф. ст.}$$

Так как аудиторская проверка стоит 80 ф. ст., ожидаемый чистый доход:

$$NE(A) = 1766 - 80 = 1686 \text{ ф. ст.}$$

Теперь можно проставить значения первого решения квадрата 1. Должен ли банк воспользоваться аудиторской проверкой? В этом "узле" максимальное математическое ожидание — 1686 ф. ст., поэтому перечеркиваем альтернативную "ветвь".

На рис.3.4 стрелками показана последовательность решений, ведущая к максимальному чистому доходу: в квадрате 1 воспользуемся аудиторской проверкой. Если выдача заема рекомендуется фирмой, тогда в квадрате 2 — выдать ссуду, если не рекомендуется, то в квадрате 3 — не выдавать ссуду, а инвестировать эти деньги под стабильные 9% годовых. "Дерево" окончательных решений для примера **3.6.** приведено на рис. 3.4.

❑ **Пример 3.7.** Фирма "Tranda plc", занимающаяся исследованием рынка, рассчитывает расширить свою деятельность, снабдив персональными компьютерами персонал, занимающийся сбором данных. Проблема состоит в том, покупать ли компьютеры или арендовывать. Предсказать рост масштабов деятельности фирмы в ближайшие четыре года нельзя, но возможно разделить его на значительный,

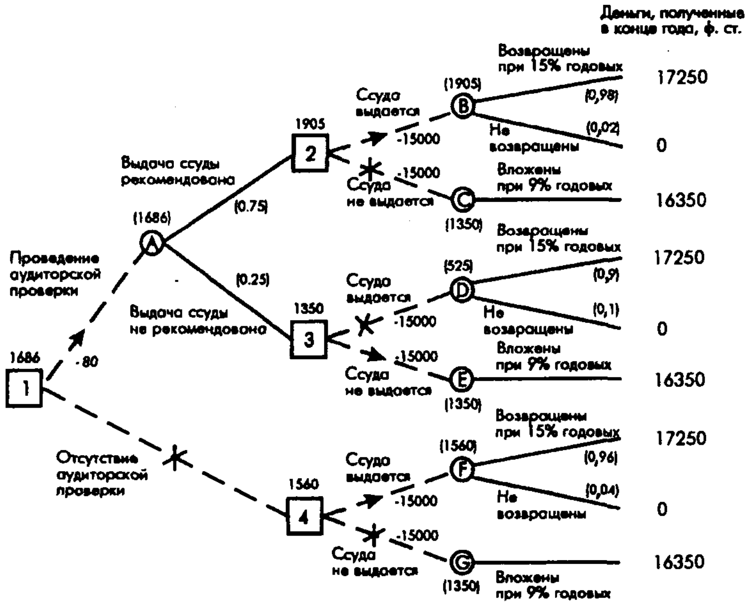


Рис. 3.4. Окончательное "дерево" решений для примера 3.6

средний и незначительный. Вероятность значительного роста масштаба деятельности в первый год после установки компьютеров составляет 0,6; среднего и незначительного — 0,3 и 0,1 соответственно. В последующие три года рост может оцениваться как значительный и незначительный. Подсчитано, что если рост значительный в первый год, то вероятность того, что он останется таким же в последующие три года, равна 0,75. Средний рост первого года изменится на незначительный в последующие годы с вероятностью 0,5. А незначительный таким же и останется с вероятностью 0,9. Чистые наличные доходы, вызванные этими изменениями, приведены в табл. 3.25.

Таблица 3.25. **Доходы наличности**

Рост	*Доход наличности на конец года, ф. ст.*
Значительный	20000
Средний	14000
Незначительный	11000

Стоимость компьютеров — 35000 ф. ст. Условия аренды: первоначальный взнос — 15000 ф. ст. плюс 25% чистой наличной выручки на конец года. Компания рассчитывает получать 12% годовой прибыли на вложенный капитал.

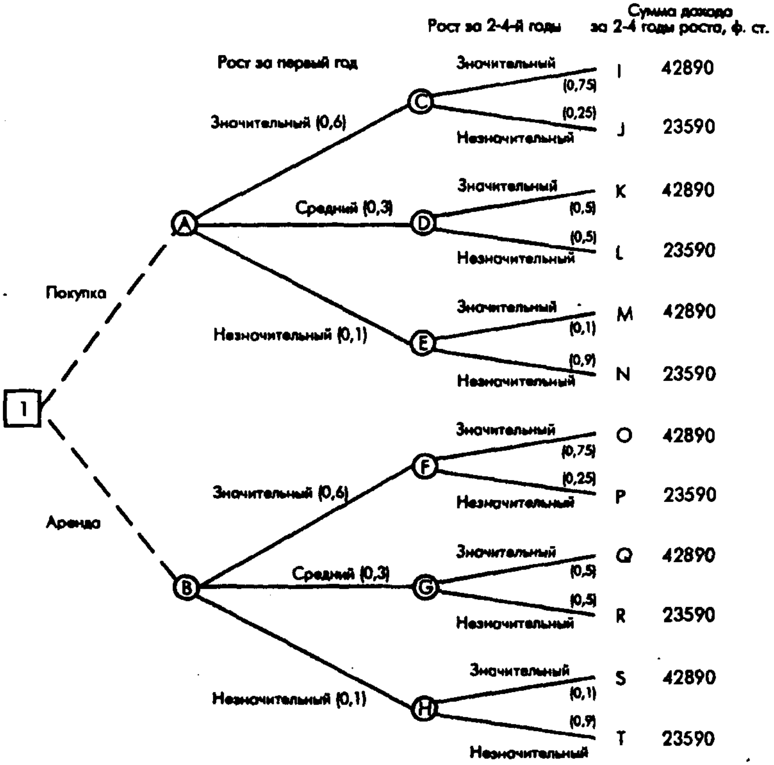


Рис. 3.5. "Дерево" решений для покупки или аренды

Для того, чтобы решить, должна ли фирма покупать или арендовывать компьютеры, воспользуемся "деревом". Критерием принятия решения является максимизация ожидаемой чистой выручки с учетом 12%-ного приращения капитала в год.

Решение.

Этап 1. Составляем "дерево" для покупки-аренды компьютеров.

Отметим, что обе половины "дерева" — покупка и аренда — не зависят от начальных затрат, а зависят только от сумм предполагаемого дохода, которые рассчитываются на конечном этапе.

Этап 2. Подсчитаем суммы, получаемые за 1–4 годы работы. Значения доходов, проставленные в крайней правой части "дерева" — это доходы за 2–4 годы, соответствующие сегодняшнему уровню доходов (табл. 3.25) с учетом годовой 12%-ной надбавки, которую предусматривает фирма.

Если в конце года компания получает А ф. ст. и рассчитывает на 12% годового прироста, то тогда сегодняшнее (текущее) значение А ф. ст. для 2–4 года работы будет равно:

$$\text{Текущее значение A} = \frac{A}{(1+0{,}12)^2} + \frac{A}{(1+0{,}12)^3} + \frac{A}{(1+0{,}12)^4} =$$

$$= A\left(\frac{1}{1{,}12^2} + \frac{1}{1{,}12^3} + \frac{1}{1{,}12^4}\right) = A \times 2{,}1445 .$$

Поэтому в "узле" I, где А (доход за год) должно быть равно 20000 ф. ст., текущее значение дохода за 2–4 годы с учетом 12% годовых:

$$PV_I = 20000 \text{ ф. ст.} \times 2{,}1445 = 42890 \text{ ф. ст.}$$

Аналогичная цифра для J:

$$PV_J = 11000 \text{ ф. ст.} \times 2{,}1445 = 23590 \text{ ф. ст.}$$

Чередуясь, эти два значения повторяются от К до Т.

Этап 3. Используя текущие значения доходов (present value), можно вычислить математическое ожидание исходов от С до Н. В исходе С несмещенная величина ожидаемого текущего дохода за годы 2–4 равна:

$$EPV_C \text{ (годы 2-4)} = (42890 \text{ ф. ст.} \times 0{,}75) + (23590 \text{ ф. ст.} \times 0{,}25) =$$

$$= 38065 \text{ ф. ст.}$$

На первом году работы этой величине соответствует доход в 20000 ф. ст., текущая величина этой суммы равна:

$$\frac{20000}{1{,}12} = 17{,}857 \text{ ф. ст.}$$

Следовательно, ожидаемая текущая стоимость "узла" С за 1–4 годы:

$$EPV_C \text{ (годы 1–4)} = 38065 + 17857 = 55922 \text{ ф. ст.}$$

Исход "узла" D, ожидаемая текущая стоимость доходов за 1–4 годы при 12% годовых, составит:

$$EPV_D = ((42890 \times 0{,}5) + (23590 \times 0{,}5)) + \frac{14000}{1{,}12} = 45740 \text{ ф. ст.}$$

Исход "узла" E, ожидаемая текущая стоимость доходов за 1–4 годы:

$$EPV_E = ((42890 \times 0{,}1) + (23590 \times 0{,}9)) + \frac{11000}{1{,}12} = 35341 \text{ ф. ст.}$$

В результате симметрии ожидаемые текущие величины доходов в кружках F, G и H составляют 55922 ф. ст., 45740 ф. ст. и 35341 ф. ст. соответственно. На этом расчеты по правой стороне "дерева" заканчиваются, и можно приступить к вычислению ожидаемых текущих доходов в "узлах" А и В. Для обоих "узлов" значения доходов одинаковые.

$$EPV_A = EPV_B = 55922 \times 0,6 + 45740 \times 0,3 +$$

$$+ 35341 \times 0,1 = 50809 \text{ ф. ст.}$$

Чистый ожидаемый текущий доход по А (если компьютеры будут куплены) составит:

$$EPV_A = \text{ожидаемая текущая стоимость} - \text{стоимость покупки} =$$

$$= 50809 - 35000 = 15809 \text{ ф. ст.}$$

Для расчета низкой ожидаемой текущей стоимости по "узлу" В вычислим стоимость аренды — 15000 ф. ст., которые выплачиваются сразу, плюс 25% чистого годового дохода наличности. Ожидаемая текущая величина чистого дохода наличности составляет 50809 ф. ст. Следовательно, ожидаемая текущая величина стоимости аренды равна:

$$15000 \text{ ф. ст.} + 25\% \text{ от } 50809 \text{ ф. ст.} = 15000 + 12702 = 27702 \text{ ф. ст.}$$

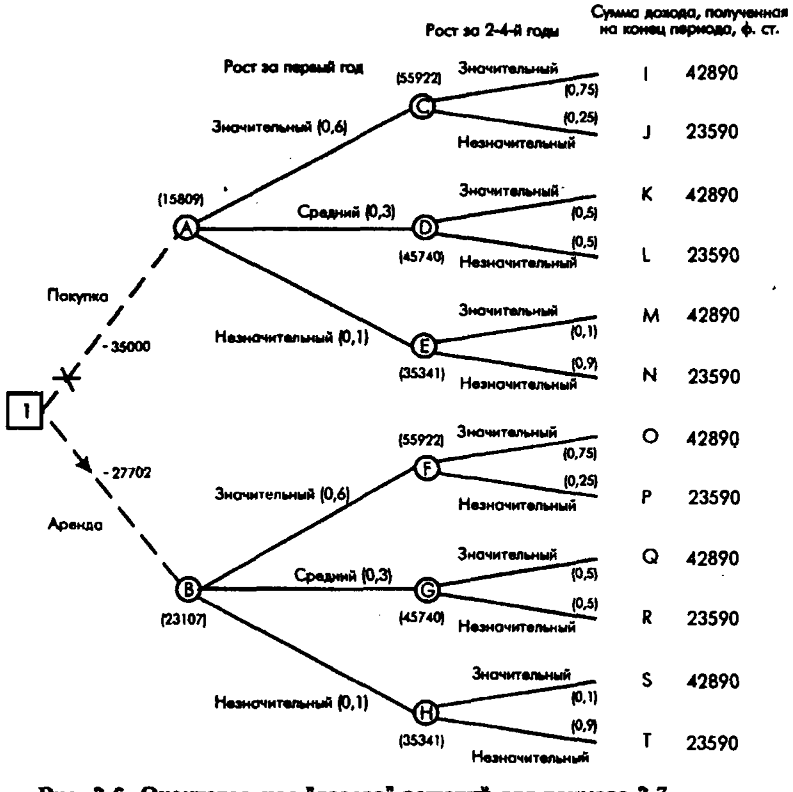


Рис. 3.6. Окончательное "дерево" решений для примера 3.7

Отсюда чистая ожидаемая текущая стоимость по исходу B (если компьютеры будут взяты в аренду) составит:

50809 ф. ст. – 27702 ф. ст. = 23107 ф. ст.

Теперь вернемся к рассмотрению квадрата 1. Максимизируя ожидаемую текущую величину чистых доходов, мы сравним значение исхода кружка A (15809 ф. ст. при покупке) со значением в кружке B (23107 ф. ст. при аренде). Из чего следует, что компания должна арендовать компьютеры. Окончательная схема для примера 3.7 приведена на рис. 3.6:

Обратим внимание на то, что в расчетах по покупке компьютеров не была учтена их остаточная стоимость через четыре года, а это может коренным образом повлиять на вектор решения.

3.5.2. "Дерево" и анализ чувствительности решений

Решения, принимаемые при помощи "дерева", зависят от вероятностей исходов. Чувствительность решения определяется размером изменений вероятности. Выбирая решение, мы должны знать, насколько оно зависит от изменений вероятностей, и, следовательно, насколько можно полагаться на этот выбор.

❐ **Пример 3.8.** Компанией "Cacus Chemical Company" был разработан новый товар. Вполне вероятно, что для него существует рынок сбыта на ближайший год. Наличие в производственном процессе высокотемпературных реакций повышает его стоимость до 2,5 млн. ф. ст. Для организации производственного процесса потребуется один год, однако, существует лишь 55%-ная вероятность, что будет обеспечена должная технологическая безопасность процесса. В связи с этим перед компанией встал вопрос о разработке компьютерной контролирующей системы (ККС), которая будет обеспечивать безопасность высокотемпературных реакций. Исследования по ККС продолжатся год и будут стоить 1 млн. ф. ст. Вероятность получения требуемой ККС – 0,75.

Разработку ККС можно начать либо немедленно, либо подождать год до выяснения технологической безопасности процесса. Если разработку начать немедленно, а производственный процесс окажется безопасным, ККС окажется бесполезной (убыток – 1 млн. ф. ст.). С другой стороны, если отложить разработку ККС, а процесс производства не будет соответствовать стандартам, то выпуск нового товара отодвигается на год до окончания исследований. И наконец, если невозможно создать безопасный процесс и работа над ККС окажется безуспешной, то альтернативного пути выпуска товара не существует, и работы по этому проекту необходимо прекратить. В случае, если продажа нового товара начинается в течение года, то прибыль составит 10 млн. ф. ст., если не принимать в расчет амортизацию по производственному процессу, или ККС. Если отложить выпуск товара на один год, прибыль упадет до 8,5 млн. ф. ст. из-за возможного появления конкурентов на рынке. Для облегчения расчетов можно не учитывать расходы на создание ККС.

1. Составьте "дерево", охватывающее все возможные варианты развития событий.
2. Как бы вы посоветовали поступить руководству компании?

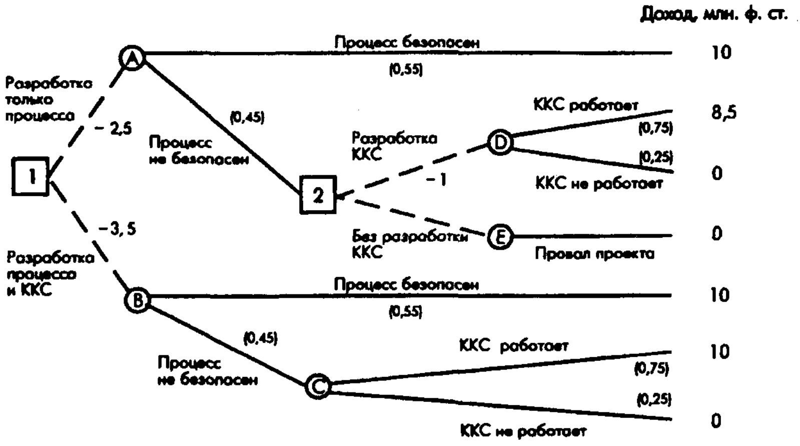


Рис. 3.7. Дерево решений для примера 3.8

3. Как должна измениться вероятность успешной разработки производственного процесса (на сегодняшний день определенная в 0,55), чтобы вы изменили свои рекомендации в вопросе 2? Имеет ли решение этого вопроса некоторый запас прочности (чувствительность) при изменениях вероятности?

Решение

1. "Дерево" решений для этой задачи представлено на рис.3.7.

2. Для того, чтобы оформить "дерево", рассчитаем ожидаемый чистый доход по "узлам". Ожидаемый доход в кружке D:

$$8{,}5 \times 0{,}75 + 0 \times 0{,}25 = 6{,}375 \text{ млн. ф. ст.}$$

Ожидаемый чистый доход:

$$6{,}375 \text{ млн. ф. ст.} - 1 \text{ млн. ф. ст.} = 5{,}375 \text{ млн. ф. ст.}$$

В кружке E ожидаемый чистый доход равен 0. Следовательно, если в квадрате 2 мы решим разрабатывать ККС, то получим чистый доход 5,375 млн. ф. ст.

В "узле" исхода A ожидаемый чистый доход:

$$(10 \times 0{,}55 + 5{,}375 \times 0{,}45) - 2{,}5 = 5{,}419 \text{ млн. ф. ст.}$$

В узле B ожидаемый чистый доход:

$$(10 \times 0{,}55 + (10 \times 0{,}75 + 0 \times 0{,}25) \times 0{,}45) - 3{,}5 = 5{,}375 \text{ млн. ф. ст.}$$

Поэтому в "узле" 1 мы выбираем разработку только производственного процесса. Если через год окажется, что он не безопасен, то приступим к разработке ККС.

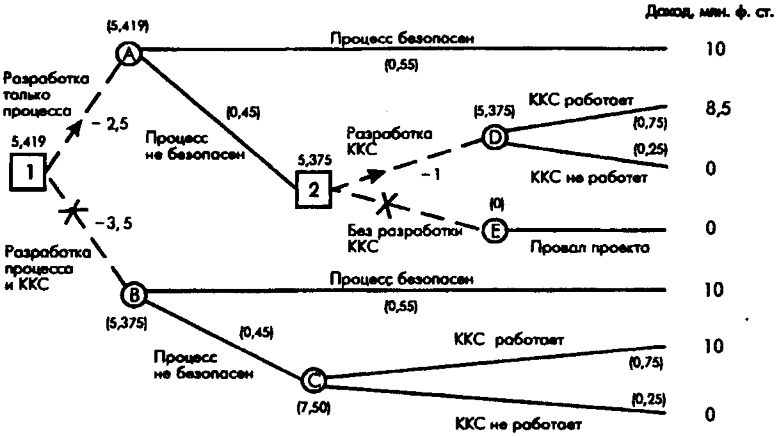


Рис. 3.8. Окончательное дерево решений для примера 3.8

Ожидаемый чистый доход составит 5,419 млн. ф. ст. Окончательный вариант "дерева" для примера **3.8** приведен на рис. 3.8.

3. Чувствительность решения. Ожидаемые чистые доходы в "узлах" А и Б почти одинаковы: 5,419 и 5,375 млн. ф. ст. Выбор решения зависит от значения вероятностей. А анализ чувствительности позволяет нам вычислить "разброс" вероятностей, которые меняют наш выбор.

В данном случае мы рассматриваем только вероятность безопасности производственного процесса, однако, на математические ожидания повлияло бы также наличие и функционирование ККС. Полный анализ чувствительности включает рассмотрение обоих вопросов. Обозначим вероятность безопасности производственного процесса через р. На данный момент р = 0,55. Ожидаемый чистый доход в "узле" А равен:

$$10 \times p + 5{,}375 \times (1 - p) - 2{,}5 = 4{,}625p + 2{,}875 \text{ (млн. ф. ст.)}.$$

Ожидаемый чистый доход в "узле" В равен:

$$10 \times p + (10 \times 0{,}75 + 0 \times 0{,}25)(1 - p) - 3{,}5 = 2{,}5p + 4{,}0 \text{ (млн. ф. ст.)}.$$

Уравнивание этих результатов дает:

$$4{,}625p + 2{,}875 = 2{,}5p + 4{,}0;$$

$$2{,}125p = 1{,}125;$$

$$p = 0{,}529.$$

Следовательно, если вероятность безопасности производственного процесса равна 0,529, то оба альтернативных решения принесут одинаковый ожидаемый чистый доход. Если вероятность меньше 0,529, то решение начать разработку процесса и ККС незамедлительно принесет больший ожидаемый чистый доход, т.е. первоначальное решение будет заменено на альтернативное.

Так как значение p = 0,529 очень близко к p = 0,55, выбор решения очень чувствителен к расчетам величины вероятности, и малейшая ошибка может привести к смене выбора, что доказывает важность анализа чувствительности в процессе принятия решений.

РЕЗЮМЕ

Эта глава посвящена процессу принятия решений в условиях неопределенности. Решение проблемы связано с количественными оценками параметров, но, как и всегда на практике, немаловажную роль играет субъективный подход.

В процессе принятия решений необходимо придерживаться следующей последовательности:

1. Определить все возможные в данной ситуации варианты решения.
2. Для каждого из них определить его возможные исходы.
3. Для каждого решения и его исходов подсчитать возможный доход.

Наиболее удобная форма представления этих этапов — таблица доходов. Принимающий решение должен выбрать правило принятия решений, которое больше всего подходит к данной ситуации. Выбранное правило и определит выбор решения.

В некоторых правилах принятия решений значения вероятностей не используются. Например, правило максимакса (для игроков) и правило максимина (для осторожных) оперируют возможными доходами. А правило минимакса использует понятие возможных потерь (или упущенных доходов). В более сложных правилах принятия решений используется вероятность для подсчета ожидаемого дохода или ожидаемых возможных потерь для каждого решения. В таких ситуациях важно проверять чувствительность правил к изменениям вероятностей исходов. Наличие дополнительной информации может привести к выбору наилучшего решения. "Стоимость достоверной информации" — это максимальная сумма денег, которую стоит заплатить за дополнительную информацию.

Правила, рассматривавшиеся до сих пор, не учитывали разницу в приносимых ими доходах, а также отношение принимающего решение к риску. Для того чтобы показать, как различаются результаты разных решений, используется стандартное отклонение доходов. Отношение к риску выражается полезностью, которая присваивается доходам. Ожидаемая полезность основывается на этих оценках и их вероятностях. Оценки полезности в этих правилах заменяют собой доходы. Обычно на практике одно решение влечет за собой другие, а для иллюстрации этой последовательности используется "дерево" решений, которое состоит из "ветвей" и "узлов" — квадратов и кружков — решений и соответствующих им исходов. В той же последовательности, в которой таблица доходов иллюстрирует единственное решение, "дерево" показывает множество решений и их исходы. Численные значения доходов (исходы) просчитываются, начиная с конца "ветвей", постепенно приближаясь к исходному вопросу. Таким образом выявляется та последовательность решений, которой нужно придерживаться. Для выявления чувствительности решения по отношению к изменениям вероятностей используется анализ чувствительности.

УПРАЖНЕНИЯ

В упражнениях 3.1–3.4 предлагаются четыре ситуации. В каждой из них используйте какое-либо из следующих правил:

1) максимакса дохода;
2) максимина дохода;
3) минимакса возможных потерь;
4) максимина ожидаемого дохода;
5) минимума ожидаемых возможных потерь.

Также рассчитайте стоимость достоверной информации.

Упражнение 3.1

Пекарня печет хлеб на продажу магазинам. Себестоимость одной булки составляет 30 пенсов, ее продают за 40 пенсов. В таблице приведены данные о спросе за последние 50 дней:

Спрос в день, тыс. шт.	10	12	14	16	18
Число дней	5	10	15	15	5

Если булка испечена, но не продана, то убытки составят 20 пенсов за штуку. Используя каждое из правил, определите, сколько булок нужно выпекать в день.

Упражнение 3.2

Администрации театра нужно решить, сколько заказать пограммок для представлений. Стоимость заказа 200 ф. ст. плюс 30 пенсов за штуку. Программки продаются по 60 пенсов за штуку, и к тому же доход от рекламы составит дополнительные 300 ф. ст. Из прошлого опыта известна посещаемость театра:

Посещаемость	4000	4500	5000	5500	6000
Ее вероятность	0,1	0,3	0,3	0,2	0,1

Ожидается, что 40% зрителей купят программки.

1. Используя каждое из правил, определите, сколько программок должна заказать администрация театра.

2. Допустим, что рекламодатели увеличат сумму с 300 до 400 ф. ст., число посетителей будет больше 5250, к тому же спрос на программки будет полностью удовлетворен. Как это повлияет на рекомендации в п.1?
3. Чувствительность: допустим, вероятность каждого варианта посещаемости равна 0,2. Как это повлияет на решение, принимаемое по правилу максимизации ожидаемого дохода?

Упражнение 3.3

Компания "Kilroy" выпускает очень специфичный безалкогольный напиток, который упаковывается в 40-пинтовые бочки. Напиток готовится в течение недели, и каждый понедельник очередная партия готова к употреблению. Однако в одно из воскресений всю готовую к продаже партию пришлось выбросить. Секретный компонент, используемый для приготовления напитка, покупается в небольшой лаборатории, которая может производить каждую неделю в течение полугода (так налажено производство) только определенное количество этого компонента. Причем он должен быть использован в кратчайший срок.

Переменные затраты на производство одной пинты напитка составляют 70 пенсов, продается она за 1,50 ф. ст. Однако компания предвидит, что срыв поставок приведет к потере части покупателей в долгосрочной перспективе, а следовательно, придется снизить цену на 30 пенсов.

За последние 50 недель каких-либо явных тенденций в спросе выявлено не было:

Спрос на бочки в неделю	3	4	5	6	7
Число недель	5	10	15	10	10

1. Для того чтобы определить, что нужно предпринять, используйте каждое из правил.
2. Исследуйте чувствительность: изменит ли увеличение продажной цены до 1,75 ф. ст. какое-либо из решений?

Упражнение 3.4

Издатель обратился в отдел маркетинга, чтобы выяснить предполагаемый спрос на книгу. Исследования отдела маркетинга показали:

Спрос на книгу в ближайшие три года, количество экз.	2000	3000	4000	5000
Вероятность	0,1	0,5	0,2	0,2

Контрибуция к капитальным затратам и прибыли составляет 9 ф. ст. за книгу. Если книга не продается, убытки составляют 4 ф. ст. за штуку. Если издатель не удовлетворяет спрос, убытки по неудовлетворенному спросу составят 1 ф. ст. (для поддержания репутации фирмы и будущего спроса).

Используя по очереди каждое из правил, определите, сколько к[illegible]жно быть издано в расчете на трехлетний период.

Упражнение 3.5

Издатель не предполагал, что для решения поставленных задач нужно поинтересоваться мнением директора по маркетингу и финансового директора относительно полезности различных сумм дохода:

Доход, тыс. ф. ст.	0	10	20	30	40	50
Полезность с точки зрения:						
директора по маркетингу	0	10	20	35	55	100
финансового директора	0	40	70	85	95	100

1. Постройте два графика полезности и определите по ним отношение к риску обоих директоров.
2. По данным упражнения 3.4, используя поочередно графики полезности каждого директора, определите полезность доходов. Пересмотрите тираж, используя правило максимизации ожидаемой полезности. Что посоветовал бы каждый директор?

Упражнение 3.6

Издательство "Buranite Publishing Company Ltd" собирается выпустить новый учебник для бухгалтеров. Отдел маркетинга определил, что средний вероятный объем продаж составит 10000 экземпляров. С вероятностью 50% можно сказать, что объем продаж будет от 8000 до 12000 экземпляров.

Книга будет продаваться по 10 ф. ст. за штуку; отчисления автору составят 10% дохода, а капитальные затраты на издание и маркетинг книги составят 25000 ф. ст. При использовании имеющихся печатных станков переменные затраты составят 4 ф. ст. на одну книгу, но издательство может арендовать специальное оборудование за 14000 ф. ст., что снизит переменные затраты до 2,50 ф. ст. на одну книгу.

1) Покажите, что стандартное отклонение вероятного объема продаж σ приблизительно равно 3000;
2) Используя $\sigma = 3000$, определите вероятность того, что компания, по крайней мере, не потерпит убытков, если:

   (i) будет использоваться имеющееся полиграфическое оборудование;
   (ii) будет арендовано специальное оборудование.

3) Сравнивая ожидаемые прибыли, решите, нужно ли издательству арендовать специальное оборудование.
4) Используя нормальное распределение, можно показать, что следующее распределение может быть применимо к продаже книг:

Продажи, тыс. экз.	0 – 5	5 – 8	8 – 10	10 – 12	12 – 15	15 – 20
Вероятность	0,05	0,20	0,25	0,25	0,20	0,05

Учитывая, что фактические объемы продаж могут оказаться только в середине интервала каждой из групп, определите ожидаемую стоимость достоверной информации и объясните ее.

Упражнение 3.7

Небольшая химическая фирма "Hetros Hetrosone Ltd" выпускает дорогой промышленный растворитель "Hetrosone", который быстро портится. Поэтому запасы "Hetros one" нельзя держать больше, чем один месяц. Объемы выпуска продукции планируются в начале каждого месяца, и под эти планы закупается необходимое сырье. К сожалению, спрос на "Hetrosone" очень сильно колеблется от месяца к месяцу, и если спрос превышает запланированный выпуск, то происходит потеря доходов, так как спрос должен быть удовлетворен сразу же. Если же, наоборот, произведено больше, чем нужно в данном месяце, предложение избыточно, снижается стоимость. Продажная цена (p) "Hetrosone" – 2400 ф. ст. за 1 т, переменные производственные расходы (v) – 1500 ф. ст. за 1 т.

Анализируя спрос за последние несколько месяцев, менеджер по сбыту установил, что спрос колеблется между 10 и 20 т в месяц. Для того чтобы упростить анализ спроса, он подразделил его на три типа – "низкий" (10 т), "средний" (15 т) и "высокий" (20 т) с соответствующими вероятностями:

Спрос, т	*Вероятность*
10	0,3
15	0,6
20	0,1

1. Учитывая уровни спроса, составьте "дерево", охватывающее все возможности, открывающиеся перед компанией, а также их исходы.
2. Предположим, уровни спроса не изменяются. Какой объем производства вы бы могли посоветовать, чтобы максимизировать прибыль в долгосрочной перспективе?
3. Величина оптимального объема выпуска продукции Q выявляется при вероятности спроса, превышающего Q, равной v/p. Учитывая, что распределение спроса более нормальное, чем простое распределение, использовавшееся в п.1 и 2, подсчитайте среднее и среднеквадратическое отклонение месячного спроса и найдите оптимальный объем выпуска продукции Q.

Упражнение 3.8

В консалтинговую фирму "Cadmus, Carna & Clytie" обратился клиент с просьбой рассмотреть варианты инвестирования 100000 ф. ст. сроком на два года. В результате маркетингового исследования были предложены два варианта (A и B), один из которых содержит небольшой риск, в то время как другой совершенно без риска.

Вариант А предусматривает несколько возможных размеров прибыли: 8%, 10%, 12%, но из-за особенностей инвестируемого объекта существует некоторая корреляция доходов первого и второго годов, что отражено в таблице:

Первый год	*Второй год*		
	8%	*10%*	*12%*
8%	0,6	0,3	0,1
10%	0,2	0,5	0,3
12%	0,1	0,2	0,7

На этом этапе все три варианта прибылей первого года равновероятны. Вариант В предусматривает стабильную прибыль 9,5% годовых. При расчетах вы можете не принимать во внимание налогообложение; а прибыль первого года реинвестируйте на второй год.

1. Вне зависимости от того, какой из вариантов будет выбран, инвестирование производится на полные два года:
   а) Нарисуйте "дерево", учитывающее все варианты и их исходы.
   б) Какой бы вариант инвестиций вы рекомендовали, основываясь на ожидаемой сумме доходов?
   в) Какова вероятность того, что прибыль по варианту В будет больше, чем по варианту А?
2. Как изменится ваше "дерево", если станет возможным перейти от варианта А к варианту В в конце первого года? Какая стратегия инвестиций в данном случае принесет максимальную ожидаемую прибыль?

Упражнение 3.9

Компания "Brownhill Manufacturing Company" собирается производить новый товар, для чего нужно будет построить новый завод. После рассмотрения нескольких вариантов были оставлены три основных.

А. Построить завод стоимостью 600000 ф. ст. При этом варианте возможны: большой спрос с вероятностью 0,7 и низкий спрос с вероятностью 0,3. Если спрос будет большим, то ожидается годовой доход в размере 250000 ф. ст. в течение следующих пяти лет; если спрос низкий, то ежегодные убытки из-за больших капиталовложений составят 50000 ф. ст.

Б. Построить маленький завод стоимостью 350000 ф. ст. Здесь также возможны большой спрос с вероятностью 0,7 и низкий спрос с вероятностью 0,3. В случае большого спроса ежегодный доход в течение пяти лет составит 150000 ф. ст., при низком спросе — 25000 ф. ст.

В. Сразу завод не строить, а отложить решение этого вопроса на один год для сбора дополнительной информации, которая может быть позитивной или негативной с вероятностями 0,8 и 0,2 соответственно. Через год, если информация окажется позитивной, можно построить большой или маленький завод по указанным выше ценам. Руководство компании может решить вообще никакого завода не строить, если информация будет негативной. Вне зависимости от типа завода вероятности большого и низкого спроса меняются на 0,9 и 0,1 соответственно, если будет получена позитивная информация. Доходы на последующие четыре года остаются такими же, какими они были в вариантах А и Б.

Все расходы выражены в текущей стоимости и не должны дисконтироваться.

а) Нарисуйте "дерево", охватывающее все возможности, открывающиеся перед компанией.

б) Определите наиболее эффективную последовательность действий руководства фирмы, основываясь на ожидаемых доходах каждого варианта.

в) Предположим, строительная компания предлагает фирме скидку, если она сразу же приступит к строительству большого завода. Какова должна быть величина этой скидки (в процентах), чтобы фирма отказалась от ранее выбранного варианта?

Часть 2

АНАЛИЗ ДАННЫХ КАК СОСТАВНАЯ ЧАСТЬ ПРИНЯТИЯ РЕШЕНИЙ

Вторая часть книги посвящена применению статистики в принятии решений. Главы 4, 5 и 6 охватывают процедуру статистического вывода и испытания гипотез. В главе 7 рассматриваются методы статистического контроля качества. В главах 8 и 9 показывается, как линейная регрессия и анализ временных рядов могут быть использованы для получения прогнозных значений.

Глава 4. ВЫБОРКА И ВЫБОРОЧНЫЕ РАСПРЕДЕЛЕНИЯ

4.1. ВВЕДЕНИЕ

Обратимся теперь к важному статистическому методу — выборке.

Термин **совокупность** обозначает общую группу людей или предметов, с которыми связано статистическое исследование. Термин **выборка** обозначает некоторую группу, отобранную из совокупности. Мы будем использовать термин **параметр**, чтобы обозначить характеристику совокупности, и термин **статистика**, чтобы обозначить соответствующую характеристику выборки. Например, если мы будем рассматривать в качестве совокупности членов Ассоциации сертифицированных бухгалтеров Великобритании (АССА), то средняя заработанная плата действительного члена АССА будет являться параметром совокупности. Но если мы произведем выборку 100 членов, то их средняя заработанная плата будет являться выборочной статистикой. Целью статистического исследования может быть измерение параметра совокупности, например такого, как средний рост профессионального бухгалтера Великобритании или же размеры доходов рабочих ручного труда в сталелитейной промышленности Франции. Или же целью может быть проверка доверия к утверждениям относительно совокупности, например, доверия к утверждению, что насилие на телевизионном экране приводит к возрастанию насилия в обществе или к другим таким же суждениям.

Использование данных, полученных по выборкам, неизбежно приводит нас в область статистического вывода ввиду того, что возникает необходимость получить заключение относительно совокупности по данным выборки. Что говорит нам выборочная статистика относительно параметра совокупности или что позволяет нам заключить случайности выборки относительно истинной характеристики совокупности? Как средний рост, определенный по выборке профессиональных бухгалтеров, соотносится со средним ростом всех профессиональных бухгалтеров? Если в выборке взрослых преобладают бегающие трусцой, можем ли мы считать бег трусцой распространенным явлением во всей совокупности? По выборке, отобранной соответствующим образом, можно оценить параметры совокупности, можно использовать выборочные данные для испытания предположений относительно характеристик совокупности.

Статистический вывод является большой и важной областью статистики, в которой рассматривается вся информация, собираемая по выборке, для того, чтобы сделать заключение относительно некоторых характеристик совокупности.

Например, аудитор может проверить выборочные данные о сделках компании и, если отобранные документы удовлетворяют всем требованиям, он делает вывод, что документы о всех сделках компании правильно оформлены. Аудитор использует выборку для своего заключения потому, что она дешевле, ее можно быстрее провести и, следовательно, она более практична, нежели проверка всех документов о сделках компании.

Первая часть этой главы посвящена различным способам отбора. Этот вопрос очень важен, так как если отбор произведен неправильно, мы не можем использовать статистический вывод для получения надежного заключения о совокупности.

4.2. ПРИЧИНЫ ПРИМЕНЕНИЯ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ

При изучении мнения студентов относительно организации обслуживания в колледже в качестве объекта исследования выступают все студенты, обучающиеся в данном колледже. Если это обследование проводится для изучения обслуживания во всех колледжах, то в качестве объекта изучения выступают все студенты всех колледжей. Так как маловероятно, чтобы в обоих случаях те, кто проводит обследование, собирали бы данные по совокупности в целом, то должна быть отобрана небольшая группа студентов, которая будет представлять всю совокупность. Тогда можно будет сделать соответствующие выводы о совокупности по тем данным, которые будут получены на основе выборки. Очень важно в самом начале исследования определить совокупность, ее границы для гарантии того, что какие-либо сделанные выводы будут иметь значение.

Совокупности, такие, как в упомянутых выше примерах, ограничены по объему и потому называются **конечными** совокупностями. Совокупности, не ограниченные по объему, называются **бесконечными** совокупностями. На практике, если совокупность является большой, т.е. такой, что устранение какого-либо одного из ее членов не существенно изменяет вероятность отбора следующего члена, то такая совокупность рассматривается статистически как бесконечная. Например, машина производит гвозди, определенная часть которых является дефектными. Если выпуск гвоздей в день составляет 100000 шт. и мы произведем малую выборку из этого объема, то вероятность попадания в нее дефектного гвоздя практически не изменяется по мере отбора последующих членов выборки. Конечная совокупность, включающая 100000 единиц, может рассматриваться как бесконечная совокупность. На первый взгляд кажется, что лучше для статистического исследования охватить всю совокупность, если это возможно. Однако практически часто оказывается, что лучше использовать выборку.

Преимущества использования выборки, нежели совокупности, состоят в следующем.

1. **Практичность**. Совокупность может быть очень большой, практически бесконечной. Следовательно, может быть физически невозможно охватить наблюдением всю совокупность.

2. **Выигрыш во времени**. Если необходимо получить данные быстро, то может быть недостаточно времени, чтобы охватить всю совокупность. Например, если мы занимаемся проверкой качества товаров массового производства, то нельзя медлить с проверкой.

3. Затраты. Затраты на сбор данных по всей совокупности могут быть невыносимо высоки. В вышеприведенном случае проверка каждого изделия массового производства (например, каждого гвоздя) будет неоправданно дорогостоящей.

4. Ошибки. Если данные собираются по большой совокупности, то условием проведения наблюдения становится включение в работу большого числа людей, и риск совершения ошибок многократно возрастет.

5. Испытания, связанные с уничтожением. Наблюдение может включать уничтожение испытуемого образца. В таком случае очевидно, что нежелательно иметь дело со всей совокупностью. Например, производитель решил сделать проверку длительности работы определенного типа батареек. Он проводит испытание нескольких батареек, для того чтобы сделать вывод относительно всех батареек.

Преимущества использования совокупности по сравнению с выборкой состоят в следующем:

1. Малые совокупности. Если совокупность мала, так что любая взятая из нее выборка будет больше относительно объема совокупности, то в этом случае издержки и точность данных при охвате совокупности будут предпочтительнее, чем при выборке, для которой они не будут существенно отличаться.

2. Точность. Если существенно то, что информация, полученная в результате наблюдений, была бы абсолютно точной, то статистический вывод из выборочных данных может быть ненадежным. Например, для магазина необходимо точно знать, сколько денег было получено помимо кассы в течение года. Для владельца не имеет смысла регистрировать выручку за дни, выбранные в этом году. Проблема ошибок остается в этом случае, но эти ошибки скорее будут арифметическими, а не ошибками статистических оценок.

4.3. СЛУЧАЙНЫЙ ОТБОР

Чрезвычайно важно, чтобы единицы выборки отбирались так, чтобы выборка была бы репрезентативной и давала правильное представление о совокупности, насколько это позволяют имеющиеся деньги и время. Ошибочно сформированная выборка даст искаженное представление о совокупности.

Имеются несколько методов проведения отбора. Эти методы могут быть подразделены на две категории — случайная выборка и неслучайная выборка.

Статистический вывод может быть применен только в том случае, если выборка спланирована как случайная. Случайный отбор означает, что каждый член совокупности имеет равный шанс попасть в выборку. Если совокупность содержит группы единиц с отличающимися характеристиками, которые имеют значение для исследователя, то тогда случайный отбор должен обеспечить представительность в выборке единиц каждой из этих групп. Чем больше группа, тем более вероятно, что ее члены будут отобраны и, следовательно, будут представлены в выборке. Таким образом размер группы связан с вероятностью того, что ее члены попадут в выборку, поэтому случайный отбор приводит к выборке, которая представляет генеральную совокупность.

Первый шаг в производстве случайной выборки из конечной совокупности состоит в том, чтобы установить **основания проведения выборки**. Это — список всех членов совокупности. В нем может не указываться полное название ("имя") перечисляемых единиц, они должны быть помечены так, чтобы можно было их идентифицировать. Каждому члену совокупности может быть присвоен номер, после чего может быть использован какой-либо метод случайного отбора номеров, затем отобранные номера идентифицируются. Представительность выборки зависит от качества основания для формирования выборки. Важно, чтобы основания проведения выборки обладали следующими свойствами:

1. **Полнота** — в основание для проведения выборки должны быть включены все единицы совокупности. Неполнота может привести к дефектам выборки, особенно, если единицы, которые не попали в выборку, принадлежали к какой-либо особой группе единиц совокупности.

2. **Точность** — информация о каждой единице должна быть точной и не дублировать единицы.

Составление списка единиц совокупности не следует путать с переписью всех единиц. Цельь основания для проведения выборки состоит только в том, чтобы обеспечить возможность идентификации единиц. Требуемые данные затем будут получены только от отобранных единиц. Случайная выборка может быть произведена вручную, например, вытаскиванием шаров из сумки или по таблице случайных чисел. Другим методом может быть использование компьютера, который генерирует случайные числа. Последние затем используются для формирования выборки, которая называется простой случайной выборкой.

4.4. ВЫБОРОЧНЫЕ РАСПРЕДЕЛЕНИЯ

Одна из целей выборки состоит в том, чтобы получить оценки различных параметров, такие, как средняя, стандартное отклонение или доля определенных единиц. Для оценки неизвестных значений этих параметров используются выборочные статистики. Нам нужно знать, как статистики могут быть использованы наилучшим образом для оценивания и какова надежность оценок, полученных таким образом.

Предположим, что выборка, полученная как простая случайная выборка, произведена много раз. Хотя каждый раз производится случайный отбор, конкретные результаты различаются. Большое значение имеют вероятностные распределения выборочных статистики, которые мы используем. Эти распределения могут быть сконструированы в принципе так, как было описано выше.

Мы произвели выборку объема n единиц из совокупности объема N. Для этой выборки мы определим значение статистики, соответствующее тому параметру, в оценке которого мы заинтересованы.

Произведем вторую выборку того же размера и определим ту же статистику, что и прежде. Скорее всего численные значения статистики в двух выборках будут различными. Если мы затем произведем третью и четвертую выборки того же размера, то получим еще одни значения статистики. Однако если мы будем продолжать производить дальнейшие выборки, то некоторые значения статистики будут повторяться. Продолжая производить все возможные выборки размером n единиц из нашей совокупности, в результате мы сможем построить частотное распределение для полученных значений статистики. Соответствующее относительное частотное распределение является вероятностным распределением статистики. Это частотное распределение выборочной статистики называется **выборочным распределением**.

Например, если мы произведем все возможные выборки размером n = 5 единиц из совокупности 50 единиц, то будем иметь 2118760 различных выборок. (Это вычисляется как число комбинаций, потому что нас не интересует порядок, в котором производится отбор).

$$^5C_5 = \frac{50!}{5!\,45!} = 2118760\,.$$

Для каждой выборки в 5 единиц мы вычисляем выборочную среднюю и выборочное стандартное отклонение. Частотное распределение всех 2118760 выборочных средних представляет выборочное распределение средних выборок по 5 единиц из этой совокупности. Точно также получение всех выборочных стандартных отклонений дает выборочное распределение выборочных стандартных отклонений для выборок размером 5 единиц из совокупности. Если 50 единиц могут быть разделены, например, на хорошие и плохие, тогда можно вычислить долю хороших единиц в каждой выборке из 5 единиц и получить графики выборочного распределения выборочных долей хороших единиц. Каждое из этих распределений будет отличным. Формы будут зависеть от совокупности, размера выборки и от статистики, которую мы измеряем. Пример, приведенный ниже, иллюстрирует сказанное для совсем малой совокупности.

❐ **Пример 4.1**. Имеется конечная совокупность, состоящая из 6 чисел: 4, 8, 12, 16, 20, 24.

Построим выборочное распределение на основе выборок по два числа, вычисляя по каждой из них выборочную среднюю в качестве статистики, получим:

$$^6C_2 = 15 \text{ различных выборок.}$$

Решение

Таблица 4.1. **Выборочные средние, n=2**

Возможные выборки размера n = 2	*Выборочная средняя $\bar{x}$ $\bar{x} = (x_1 + x_2)/2$*	*Возможные выборки размера n = 2*	*Выборочная средняя $\bar{x}$ $\bar{x} = (x_1 + x_2)/2$*
4, 8	6	8, 24	16
4, 12	8	12, 16	14
4, 16	10	12, 20	16
4, 20	12	12, 24	18
4, 24	14	16, 20	18
8, 12	10	16, 24	20
8, 16	12	20, 24	22
8, 20	14		

Получаем выборочное распределение средних этих выборок:

Таблица 4.2. **Выборочное распределение**

Выборочная средняя, x размер выборки n = 2	*Частота, f*
6	1
8	1
10	2
12	2
14	3
16	2
18	2
20	1
22	1
Итого	15

Хотя выборочное распределение могло бы быть построено для любой статистики, ниже будет показано, что два из них наиболее полезны — выборочное распределение выборочных средних и выборочное распределение выборочных дисперсий.

Статистическая процедура, которая рассматривается в следующих главах, направлена на установление связей между выборочным распределением и генеральной совокупностью. В этой книге мы не пытаемся дать полное теоретическое обоснование этих соотношений, однако, выборка позволяет сделать те заключения, которые нам необходимо использовать. Как только соотношение между выборочным распределением и генеральной совокупностью установлено в целом, мы можем взять одну единственную выборку и использовать выведенное соотношение для того, чтобы сделать заключение о неизвестной генеральной совокупности, из которой эта выборка была взята. Во всех случаях мы предполагаем, что генеральная совокупность является нормальной или приблизительно нормальной и что требуемые выборочные статистики известны. Таким образом мы можем, например, определить, что генеральная средняя μ вероятно равна какому-то значению или то, что она лежит в определенных пределах, которые мы ожидаем с учетом ее дисперсии σ^2.

4.4.1. Выборочное распределение выборочных средних

Предположим, что мы хотим узнать среднее значение некоторой характеристики генеральной совокупности, например, средний рост десятилетних детей, среднюю заработную плату конторских работников в текстильной промышленности или средний диаметр производимых стальных шайб. Эти генеральные средние могут быть оценены по выборке. Если исходная, или генеральная совокупность нормально распределена, то выборочное распределение выборочных средних также будет иметь нормальное распределение. Даже для ненормального распределения генеральной совокупности, если выборка большая по размеру ($n \geq 30$), выборочное распределение выборочных средних будет иметь приблизительно нормальное распределение. Этот очень важный вывод основан на **центральной предельной теореме**.

Это позволяет нам использовать здесь все идеи о нормальном распределении, стандартизованных таблицах и z величинах, сформулированных в 2.7.

По выборочному распределению мы можем вычислить среднее значение всех выборочных средних. Оно представляет собой математическое ожидание выборочной средней:

$$E(\bar{x}) = \frac{\sum f\bar{x}}{\sum f}.$$

Если генеральная совокупность является нормальной, то математическое ожидание выборочной средней есть ни что иное как генеральная средняя μ, т.е.

$$E(\bar{x}) = \mu.$$

Это равенство справедливо только в том случае, если формирование выборки производилось случайно. В этом случае средняя, полученная по выборочному распределению, называется **несмещенной оценкой** генеральной средней (средней по совокупности). Пример **4.2** показывает, что это может быть верно и для ненормальной совокупности.

❐ **Пример 4.2**. Обратимся к примеру **4.1**. Выборочное распределение выборочных средних при n=2 следующее:

Таблица 4.3. **Вычисление E ($\bar{x}$)**

Выборочная средняя, $\bar{x}$	*Частота,* f	$\bar{x}f$
6	1	6
8	1	8
10	2	20
12	2	24
14	3	42
16	2	32
18	2	36
20	1	20
22	1	22
Итого	15	210

Математическое ожидание, полученное по выборочному распределению, составляет:

$$E(\bar{x}) = \frac{\sum f\bar{x}}{\sum f} = \frac{210}{15} = 14{,}0.$$

Данные совокупности: 4, 8, 12, 16, 20 и средняя по совокупности равна:

$$\mu = (4 + 8 + 12 + 16 + 20 + 24)/6 = 84/6 = 14{,}0.$$

Следовательно, в нашем примере $E(\bar{x}) = \mu$.

Однако, как уже упоминалось, на практике мы бы не стали действительно строить выборочное распределение на основе многократного проведения выборок из одной и той же совокупности. Следовательно, $E(\bar{x})$ не может быть вычислено таким образом. Обычно мы располагаем данными только по одной единственной выборке. Но ввиду того, что нам известно, что $E(\bar{x}) = \mu$, мы можем использовать единственную выборочную среднюю как несмешенную оценку генеральной средней:

$$\bar{x} = \hat{\mu},$$

где знак «ˆ» обозначает оцениваемую величину.

Надежность оценки будет детально обсуждаться ниже, но она может быть выражена через дисперсию выборочного распределения. Стандартное отклонение выборочного распределения представляет собой стандартную ошибку выборочного распределения, которое обозначается SE (Standard error). Стандартная ошибка выборочных средних обозначается как $SE_{(\bar{x})}$, где:

$$SE_{(\bar{x})} = \sqrt{\frac{\sum f(\bar{x} - E(\bar{x}))^2}{\sum f}} = \sqrt{\left(\frac{\sum f\bar{x}^2}{\sum f} - E(\bar{x})^2\right)}$$

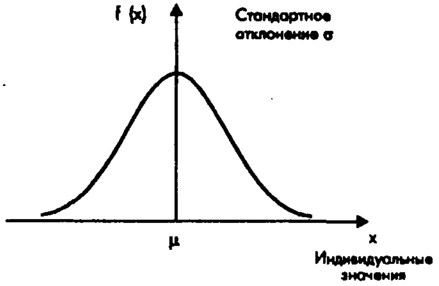


Рис. 4.1. Нормальная генеральная совокупность

Для нормально распределенной генеральной совокупности стандартная ошибка выборочного распределения выборочных средних определяется по формуле:

$$SE_{(\bar{x})} = \sqrt{\frac{(N-n)\,\sigma^2}{(N-1)\,n}},$$

где σ^2 — генеральная дисперсия.

Если генеральная совокупность велика по сравнению с размером выборки (обычно, если это соотношение $n/N \le 0{,}05$), то:

$$\sqrt{\frac{(N-n)}{(N-1)}} \approx 1,$$

и стандартная ошибка становится равной:

$$SE_{(\bar{x})} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\textit{генеральная дисперсия}}{\textit{размер выборки}}},$$

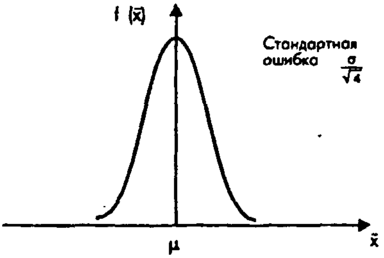


Рис. 4.2. Выборочное распределение средних $\bar{x}$ при размере выборок n = 4

Если мы будем изменять размер выборки, то увидим, что средняя выборочного распределения не изменяется, так что $E(\bar{x}) = \mu$, т.е. несмещенная оценка не зависит от размера выборки, тогда как $SE_{(\bar{x})}$ уменьшается при возрастании объема выборки (рис. 4.4).

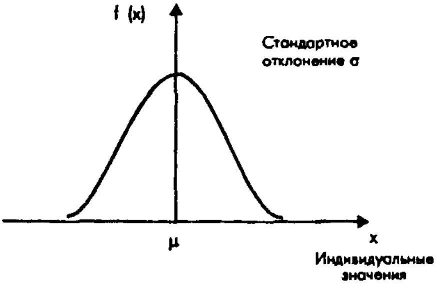


Рис. 4.3. Нормальная генеральная совокупность индивидуальных значений

При вычислении стандартной ошибки выборочных средних мы предполагаем, что нам известна σ^2 (т.е. что известна генеральная дисперсия). Фактически же ее величина неизвестна, и нам необходимо как-то получить оценку генеральной дисперсии по выборке.

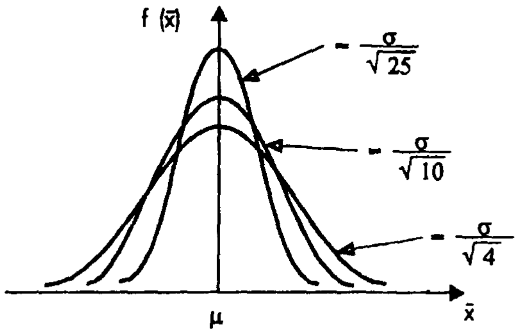


Рис. 4.4. Выборочные распределения средних $\bar{x}$ для выборок объема n = 4, 10, 25

4.4.2. Выборочное распределение выборочной дисперсии

Выборочное распределение выборочной дисперсии может быть рассмотрено с помощью метода, описанного выше, при этом дисперсия каждой выборки должна быть зафиксирована. Однако в отличие от выборочного распределения выборочных средних выборочное распределение выборочных дисперсий не является нормальным. Если генеральная совокупность является нормальной, то выборочное распределение выборочной дисперсии имеет распределение χ^2 (хи-квадрат).

КОНЕЧНАЯ ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

Если генеральная совокупность является конечной, то математическое ожидание выборочной дисперсии имеет вид:

$$E(s^2) = \frac{N}{(N-1)} \cdot \frac{(n-1)\sigma^2}{n}.$$

Следовательно,

$$\sigma^2 = \frac{(N-1)\,n}{N\,(n-1)} E(s^2).$$

Поэтому, если математическое ожидание выборочной дисперсии известно, то может быть определена генеральная дисперсия σ^2. $E(s^2)$ является несмещенной оценкой генеральной дисперсии. В этом случае, как и при оценке генеральной средней, наиболее вероятно, что мы будем располагать дисперсией только одной выборки и использовать ее для получения оценки генеральной дисперсии. Тогда:

$$\hat{\sigma}^2 = \frac{(N-1)}{N} \cdot \frac{n s^2}{(n-1)}, \quad \text{где } s^2 = \frac{\sum (\bar{x} - \bar{x})^2}{n}.$$

Это дает нам несмещенную оценку генеральной дисперсии.

❐ **Пример 4.3.** Обратимся к примеру **4.1**, где мы имели дело с конечной совокупностью, состоящей из чисел: 4, 8, 12, 16, 20, 24. Построим выборочное распределение, формируя выборки, объем которых равен двум наблюдениям, и рассчитаем дисперсию в качестве статистики для каждой выборки. Выборочная дисперсия:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n}.$$

Решение

Таблица 4.4. **Выборочные дисперсии, n = 2**

Реализованные выборки	*Выборочная дисперсия, s^2*	*Реализованные выборки*	*Выборочная дисперсия, s^2*
4, 8	4	8, 24	64
4, 12	16	12, 16	4
4, 16	36	12, 20	16
4, 20	64	12, 24	36
4, 24	100	16, 20	4
8, 12	4	16, 24	16
8, 16	16	20, 24	4
8, 20	36		

Выборочное распределение дисперсий этих выборок следующее:

Таблица 4.5. **Вычисление E (s^2)**

Выборочная дисперсия, s^2	*Частота, f*	$f s^2$
4	5	20
16	4	64
36	3	108
64	2	128
100	1	100
Итого	15	420

Математическое ожидание выборочной дисперсии составляет:

$$E(s^2) = \frac{\sum f s^2}{\sum f} = \frac{420}{15} = 28{,}0.$$

Отсюда

$$\sigma^2 = \frac{(N-1)}{N} \cdot \frac{n}{(n-1)} \cdot E(s^2) = \frac{(6-1)}{6} \cdot \frac{2}{(2-1)} \cdot 28{,}0 = 46{,}6.$$

Так как нам известны данные генеральной совокупности, то можем вычислить значение генеральной совокупности непосредственно. Нам известно, что генеральная средняя $\mu = 14$, следовательно:

Таблица 4.6. **Вычисление σ^2**

Значение переменной, x	$(x-\mu)^2$
4	100
8	36
12	4
16	4
20	36
24	100
Итого	280

Отсюда

$$\sigma^2 = \frac{\sum (x-\mu)^2}{n} = \frac{280}{6} = 46{,}6.$$

Это значение совпадает с тем, которое было получено по выборочному распределению выборочных дисперсий.

БОЛЬШАЯ (БЕСКОНЕЧНАЯ) ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

Если генеральная совокупность большая, то приблизительно:

$$\sqrt{\frac{N}{(N-1)}} = 1\ .$$

Тогда выражение несмещенной оценки генеральной дисперсии имеет вид:

$$E(s^2) = \frac{(n-1)\,\sigma^2}{n} \qquad \text{или} \qquad \sigma^2 = \frac{n}{(n-1)}\,E(s^2)$$

и несмещенная оценка генеральной дисперсии может быть выражена

$$\hat{\sigma}^2 = \frac{n}{(n-1)}\,s^2, \quad \text{где } s^2 = \frac{\sum (x-\bar{x})^2}{n}\ .$$

Следовательно, $\hat{\sigma}^2$ может быть записана как:

$$\hat{\sigma}^2 = \frac{n}{(n-1)}\,\frac{\sum (x-\bar{x})^2}{n} = \frac{\sum (x-\bar{x})^2}{(n-1)}\ .$$

Последнее выражение дает несмещенную оценку генеральной дисперсии. Вы, должно быть, заметили, что несмещенная оценка генеральной дисперсии обычно обозначается:

$$s^2 = \frac{\sum (x - \bar{x})^2}{(n - 1)}$$

(обозначение $\hat{\sigma}^2$ не используется).

Важно не путать оценку генеральной дисперсии с выборочной дисперсией:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n}.$$

Вы должны понимать, какую статистику вы рассчитываете. В этой книге мы будем обозначать несмещенную оценку генеральной дисперсии как $\hat{\sigma}^2$, чтобы избежать путаницы.

Если мы имеем дело с выборками из нормальной совокупности, то используем таблицы χ^2 – распределения, так как выборочное распределение ns^2/σ^2 подчиняется χ^2 – распределению с $(n - 1)$ степенями свободы (см. 4.4.4).

4.4.3. Оценка стандартной ошибки выборочного распределения выборочных средних

В 4.4.1 стандартная ошибка выборочного распределения выборочных средних при больших совокупностях определялась как:

$$SE_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}.$$

Чтобы получить стандартную ошибку из этого выражения, мы должны знать генеральную дисперсию σ^2. Если нам неизвестна ее величина, то можно использовать в качестве ее оценки выражение:

$$\hat{\sigma}^2 = \frac{n\,s^2}{(n - 1)}.$$

Следовательно, наилучшей оценкой стандартной ошибки является:

$$\hat{SE}_{\bar{x}} = \sqrt{\frac{(n\,s^2)/n}{(n - 1)}} = \sqrt{\frac{s^2}{(n - 1)}}, \text{ где } s^2 = \frac{\sum (x - \bar{x})^2}{n}$$

или

$$\hat{SE}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}^2}{n}} = \frac{\hat{\sigma}}{\sqrt{n}}, \text{ где } \hat{\sigma}^2 = \frac{\sum (x - \bar{x})^2}{(n - 1)}.$$

4.4.4. Стандартные выборочные распределения — z, t, χ^2, F

Имеются четыре стандартных распределения, к которым мы будем часто обращаться в последующих главах. Это нормальное распределение (z), t, χ^2 и F распределения. В этом разделе будут рассмотрены основные особенности каждого из этих распределений в связи с их использованием для проведения статистического вывода.

СТАНДАРТНОЕ НОРМАЛЬНОЕ (z) РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНЫХ СРЕДНИХ

В разделе **2.7** мы рассматривали нормальное распределение. Было показано, как можно преобразовать любое нормальное распределение в стандартное нормальное распределение, для которого среднее значение $\mu = 0$ и дисперсия $\sigma^2 = 1$. Значения переменной для такого стандартного нормального распределения обозначаются z и определяются как:

$$z = \frac{x - \mu}{\sigma}.$$

Рассчитанные таким образом значения z используются для нахождения требуемых вероятностей по таблице стандартного нормального распределения.

Выборочное распределение выборочных средних является нормальным распределением, если выборки получены как простые случайные из нормальной совокупности. Такое распределение описывается теми же характеристиками, что и любое нормальное наблюдение, только лишь следует иметь в виду, что z в этом случае:

$$z = \frac{\bar{x} - \mu}{SE_{\bar{x}}}$$

ввиду того, что выборочное распределение выборочных средних является распределением значений $\bar{x}$ (а не x), для которых средняя есть μ , а стандартное отклонение или стандартная ошибка обозначается как $SE_{\bar{x}}$. Значение z измеряется числом стандартных ошибок, которые отделяют выборочную среднюю от генеральной средней.

Ввиду того, что $SE_{\bar{x}} = \sigma/\sqrt{n}$ для больших совокупностей, z может быть выражено как:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

Чтобы использовать это равенство, мы должны знать генеральную дисперсию (σ^2). Если мы не знаем σ^2, то мы оцениваем ее, используя ее выборочную дисперсию:

$$\hat{\sigma}^2 = \frac{ns^2}{(n - 1)}.$$

Стандартизованная переменная в этом случае запишется как:

$$z = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}},$$

но ее распределение не всегда нормально. Стандартное нормальное (z) распределение может заменяться t-распределением.

t-РАСПРЕДЕЛЕНИЕ ДЛЯ ВЫБОРОЧНЫХ СРЕДНИХ

Если простая случайная выборка произведена из нормальной совокупности, дисперсия которой неизвестна, стандартное распределение выборочных средних — t-распределение, где:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{(n-1)}} = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}}.$$

t-распределение является симметричным относительно генеральной средней μ, но в отличие от нормального t-распределения форма t-распределения зависит от n, т.е. от объема выборки. Когда n мало, t-распределение является более пологим по сравнению с z-распределением. По мере того, как возрастает объем выборки, t-распределение приближается к стандартному нормальному распределению; следовательно, нормальное распределение можно использовать в качестве приближения t-распределения для выборки большого объема. Размер выборки считается большим, если $n \geq 30$.

Зависимость t-распределения от объема выборки не является однозначной. В действительности t-распределение варьирует с изменением числа **степеней свободы** для каждого конкретного случая. Например, если мы имеем дело со средней единственной выборки размером n, число степеней свободы будет равно $(n - 1)$, но если мы рассматриваем средние двух выборок, которые имеют размеры n_1 и n_2, то число степеней свободы будет равно $(n_1 + n_2 - 2)$.

Понятие числа степеней свободы может быть проиллюстрировано с помощью следующего простого примера. Если мы вычисляем среднюю из пяти чисел, то при этом мы свободны в выборе четырех из них, но значение пятого числа предопределено величиной данной средней. Например, если средняя из пяти чисел равна 6, мы можем выбрать 2, 7, 9 и 3 в качестве первых четырех чисел. Пятое число y является определенным, потому что средняя равна: $6 = (2 + 7 + 9 + 3 + y/5) = (21 + y/5)$, т.е. y должно быть равно 9. У нас нет свободы выбора последнего значения и поэтому мы имеем четыре степени свободы.

Если значения величины t рассчитаны, то могут использоваться стандартные выроятностные таблицы (t-таблицы), которые используются так же, как таблицы стандартного нормального распределения. Однако поскольку таблицы t-распределения должны содержать значения числа степеней свободы так же, как и различные значения t, то необходимо соединить всю необходимую информацию. Фактически эти таблицы организованы обычно таким образом, чтобы значения t были связаны с конкретными вероятностями для различных степеней свободы. (см. таблицу в Приложении 2). Использование таких таблиц будет более детально поясняться в гл. 5.

ВЫБОРОЧНОЕ РАСПРЕДЕЛЕНИЕ χ^2 ДЛЯ ВЫБОРОЧНОЙ ДИСПЕРСИИ

Те же самые предположения делаются для распределения χ^2, т.е. прежде всего то, что выборка произведена из нормальной совокупности. Статистика:

$$\chi^2 = n\frac{s^2}{\sigma^2}$$

подчиняется распределению χ^2 с $(n-1)$ степенями свободы. Как и t-распределение, форма этого распределения зависит от числа степеней свободы.

На рис. 4.5. приведены примеры отдельных распределений χ^2 при разном числе степеней свободы. Распределение не симметрично и изменяется по мере увеличения объема выборки.

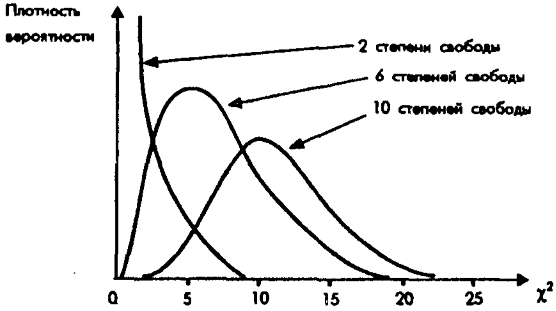


Рис. 4.5. Распределение χ^2 при различном числе степеней свободы

Значения χ^2 представлены так же, как и t, в специальных таблицах, где конкретные значения χ^2 даются для тех или иных степеней свободы.

ВЫБОРОЧНОЕ РАСПРЕДЕЛЕНИЕ F ДЛЯ ДВУХ ВЫБОРОЧНЫХ ДИСПЕРСИЙ

Если мы имеем две выборки, которые были отобраны случайно из нормальных совокупностей, то для того, чтобы сравнить две выборочные дисперсии, нам потребуется новое выборочное F-распределение. Статистика:

$$F = \frac{n_1 s_1^2}{(n_1-1)\sigma_1^2} \Big/ \frac{n_2 s_2^2}{(n_2-1)\sigma_2^2}$$

подчиняется F-распределению. Точно так же, как таблицы t-распределения, таблицы F показывают значения F, соответствующие вероятности. Таблицы содержат значения F статистики для комбинации числа степеней свободы в двух выборках. Более детально F-распределение рассматривается в гл. 6.

КАК БЫТЬ, ЕСЛИ СОВОКУПНОСТЬ НЕ ЯВЛЯЕТСЯ НОРМАЛЬНОЙ

Все стандартные распределения z, t, χ^2 и F предполагают, что выборка представляет собой случайную выборку из нормально распределенной генеральной совокупности. Мы можем обеспечить случайность отбора, но не можем контролировать нормальность распределения генеральной совокупности. Самый простой способ проверки приближения к нормальности состоит в использовании для этой цели выборочных данных. Если графическое изображение выглядит симметрично, то тогда можно предположить соответствие нормальному распределению. Существует и более формальные статистические тесты для проверки нормальности, но нет необходимости излагать их в этой книге. Если выборочное распределение очевидно асимметричное, то нужно быть осторожным с его данными. Имеются специальные подходы к работе с такими выборками.

Что касается средней величины, центральная предельная теорема позволяет нам пользоваться z-распределением, если размер выборки равен по крайней мере 30-ти единицам и более. Согласно центральной предельной теореме, если мы берем достаточно большую выборку из совокупности, независимо от ее распределения, со средней μ и стандартным отклонением σ, то распределение выборочных средних будет приблизительно нормальным. Чем больше размер выборки, тем ближе к нормальному будет это распределение. Общее правило таково, что выборка должна быть объемом 30 единиц и более. В этом случае можно не поднимать вопрос о нормальности распределения генеральной совокупности.

Альтернативный подход состоит в трансформации переменной: переменная, которая не имеет нормального распределения, может быть трансформирована каким-либо образом, например, путем перехода к логарифмам значений, и таким образом может быть обеспечено соответствие нормальному распределению. Третий путь состоит в том, чтобы использовать непараметрическую статистику, которая не требует предположения о нормальности.

РЕЗЮМЕ

Генеральная совокупность включает все единицы, которые составляют объект исследования. Выборка включает гораздо меньшее число единиц, отобранных из генеральной совокупности. Статистика вычисляется по данным выборки. На ее основе делается вывод относительно соответствующего генерального параметра.

Чтобы применить статистические методы анализа, выборка должна быть случайной. Это означает, что каждая единица должна иметь равный шанс попасть в выборку.

Существуют разные процедуры, обеспечивающие случайность отбора. Простой случайный отбор является основным. Члены генеральной совокупности нумеруются, тем самым создается основа для проведения отбора. Номера выбираются или по таблице случайных чисел или отбираемые номера генерируются случайным образом компьютером. Эти случайные числа используются для идентификации тех единиц, которые попали в выборку. Другие методы отбора используются, когда необходимо уменьшить размер выборки и вместе с тем гарантировать репрезентативность генеральных характеристик, или для того, чтобы упростить процедуру отбора.

Если мы произведем все возможные выборки объема n из нормально распределенной генеральной совокупности и вычислим выборочное статистики для каждой из них, то сможем получить выборочное распределение для этой статистики. Это позволяет увидеть, как выборочная статистика связана с генеральным параметром. Распределение всех выборочных средних, например, для выборки объемом n единиц называется выборочным распределением выборочных средних.

Для нормальной совокупности выборочное распределение выборочных средних (выборка объемом n) имеет математическое ожидание:

$$E(\bar{x}) = \mu \text{ — генеральную среднюю}$$

и стандартную ошибку:

$$SE_{\bar{x}} = \sqrt{\frac{(N-n)\,\sigma^2}{(N-1)\,n}},$$

где σ^2 — генеральная дисперсия в том случае, если объем выборки сравнительно большой по отношению к объему генеральной совокупности N.

Если генеральная совокупность бесконечна или $(n/N) \leq 0{,}05$, то:

$$SE_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}.$$

На основе обобщения выборочных распределений мы можем по одной выборке получить оценки генеральных параметров: $\bar{\mu} = \bar{x}$ и

$$\hat{\sigma} = \sqrt{\frac{n}{n-1}}\,s,$$

где

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \quad \text{или} \quad \hat{\sigma} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}.$$

В последующих главах используются четыре главных распределения. Все они основаны на предположении случайной выборки. Нормальное распределение z описывает распределение выборочных средних, если генеральная совокупность является нормальной:

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}.$$

Это распределение может использоваться и в том случае, если совокупность не является нормальной, но выборка большего размера $n \geq 30$ в соответствии с центральной предельной теоремой.

Если σ^2 не известна, но известно, что совокупность нормальна, стандартизованное распределение выборочных средних описывается t-распределением, где t равно:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{(n-1)}} = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}},$$

а t-распределение зависит от числа степеней свободы, которое определяется для конкретной задачи.

Распределение χ^2 представляет распределение выборочных дисперсий в том случае, если совокупность является нормальной:

$$\chi^2 = n\frac{s^2}{\sigma^2}$$

с (n-1) степенями свободы.

F—распределение описывает отношение двух выборочных дисперсий из двух нормально распределенных совокупностей:

$$F = \frac{n_1 s_1^2}{(n_1 - 1)\sigma_1^2} \Big/ \frac{n_2 s_2^2}{(n_2 - 1)\sigma_2^2}.$$

Эти распределения табулируются различными способами и будут рассмотрены в последующих главах.

УПРАЖНЕНИЯ

Упражнение 4.1

Являются ли данные следующих совокупностей конечными или бесконечными?

1. Заказы, выполненные прачечной за последнюю неделю.
2. Число сигналов опасности, которые были приняты пожарной бригадой.
3. Число голосов, зарегистрированных на выборах в городе.
4. Число заказов, которое могло быть выполнено с помощью инженерных работ.
5. Все компьютеры, которые могли бы быть произведены в Соединенном Королевстве.

Упражнение 4.2

Укажите, что будет являться объектом исследования:

1. При изучении отношения общества к курению.
2. При изучении часовой заработной платы рабочих в обувной промышленности.
3. При изучении счетов компании в процессе аудиторской проверки года, окончившегося в апреле 19... года.
4. При определении среднего срока службы электронного элемента.

Упражнение 4.3

Опишите, как могли быть получены основания для проведения выборки для следующих обследований:

1. Домохозяйств в Отауне, которые обслуживаются почтовым отделением.
2. Студентов, которые в настоящее время обучаются в Университете Битон.
3. Адвокатов в Цитауне.
4. Рынков сбыта для розничной торговли электрическими товарами в Дитоне.

Упражнение 4.4

Объясните, почему следующие методы отбора данных могли бы привести к ошибкам и нарушению случайности отбора в обследованиях, проводимых с целью:

1. Получить представление о взглядах общества на вивисекцию с помощью ответов на вопросник, напечатанный в Журнале о природе.
2. Получить реакцию общества на определенный метод налогообложения проведением опроса при посещении домов в течение дня.
3. Выяснить отношение общества к проблеме проживания в центральной части города по телефону, случайно выбирая абонентов.

Упражнение 4.5

В 4.4 этой книги идея выборочного распределения была проиллюстрирована на примере выборок по две единицы из небольшой совокупности, состоящей из 6 чисел: 4, 8, 12, 16, 20, 24. Продолжим этот пример следующим образом.

1. Произведите 5 выборок по 4 единицы из этой совокупности.
2. Для каждой выборки из 4 единиц вычислите среднюю и стандартное отклонение.
3. Представьте в виде частотной диаграммы выборочные средние для всех выборок объемом 2 единицы. Их значение возьмите 4.4.1. На тот же график нанесите распределение выборочных средних для выборок по 4 единицы. Сравните полученные диаграммы.
4. Постройте частотную диаграмму для стандартных отклонений для выборок объемом по 2 единицы. Данные приведены в 4.4.2. На том же графике представьте распределение выборочных стандартных отклонений для выборок по 4 единицы. Сравните полученные диаграммы.

Глава 5. СТАТИСТИЧЕСКИЙ ВЫВОД 1: ОЦЕНИВАНИЕ — ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

5.1. ВВЕДЕНИЕ

В главе 4 рассматривалось, как выборка может быть использована для оценки параметра генеральной совокупности. Мы также отметили, что в любой ситуации, в которой используется такая оценка, необходимо иметь представление о её надежности. Такую характеристику дает стандартная ошибка. Чем меньше стандартная ошибка, тем меньше дисперсность выборочного распределения, и, следовательно, менее изменчива выборочная статистика. Лучшим подходом к оценке будет установление интервала значений, в пределах которого, как мы можем быть уверены, лежит параметр генеральной совокупности. Этот предел значений называется **доверительным интервалом**. Доверительные интервалы могут быть установлены для любого параметра генеральной совокупности. Чаще всего они определяются для средних и относительных величин.

5.2. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ГЕНЕРАЛЬНОЙ СРЕДНЕЙ μ

5.2.1. Генеральная дисперсия σ^2 известна

Если исходная генеральная совокупность нормальная, то выборочное распределение выборочных средних также будет нормальным. Если генеральная совокупность имеет среднюю величину μ и стандартное отклонение σ , то выборочное распределение средних будет иметь среднюю величину, $E(\bar{x}) = \mu$ и стандартную ошибку $SE_{\bar{x}} = \sigma/\sqrt{n}$. Из центральной предельной теоремы известно, что данные утверждения справедливы для ненормальной генеральной совокупности, если объем выборки n не меньше 30.

Если мы отобрали n единиц из генеральной совокупности N и нашли среднюю величину по выборке $\bar{x}$, то $\bar{x}$ может быть использовано для оценки генеральной средней μ. Насколько надежна эта оценка?

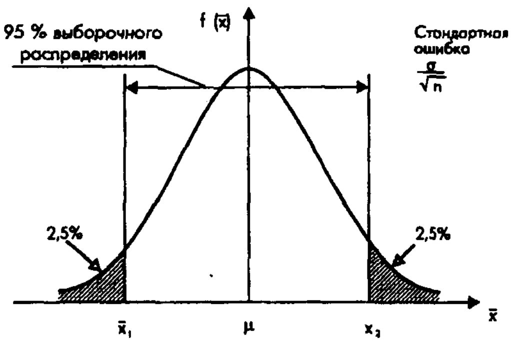


Рис. 5.1. Выборочное распределение выборочных средних $\bar{x}$ для выборок объема n, взятых из нормальной генеральной совокупности со средней величиной μ и дисперсией σ^2

На рис. 5.1 показано выборочное распределение выборочных средних. Величины $\bar{x}_1$ и $\bar{x}_2$ расположены симметрично относительно генеральной средний величины. Площадь под кривой, ограниченная этими пределами, включает 95% выборочного распределения. Область ниже изгиба кривой до величины $\bar{x}_1$ включает 2,5% и область ниже изгиба кривой после величины $\bar{x}_2$ – 2,5% распределения. Следовательно, мы можем сказать, что выборочная совокупность n единиц, взятая из нашей исходной генеральной совокупности, с вероятностью 95% будет иметь среднюю величину, лежащую между $\bar{x}_1$ и $\bar{x}_2$.

Средняя величина $\bar{x}_1$ со стандарной ошибкой $(\bar{x}_1 - \mu)/\ SE_{\bar{x}}$ находится на расстоянии ниже генеральной средней μ, а средняя величина $\bar{x}_2$ находится на таком же расстоянии выше μ.

Поскольку интервал между $\bar{x}_1$ и $\bar{x}_2$ включает 95% распределения, мы можем определить из таблицы нормального распределения, что $\bar{x}_2$ соответствует 1,96 стандартных ошибок выше средней μ и $\bar{x}_1$ соответствует 1,96 стандартных ошибок ниже средней μ:

$$\bar{x}_1 = \mu - 1{,}96 \cdot SE_{\bar{x}}\ ;$$

$$\bar{x}_2 = \mu + 1{,}96 \cdot SE_{\bar{x}}\ .$$

Поэтому расстояние от $\bar{x}_1$ до $\bar{x}_2$ может быть записано как:

$$\mu \pm 1{,}96 \cdot SE_{\bar{x}}\ .$$

Обычно мы проводим всего одну выборку из генеральной совокупности; рассчитываем среднее значение выборки $\bar{x}$ и используем его для вывода о среднем значении генеральной совокупности μ, из которой была взята выборка. На 95% мы уверены, что наше единственное значение $\bar{x}$ лежит между $\bar{x}_1$ и $\bar{x}_2$. Если $\bar{x}$ действительно попадает в интервал между $\bar{x}_1$ и $\bar{x}_2$, то тогда μ должно находиться где-нибудь в пределе:

$$\bar{x} \pm 1{,}96 \cdot SE_{\bar{x}}.$$

Мы можем сказать, что уверены в этом на 95%. Следовательно, $\bar{x} \pm 1{,}96 \cdot SE_{\bar{x}}$ является доверительным интервалом для среднего значения генеральной совокупности с вероятностью 95%. Если, например, $\bar{x}$ точно равно $\bar{x}_1$, то μ находится правее точки $\bar{x}_1$. Если $\bar{x}$ меньше $\bar{x}_1$, то μ не лежит в доверительном интервале. В этом случае мы выбрали *одну из* 5% выборочных совокупностей, *для которой* вывод, сделанный выше, неверен. Мы ограничились 95% выборочного распределения. Это был совершенно субъективный выбор. Может быть использован любой размер интервала и любая степень уверенности, что мы в нее попадем в зависимости от того, насколько мы хотим быть уверены, что среднее значение генеральной совокупности лежит внутри указанного интервала. Типичными являются 90%, 95% или 99%-ный доверительные интервалы. Какую бы величину мы не выбрали, построение доверительного интервала остается тем же. Единственная разница возникает в значении стандартизованной нормальной переменной z. Следовательно, общая формула доверительного интервала для генеральной средней имеет вид:

$$\bar{x} \pm z_{\alpha/2} \cdot SE_{\bar{x}},$$

где $z_{\alpha/2}$ — величина стандартизованной нормальной переменной, выше которой лежит $(\alpha/2)100\%$ значений. Это дает $(1-\alpha)100\%$ доверительный интервал.

Например, если нам требуется найти доверительный интервал с вероятностью 95%, то тогда $\alpha = 0{,}05$ и доверительный интервал может быть записан как:

$$\bar{x} \pm z_{0{,}025} \cdot SE_{\bar{x}}.$$

Если известно стандартное отклонение генеральной совокупности то, тогда стандартная ошибка распределения выборочных средних находится по формуле:

$$SE_{\bar{x}} = \sigma/\sqrt{n},$$

и $(1-\alpha)100\%$ доверительный интервал для генеральной средней может быть записан как:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

❐ **Пример 5.1.** Импортер упаковывает чай в пакеты по 125 г. Известно, что наполняющая машина работает со стандартным отклонением σ, равным 10 г. Выборка 50 пакетов n показала средний вес $\bar{x} = 128{,}5$ г.

Найти доверительный интервал для среднего веса σ в генеральной совокупности с вероятностью 95%. Предположим, что пакеты чая распределены по весу.

Решение.

Доверительный интервал с вероятностью 95% для среднего значения генеральной совокупности находится по формуле:

$$\bar{x} \pm 1{,}96\ \frac{\sigma}{\sqrt{n}},$$

где 1,96 является числом стандартных ошибок выше и ниже среднего значения для интервала, включающего 95% нормального распределения. Следовательно, доверительный интервал для генеральной средней находится как:

$$128{,}5 \pm 1{,}96 \times \frac{10}{\sqrt{50}} = 128{,}5 \pm 2{,}77 \text{ г.}$$

Мы на 95% уверены, что средний вес пачки чая в генеральной совокупности μ находится между двумя значениями: 125,73 г. и 131,27 г. Интервал в ± 2,77 г составляет примерно ± 2% среднего веса пачки чая в выборке, который равен 128,5 г. Это не очень большое отклонение для данного примера. Следовательно, среднее значение выборки может считаться надежной оценкой среднего значения генеральной совокупности. Однако необходимо помнить, что в 5% случаев мы можем ошибиться и получить значение вне доверительного интервала.

☐ **Пример 5.2.** Машина, которая упаковывает сахар, долгое время обеспечивала нормальное распределение веса в наполняемых пакетах. Стандартное отклонение веса σ равнялось ± 2,5 г. Был установлен новый размер упаковок. Для контроля была проведена случайная выборка 20 новых пакетов. Средний вес пакетов в выборке $\bar{x}$ = 1002 г. Предполагая, что переход на новую упаковку не повлиял на колебаемость наполняемости пакетов, найдем доверительный интервал для среднего веса упаковки в генеральной совокупности с вероятностью 99%.

Решение

Доверительный интервал для генеральной средней с вероятностью 99% находится следующим образом:

$$\bar{x} \pm 2{,}576\ \frac{\sigma}{\sqrt{n}},$$

где 2,576 = z — число стандартных ошибок выше и ниже среднего значения соответствует интервалу, который включает 99% нормального распределения. Тогда доверительный интервал с вероятностью 99% составит:

$$1002 \pm 2{,}576 \times \frac{2{,}5}{\sqrt{20}} = 1002 \pm 1{,}44 \text{ г.}$$

Мы на 99% уверены, что средний вес упаковки сахара μ генеральной совокупности находится в пределах от 1000,56 г до 1003,44 г. Размах в ±1,44 г составляет примерно ±0,1% среднего значения наполняемости в выборке (1002). 0,1% колеблемости в весе упаковки невелико, следовательно, среднее значение выборки может считаться точной оценкой среднего значения генеральной совокупности. Следовательно, мы можем ошибиться в одном из 100 случаев и μ может оказаться вне доверительного интервала.

5.2.2. Генеральная дисперсия неизвестна

Если дисперсия в генеральной совокупности неизвестна, то тогда стандартная ошибка выборочного распределения средних значений должна быть оценена следующим образом (см. гл. 4.):

$$\hat{SE}_{\bar{x}} = \frac{s}{\sqrt{n-1}} = \frac{\hat{\sigma}}{\sqrt{n}},$$

где

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$$

выборочное стандартное отклонение;

$$\hat{\sigma} = \sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}}$$

— несмещенная оценка стандартного отклонения генеральной совокупности.

Распределение соответствующей стандартизованной переменной далеко не нормальное. В этом случае взамен нормального распределения используется t-распределение Стьюдента. $(1 - \alpha)$ 100% доверительный интервал для среднего значения генеральной совокупности запишется как:

$$\bar{x} \pm t_{\alpha/2,(n-1)} \frac{s}{\sqrt{(n-1)}} \quad \text{или} \quad \bar{x} \pm t_{\alpha/2,(n-1)} \frac{\hat{\sigma}}{\sqrt{n}},$$

где $t_{\alpha/2,(n-1)}$ — стандартизованная t-переменная для (n-1) степеней свободы, выше которой лежит $(\alpha/2)$ 100% площади t-распределения.

В главе 4 отмечалось, что если объем выборки n, по крайней мере равняется 30, то t-распределение можно считать тождественным нормальному распределению.

❒ **Пример 5.3.** Случайная выборка $n = 25$ пакетов яблок показала, что средний вес пакета $\bar{x}$ равен 1020 г со стандартным отклонением $s = 12$ г. Найти доверительный интервал для среднего веса яблок генеральной совокупности с вероятностью 95%. Предполагается, что генеральная совокупность нормальная.

Решение

Поскольку стандартное отклонение генеральной совокупности неизвестно, то может быть оценена только стандартная ошибка выборки. t-распределение является подходящим выборочным распределением. Доверительный интервал с вероятностью 95% для среднего значения генеральной совокупности находится следующим образом:

$$\bar{x} \pm t_{0,025,\,24} \frac{s}{\sqrt{(n-1)}},$$

где $t_{0,025,\,24}$ — стандартная t-переменная, выше которой находится 2,5% площади t-распределения с (25 – 1)-степенями свободы. Используя таблицу t-распределения (см. приложение 2), находим, что:

$$t_{0,025,\,24} = 2{,}064.$$

Следовательно, доверительный интервал с вероятностью 95% составит:

$$1020 \pm 2{,}064 \times \frac{12}{\sqrt{24}} = 1020 \pm 5{,}06 \text{ г}.$$

На 95% мы уверены, что средний вес пакетов с яблоками в генеральной совокупности μ находится в пределах от 1015 г до 1025 г.

Отклонение в ±5,06 г составляет приблизительно ±0,5% среднего веса пакета в выборке, равного 1020 г. Полученное выборочное среднее значение может считаться надежной оценкой генеральной средней.

❐ **Пример 5.4.** Производитель автомобильных шин заинтересован в получении оценки средней износоустойчивости шин одной особой модели. Он провел случайную выборку объемом 10 шин и подверг их специальному испытанию. Средняя износоустойчивость по данным выборки оказалась равной 22500 миль со стандартным отклонением s = 3000 миль. Найдем доверительный интервал с вероятностью 99% для средней износоустойчивости всего выпуска шин этого типа. Как обычно предполагается, что генеральная совокупность нормальная.

Решение

Поскольку стандартное отклонение генеральной совокупности неизвестно, то возможна лишь оценка стандартной ошибки выборочного распределения. Принимаем, что t-распределение является подходящим для использования стандартного распределения. Доверительный интервал с вероятностью 99% для среднего значения генеральной совокупности находится следующим образом:

$$\bar{x} \pm t_{0,005,\,9} \frac{s}{\sqrt{(n-1)}},$$

где $t_{0,005,\,9}$ — стандартизованная t-переменная, выше которой находится 0,5%-ное t-распределение с (10 — 1)-степенями свободы.

Используя таблицу t-распределения (см. приложение 2), находим:

$$t_{0,005,\,9} = 3{,}25.$$

Следовательно, доверительный интервал с вероятностью 99% составит:

$$22500 \pm 3{,}25 \times \frac{3000}{\sqrt{9}} = 22500 \pm 3250 \text{ миль.}$$

Мы на 99% уверены, что средняя износоустойчивость этого типа шин в генеральной совокупности μ находится в пределах между 19250 миль и 27750 миль.

Отклонение в $\pm$ 3250 миль составляет приблизительно $\pm$ 14% выборочной средней износоустойчивости (22500 миль). Такое отклонение, по-видимому, является слишком большим, и, следовательно, выборочное среднее значение износоустойчивости может быть недостоверным для надежной оценки среднего значения генеральной совокупности.

5.3. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ГЕНЕРАЛЬНОЙ ДОЛИ (ОТНОСИТЕЛЬНОЙ ВЕЛИЧИНЫ) p

Нас обычно интересует оценка генеральной доли (пропорции) событий с некоторым характерным свойством. Например, аудиторов интересует доля ошибок в совокупности счетов. Из главы 2 мы знаем, что доля событий, в которых возникает характерное свойство, соответствует биномиальному вероятностному распределению. Однако если мы рассматриваем только большие выборочные совокупности и случаи, в которых доля особых событий ни маленькая, ни большая, то мы можем предположить, что выборочное распределение выборочной доли приблизительно нормальное. Это означает, что можно найти доверительные интервалы для генеральной доли, основываясь на нормальном распределении. Можно использовать те же положения для этой аппроксимации, что и в разделе 2.8, где обсуждалась нормальная аппроксимация биномиального распределения, при условии, что

$$np \ge 5 \quad \text{и} \quad n(1-p) \ge 5.$$

Обозначим генеральную долю событий p и $\hat{p}$-выборочную долю. Стандартная ошибка выборочного распределения доли особых событий в выборке находится по следующей формуле:

$$SE_{\bar{x}} = \sqrt{\frac{p(1-p)}{n}}.$$

Как было указано в разделе 2.3.3, это выражение является стандартным отклонением биномиального распределения. Поскольку генеральная доля p обычно неизвестна, стандартная ошибка будет оцениваться при использовании $\hat{p}$ как оценки p:

$$\hat{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Теперь мы можем найти $(1 - \alpha)$ 100%-ный доверительный интервал для генеральной доли:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

где $z_{\alpha/2}$ — величина стандартизованной нормальной переменной, выше которой лежит $(\alpha/2)$ 100% выборочного распределения.

❒ **Пример 5.5.** В ходе аудиторской проверки фирмы была проведена случайная выборка записей по счетам. Из выборки $n = 500$ записей 10 содержали некоторые ошибки в самой записи или процедуре. Требуется оценить с вероятностью 95% доверительный интервал для доли ошибок во всей генеральной совокупности записей.

Решение

Поскольку $np = 10$ и $n(1-p) = 490$ больше 5, то может быть использована нормальная аппроксимация для биномиального распределения. Так как доля ошибок генеральной совокупности p неизвестна, то стандартная ошибка выборочного распределения должна быть оценена с помощью выборочной доли p. Доверительный интервал с вероятностью 95% для генеральной доли может быть записан как:

$$\hat{p} \pm 1{,}96 \sqrt{\frac{\hat{p}\ (1-\hat{p})}{n}},$$

где 1,96 — величина стандартизованной переменной z, которая включает 95%-ное стандартное нормальное распределение

$$\hat{p} = 10/500 = 0{,}02.$$

Поэтому

$$SE_{\hat{p}} = \sqrt{0{,}02 \times 0{,}98/500} = 0{,}0063.$$

Следовательно, доверительный интервал с вероятностью 95% составит:

$$0{,}02 \pm 1{,}96 \times 0{,}0063 = 0{,}02 \pm 0{,}012.$$

На 95% мы уверены, что доля ошибок в записях на бухгалтерских счетах в генеральной совокупности p находится в пределах между 0,008 и 0,032.

Отклонение $\pm 0{,}012$ в величине доли составляет приблизительно 60% выборочной доли, равной 0,02. Это много и, следовательно, выборочная доля является ненадежной оценкой генеральной доли.

❒ **Пример 5.6.** Проведена случайная выборка личных заемных счетов в банке. Из $n = 1000$ отобранных счетов 60 оказались с задолженностью по возврату ссуды сроком до трех месяцев.

Найдем доверительный интервал с вероятностью 90% для доли счетов в генеральной совокупности p, которые имеют задолженность до трех месяцев.

Если банк насчитывает 30000 личных заемных счетов, каков доверительный интервал с вероятностью 90% для числа счетов, имеющих задолженность до трех месяцев?

Решение

Доверительные границы — это значения границ доверительного интервала. Так как $np = 60$ и $n\ (1 - p) = 940$ (больше 5), может быть использована нормальная аппроксимация. Поскольку соответствующая генеральная доля p неизвестна, то стандартная ошибка выборочного распределения должна быть оценена на основе выборочной доли $\hat{p}$. Выборочная доля интересующих нас счетов равна:

$$\hat{p} = 60/1000 = 0{,}06;$$

соответственно,

$$\hat{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{0{,}06 \times 0{,}94/1000} = 0{,}0075.$$

Доверительный интервал с вероятностью 90% для генеральной доли находится следующим образом:

$$\hat{p} \pm 1{,}645\, \hat{SE}_{\hat{p}},$$

где $\pm 1{,}645$ — величина стандартизованной нормальной переменной z, которая включает 90%-ное распределение. Следовательно,

$$0{,}06 \pm 1{,}645 \times 0{,}0075 = 0{,}06 \pm 0{,}012.$$

Пределы доверительного интервала (доверительные границы) составили:

$$0{,}06 + 0{,}012 = 0{,}075;$$

$$0{,}06 - 0{,}012 = 0{,}048.$$

Если общее числа счетов равняется 30000, то доверительный интервал с вероятностью 90% для числа интересующих нас счетов:

$$30000\,(0{,}06 \pm 0{,}012) = 1800 \pm 360.$$

Следовательно, на 90% мы уверены, что число счетов, имеющих задолженность до трех месяцев, лежит в пределах между 1440 и 2160.

5.4. ОПРЕДЕЛЕНИЕ СООТВЕТСТВУЮЩЕГО ОБЪЕМА ВЫБОРКИ

При рассмотрении предыдущих вопросов в этой главе мы предполагали, что размер выборки задан. Формула доверительного интервала может быть использована для определения подходящей величины выборки. Но сначала мы должны принять несколько решений, например, какой уровень доверительной вероятности должен быть использован: 90%, 95 или 99%? Нам также необходимо знать, выбрав доверительную вероятность и уровень значимости, какую ширину доверительного интервала мы хотим принять: ± 2 г — для средней, ± 3% для доли или 5 мм для стандартного отклонения, например.

Если σ или p нам уже известны, то тогда можно рассчитать требуемую величину объема выборки. Однако обычно этого не бывает и мы должны оценить

σ или p. Для этого мы либо указываем их величины, используя некоторые прежние знания, либо проводим пробную выборку и используем ее для оценки σ или p. В следующих разделах будем использовать выборочные данные прошлых примеров главы, рассматривая их как предварительные выборки.

5.4.1. Объем выборки, необходимый для оценки генеральной средней

❐ **Пример 5.7.** В примере **5.1** по выборке n = 50 упаковок чая был определен средний вес $\bar{x}$ = 128,5 со стандартным отклонением в генеральной совокупности σ = 10 г. Доверительный интервал при 95%-ной вероятности: 128,5 + 2,77 г. Найдем объем выборки n с вероятностью·95% при доверительном интервале, равном ± 2,0 г для генеральной средней.

Решение

Размах доверительного интервала:

$$\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Для решения задачи, в которой доверительная вероятность равна 95%, требуется, чтобы:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le 2 \text{ г}.$$

Следовательно,

$$1{,}96 \times \frac{10}{\sqrt{n}} \le 2 \text{ г}.$$

Отсюда

$$\sqrt{n} \ge \frac{1{,}96 \times 10}{2} = 9{,}8 ,$$

тогда n ≥ 96.

Объем выборки n должен быть увеличен с 50 до 96 упаковок чая для того, чтобы сузить доверительный интервал при постоянной вероятности 95% от ± 2,77 г до ± 2,0 г.

❐ **Пример 5.8.** В примере **5.3** мы имели n = 25 пакетов яблок со средним значением веса пакета в выборке $\bar{x}$ = 1020 г и стандартным отклонением s = 12 г.

Доверительный интервал с вероятностью 95% равнялся ± 5,06. Каким должен быть объем выборки, если мы хотим определить доверительный интервал среднего веса пакета с вероятностью 99% и отклонением, равным ± 5,0 г.

Решение

Размах доверительного интервала:

$$\bar{x} \pm t_{\alpha/2,\,(n-1)}\frac{s}{\sqrt{(n-1)}}$$

должен быть меньше, чем ± 5 г.

Мы используем стандартное отклонение предварительной (или пробной) выборки, равное 12 г для оценки s, т.е. стандартное отклонение планируемой выборки.

К сожалению, величина t, которую мы хотим использовать, $t_{0,005,\,(n-1)}$, зависит от объема выборки, и мы не можем знать n, пока не рассчитаем его величину. Мы игнорируем эту проблему на данный момент, просто отметив, что наша оценка n вероятно является большой, и мы используем $t_{0,005,\,(n-1)} = 2,797$. Сделав подстановку, имеем:

$$\frac{2,797 \times 12}{\sqrt{(n-1)}} \le 5,0 .$$

Отсюда

$$\sqrt{(n-1} \ge (12,797 \times 12)/5 = 6,71.$$

Следовательно,

$$n - 1 \ge 45,1,$$

тогда: $n \ge 46,1$.

Оцененный объем выборки будет по крайней мере равняться 47 с вероятностью 99% для величины доверительного интервала ± 5 г.. Если мы хотим быть более точными, то можем теперь пересчитать n с учетом числа степеней свободы, или, поскольку мы теперь знаем, что размер выборки, вероятно, больше 30, можем использовать величину z вместо величины t. Заменив t величиной $z_{0,005} = 2,576$, получим:

$$\frac{2,576 \times 12}{\sqrt{(n-1)}} \le 5 .$$

Отсюда

$$n - 1 \ge 38,2.$$

Следовательно, $n \ge 39,2$.

Итак, требуемый объем выборки после корректировки равен 40 единицам и 47 единицам — без корректировки.

5.4.2. Объем выборки, необходимый для оценки генеральной доли

☐ **Пример 5.9.** Используя данные примера 5.5, в котором аудитор провел выборку 500 записей для оценки доли ошибок, определим,какую выборку проведет аудитор, если он хочет быть в пределах 0,005 генеральной доли с доверительной вероятностью 95%. Доля ошибок в первоначальной выборке была равна 0,02 с доверительным интервалом +0,012 с вероятностью 95%.

Решение

Размах доверительного интервала для доли:

$$\pm z_{\alpha/2} \cdot SE_{\hat{p}} .$$

Его величина может быть найдена по следующей формуле:

$$\pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} .$$

Аудитору требуется, чтобы размах доверительного интервала с вероятностью 95% для доли максимально достигал ± 0,005. Поэтому необходимо, чтобы:

$$1,96 \sqrt{\frac{\hat{p}\ (1-\hat{p})}{n}} \leq 0,005 .$$

Используя данные предыдущей выборки, мы можем установить, что $\hat{p} = 0,02$ и решить неравенство для n:

$$1,96 \times \sqrt{\frac{0,02 \times 0,98}{n}} \leq 0,005 .$$

Отсюда

$$\frac{1,96^2 \times 0,02 \times 0,98}{n} \leq 0,005^2 .$$

Следовательно,

$$\frac{1,96^2 \times 0,02 \times 0,98}{0,005^2} \leq n ,$$

тогда $n \geq 3012$.

Если аудитор хочет объяснить пределы генеральной доли 0,005 с доверительной вероятностью 95%, объем его выборки должен быть не меньше 3012.

Если мы не учитываем данные первоначальной выборки, то должны предположить разумную величину выборочной доли $\hat{p}$. Если оценить ее равной 0,03, то тогда требуемый размер выборки изменится: n = 4500. Поскольку наблюдаемая доля ошибок очень мала, то маленькое абсолютное изменение от 0,02 до 0,03 приводит к большой разнице в требуемом объеме выборки.

РЕЗЮМЕ

Информация, полученная по простой случайной выборке, произведенной из нормальной генеральной совокупности, используется для вывода о среднем значении или доли исходной генеральной совокупности. Доверительный интервал с вероятностью 95% позволяет нам быть на 95% уверенными, что параметр генеральной совокупности лежит в пределах доверительного интервала, соответственно. В 5% случаев мы можем ошибаться, и параметр лежит вне интервала. $(1 - \alpha)$ 100% доверительные интервалы рассчитываются по следующей формуле (используем стандартное z и t-выборочные распределения, приведенные в таблицах Приложения 2):

1. Доверительный интервал для генеральной средней μ при известной σ^2:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} .$$

2. Доверительный интервал для генеральной средней μ при неизвестной σ^2:

$$\bar{x} \pm t_{\alpha/2,(n-1)} \frac{s}{\sqrt{(n-1)}} \quad \text{или} \quad \bar{x} \pm t_{\alpha/2,(n-1)} \frac{\hat{\sigma}}{\sqrt{n}}.$$

В соответствии с центральной предельной теоремой эта формула также применима в случае ненормальной генеральной совокупности при объеме выборки $n \geq 30$.

3. Доверительный интервал для генеральной доли p, когда и $np \geq 5$ и $n(1-p) \geq 5$:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

На основе этих формул может быть рассчитан объем выборки n, требуемый для построения доверительного интервала определенной ширины; также может быть найден уровень значимости. Решение этих задач предполагает, что σ^2 известна или мы можем оценить её с помощью s или p из первоначальной выборки.

УПРАЖНЕНИЯ

Упражнение 5.1

Машина распределяет жидкий шоколад в формы для получения шоколадных плиток. В течение длительного периода времени вес шоколада в формах соответствовал нормальному распределению со стандартным отклонением 2,5 г. В порядке качественного контроля была проведена случайная выборка 15 плиток из готовой продукции и произведено их взвешивание. Средний вес плитки в выборке оказался равным 99,5 г. Найдите доверительный интервал с вероятностью 95% для истинного среднего значения веса шоколадной массы, распределенной в формы.

Упражнение 5.2

Определенный компонент в цепи транзистора имеет срок службы, который придерживается приблизительно нормального распределения. Случайная выборка 50 компонентов из недельного выпуска показала, что средний срок службы равен 840 ч со стандартным отклонением 22 ч. Найдите доверительный интервал с вероятностью 99% для среднего срока службы генеральной совокупности элементов.

Упражнение 5.3

Управляющий по качественному контролю в фирме, выпускающей электрические лампочки, хочет узнать средний срок службы (время горения) отдельного вида лампочек. С этой целью случайная выборка 75 лампочек была испытана на продолжительность горения. Средняя выборочная равняется 2920 ч со стандартным отклонением 400 ч.

Определите:

1. Репрезентативна ли выборка (дает ли среднее значение выборки хорошую оценку среднего значения генеральной совокупности)?

2. Что нужно сделать управляющему для оценки среднего срока службы лампочки генеральной совокупности в пределах ± 50 ч при доверительной вероятности 95%?

Упражнение 5.4

Случайная выборка 800 домохозяек в центре города, проведенная утром, показала, что 480 из них хотели бы, чтобы торговый центр города был свободен от транспорта.

Определите доверительные пределы с вероятностью 90% для доли всех домохозяек в городе, кто хотел бы, чтобы торговый центр был свободен от транспорта.

Упражнение 5.5

Мисс Сэлли Бригс работает менеджером по продаже кондитерских изделий. При изучении случайной выборки 200 выпусков в Уэльсе она обнаружила, что для 50 из них желательно изменение ассортимента продукции.

Требуется:

1. Найти 95%-ный доверительный интервал для доли потребителей в Уэльсе, которые будут брать новый вид продукции.
2. Мисс Бригс наметила провести такое же обследование в Шотландии и решила достичь оценки доли потребителей новой продукции в пределах ± 4%. Насколько большой должна быть выборка в Шотландии? Предполагается, что она определяет доверительный интервал с вероятностью 95%.

Упражнение 5.6

Компания "Jones" торгует в розницу модной одеждой для мужчин и женщин через сеть 20-ти фешенебельных магазинов в Англии и Уэлсе.

"Jones" собирается выпустить свои собственные кредитные карточки. Как часть осуществляемых предварительных работ компания провела опрос своих покупателей в магазине на Оксфорд-стрит в Лондоне в порядке случайной выборки. Из выборки 80 покупателей 32 покупателя изъявили желание приобрести такую кредитную карточку. Эти 32 чел. были затем снова опрошены, чтобы определить, сколько бы они приобрели, используя кредитную карточку. По 32 ответам средняя сумма составила 450 ф. ст. со стандартным отклонением 150 ф. ст.

Требуется:

1. Найти 95%-ный доверительный интервал для доли всех покупателей, кто бы хотел приобрести кредитную карточку.
2. Найти 95%-ный доверительный интервал для средней суммы возможных приобретений по кредитной карточке.

   После такого локального опроса компания намеревалась провести его вторично по всей Англии и Уэльсу.
3. Какого размера должна быть выборка, если компания хочет оценить с ошибкой не более 5% долю покупателей, которые приобрели бы кредитную карточку, и оценить среднюю сумму кредита в пределах ±20 ф. ст. (в обоих случаях принимается 95%-ный доверительный интервал).
4. Укажите причины того, что результаты всеобщего опроса будут отличаться от результатов опроса в магазине на Оксфорд-стрит.

Упражнение 5.7

Производитель электрических и других комплектующих изделий компании "Tauras Ltd" в течение определенного периода исследовал стоимость сборки одного из ведущих видов продукции. По данным прошлого опыта, было установлено, что среднее время сборки этого изделия составляет 90 мин. Однако главный инженер утверждал, что процедура сборки может быть улучшена. Он изобрел новый метод сборки, назначил случайно отобранных рабочих для проведения испытания. Продолжительность сборки 10 изделий новым методом составила (мин):

79 74 112 95 83 96 77 84 70 90

1. Рассчитайте несмещенную оценку средней и стандартного отклонения для времени сборки продукции при использовании нового метода.
2. Ободренный результатами главный инженер зафиксировал время сборки следующих 30 изделий. Несмещенная оценка средней и стандартного отклонения для этих 30 наблюдений составила 84,0 и 14,65 соответственно. Объединяя первую и вторую выборки, по 40 изделиям найдите несмещенную оценку генеральной средней и стандартного отклонения и определите доверительный интервал с вероятностью 95% для среднего времени сборки.
3. Рассчитайте наименьший общий размер выборки, необходимый для получения среднего значения генеральной совокупности в пределах трех минут с доверительной вероятностью 95%.
4. Прямая оплата труда, включаемая в затраты, составляет 8,70 ф. ст. в час. Если стоимость сборки оценивается в 20000 ф. ст., то определите на основе данных выборки сумму экономии компании за счет использования нового метода сборки при условии, что компания будет продавать 50000 изделий этого вида.

Упражнение 5.8

Компания "Martin Electronic Controls Ltd" является мелким производителем электронных компонентов и специализируется на промышленном контрольном оборудовании. Используя внешний аудит, главный бухгалтер компании решил предпринять выборочную проверку и выбрал 18 из 1200 компонентов продававшихся в прошлом месяце для того, чтобы иметь представление об общем объеме реализаций. Стоимость отобранных компонентов следующая (ф. ст.):

82	30	98	116	80	150	200	88	70
90	160	100	86	76	90	140	76	68

(Заметим, что $\sum x = 1800$ и $\sum x^2 = 207200$).

1. На основе выборки n найдите оценку общей величины стоимости всех компонентов.
2. Найдите пределы доверительного интервала с вероятностью 95% для общей стоимости всех компонентов.
3. Сколько компонентов должен исследовать внешний аудитор, если он захочет, чтобы его оценка общей стоимости 1200 компонентов была в пределах ± 5000 ф. ст.

Глава 6. СТАТИСТИЧЕСКИЙ ВЫВОД 2: ИСПЫТАНИЕ ГИПОТЕЗ

6.1. ВВЕДЕНИЕ

В соответствии с правилами фасовочная машина должна быть установлена так, чтобы при расфасовке каждый пакет в среднем имел определенный вес. Производитель понимает, что с момента установки машины, средняя наполняемость пакета вряд ли останется постоянной — вполне вероятны изменения в весе расфасовок. Отклонения от нормы будут возрастать по мере износа машины. Для сохранения контроля над процессом расфасовки может быть взята выборочная совокупность фасовки и определена средняя наполняемость. Как мы уже видели в гл. 5, одним из направлений анализа является использование выборочной средней для построения доверительного интервала для генеральной средней наполняемости.

Можно также использовать испытание гипотез для оценки того, удовлетворяет ли производителя работа машины. В этом случае производитель придерживается мнения или гипотезы о среднем достаточном количестве. Тогда он использует среднюю выборки как основание для подтверждения или не подтверждения его мнения. На основании выборки производитель решает, насколько это мнение или гипотеза верна.

Если кость была брошена 102 раза, то в среднем 17 результатов будут "шестерками" (102:6=17). Предположим, что мы бросили кость 102 раза и выпало 20 "шестерок". Можем ли мы автоматически предположить, что на таком выпадении "шестерок" сказалось влияние какого-то фактора?

Поскольку при бросании кости результат является случайным, 17 "шестерок" не будут получены в каждом испытании из 102 раз. Поэтому вопрос состоит в том, сколько "шестерок" мы можем получить, чтобы с достаточной уверенностью сказать, что этот результат не случаен. Ясно, что если все 102 выбрасывания дали "шестерки", мы могли бы быть уверены, что этот результат не случаен. Где можно провести черту между результатом, свойственным случайной природе эксперимента с правильной костью, и результатом не случайным, свойственным неправильной кости? Соответствующее испытание гипотез или, как часто называют, испытание значимости, даст нам возможность прийти к решению такого рода проблем. Такие испытания проводятся на основе определенных правил и базируются на суждении лица, принимающего решения. Мы никогда не бываем абсолютно уверены в подтверждении правильности выдвинутой гипотезы, мы просто делаем оценку, насколько вероятно то, что гипотеза верна.

6.2. ПРОЦЕДУРА ИСПЫТАНИЯ ГИПОТЕЗ

Для оценки доказательств выборки мы должны формулировать наши гипотезы так, чтобы можно было использовать известное вероятностное распределение. Такая исходная гипотеза называется нулевой гипотезой и обозначается H_0. Нулевая гипотеза всегда формулируется для утверждения того, что выборочная статистика согласуется с принятым параметром генеральной совокупности. Сформулировав нулевую гипотезу, мы исследуем выборку для того, чтобы увидеть, согласуется ли она с этой гипотезой. Заметим, что для обеспечения как можно большей объективности важно, чтобы гипотеза формулировалась до того, как собираются данные. Весь спектр возможных результатов, обычно подразделяется на три категории:

1) доказательство согласуется с нулевой гипотезой;
2) доказательство не согласуется с нулевой гипотезой;
3) доказательство является неубедительным, поэтому требуется больше данных для принятия решения.

Если результат соответствует категории 1, то решением будет принятие нулевой гипотезы как наиболее верной. Предполагается, что различие между величиной выборочной статистики и параметром генеральной совокупности объясняется случайной вариацией, свойственной выборочному исследованию.

Если результат соответствует категории 2, то решением будет отклонение нулевой гипотезы, как, вероятно, неверной. Предполагается, что различие между выборочной статистикой и параметром генеральной совокупности не объясняется случайной выборочной вариацией. В этом случае принято применять альтернативную гипотезу. Испытание гипотез не включает доказательство согласования выборки с альтеративной гипотезой. При применении альтернативной гипотезы, мы можем предположить, поскольку нулевая гипотеза оказалась неприемлемой, что взамен нулевой гипотезы следует использовать альтернативную гипотезу. Мы не приводим какое-либо статистическое доказательство правильности нашего предположения.

Альтернативная гипотеза обычно обозначается H_1, как и H_0, она должна быть сформулирована в самом начале исследования.

Например, машина изготавливает металлические диски. Она установлена так, что средний диаметр дисков равен 2,0 см. Выборка из партии дисков показала средний диаметр, равный 2,3 см.

Вопрос состоит в том, правильно ли все еще настроена машина?

Нулевая гипотеза предполагает, что машина настроена все еще правильно и выборочная средняя согласуется с выборкой, взятой из нормальной генеральной совокупности со средним значением, равным 2,0 см. Если при испытании гипотезы мы обнаруживаем, что данные выборки не согласуются с нулевой гипотезой, то тогда мы должны решить, какое примем альтернативное заключение. Альтернативной гипотезой может быть просто предположение, что среднее значение генеральной совокупности не равняется 2,0 см, а также альтернативной гипотезой может быть предположение, что генеральная средняя больше, чем 2,0 см. Альтеранативная гипотеза определяет точные условия испытания нулевой гипотезы. Отмеченные две формулировки H_1 можно записать следующим образом:

Случай 1:

$$H_0: \mu = 2{,}0 \text{ см},$$

$$H_1: \mu \neq 2{,}0 \text{ см}.$$

Случай 2:

$$H_0: \mu = 2{,}0 \text{ см},$$

$$H_1: \mu > 2{,}0 \text{ см}.$$

Если результат относится к категории 3, то никакое решение не может быть принято до тех пор, пока не будет получено больше данных, и испытание гипотезы будет проведено вновь. Однако следует отметить, что разграничение этих трех категорий проводится на субъективной основе лицом, принимающим решение. Какой бы результат ни был получен, мы никогда не можем определенно (на 100%) доказать или опровергнуть нулевую гипотезу. Все, что мы можем сделать,— это или признать то, что нулевая гипотеза почти наверняка верна, или, что правильность нулевой гипотезы маловероятна.

Следующий пример поможет прояснить, почему это так.

❐ **Пример 6.1.** Рассмотрим еще раз эксперимент с бросанием кости 102 раза. Выпало 20 "шестерок". Очевидно ли то, что на выпадении "шестерок" сказалось влияние какого-то неслучайного фактора?

Решение

Если кость без смещения, то число выпадения "шестерок" при 102 бросаниях подчиняется биномиальному распределению.

Нулевая гипотеза предполагает, что число "шестерок" биномиально распределено с $p=1/6$, то есть на выпадение "шестерок" не оказывается какого-либо воздействия.

Альтернативная гипотеза предполагает, что на кость было оказано влияние для более частого выпадения "шестерок" — число "шестерок" не подчиняется биномиальному распределению с $p=1/6$.

Вероятность достижения r "шестерок" из 102 бросаний, если на кость не оказано влияние, находится следующим образом:

$$P(r) = {}^{102}C_r \times (1/6)^r \times (5/6)^{102-r}, \qquad r = 0, 1, 2, 3, \ldots, 102.$$

Поэтому вероятность наблюдаемого числа "шестерок" равна:

$$P(r = 20) = {}^{102}C_{20} \times (1/6)^{20} \times (5/6)^{82} = 0{,}073,$$

т. е. это событие маловероятно.

Вероятность ожидаемого числа "шестерок" равна:

$$P(r = 17) = {}^{102}C_{17} \times (1/6)^{17} \times (5/6)^{85} = 0{,}105,$$

что также маловероятно.

Как же мы можем оценить, важна ли разница между этими вероятностями или нет? Необходимы стандартные критерии, на основе которых мы можем сравнить вероятности. Одним из таких критериев является общая вероятность всех исходов, которая равна 1, то есть:

$$P\ (\text{все возможные исходы}) = 1.$$

Поскольку вероятность выпадения любой величины будет мала по сравнению с 1, то процедурой, используемой на практике является сравнение P (20 или более "шестерок") с 1. Если P (20 или более "шестерок") меньше 1, то мы утверждаем, что если нулевая гипотеза верна, это событие маловероятно. В этом случае мы не относим событие к статистической колеблемости или к случайности. Мы предполагаем, что имеется причина, которая привела к такому результату. Мы не знаем, что это за причина, но самым резонным предположением является то, что на кость оказывается влияние. Следовательно, мы отклоняем нулевую гипотезу, поскольку нет оснований для ее подтверждения. Если P(20 или более "шестерок") является не маленькой по сравнению с 1, то событие могло произойти случайно. Нулевая гипотеза не противоречит данным и, следовательно, не может быть отклонена. Этот результат не доказывает, что нулевая гипотеза верна; мы просто говорим, что нет оснований для ее отклонения.

Остается определить, что мы имеем в виду под малой вероятностью. Эта та отправная точка, на которую опирается лицо, принимающее решение, основываясь на уверенности, требуемой для данного решения. Лицо, принимающее решение, выбирает граничную величину для P(20 или более "шестерок"). Если действительная величина вероятности больше, чем эта величина, то тогда нулевая гипотеза принимается. Если действительная вероятность меньше, чем принятая граничная величина, решением будет отклонение нулевой гипотезы. Лицо, принимающее решение, может выбрать любую граничную величину. Однако обычно используются следующие значения:

1. Граничная величина для вероятности $p=0{,}05$. Если вероятность меньше этого значения, то нулевая гипотеза может быть явно неверной. Мы не доверяем этой гипотезе и отклоняем ее.
2. Граничная величина для вероятности $p=0{,}01$. Если вероятность меньше этого значения, то более вероятно, что нулевая гипотеза неверна. Гипотеза подвергается серьезному сомнению и отклоняется.
3. Граничная величина для вероятности $p=0{,}001$. Если вероятность меньше этого значения, то нулевая гипотеза почти наверняка неверна и отклоняется.

Когда принимается решение, используется термин значимость. Например, если решение необходимо сделать на границе 0,05 и действительная вероятность меньше 0,05, то результат значим на уровне 5%. Это означает, что вероятность случайного получения выборочной статистики большей или равной наблюдаемой величины меньше 5%, следовательно, нулевая гипотеза находится под сомнением. Однако, если действительная вероятность больше 0,05, то результат не значим для уровня 5%. Вероятность случайного достижения выборочной статистики или большей величины больше чем 5%, и нет причины для отклонения нулевой гипотезы.

Следует подчеркнуть, что несостоятельность отклонения нулевой гипотезы не доказывает, что нулевая гипотеза верна. Существует много способов наблюдения за выборкой статистических критериев, которые могут быть рассчитаны.

Некоторые из этих статистик будут согласовываться с нулевой гипотезой, некоторые – нет. Одни тесты дадут подтверждение, другие – опровержение. Лицо, принимающее решения, должно оценить все доказательства и вынести свое личное суждение.

Вернемся к примеру с костью. Примем уровень значимости решения равным 5%. Следующим шагом будет расчет вероятности достижения 20 или более "шестерок". В результате мы получили P (20 или более "шестерок")=0,2480. Теперь сравним P (20 или более "шестерок") с граничной величиной 0,05. P (20 или более "шестерок") больше 0,05, следовательно, можно сказать, что результат не существенен на уровне 5%. Доказательство согласуется с нулевой гипотезой. У нас нет причины отклонить нулевую гипотезу. Вероятность достижения 20 или более "шестерок" при бросании правильной кости равняется примерно 25%. У нас нет оснований предполагать, что на кость оказывают влияние неслучайные факторы.

6.2.1. Правила испытания гипотез

В принципе мы можем испытывать значимость любой статистики, имеющей любое вероятностное распределение. Однако обычно мы сталкиваемся с несколькими стандартными случаями. Выборочная статистика – средняя, доля и дисперсия подчиняются либо нормальному, либо t, F или хи-квадрат-распределениям.

1. **Проверка на основе нормального распределения.** Этот критерий используется для испытания среднего значения выборки ($\bar{x}$) по отношению к среднему значению генеральной совокупности (μ). Такой критерий применяется при любой величине выборочной совокупности (n), когда дисперсия генеральной совокупности (σ^2) известна. Кроме того, если мы хотим тестировать выборочную долю p, то можно использовать нормальное распределение, если величина выборки большая, $np > 5$ и $n(1-p) > 5$, поскольку в этом случае нормальное распределение дает хорошее приближение к биномиальному распределению.

2. **t-критерий.** Используется для испытания гипотезы о среднем значении при любой величине выборочной совокупности при неизвестной генеральной дисперсии. Для больших выборок t-распределение приближается к нормальному распределению.

3. **F-критерий.** Используется для сравнения генеральных дисперсий. Размер выборки может быть любым при условии, что выборка взята из нормальной генеральной совокупности.

4. **Критерий χ-квадрат.** Это непараметрический критерий, то есть значения выборочной статистики не требуются. Этот Критерий основан на частоте появлений значений случайных переменных. Может быть использован для испытания гипотезы о связи между характеристиками или о согласии наблюдаемого частотного распределения с некоторым стандартным распределением.

6.2.2. Одно- и двусторонние тесты

Как было уже отмечено, выбор уровня значимости, на котором принимается решение, делается лицом, принимающим решения. Одним из важных аспектов, который должен приниматься во внимание, является природа альтернативной гипотезы. То, как задана альтернативная гипотеза, влияет на выбор границы между критической областью и областью доверительных значений.

Вернемся к примеру, в котором машина производит металлические диски со средним диаметром 2,0 см. Случайная выборка изготовленных дисков показала, что средний диаметр равен 2,3 см. Если лицо, принимающее решения, просто интересуется, правильно ли его машина настроена, то не важно, больше или меньше выборочная средняя, чем предполагаемая средняя генеральной совокупности. Следовательно, нулевая и альтернативная гипотезы будут:

$$H_0:\ \mu = 2{,}0 \text{ см};$$

$$H_1:\ \mu = 2{,}0 \text{ см}.$$

Если принимается решение с 5%-ным уровнем значимости, то можно предположить, что границы расположены симметрично по выборочному распределению, как это показано на рис. 6.1.

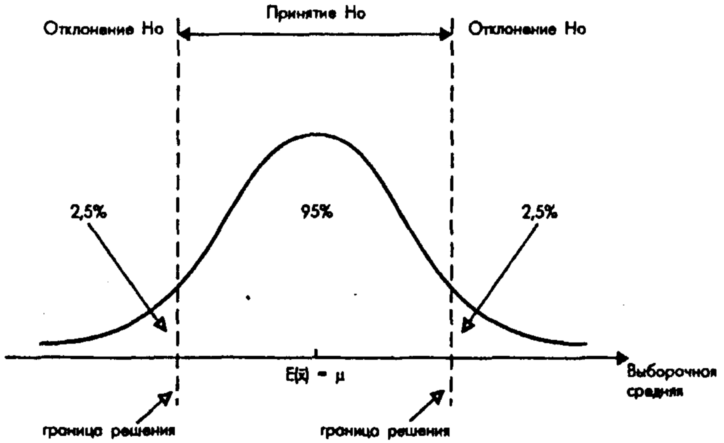


Рис. 6.1. 5%-ная двусторонняя проверка

В этом случае применяют двусторонний тест. Принимающему решение безразлично, будет ли среднее значение выборки в действительности больше или меньше предполагаемой генеральной средней. Он просто хочет знать, произошли ли какие-нибудь изменения или же нет. Однако если лицо, принимающее решение, беспокоится, что средняя действительно стала больше, то он должен принять другую альтернативную гипотезу:

$$H_0:\ \mu = 2{,}0 \text{ см},$$

$$H_1:\ \mu > 2{,}0 \text{ см}.$$

В этом случае, лицо принимающее решение, исследует существенность различий в определенном направлении. Если опять-таки решение должно быть принято с 5%-ным уровнем значимости, то на этот раз выбирается единственная граница по выборочному распределению, как показано на рис. 6.2.

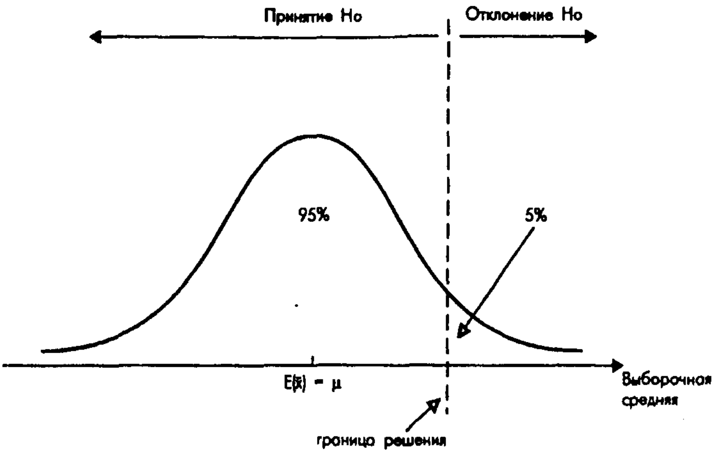


Рис. 6.2. 5%-ная односторонняя проверка

Такая постановка задачи приводит к проверке с одной границей.

Отличие между проверкой с одной границей и с двумя границами состоит в изменении уровня значимости при принятии решения. Это обстоятельство является основным, поскольку мы будем уверены в решении, только если результаты явно отклоняют тот или иной путь. Рассмотрим особые виды испытаний, помня, что основные принципы остаются одними и теми же во всех случаях.

6.3. ИСПЫТАНИЕ ГИПОТЕЗЫ НА ОСНОВЕ ВЫБОРОЧНОЙ СРЕДНЕЙ: ГЕНЕРАЛЬНАЯ ДИСПЕРСИЯ ИЗВЕСТНА

Сразу заметим, что конкретный способ проведения испытания гипотез может изменяться. Метод, описанный ниже является единственным, в котором используются стандартные таблицы вероятностного распределения.

❐ **Пример 6.2.** Рафинированный сахарный песок упаковывается в пакеты весом в среднем 1,0 кг (μ) со стандартным отклонением (σ), равным 0,01 кг. Случайная выборка $n = 16$ пакетов готовой продукции выявила средний вес $\bar{x} = 1{,}01$ кг. Имеется ли какое-нибудь основание предполагать, что фасовочная машина работает без нарушений в настройке?

Решение

Можно предположить, что вес пакетов, наполненных машиной, соответствует приблизительно нормальному вероятностному распределению. Нулевой гипотезой является то, что настройка машины не отклоняется от нормального состояния, т.е. H_0: выборочная средняя согласуется с выборкой, взятой из нормальной генеральной совокупности со средней, равной 1,0 кг, то есть $\mu = 1{,}0$ кг.

Есть ли основание полагать, что настройка машины осталась на надлежащем уровне или она изменилась? Если данные, взятые из выборки, заставляют нас отклонить нулевую гипотезу, логично предположить, что машина работала при неправильной настройке и, следовательно, альтернативной гипотезой является:

H_1: выборка была взята не из нормального распределения со средней равной 1,0 кг, то есть $\mu \neq 1,0$ кг.

Из H_0 следует, что выборочное распределение выборочных средних является тоже нормальным распределением со средней, равной 1,0 кг, и стандартной ошибкой, равной $(0,01/\sqrt{16})$ кг. Проверим нулевую гипотезу при 5%-ной значимости, используя нормальное распределение с двумя границами (см. рис. 6.1.).

Используя таблицы стандартного распределения, находим, что $\bar{x}_1$ и $\bar{x}_2$ равны 1,96 стандартным ошибкам расстояния от генеральной средней, и тогда рис. 6.1. может быть представлен следующим образом:

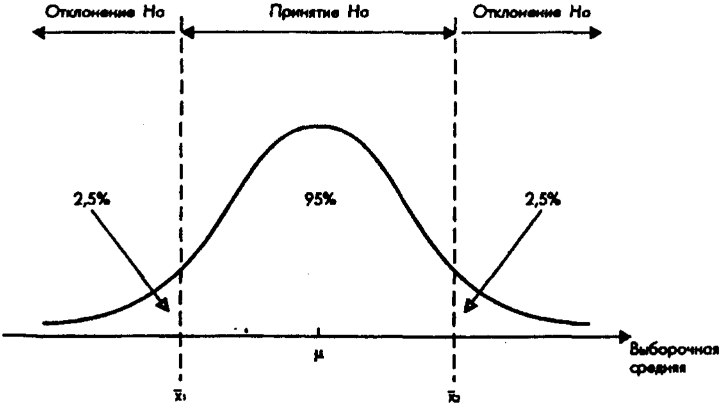


Рис. 6.3. Критические значения выборочного распределения для 5%-ого уровня значимости

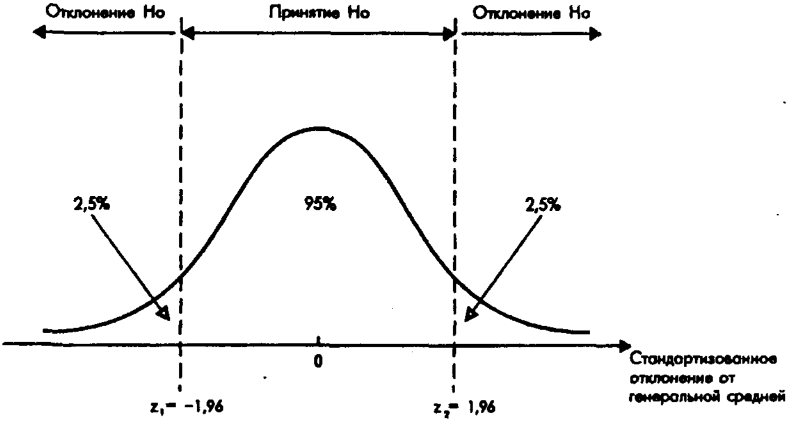


Рис. 6.4. Критические значения стандартного нормального распределения для 5%-ого уровня значимости

Если теперь мы имеем значения выборочной средней $\bar{x}$, равное 1,01 кг, то можем выразить его отклонение от генеральной средней через количество стандартных ошибок:

$$z = \frac{\bar{x} - \mu}{SE_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

Это то же самое выражение, которое использовалось в гл. 2, но стандартное отклонение генеральной совокупности заменяется стандартной ошибкой выборочного распределения. Следовательно,

$$z = \frac{1{,}01 - 1{,}0}{0{,}01/4} = 4{,}0$$

стандартным ошибкам, которые лежат выше средней μ.

Значение стандартизованной переменной 4,0 больше, чем граничная величина 1,96. Это означает, что:

$$P\ (\text{стандартизованная переменная} \geq 4{,}0) < 0{,}025$$

и результат существенен на 5%-ном уровне значимости. Если мы нанесем на *диаграмму значение z=4,0, то оно попадает в область отклонения H_0. Поскольку* результат существенен на 5%-ном уровне, мы приходим к заключению, что имеется резонное основание, что выборочная средняя не согласуется с нулевой гипотезой. Мы отклоняем эту гипотезу в пользу альтернативной. Вероятность появления выборочной средней, равной 1,01 кг или больше, из-за случайных колебаний результатов выборочного исследования в случайной выборке величиной 16 единиц, взятой из нормальной генеральной совокупности, меньше чем 5%. Таким образом, мы считали, что выборка была взята из генеральной совокупности, средняя которой была не 1,0 кг. Мы делаем вывод, что машина работала в условиях нарушения нормальной настройки. Рассмотрим другой пример процедуры принятия решения.

☐ **Пример 6.3.** Штамповочный пресс делает отверстия в металлических шайбах. В среднем величина отверстия μ = 4,00 мм со стандартным отклонением σ = 0,20 мм. Случайная выборка n = 25 шайб показала, что средний диаметр $\bar{x}$ = 3,88 мм.

Сохраняется ли нормальная настройка пресса?

Решение

Можно предположить, что диаметр отверстий, сделанных штамповкой, следует приблизительно нормальному распределению.

Нулевой гипотезой является то, что машина работает при правильной настройке.

H_0: Выборочная средняя согласуется с выборкой, взятой из нормальной совокупности со средней 4,0 мм, то есть μ =4,0 мм.

Если данные выборки приводят к отклонению нулевой гипотезы, логично предположить, что машина работает при неправильной настройке, то есть альтернативной гипотезой является:

H_1: Выборка не была взята из нормального распределения со средней 4,0 мм, то есть $\mu \neq 4{,}0$ мм.

Из H_0 следует, что выборочное распределение выборочных средних также является нормальным распределением со средней 4,0 мм и стандартной ошибкой $(0{,}2/\sqrt{25})$ мм. Из H_1 следует, что мы будем проводить двустороннюю проверку.

Примем решение относительно нулевой гипотезы на 1%-ном уровне значимости. Используя таблицы стандартного нормального распределения, находим, что граничные величины (обе) равны 2,576 стандартным ошибкам расстояния от средней μ. Это показано на рис. 6.5.:

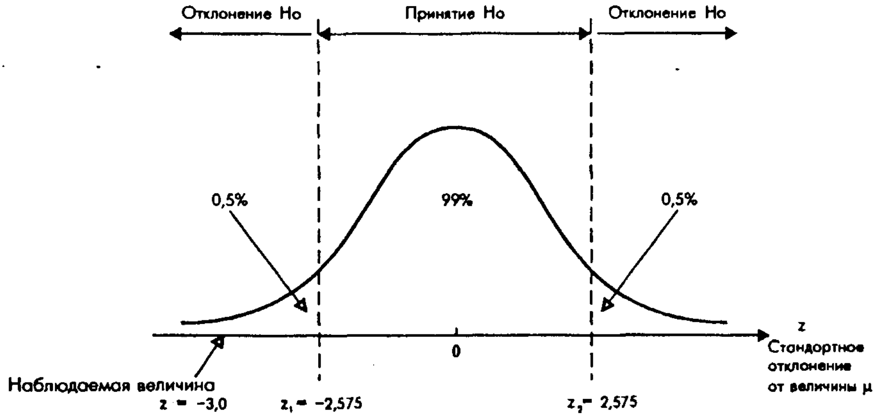


Рис. 6.5. Критические значения величины для 1%-ного уровня значимости

Рассчитаем по нашим данным величину z:

$$z = \frac{\bar{x} - \mu}{SE_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

Следовательно:

$$z = \frac{3{,}88 - 4{,}0}{0{,}20/5} = -3{,}0$$

(число стандартных отклонений от μ, лежащих ниже μ).

Величина z является критериальной статистикой. Значение 3,0 меньше критической величины 2,575. Это означает, что

$$P\ (\text{стандартизованная переменная} \le -3{,}0) < 0{,}005.$$

Результат существенен на уровне 1%. Мы заключаем, что имеется основание предполагать что выборка не согласуется с нулевой гипотезой. Мы отклоняем эту гипотезу в пользу альтернативной. Вероятность того, что средняя по выборке 3,88 мм или меньше из-за случайностей выборки 25 единиц, взятых из нормальной генеральной совокупности со средней 4,0 мм, меньше, чем 1%. Следовательно, мы делаем вывод, что настройка штамповочного пресса отклоняется от нормальной.

❑ **Пример 6.4.** Высота отдельных ростков рассады распределена нормально со средней $\mu = 53$ см и дисперсией $\sigma^2 = 12$ см^2. В прошлом году в ящик, в котором были высажены $n = 15$ таких растений была внесена по ошибке двойная норма удобрения. Средняя высота рассады в этом ящике достигла $\bar{x} = 55$ см. Есть ли какое-либо основание предполагать, что повышенное внесение удобрений дало положительный эффект?

Решение.

Нулевой гипотезой является то, что дополнительное внесение удобрений не дало эффекта.

H_0: Выборочная средняя согласуется с выборкой, взятой из нормальной генеральной совокупности со средней $\mu = 53$ см. Есть ли основание предполагать, что сверхудобрение не дало эффекта или, наоборот, способствовало росту растений? Если мы решаем отклонить нулевую гипотезу, логически предположим что сверхудобрение дает положительный эффект и растения становятся очень высокими. Таким образом, альтернативной гипотезой является следующая:

H_1: Выборка не была взята из нормального распределения со средней 53 см, а из генеральной совокупности со средней большей, чем 53 см, то есть $\mu > 53$ см.

Из H_0 следует, что распределение выборочных средних также является нормальным распределением со средней 53 см и стандартной ошибкой $\sqrt{12/15}$ см. Из H_1 следует, что мы используем проверку с одной границей. Примем решение нулевой гипотезы на 0,1%-ном уровне значимости. Используя таблицы стандартного нормального распределения, находим, что граничная величина $z = 3{,}09$ стандартных ошибок выше средней. Граничное значение выборочного распределения показано на рис. 6.6.

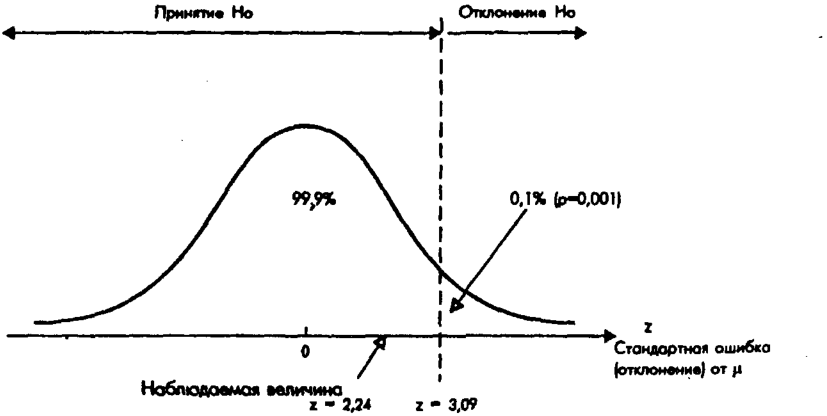


Рис. 6.6. Граничное значение для 0,1%-ного уровня значимости

Рассчитываем проверочную статистику z:

$$z = \frac{\bar{x} - \mu}{SE_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

Следовательно,

$$z = \frac{55 - 53}{\sqrt{12/15}} = 2{,}24 .$$

Рассчитанное значение z=2,24, меньше, чем граничная величина z = 3,09. Это означает, что:

$$P\ (\text{стандартизованная переменная} \geq 2{,}24) > 0{,}001.$$

Результат не существенен при 0,1%-ном уровне значимости. Поэтому можно заключить, что наблюдаемое явление согласуется с нулевой гипотезой, которую мы принимаем. Вероятность появления выборочной средней 55 см или более из-за случайностей выборки 15 наблюдений, взятых из нормальной генеральной совокупности со средней 53 см, больше чем 0,1%. Мы верим, что выборка взята из генеральной совокупности, для которой средняя составляет 53 см, т.е. что вспомогательное внесение удобрений не дало положительного эффекта.

6.4. ИСПЫТАНИЕ ГИПОТЕЗЫ НА ОСНОВЕ ВЫБОРОЧНОЙ СРЕДНЕЙ — ГЕНЕРАЛЬНАЯ ДИСПЕРСИЯ НЕИЗВЕСТНА

В главах 4 и 5 мы встречались с положением, когда дисперсия генеральной совокупности σ^2 неизвестна, в этом случае мы можем произвести её оценку, используя выборочное стандартное отклонение s. Тогда соответствующее стандартизованное распределение становится t-распределением с (n1) степенями свободы.

❐ **Пример 6.5.** Компания "Britelite plc" производит электрические лампочки. Для определенного типа лампочек установлен нормативный срок использования μ =1500 ч. Для испытания новой партии была взята выборка n = 10 лампочек. Среднее время пользования лампочкой в выборке равна 1410 ч $\bar{x}$ со стандартным отклонением s = 90 ч. Свидетельствуют ли эти данные о том, что ожидаемый срок использования изменился по сравнению с 1500 ч?

Решение

Нулевой гипотезой является предположение о том, что выборка была взята из генеральной совокупности со средней 1500 ч.

H_0: Выборочная средняя согласуется с выборкой, взятой из нормальной генеральной совокупности со средней 1500 ч, то есть μ =1500 ч.

H_1: Выборка не была взята из нормально распределенной совокупности со средней 1500 ч, то есть $\mu \neq$ 1500 ч. Из H_1 следует, что мы будем использовать испытание с двумя границами, из H_0 — что выборочное распределение выборочных средних также является нормальным распределением со средней 1500 ч и стандартной ошибкой ($\sigma/\sqrt{10}$) ч. Поскольку σ неизвестна, то для испытания гипотезы используем стандартное t-распределение с числом степеней свободы, равным (10 — 1) 9. Примем решение при 5%-ном уровне значимости. Используя таблицы t-распределения (Приложение 2), находим, что $t_{0,05/2,9}$ равняется ± 2,26. Граничные величины стандартного распределения показаны на рис. 6.7.

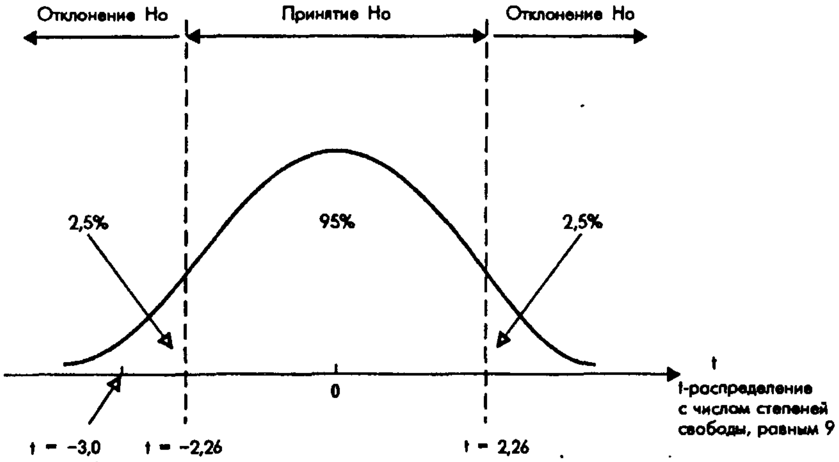


Рис. 6.7. Граничные значения t при 5%-ном уровне значимости

Теперь проверочной статистикой является t:

$$t = \frac{\bar{x} - \mu}{\hat{SE}_{\bar{x}}} = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} .$$

В разделе 4.4.3 было показано, что:

$$\hat{\sigma} = \sqrt{\frac{n}{n-1}}\, s .$$

Поэтому:

$$\hat{SE}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \sqrt{\frac{n}{n-1}}\, \frac{s}{\sqrt{n}} = \frac{s}{\sqrt{n-1}} .$$

Следовательно,

$$t = \frac{1410 - 1500}{90\ /(10 - 1)} = -3{,}0 .$$

Проверочная статистика −3,0, меньше граничной величины −2,26. Это означает, что:

$$P\ (\text{стандартизованная переменная} < -3{,}0) < 025.$$

Результат значим на 5%-ном уровне. Поскольку результат является значимым, мы заключаем, что имеется резонное основание считать, что выборка не согласуется с нулевой гипотезой. Мы отклоняем эту гипотезу. Вероятность появления выборочной средней равной 1410 ч или менее, из-за случайностей отбора при выборке величиной 10 единиц, взятой из нормальной генеральной совокупности со средней 1500 ч, меньше чем 5%. Мы верим, что выборка не была взята из такой генеральной совокупности. Средний срок использования лампочек изменился.

❐ **Пример 6.6.** Компания упаковывает сахарный песок в пакетики, на которых указан вес 250 г. Компания утверждает, что среднее содержимое пакетиков не меньше 250 г (μ). Служащий Министерства торговых стандартов посетил фабрику компании с целью проверки этого утверждения. Он отобрал в случайном порядке 8 пакетов n и взвесил содержимое. Результаты оказались следующими:

Таблица 6.1. **Вес случайно выбранных 8 пакетов**

Номер пакета	1	2	3	4	5	6	7	8
Вес, г	247	252	248	253	246	245	248	250

Какое заключение может сделать представитель Министерства торговых стандартов на основе этой выборки? (Предполагается, что вес пакетиков с сахаром нормально распределен).

H_0: Выборочная средняя согласуется с выборкой, взятой из нормальной генеральной совокупности со средней $\mu = 250$ г.

H_1: Выборка взята из распределения со средней меньшей, чем 250 г, то есть $\mu < 250$ г.

Из H_0 следует, что выборочное распределение выборочных средних является нормальным распределением со средней 250 г и стандартной ошибкой $(\sigma/\sqrt{8})$ г. Из H_1 следует, что мы будем использовать испытание с одной границей. Рассчитаем выборочную среднюю $\bar{x}$ и стандартное отклонение s:

$$\bar{x} = \frac{\sum x}{n} = \frac{1989}{8} = 248{,}6 \text{ г},$$

$$s = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2} = 2{,}643 \text{ г}.$$

Поскольку σ неизвестна, для испытания гипотезы используем стандартное t-распределение с числом степеней свободы (8 1)=7. Примем решение относительно нулевой гипотезы на 1%-ном уровне значимости. Используя таблицы t-распределения из Приложения 2, находим, что $t_{0,01,7} = -3{,}00$. Граничные значения показаны на рис. 6.8.

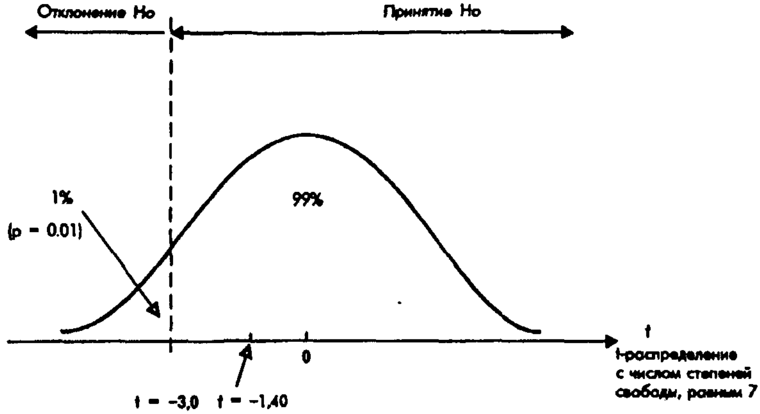


Рис. 6.8. Критическое значение t на 1%-ном уровне значимости

Проверочная статистика t рассчитывается также, как и раньше:

$$t = \frac{\bar{x} - \mu}{\hat{SE}_{\bar{x}}} = \frac{\bar{x} - \mu}{s/\sqrt{n}},$$

Следовательно:

$$t = \frac{248{,}6 - 250}{2{,}643/\sqrt{(8-1)}} = -1{,}40\,.$$

Проверочная статистика, $-1{,}40$, больше граничной (критической) величины $-3{,}00$. Это означает, что

$$P\ (\text{стандартизованная переменная} < 1{,}40) > 0{,}01.$$

Результат не существенен на 1%-ном уровне. Поскольку результат не существенен, мы принимаем нулевую гипотезу на этом уровне . Вероятность появления величины выборочной средней 248,6 г или меньше из-за колебаний случайной выборки объемом 8 единиц, взятой из нормальной генеральной совокупности со средней 250 г, больше чем 1%. И, таким образом, мы утверждаем, что выборка была взята из такой генеральной совокупности. Утверждение компании, что сахар расфасован в пакетики по 250 г, верно.

6.5. ИСПЫТАНИЕ ГИПОТЕЗЫ НА ОСНОВЕ ВЫБОРОЧНОЙ ДОЛИ

Рассмотренная процедура испытания гипотез может быть также использована для проверки гипотезы о выборочной доле. Доля имеет биномиальное распределение, но при большом объеме выборки может быть использовано нормальное распределение в качестве аппроксимации биномиального.

☐ **Пример 6.7.** Поставщик электронных компонентов производит продукцию, которая иногда сразу отказывает. Он попытался контролировать производственный процесс так, чтобы доля неисправной продукции была меньше, чем 4%. Из поставляемой партии 500 компонентов 28 оказались неисправными. Имеется ли какое-нибудь основание предполагать, что производственный процесс вышел из под контроля и производится много неисправных изделий?

Решение.

H_0: Доля неисправных компонентов в произведенных изделиях равна 4%, то есть $p=0{,}04$.

H_1: Доля произведенных неисправных компонентов возросла, то есть $p > 0{,}04$.

Из H_1 следует, что мы будем проводить испытание с одной границей. Объем выборки из 500 компонентов большой, поэтому мы аппроксимируем биномиальное распределение посредством нормального распределения. Выборочное распределение выборочных долей будет приблизительно нормальным со средней долей $p = 0{,}04$. Стандартная ошибка выборочного распределения равняется:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}04 \times 0{,}96}{500}} = 0{,}00876.$$

Доля дефектов в выборке равняется: $\hat{p} = 28/500 = 0{,}056$.

Будем испытывать H_0 на 1%-ном уровне значимости. Критическое значение стандартизованной нормальной переменной $z = 2{,}33$.

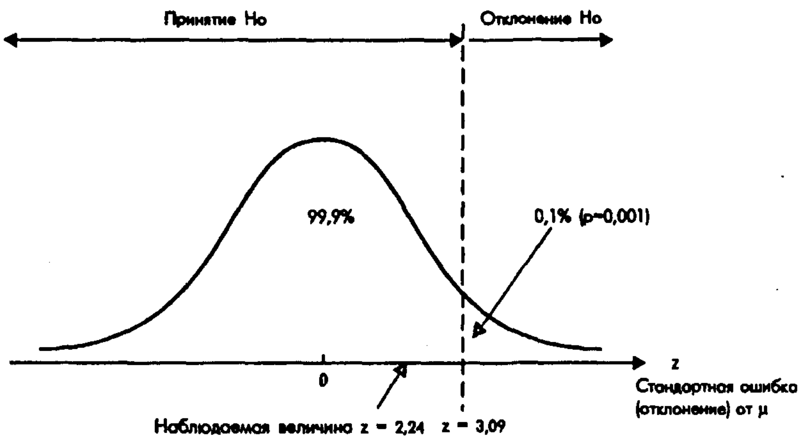


Рис. 6.9. Критическое значение z на 1%-ном уровне значимости

Проверочная статистика:

$$z = \frac{\hat{p} - p}{SE_{\hat{p}}},$$

Следовательно:

$$z = \frac{0{,}056 - 0{,}04}{0{,}00876} = 1{,}83\ .$$

Проверочная статистика меньше критической величины на 1%-ном уровне. Результат не существенен на уровне 1%: $P\ (z > 1{,}83) > 0{,}01$. Выборочная доля дефектных компонентов 0,056 может быть получена в результате случайностей выборки.

Следовательно, мы принимаем нулевую гипотезу на 1%-ном уровне значимости. Нет оснований предполагать, что производственный процесс вышел из под контроля и дает дефектов больше 4%.

Мы рассмотрели последовательность статистического вывода по одной выборке — сравнение выборочной статистики с предполагаемым параметром генеральной совокупности. Теперь мы должны рассмотреть ситуации, в которых имеются две выборочные совокупности, которые необходимо сравнить. В следующих разделах рассмотрим применение испытания гипотез при сравнении двух выборочных статистик.

6.6. ИСПЫТАНИЕ ГИПОТЕЗ О ДВУХ ГЕНЕРАЛЬНЫХ ДИСПЕРСИЯХ

Существует много ситуаций, в которых вариация данных важна не менее, чем средняя величина. Когда мы оцениваем портфель инвестиций, то исходим из ожидаемой прибыли, но, в то же время нельзя сбрасывать со счетов риск инвестирования. Такой риск может быть оценен на основе дисперсии возможной прибыли инвестиций (см. раздел 3.3). Предположим, что у нас имеются две независимые выборки и мы хотим знать, взяты ли они из нормальных генеральных совокупностей с одинаковой дисперсией. Например, предположим, что компания производит определенный элемент на двух автономных производственных линиях — А и В. Характеристики обеих линий одинаковые. Как определить, одинакова ли вариация продукции на этих линиях? Ответ на этот вопрос можно получить сравнив дисперсии случайных выборок, взятых из продукции первой и второй линий, используя соответствующую процедуру испытания гипотез. Так же можно сравнить риск двух различных инвестиционных портфелей. Сравнение дисперсий фактической прибыли, полученной в прошлые годы, даст возможность принять решение.

6.6.1. Отношение дисперсий или F-критерий

В разделе 4.4. было показано, что отношение двух дисперсий подчиняется распределению F-статистики:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}.$$

Поскольку лучшая оценка дисперсии генеральной совокупности вычисляется по формуле:

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2,$$

то

$$F = \frac{n_1 s_1^2}{(n_1 - 1)} \frac{(n_2 - 1)}{n_2 s_2^2}.$$

Нулевая гипотеза предполагает, что две выборки независимы и взяты из нормальных генеральных совокупностей с одинаковыми дисперсиями, $\sigma_1^2 = \sigma_2^2$. В этом случае F равно 1. Из теории испытания гипотез мы знаем, что если даже нулевая гипотеза верна, то маловероятно, что σ_1^2 имеет точно такое же значение, что и σ_2^2 из-за колебаний отбора. Следовательно, маловероятно, что F-статистика будет равна 1. Решением, которое мы собираемся принять, используя испытание гипотез, является то, будет ли истинная величина F достаточно близка к 1 для того, чтобы подтвердить вероятность, что выборочные совокупности были взяты из нормальных генеральных совокупностей с одинаковой дисперсией. В этом случае различие в значениях $\hat{\sigma}_1^2$ и $\hat{\sigma}_2^2$ может быть отнесено к случайностям.

Как отмечалось в разделе 4.4.4, F-распределение зависит от числа степеней свободы в обеих сравниваемых выборках. Когда мы производим оценку единственного генерального параметра по выборке, то теряем одну степень свободы. Таким образом, для каждой выборки остаются (n 1) степени свободы.

Для того, чтобы привести стандартную таблицу для F-распределения к более удобному виду, даны только значения $F \geq 1$, то есть это – таблицы с одной границей. Чтобы использовать эти таблицы при расчете F, делим большую дисперсию на меньшую. Проверочной статистикой является:

F = (большая оцененная дисперсия) / (меньшая оцененная дисперсия).

❑ **Пример 6.8.** Биржевой маклер исследует две инвестиции – А и В – от имени клиента. Инвестиция А предполагается на срок 10 лет с ожидаемой ежегодной прибылью в течение этого периода 17,8%. Инвестиция В рассчитана на срок 8 лет также с ожидаемой готовой прибылью 17,8%. Дисперсии ежегодных прибылей от двух инвестиций составляют 3,21%² и 7,14%². Есть ли какое-либо основание считать, что риски инвестиций А и В неравны? Предполагается, что распределения ежегодных прибылей на инвестиции подчиняются нормальному распределению.

Решение.

Дисперсии ежегодных прибылей могут быть использованы для определения риска. Мы хотим знать, взяты ли эти две выборочные совокупности ежегодных прибылей от двух инвестиций, из нормальных генеральных совокупностей с равными дисперсиями. Поэтому:

$$H_0\colon \sigma_A^2 = \sigma_B^2,$$

$$H_1\colon \sigma_A^2 \neq \sigma_B^2.$$

Будем испытывать нулевую гипотезу, используя F-критерий с двумя границами, на 5%-ном уровне значимости. Это эквивалентно 2,5%-ному уровню значимости с одной границей, поэтому мы используем $\alpha = 0{,}025$ для определения критического значения в таблице F. Лучшие оценки двух генеральных дисперсий могут быть получены на основе выборочных дисперсий следующим образом:

$$\hat{\sigma}_A^2 = \frac{n_A}{n_A - 1} s_A^2 = \frac{10}{9} \times 3{,}21^2 = 11{,}449 \quad \text{с 9 степенями свободы;}$$

$$\hat{\sigma}_B^2 = \frac{n_B}{n_B - 1} s_B^2 = \frac{8}{7} \times 7{,}14^2 = 58{,}2624 \quad \text{с 7 степенями свободы.}$$

Поскольку:

$$\hat{\sigma}_B^2 > \hat{\sigma}_A^2,$$

$$F = (\text{большая оцененная дисперсия})/(\text{меньшая оцененная дисперсия}) = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_A^2}.$$

Следует помнить, что при использовании стандартных таблиц значения $F \geq 1$.

F-таблицы построены так, что степени свободы большей дисперсии ($v_1=7$ степеней свободы) приводятся вверху таблицы, а степени свободы меньшей дисперсии ($v_2=9$ степеней свободы) — по столбцу вниз. Используя 2,5%-ное табличное значение F из Приложения 2, что эквивалентно 5%-ному уровню значимости с двумя границами, с 7 и 9 степенями свободы, критическая величина равна:

$$F_{0,05/2,7,9} = 4{,}197.$$

По данным выборки, проверочная статистика равна:

$$F = \frac{58{,}2624}{11{,}449} = 5{,}09.$$

Поскольку:

$$5{,}09 > F_{0,05/2,7,9},$$

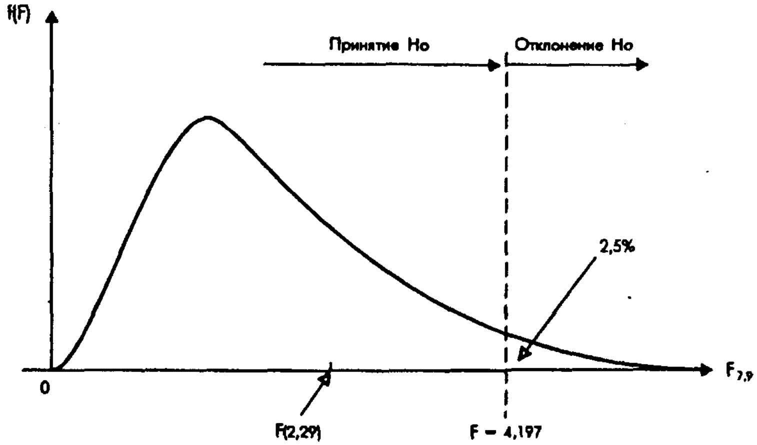


Рис. 6.10. Критическая величина F-критерия на 5%-ном уровне значимости с двумя границами

результат существен на 5%-ном уровне значимости. У нас есть основания предполагать, что риски (определенные дисперсиями ежегодных прибылей) двух инвестиций не равны.

Пример 6.9. Компания "POP" производит химический продукт X, используя серийное производство. Имеются два разных способа производства, и дирекция компании решила использовать тот способ, который дает более твердый продукт. Твердость химического продукта X может быть оценена через его плотность. Чем дольше процесс затвердевания, тем меньше будет вариация плотности.

Управляющий производством компании провел $n_1 = 10$ испытаний изделий, изготовленных первым способом и получил стандартное отклонение для плотности химического продукта, равное 4,1 единиц (s_1). Затем он провел $n_2 = 12$ испытаний изделий, изготовленных вторым способом, и получил стандартное отклонение для плотности химического продукта, равное 2,0 единицы (s_2). Дает ли второй способ действительно более твердое вещество, нежели первый способ? Предполагается, что плотность химического вещества нормально распределена.

Решение

Нулевая гипотеза: выборки взяты из нормальных генеральных совокупностей с равными дисперсиями. Твердость продукции при обоих методах производства одинакова:

$$H_0: \sigma_1^2 = \sigma_2^2,$$

$$H_1: \sigma_1^2 > \sigma_2^2,$$

H_1 предполагает, что при втором способе – продукция более твердая при наименьшей дисперсии. H_1 показывает, что мы должны применить испытание с одной границей. Будем испытывать нулевую гипотезу на 1%-ном уровне значимости. Проверочной статистикой является:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\ .$$

Лучшие оценки генеральных дисперсий — это выборочные дисперсии:

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1}\, s_1^2 = \frac{10}{9} \times 4{,}1^2 = 18{,}68 \text{ с 9 степенями свободы;}$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1}\, s_2^2 = \frac{12}{11} \times 2{,}02^2 = 4{,}36 \text{ с 7 степенями свободы.}$$

Используя табличное значение F-критерия из F-таблицы с одной границей при уровне значимости 1% с 9 степенями свободы v_1 по столбцу (см. Приложение 2) и 11 степенями свободы v_2 по строке, находим критическое значение:

$$F_{0,01,9,11} = 4{,}632$$

По данным выборки находим F-статистику:

$$F = 18{,}68/4{,}36 = 4{,}28\ .$$

Поскольку:

$$4{,}28 < F_{0,01,9,11} = 4{,}632\ ,$$

то результат не существенен на 1%-ном уровне, т.е. на этом уровне значимости у нас нет причины предполагать, что второй способ даст более твердое вещество, чем первый.

6.7. СРАВНЕНИЕ СРЕДНИХ ВЕЛИЧИН ДВУХ ВЫБОРОК ПРИ ИЗВЕСТНЫХ ГЕНЕРАЛЬНЫХ ДИСПЕРСИЯХ

В разделах 6.3—6.5 мы имели дело с одной выборкой и оценивали, взята ли она из нормальной генеральной совокупности. Теперь мы будем рассматривать ситуации, в которых имеются две выборочные совокупности. Нужно определить, взяты ли они из нормальных генеральных совокупностей с равными средними. Например, если аудитора удовлетворяет качество бухгалтерского учета в компании, он может взять выборку счетов для оценки величины ошибки в генеральной совокупности. Если система учета продолжает действовать исправно, то вторая выборка должна дать оценку величины ошибки генеральной совокупности, которая незначительно отличается от первой.

Подобным образом случайная выборка даст возможность оценить среднюю наполняемость соусом бутылок в генеральной совокупности. Если процесс разлива и дальше функционирует исправно, то последующие выборки не должны дать оценки средней наполняемости, которые бы значительно отличались от предыдущих. Это очень важный аспект для контроля качества.

В обоих примерах, резонно заключить, что дисперсия генеральной совокупности остается той же. Однако, если мы рассматриваем две различные производственные линии наполнения соусом бутылок, то можем взять выборку из каждой линии для того, чтобы увидеть имеется ли значительная разница между средней наполняемостью генеральных совокупностей одной и другой линий. В этом случае нет основания предполагать, что две генеральных дисперсии равны между собой.

При неизвестных дисперсиях генеральных совокупностей процедура испытания гипотез зависит от того, предполагается ли равенство дисперсий или нет. Однако форма нулевой гипотезы остается той же во всех случаях. Для испытания гипотезы по двум выборочным средним нулевая гипотеза предполагает, что две выборочные совокупности взяты из генеральных совокупностей с равными средними.

H_0: $\mu_1 = \mu_2$, т.е. генеральные средние равны между собой.

Создается новая переменная, которая является разницей между выборочными средними $(\bar{x}_1 - \bar{x}_2)$ и сравнивается с предполагаемой разницей между генеральными средними, то есть $\mu_1 - \mu_2 = 0$. Если разница между выборочными средними незначительно отличается от нуля, то мы можем предположить, что нулевая гипотеза приемлема. Если разница значительно отличается от нуля, то можем предположить, что нулевая гипотеза не приемлема.

Если σ_1^2 и σ_2^2 известны, то проверочная статистика следует нормальному распределению и находится следующим образом:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE_{\bar{x}_1 - \bar{x}_2}},$$

где:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

❑ **Пример 6.10.** Компания по производству сахарного песка имеет две производственные линии для наполнения мешочков сахарным песком по 1 кг (μ_1 и μ_2). Используя данные, собранные в течение долгого периода времени, управляющий оценивает генеральное стандартное отклонение веса мешочков, поставляемых с линии 1 в 0,02 кг (σ_1) и с линии 2 в 0,04 кг (σ_2). Из линии 1 была взята случайная выборка объемом $n_1 = 10$ мешочков и найден средний вес содержимого в мешочках $\bar{x}_1 = 1{,}018$ кг. Подобная выборка объемом $n_2 = 12$ мешочков была взята из линии 2 и найден средний вес $\bar{x}_2 = 0{,}989$ кг. Имеется ли какое-нибудь основание предположить, что две производственные линии развешивают сахарный песок по мешочкам, средний вес которых отличается?

Решение

Нулевая гипотеза предполагает, что две выборочные средние согласуются с выборочными совокупностями, взятыми из нормальных генеральных совокупностей с одинаковой генеральной средней.

$$H_0: \mu_1 = \mu_2, \text{ то есть } \mu_1 - \mu_2 = 0\ ,$$

$$H_1: \mu_1 \neq \mu_2, \text{ то есть } \mu_1 - \mu_2 \neq 0\ .$$

Из H_1 следует выбор испытания с двумя границами.

Поскольку генеральные дисперсии (σ_1^2 и σ_2^2) известны, мы испытываем разницу между выборочными средними, используя нормальное распределение. Проведем испытание на 1%-ном уровне значимости. Из таблиц стандартного нормального распределения в Приложении 2 находим граничную величину для z: ± 2,576.

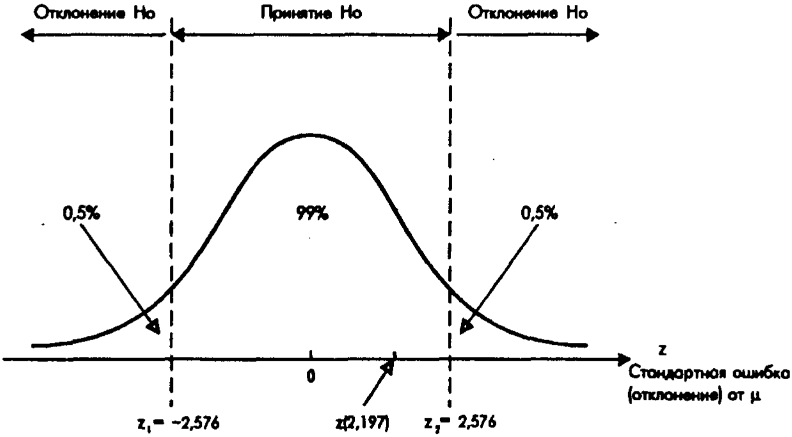


Рис. 6.11. Критические значения z для 1%-ного уровня значимости

Проверочная статистика равна:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE_{\bar{x}_1 - \bar{x}_2}}\ ,$$

где

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0{,}02^2}{10} + \frac{0{,}04^2}{12}} = 0{,}0132 \text{ г}.$$

Отсюда

$$z = \frac{(1{,}018 - 0{,}989) - 0}{0{,}0132} = 2{,}197 .$$

Поскольку

$$2{,}197 < z_{0{,}01} = 2{,}576,$$

результат не существенен на 1%-ном уровне, т.е. нет основания отклонять H_0. Мы предполагаем, что две производственные линии наполняют мешочки сахаром с одинаковым средним весом.

❒ **Пример 6.11.** Несколько сезонов садовник выращивал два сходных сорта крыжовника. Сбор урожая фактически был одинаков для обоих сортов с дисперсией 1,26 кг^2 (σ_A^2) для сорта А и дисперсией 1,20 кг^2 (σ_B^2) для разновидности В. Затем он решил использовать новый участок для выращивания крыжовника, но он не знает, будут ли почвенные условия воздействовать одинаково на оба сорта. Для эксперимента садовник посадил 30 кустов (n_A и n_B) каждого сорта на новом участке. Сорт А дал урожай в среднем 3,0 кг ($\bar{x}_A$) с куста, в то время, как урожай сорта В составил в среднем 3,5 кг ($\bar{x}_B$) с куста. Есть ли какое-нибудь основание предполагать, что на новом участке сорт В имеет больший, в среднем, урожай, чем сорт А?

Решение.

Нулевая гипотеза предполагает, что две выборочные совокупности взяты из нормальных распределений с одинаковой генеральной средней:

H_0: $\mu_A = \mu_B$, то есть $\mu_A - \mu_B = 0$,

H_1: $\mu_A < \mu_B$, то есть сорт В имеет в среднем больший урожай.

Это означает, что мы должны провести испытание с одной границей.

Поскольку генеральные отклонения известны, мы испытываем разницу между выборочными средними, используя нормальное испытание. Будем испытывать на 5%-ном уровне значимости. Из таблицы нормального распределения в Приложении 2 находим граничную величину для $Z_{0,05}$ равную 1,645. Проверочная статистика находится по формуле:

$$z = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{SE_{\bar{x}_A - \bar{x}_B}} ,$$

где

$$SE_{\bar{x}_A - \bar{x}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{1{,}26}{30} + \frac{1{,}20}{30}} = 0{,}286 \text{ кг}.$$

Отсюда

$$z = \frac{(3,0 - 3,5) - 0}{0,286} = -1,746 \ .$$

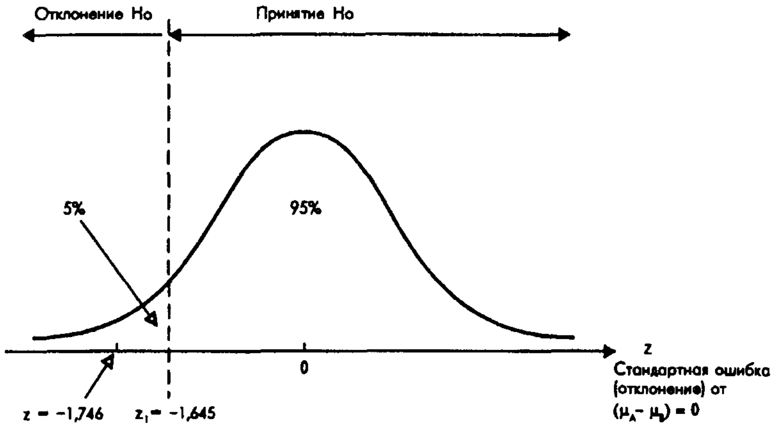


Рис. 6.12. Критическое значение величины z на 5%-ном уровне значимости

Поскольку:

$$1,746 < Z_{0,05} = 1,645,$$

результат существенен на 5%-ном уровне. Имеется достаточно сильное основание для утверждения, что выборочные характеристики не согласуются с H_0. Мы отклоняем H_0 и принимаем H_1. Вероятность получения разницы в значениях выборочных средних в 0,05 кг или более из-за случайностей выборки, меньше, чем 5%. Мы делаем вывод, что сорт В дает на новом участке больший урожай, чем сорт А.

6.8. ИСПЫТАНИЕ ГИПОТЕЗЫ ПО ВЫБОРОЧНЫМ СРЕДНИМ — ГЕНЕРАЛЬНЫЕ ДИСПЕРСИИ НЕИЗВЕСТНЫ

В этом случае стандартная ошибка зависит от того, можем ли мы предположить, что две генеральных дисперсии равны между собой?

Стандартная ошибка разницы между двумя выборочными средними находится по формуле:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \ .$$

Если σ_1^2 и σ_2^2 неизвестны, то они могут быть оценены посредством выборочных дисперсий. Возможны два случая:

1. Если генеральные дисперсии равны между собой, то $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

   Тогда

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Лучшая оценка дисперсии достигается сложением двух выборочных дисперсий (s_1^2 и s_2^2) по сравнению с использованием одной или другой по отдельности. Лучшая оценка генерального стандартного отклонения вычисляется по формуле:

$$\hat{\sigma} = \sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{(n_1 + n_2 - 2)}}.$$

Поэтому лучшей оценкой требуемой стандартной ошибки является:

$$\hat{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

где

$$s^2 = \frac{\sum (x - \bar{x})^2}{n}.$$

С другой стороны, можно написать:

$$\hat{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)\,\hat{\sigma}_1^2 + (n_2 - 1)\,\hat{\sigma}_2^2}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

где.

$$\hat{\sigma}^2 = \frac{\sum (x - \bar{x})^2}{n - 1}.$$

Проверочная статистика для испытания гипотез на двух выборочных средних не относится к нормальному распределению, но следует стандартному t-распределению с $(n_1 + n_2 - 2)$ степенями свободы. Это может быть записано следующим образом:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

Поэтому мы используем тот же метод, что и в разделе 6.4.

2. Если генеральные дисперсии не равны друг другу, то каждая генеральная дисперсия должна быть оценена соответствующей выборочной дисперсией:

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2,$$

Следовательно:

$$\hat{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}},$$

где.

$$s^2 = \frac{\sum (x - \bar{x})^2}{n}.$$

Или

$$\hat{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}},$$

где

$$\hat{\sigma}^2 = \frac{\sum (x - \bar{x})^2}{n - 1}.$$

Проверочная статистика для испытания гипотезы по двум выборочным средним находится по формуле:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}.$$

Эта статистика не подчиняется ни нормальному распределению, ни t-распределению. Можно исполььзовать в качестве приближения t-распределения, но зависимость от числа степеней свободы более сложная. Если размеры выборки большие, $n \geq 30$, распределение этой новой статистики приблизительно нормальное, как описано в центральной предельной теореме. Мы будем обсуждать задачи этого вида только для случая больших выборок.

Для выбора подходящей проверочной статистики в случае, когда генеральные дисперсии не известны, мы должны знать, какое предположение принимается. Прежде всего нужно решить, можно ли считать неизвестные генеральные дисперсии равными или нет. Для принятия решения используется F-критерий.

❑ **Пример 6.12.** Для исследования качества определенного вида производимого полимера были сделаны выборки по 10 единиц из каждой последовательной серии

(n_1 и n_2) и определен процент химического вещества x в каждой выборке. Для первой серии средний процент составил 68,2% ($\bar{x}_1$) со стандартным отклонением s_1 = 0,70%. Для второй серии среднее содержание химического вещества x составило 67,0% ($\bar{x}_2$) со стандартным отклонением s_2 = 0,74%.

Имеется ли основание предполагать, что две серии содержат разный процент химического вещества x?

Решение

Нулевая гипотеза предполагает, что выборочные средние согласуются с двумя выборками, взятыми из нормальных генеральных совокупностей с одинаковой генеральной средней.

$$H_0: \mu_1 = \mu_2 ,$$

$$H_1: \mu_1 \neq \mu_2 .$$

Альтернативная гипотеза состоит в том, что две серии взяты не из одной и той же генеральной совокупности. Следовательно, должна быть проведена двусторонняя проверка.

Поскольку генеральные дисперсии неизвестны, мы должны использовать F-критерий для предположения, что две генеральных дисперсии равны друг другу. Для испытания с помощью F-критерия формулируем гипотезы:

$$H_0: \sigma_1^2 = \sigma_2^2 ,$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 ,$$

Будем испытывать нулевую гипотезу на 5%-ном уровне значимости, используя испытание с двумя границами. Это означает, что мы используем строки 0,025 F-таблицы в Приложении 2 с 9 и 9 степенями свободы.

$$F_{0,05/2,9,9} = 4,026 ,$$

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} s_1^2 = \frac{10}{9} \times 0,70^2 = 0,544,$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} s_1^2 = \frac{10}{9} \times 0,74^2 = 0,608.$$

Поскольку σ_2^2 является большей, то F-статистика равна:

$$F = 0,608/0,544 = 1,12 ,$$

Поскольку

$$1,12 < F_{0,05/2,9,9} = 4,026,$$

различия между дисперсиями не существенны на 5%-ном уровне. Наблюдаемые значения согласуются с нулевой гипотезой. Мы можем предположить, что две генеральных дисперсии равны друг другу, и использовать t-критерий для проверки гипотезы по выборочным средним.

Теперь мы продолжим испытание гипотез на двух выборочных средних на 5%-ном уровне значимости, используя t-критерий с двумя границами с числом степеней свободы: 10 + 10 − 2 = 18.

Из таблицы в Приложении 2 находим, что $t_{0,05/2,18} = 2,10$.

Поскольку мы предположили, что

$$\sigma_1^2 = \sigma_2^2 .$$

$$\hat{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{(10 \times 0,70^2 + 10 \times 0,74^2)}{10 + 10 - 2}\left(\frac{1}{10} + \frac{1}{10}\right)} =$$

$$= \sqrt{0,1153} = 0,3395 .$$

Проверочной статистикой является:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{SE}_{\bar{x}_1 - \bar{x}_2}} = \frac{68,2 - 67,0}{0,3395} = 3,53 .$$

Поскольку

$$3,53 > t_{0,05/2,18} = 2,10,$$

результат существенен на 5%-ном уровне. Очевидно, что наблюдения не согласуются с нулевой гипотезой. Мы отклоняем H_0 и принимаем H_1 как верную: две серии имеют разное содержание химического вещества X.

☐ **Пример 6.13.** Компания "JBO", производящая батарейки, утверждает, что в среднем период использования их батареек более длителен, чем батареек их конкурента — компании "Sparky". Ассоциация потребителей взяла случайную выборку $n_1 = 35$ батареек, производства компании "JBO" и испытала их на разрушение. Средний срок службы оказался равным 198 ч ($\bar{x}_1$) со стандартным отклонением 8,7 часа (s_1). Такая же проверка была сделана по продукции компании "Sparky" ($n_2 = 30$ батареек). Средний срок службы оказался равен 193,8 часа ($\bar{x}_2$) со стандартным отклонением 5,8 часа (s_2). Подтверждают ли эти результаты заявление "JBO"?

Нулевая гипотеза состоит в том, что две выборки взяты из нормальной совокупности с общей средней.

$$H_0: \mu_1 = \mu_2 ,$$

$$H_1: \mu_1 > \mu_2 .$$

Альтернативная гипотеза состоит в том, что батарейки "JBO" служат дольше, поэтому будем проводить одностороннюю проверку. Поскольку генеральные дисперсии неизвестны, нам придется использовать F-критерий, чтобы определить можно ли допустить что две генеральные дисперсии равны, прежде чем проводить проверку гипотезы по средним.

$$H_0\colon \sigma_1^2 = \sigma_2^2\,,$$

$$H_1\colon \sigma_1^2 \neq \sigma_2^2\,,$$

Произведем проверку нулевой гипотезы на 5%-ном уровне значимости, используя двустороннюю проверку:

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1}\, s_1^2 = \frac{35}{34} \times 8{,}7^2 = 77{,}92,$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1}\, s_1^2 = \frac{30}{29} \times 5{,}8^2 = 34{,}80.$$

Поскольку $\hat{\sigma}_1^2$ является большей, то F-статистика равна:

$$F = 77{,}92/43{,}80 = 2{,}24\,.$$

Из F-таблиц (Приложение 2) для уровня значимости 0,025 с одной границей при 34 степенях свободы по столбцу и 29 степенях свободы по строке таблицы $F_{0,025,34,29}$ лежит между $F_{0,025,30,29}$=2,092 и $F_{0,025,50,29}$ = 1,987.

Наша F-статистика больше этих двух значений, поэтому результат существенен на 5%-ном уровне и не согласуется с нулевой гипотезой. Мы не можем предположить, что генеральные дисперсии равны друг другу, следовательно и мы должны предположить:

$$\sigma_1^2 \neq \sigma_2^2\,.$$

Испытание гипотезы с помощью t-статистики не подходит. Поскольку размеры выборки большие (n_1 и n_2) ≥ 30, мы можем использовать нормальное распределение как аппроксимацию к правильной статистике для двух выборочных средних. Будем испытывать на 5%-ном уровне, используя нормальную проверку с одной границей. Из таблицы (Приложение 2) находим, что $z_{0,05}$ = 1,645. Поскольку мы предположили

$$\sigma_1^2 \neq \sigma_2^2,$$

то

$$\hat{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{(n_1 - 1)} + \frac{s_2^2}{(n_2 - 1)}} = \sqrt{\frac{8{,}7^2}{34} + \frac{5{,}8^2}{29}} = 1{,}84\,.$$

Проверочная статистика равна:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\hat{SE}_{\bar{x}_1 - \bar{x}_2}} = \frac{198 - 193,8}{1,84} = 2,28 .$$

Поскольку

$$2,28 > z_{0,05} = 1,645 .$$

Результат существенен на 5%-ном уровне. Очевидно, что факты не согласуются с нулевой гипотезой. Мы отклоняем H_0 на этом уровне. Следовательно, предполагаем, что альтернативная гипотеза H_1 верна. Средний срок службы батареек "JBO" больше, чем батареек "Sparky". Таким образом, наше заключение подтвердило правдивость рекламы компании "JBO".

❒ **Пример 6.14.** Компания "Electra plc" производит электрические компоненты. Выполнение работ по их сборке с максимальной производительностью требует обучения новых работников в течение месяца. Поскольку обучающая программа стоит немалых средств, руководство компании стремится сократить время, отведенное на обучение. Руководитель программы изобрел новый метод обучения. Руководство фирмы заинтересовано в проведении специального исследования для того, чтобы определить, сокращает ли новый метод период обучения.

Две группы по 10 новых работников n_1 и n_2 обучались три недели. Одна группа использовала новый метод, другая — стандартную процедуру. По окончании трехнедельного периода, каждый новый работник проводил сборку элемента и записывалось время сборки.

Таблица 6.2. **Время, затраченное на сборку единицы продукции, мин.**

Традиционное обучение группа 1	*Обучение новым методом группа 2*
32	35
37	31
35	29
28	25
41	34
44	40
35	27
31	32
34	31
30	33

Доказывают ли эти данные эффективность нового метода обучения?

Решение

Нулевая гипотеза состоит в предположении, что две выборки были взяты из нормальных генеральных совокупностей с одинаковой средней:

$$H_0\colon \mu_1 = \mu_2\ ,$$

$$H_1\colon \mu_1 > \mu_2\ .$$

Предположим, что традиционное обучение требует больше времени, поэтому испытание будет с одной границей.

Для сравнения выборочных средних, мы сначала должны испытать дисперсии с помощью F-критерия:

$$H_0\colon \sigma_1^2 = \sigma_2^2\ ,$$

$$H_1\colon \sigma_1^2 \neq \sigma_2^2\ ,$$

Мы будем испытывать нулевую гипотезу на 5%-ном уровне, используя испытание с двумя границами.

Вычислим по данным табл. 6.2. средние значения и дисперсии:

$$\bar{x}_1 = 34{,}7 \text{ мин. } s_1 = 4{,}691 \text{ мин.},$$

$$\bar{x}_2 = 31{,}7 \text{ мин. } s_2 = 4{,}026 \text{ мин.}$$

Поэтому

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1}\, s_1^2 = \frac{10}{9} \times 4{,}691^2 = 24{,}45\ ;$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1}\, s_1^2 = \frac{10}{9} \times 4{,}026^2 = 18{,}01\ .$$

Из строки 0,025 таблицы распределения с одной границей в Приложении 2 находим:

$$F_{0,05/2,9,9} = 4{,}026;$$

$$1{,}36 < F_{0,05/2,9,9}.$$

Результат не существенен на уровне 5%. Данные согласуются с нулевой гипотезой. Поэтому мы предполагаем, что две генеральные дисперсии равны друг другу. Теперь мы продолжим испытание гипотезы на двух выборочных средних.

Мы выбираем для испытания 1%-ный уровень значимости, используя t-критерий с одной границей с 10 + 10 − 2 = 18 степенями свободы.

Из таблицы в Приложении 2 находим, что:

$$t_{0,01,18} = 2{,}552.$$

Поскольку мы предположили, что

$$\sigma_1^2 = \sigma_2^2\ ,$$

$$\hat{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{(10 \times 4{,}691^2 + 10 \times 4{,}026^2)}{(10 + 10 - 2)}\left(\frac{1}{10} + \frac{1}{10}\right)} =$$

$$= \sqrt{4{,}246} = 2{,}06 .$$

Проверочной статистикой является:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\hat{SE}_{\bar{x}_1 - \bar{x}_2}} = \frac{34{,}7 - 31{,}7}{2{,}06} = 1{,}46 .$$

Поскольку

$$1{,}46 < t_{0{,}01{,}18} = 2{,}552,$$

т.е. различие в среднем времени сборки не существенно на 1%-ном уровне.

Факты согласуются с нулевой гипотезой. Мы принимаем Но, следовательно, нет оснований предполагать, что новый метод сократит в среднем время обучения.

6.9. ИСПЫТАНИЕ ГИПОТЕЗЫ ПО ДВУМ ВЫБОРОЧНЫМ ДОЛЯМ

В разделе **6.5** мы утверждали, что, если мы берем большую случайную выборку из генеральной совокупности, в которой доля случаев p соответствующей характеристики, следует биномиальному распределению, то выборочное распределение выборочной доли $\hat{p}$ приближается к нормальному распределению. Таким же образом мы находим, что если две большие выборки взяты независимо из двух биномиальных генеральных совокупностей, то статистика $(\hat{p}_1 - \hat{p}_2)$ нормально распределена со средней $(p_1 - p_2)$ и стандартной ошибкой:

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}} ,$$

где $\hat{p}$ — выборочная статистика, p — параметр генеральной совокупности и обе выборки большие, то есть (n_1 и n_2) больше или равны 30.

Нас обычно интересует, взяты ли или нет две выборки из биномиальных генеральных совокупностей с одинаковой долей случаев, то есть $p_1 = p_2$. Проверочная статистика приблизительно нормально распределена при больших размерах выборки:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\hat{SE}_{\hat{p}_1 - \hat{p}_2}} .$$

❒ **Пример 6.15.** Внутренние аудиторы большой компании интересуются системой обработки счетов доходов. Они взяли случайную выборку объемом $n_1 = 50$ законченных счетов и проверяют их подробно. Четыре из них оказались дефектными.

Тогда аудиторы предложили некоторые модификации в процедуре. Дав определенное время клеркам для приспособления к новым процедурам, аудиторы затем провели случайную выборку объемом $n_2 = 60$ завершенных счетов. Теперь они обнаружили три неисправных счета. Имеется ли какое-либо основание предполагать, что новые процедуры уменьшают ошибки?

Решение

Нулевая гипотеза предполагает, что две выборки случайно взяты из двух биномиальных генеральных совокупностей с равными долями дефектов:

$$H_0: p_1 = p_2 = p;$$

$$H_1: p_1 > p_2,$$

то есть предполагается, что новые процедуры сократили долю ошибок, поэтому здесь приемлемо испытание с одной границей.

Будем принимать наше решение на 5%-ном уровне значимости. Здесь подходит нормальное распределение, поскольку размеры обеих выборок большие. По таблице нормального распределения в Приложении 2, находим:

$$z_{0,05} = 1,645;$$

$$\hat{p}_1 = 4/50 = 0,08 \text{ и } \hat{p}_2 = 3/60 = 0,05.$$

Предполагая, что H_0 — верна, лучшая оценка доли дефектных счетов в генеральной совокупности достигается комбинированием долей двух выборок. В общем оказывается 7 дефектов из 110 случаев. Поэтому, лучшей оценкой генеральной доли является:

$$\bar{p} = 7/110 = 0,0636.$$

Поэтому

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}},$$

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{0,0636 \times 0,9364}{50} + \frac{0,0636 \times 0,9364}{60}} = 0,0467.$$

Проверочной статистикой является:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\hat{SE}_{\hat{p}_1 - \hat{p}_2}} = \frac{0,08 - 0,05}{0,0467} = 0,64.$$

Поскольку

$$0,64 < Z_{0,05} = 1,645,$$

результат не существенен на 5%-ном уровне. Факты согласуются с H_0 на данном уровне значимости. У нас нет причины предполагать, что модификации в системе обработки счетов сократили ошибки.

6.10. ИСПЫТАНИЕ ГИПОТЕЗЫ ПО СПАРЕННЫМ ДАННЫМ — ЗАВИСИМЫЕ ВЫБОРКИ

В некоторых ситуациях прямое испытание гипотез по двум выборочным средним не будет действенным, потому что выборки не являются независимыми. Имеются факторы, которые влияют на выборки каким-то неизвестным путем. В этих случаях возможно решение проблемы посредством соединения членов одной выборки с членами другой выборки.

После этого может быть проведено испытание гипотезы на среднюю разность между парными измерениями.

Поясним этот подход на следующем примере.

❑ **Пример 6.16.** Потребительская организация хочет сравнить износ определенного типа автомобильных шин, которые производятся заводами X и Y. Шины были установлены на различные автомобили. Каждый автомобиль имел или все X шины, или все Y шины. К сожалению, только позже было доказано, что любая разница между X и Y возникала из-за автомобиля или шофера, а не от шин. Во избежание этой путаницы, организация решила выбрать одну шину производителя X и одну шину — Y случайным образом и установить их на задние колеса шести машин. Результаты замены шин представлены в табл.6.3.

Таблица 6.3. **Результаты испытаний — расстояние в милях до замены шин**

Номер машин	*1*	*2*	*3*	*4*	*5*	*6*
Расстояние для X шин, тыс. миль	50,1	47,0	48,6	48,8	50,2	48,0
Расстояние для Y шин, тыс. миль	53,9	50,3	48,5	51,3	49,7	51,0

Имеется ли основание предполагать, что два выпуска шин имеет разную износоустойчивость?

Решение.

Поскольку износ шин будет зависеть от ряда факторов, таких как квалификация шофера, качество машины, дорожные условия, две выборки изначально не являются независимыми. Испытание гипотез на двух выборочных средних в этом случае не подходит, поскольку мы не будем знать исходит ли результат от шин или от других факторов. Мы можем исключить последнее, используя парные данные для вывода по одной выборке. Для каждой машины мы рассчитываем следующие данные:

Таблица 6.4. **Различие в расстоянии пробега в милях между шинами X и Y**

Номер машин	*1*	*2*	*3*	*4*	*5*	*6*
Расстояние для X шин, тыс. миль	50,1	47,0	48,6	48,8	50,2	48,0
Расстояние для Y шин, тыс. миль	53,9	50,3	48,5	51,3	49,7	51,0
Разница, d = X − Y	− 3,8	− 3,3	0,1	− 2,5	0,5	− 3,0

Средняя величина различий $x_d = -2,0$ тыс. миль.

Стандартное отклонение, $s_d = 1,675$ тыс. миль.

Нулевой гипотезой является:

H_0: $\mu_d = 0$: нет разницы между марками шин,

H_1: $\mu_d \neq 0$: марки шин имеют различную прочность.

Из H_1 следует, что необходимо проводить испытание с двумя границами. Будем принимать решение на 5%-ном уровне значимости, используя t-распределение с $(n-1)=5$ степенями свободы. Из таблицы t-распределения в Приложении 2 находим:

$$t_{0,05/2,5} = \pm\, 2{,}571.$$

Проверочной статистикой является:

$$t = \frac{\bar{x}_d - 0}{\hat{SE}_d} = \frac{2{,}0}{1{,}675\,\sqrt{5}} = -2{,}67\,.$$

Поскольку

$$-2{,}67 < -\, t_{0,05/2,5} = -\, 2{,}571,$$

результат существенен на 5%-ном уровне. Мы отклоняем H_0 и принимаем H_1. Таким образом резонно считать, что существует разница между средней износоустойчивостью шин, производимых двумя производителями.

6.11. НЕПАРАМЕТРИЧЕСКИЕ ИСПЫТАНИЯ ГИПОТЕЗ – КРИТЕРИЙ ХИ-КВАДРАТ

В прошлых разделах испытание касалось сравнения выборочной статистики с соответствующими генеральными параметрами. Для больших выборочных совокупностей мы предпологали, что генеральные совокупности нормальны или приблизительно нормальны. Теперь мы будем рассматривать примеры испытаний гипотез, которые не требуют ни этого предположения, ни использования генеральных параметров. Эта группа испытаний относится к **непараметрическим испытаниям**. Общая процедура испытания гипотез та же, что и для параметрических испытаний. Но расчет проверочной статистики другой.

Рассмотрим самый общий непараметрический **критерий хи-квадрат**. Это — метод сравнения ряда наблюдаемых частот с ожидаемыми частотами, если верна нулевая гипотеза. Мы будем использовать этот метод для проверки взаимосвязи признаков.

Признак — это характеристика переменной. Характеристики обычно относят к категории. Например цвет глаз — это признак человека, может быть отнесен к категориям: карие, голубые, серые или зеленые. Положение счетов клиентов в банке может быть отнесено к категориям: "всегда в кредите", "обычно в кредите", "часто превышает кредит", "постоянно в долгу". Месячные суммы выручки от продажи товаров могут быть описаны как "высокие", "средние", "низкие".

Предположим, нас интересуют две разные характеристики переменной и мы хотим знать существует ли между ними какие-либо связи. Например, у нас имеются данные по оценкам, полученным группой студентов на экзамене по бухгалтерскому учету и на экзамене по математике. Нас интересует, существует ли связь между оценками, полученными на экзамене по бухучету и тем, сдали ли студенты или провалили экзамен по математике.

Могут быть следующие категории:

Таблица 6.5. **Пример таблицы сопряженности**

Экзамен по математике	*Оценки по экзамену бух. учета*			
	A	*B*	*C*	*Не сдали*
Сдали				
Не сдали				

Число или частота студентов, которые сдали экзамен по математике и получили оценку А по бухгалтерскому учету, записано в верхней левой части таблицы. Число студентов, не сдавших математику и получивших оценку А по бухгалтерскому учету, записывается в нижней левой части таблицы и т.д. Такой тип таблицы называется **таблицей сопряженности**.

Таблица 6.5. имеет две строки и четыре столбца, т.е. является таблицей 2 × 4 (два на четыре). Используя соответствующую нулевую гипотезу, мы можем рассчитать число студентов, которое ожидается в каждой клетке. Если нулевая гипотеза верна, различия между наблюдаемыми и ожидаемыми частотами будут небольшие. Будем использовать те же правила для решения, как и в прошлом испытании. Проверочная статистика рассчитывается на основе разницы между наблюдаемыми и ожидаемыми частотами для всех клеток таблицы.

Если обозначить наблюдаемую частоту события f_o и ожидаемую частоту — f_E, то $(f_o - f_E)$ — различия между наблюдаемой и ожидаемой частотами. Проверочной статистикой будет служить:

$$\sum \left[\frac{(f_o - f_E)^2}{f_E}\right].$$

Возведение в квадрат разницы $(f_o - f_E)$ необходимо для того, чтобы избежать нулевого эффекта при суммировании отрицательных и положительных величин. К тому же, чтобы достичь независимости от значения фактических частот, квадраты отклонений делятся на ожидаемые частоты. Это нормализует все величины. Получаемая статистика подчиняется χ^2-распределению при достаточно больших значениях ожидаемых частот. Ориентиром обычно служит условие:

$$f_E \geq 5 .$$

Если одна или более ожидаемых частот меньше, чем 5, то категории должны быть скомбинированны до тех пор, пока частота не превысит установленного значения.

Для таблиц сопряженности 2 × 2, в которых сумма частот меньше или равна 100, иногда применяется корректировка — **поправка Йетса**. Тогда проверочная статистика вычисляется по следующей формуле:

$$\chi^2 = \sum \left[\frac{(|f_o - f_E| - 0{,}5)^2}{f_E}\right].$$

Такая поправка проводится потому, что χ^2 является непрерывным распределением, а данные выборки — дискретные. В гл. 2 мы обсуждали необходимость такой корректировки при использовании нормального распределения для аппроксимации дискретного распределения. Для больших выборок разница между исправленными и неисправленными величинами χ^2 является небольшой и в таких случаях корректировка не требуется.

Как мы установили в гл. 4, форма χ^2-распределения зависит от числа степеней свободы в данной задаче. При использовании таблиц сопряженности число степеней свободы равняется:

$$(r - 1)(c - 1),$$

где r и c — число строк и столбцов в таблице сопряженности. Если таблица имеет только одну строку, то число степеней свободы: (c − 1).

❐ **Пример 6.17.** Компания "Autosure plc" (товарищество с ограниченной ответственностью) является крупной страховой компанией, специализирующейся на страховании автомобилей. Обычной политикой компании является начисление различных премий в соответствии с размером машины, которая страхуется. Чем больше машина, тем больше выплаты. Однако такая политика оказывается неправильной, поскольку руководители отделов сообщают о большой частоте заявлений о случаях личного ущерба для машин меньших размеров. Один из аналитиков компании исследовал данные из 566 недавно поступивших заявлений. Собранные данные представлены в таблице 6.6.

Таблица 6.6. **Данные 566 заявителей**

Тип заявления	*Размер страхуемой машины*		
	маленький	*средний*	*большой*
Есть личный ущерб	120	57	42
Нет личного ущерба	149	105	93

Указывают ли данные на то, что частота заявлений о личном ущербе связана с размером страхуемой машины?

Решение

Для начала мы должны установить нулевую гипотезу. Если нет связи между типом страхового случая и размером машины, то будем предполагать, что частота заявлений в таблице зависимости будет пропорциональна итоговым данным по каждой категории:

H_0: нет связи между типом заявления и размером страхуемой машины;

H_1: есть связь между заявлением и размером машины.

Будем испытывать гипотезу на 5%-иом уровне значимости используя χ^2 критерий с (2–1)(3–1)=2 степенями свободы. Из таблиц в Приложении 2 находим, что: $\chi^2_{0,05,\,2} = 5{,}991$.

Для расчета проверочной статистики χ^2 мы должны определить ожидаемые частоты из итоговых данных по каждой категории.

Таблица 6.7. **Наблюдаемые частоты**

Тип заявления	*Размер страхуемой машины*			
	маленький	*средний*	*большой*	*итого*
Есть личный ущерб	120	57	42	219
Нет личного ущерба	149	105	93	347
Итого	269	162	135	566

Имеется 566 заявлений, из которых в 219 фигурирует личный ущерб; доля таких заявлений составляет: 219/566. Охвачено всего 269 маленьких машин, и если нет связи между двумя факторами, то можно ожидать, что 219/566 из 269 относится к категории "маленький автомобиль". Таким образом, ожидаемая частота в первой клетке таблицы равна:

$$(219/566) \times 269 = 104{,}08.$$

Подобно этому можно рассчитать ожидаемое число заявлений в других категориях. Все результаты показаны в ниже представленных таблицах. Ожидаемые частоты записаны слева в десятичных дробях. Поскольку они являются средними величинами, то не могут быть округлены до целого значения.

Таблица 6.8. **Расчет ожидаемых частот**

Тип заявления	*Размер страхуемой машины*			
	маленький	*средний*	*большой*	*итого*
Есть личный ущерб	269× 219/566	162 × 219/566	135 × 219/566	219
Нет личного ущерба	269 × 347/566	162 × 347/566	135 × 347/566	347
Итого	269	162	135	566

Ожидаемые частоты представлены в табл. 6.9.

Таблица 6.9. **Ожидаемые частоты**

Тип заявления	*Размер страхуемой машины*			
	маленький	*средний*	*большой*	*итого*
Есть личный ущерб	104,1	62,7	52,2	219
Нет личного ущерба	164,9	99,3	82,8	347
Итого	269	162	135	566

Критерий χ^2 находится по формуле:

$$\chi^2 = \sum \left[\frac{(f_o - f_E)^2}{f_E}\right].$$

Расчет χ^2 приведен в табл. 6.10

Таблица 6.10. **Расчет χ^2**

f_O	f_E	$(f_O - f_E)$	$(f_O - f_E)^2$	$(f_O - f_E)^2/f_E$
120	104,1	15,9	252,81	2,43
57	62,7	− 5,7	32,49	0,52
42	52,2	−10,2	104,04	1,99
149	164,9	−15,9	252,81	1,53
105	99,3	5,7	32,49	0,33
93	82,8	10,2	104,04	1,26

Найденное значение χ^2 = 8,06 показано на рис. 6.13.

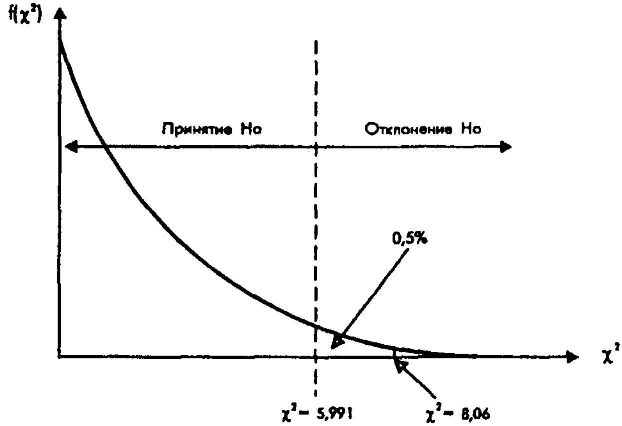


Рис.6.13. χ^2 критическое значение на 5%-ном уровне значимости при двух степенях свободы

Поскольку

$$8{,}06 > \chi^2_{0{,}05,2} = 5{,}991,$$

результат статистически значим на 5%-ном уровне. Мы отклоняем H_0 на этом уровне и принимаем H_1. Мы можем быть вполне уверены, что данные указывают на связь между заявлениями, в которых отмечается личный ущерб, и размером страхуемой машины. На этой ступени мы не знаем, какой это вид связи. Компания считает, что она получает больше заявлений о личном ущербе при страховке машин маленьких размеров. Чтобы убедиться так ли это, покажем составляющие χ^2 (табл.6.11.).

Таблица 6.11. **Различия между наблюдаемыми и ожидаемыми частотами**

Тип заявления	*Размер страхуемой машины*		
	маленький	*средний*	*большой*
Есть личный ущерб	+ 15,9	− 5,7	− 10,2
Нет личного ущерба	− 15,9	+ 5,7	+ 10,2

Таблица подтверждает подозрения компании: число заявлений о личном ущербе больше, от владельцев маленьких машин. Учитывая сверхзатраты на заявления о личном ущербе, компания должна пересмотреть свою политику начислений.

❒ **Пример 6.18.** Международная фирма подготовки бухгалтеров принимает 150 выпускников школ для обучения бухгалтерским методам по результатам персональной беседы с каждым кандидатом. Управляющий хочет сравнить результаты обучения во время первого года обучения со школьным аттестатом, чтобы выяснить, есть ли между ними связь. Собранные данные приведены ниже:

Таблица 6.12. **Наблюдаемые частоты**

Результаты обучения	*Школьный аттестат*		
	хорошо	*средне*	*плохо*
Хорошо	18	12	5
Средне	39	34	18
Плохо	6	11	7

Решение

Для начала мы должны установить подходящую нулевую гипотезу. H_0 должна быть так выбрана, что бы мы смогли рассчитать ожидаемые частоты. Если мы предполагаем, что нет связи между результатами обучения в первом году и школьным аттестатом, то будем ожидать частоты обучающихся в таблице сопряженности пропорционально общим числам в каждой категории.

H_0: Нет связи между результатами обучающихся первый год и их конечными оценками в школе.

H_1: Имеется некоторая связь между результатом обучения и аттестатом.

Будем испытывать гипотезу на 5%-ном уровне значимости, используя χ^2 критерий с $(3 - 1)(3 - 1) = 4$ степенями свободы.

Причем все наблюдаемые частоты не менее 5.

Для расчета проерочной величины χ^2 мы определяем общее число обучающихся каждой категории и используем это для нахождения ожидаемых частот.

Таблица 6.3. **Общее число обучающихся в каждой категории**

Результаты обучения	*Школьный аттестат*			
	хороший	*средний*	*плохой*	*итого*
Хорошие	18	12	5	35
Средние	39	34	18	91
Плохие	6	11	7	24
Итого	63	57	30	150

Имеется 150 обучающихся, 35 из них имеют хороший результат обучения во время первого года. Поэтому доля получения хорошего результата 35/150.

Мы используем это для расчета ожидаемых частот для верхней строки таблицы.

Имеется 63 обучающихся с хорошим дипломом об окончании школы, и если нет связи между двумя факторами, то будем предполагать, что 35/150 из 63 относится к первой категории в среднем.

Ожидаемой частотой в первой клетке является:

$$(35/150) \times 63 = 14{,}7.$$

Подобно этому, ожидаемое число обучающихся с хорошими результатами среди получивших средний школьный диплом равно:

$$(35/150) \times 57 = 13{,}3,$$

в то время как ожидаемое число обучающихся с хорошими результатами среди тех, кто имеет плохой школьный диплом равно:

$$(35/150) \times 30 = 7{,}0.$$

Следует отметить, что эти три ожидаемые частоты при суммировании дают итог строки:

$$14{,}7 + 13{,}3 + 7{,}0 = 35.$$

Ожидаемые частоты в других строках таблицы вычисляются подобным образом. Доля обучающихся со средними результатами первого года обучения равняется 91/150, поэтому ожидаемое число обучающихся среди имеющих хороший школьный аттестат равняется:

$$(91/150) \times 63=38,2.$$

Остающиеся ожидаемые частоты рассчитываются таким же образом. Окончательное распределение показано в табл. 6.14.

Таблица 6.14. **Ожидаемые частоты**

Результаты обучения	*Школьный аттестат*			
	хороший	*средний*	*плохой*	*итого*
Хорошие	14,7	13,3	7	35
Средние	38,2	34,6	18,2	91
Плохие	10,1	9,1	4,8	24
Итого	63	57	30	150

Все строки и столбцы таблицы ожидаемых частот должны иметь такие же итоги, как и в исходной таблице сопряженности. Мы имеем только одну ожидаемую частоту, которая меньше 5 в клетке (3,3). Для того, чтобы использовать χ^2 распределение, мы должны соединить две категории. С точки зрения испытания не важно, сокращаем ли мы число категорий школьного аттестата или результатов обучения. Нужно выбрать самое значительное для задачи. Предположим в этом случае, мы определим школьный аттестат как "хорошо" или "не хорошо", т.е. объединим столбцы "средне" и "плохо" в таблице. Тогда таблица сопряженности будет следующей:

Таблица 6.15. **Исправленные наблюдаемые частоты.**

Результаты обучения	*Школьный аттестат*		
	хорошо	*не хорошо*	*итого*
Хорошие	18	17	35
Средние	39	52	91
Плохие	6	18	24
Итого	63	87	150

Соответствующие ожидаемые частоты представлены в табл. 6.16:

Таблица 6.16. **Исправленные ожидаемые частоты**

Результаты обучения	*Школьный аттестат*		
	хорошо	*не хорошо*	*итого*
Хорошие	14,7	20,3	35
Средние	38,2	52,8	91
Плохие	10,1	13,9	24
Итого	63	87	150

Ожидаемые частоты теперь превышают 5 и χ^2 испытание может быть продолжено как и раньше, но с

$$(3-1)(2-1)=2$$

степенями свободы вместо 4.

Из таблицы в Приложении 2 находим:

$$\chi^2_{0,05,\,2}=5,991\;,$$

$$\chi^2=\sum\left[\frac{(f_o-f_E)^2}{f_E}\right].$$

Это значение рассчитано в табл. 6.17:

Таблица 6.17. **Расчет χ^2**

f_O	f_E	(f_O-f_E)	$(f_O-f_E)^2$	$(f_O-f_E)^2/f_E$
18	14,7	3,3	10,89	0,74
17	20,3	−3,3	10,89	0,54
39	38,2	0,8	0,64	0,02
52	52,8	−0,8	0,64	0,01
6	10,1	4,1	16,81	1,66
18	13,9	−4,1	16,81	1,21

Проверка = 0 $\qquad\chi^2=4,18$

Поскольку:

$$4,18<\chi^2_{0,05,\,2}=5,991\;,$$

результат не значим на уровне 5%. Мы вполне уверены, что наши наблюдения согласуются с H_0 и мы принимаем ее на этом уровне. Мы заключаем, что не существует связи между результатами обучения во время первого года и школьным аттестатом.

РЕЗЮМЕ

Процедура всех испытаний гипотез одна и та же. Мы решаем, на какой вопрос хотим ответить. Мы устанавливаем нулевую и альтернативную гипотезы H_0 и H_1. Это объясняет заключения, которые мы можем сделать, и указывает статистику, используемую для испытания. Мы выбираем уровень значимости. Вместе с гипотезами устанавливаются критические величины или для испытания с одной границей, или для испытания с двумя границами. Мы производим случайную выборку данных, рассчитываем проверочную статистику, сравниваем её с критическими величинами и делаем наши заключения. Основные критерии кратко изложены ниже.

Имеются два критерия для сравнения одной выборочной средней $\bar{x}$ с генеральной средней μ. Если мы знаем генеральную дисперсию σ^2, то проверочной статистикой является:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Если мы не знаем σ^2, то оцениваем её, используя выборочную дисперсию s^2, и проверочной статистикой является:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{(n - 1)}}$$

с $(n-1)$ степенями свободы.

Для испытания одной выборочной доли $\hat{p}$ по сравнению с генеральной долей p, если объем выборки не больше 30 и np и n $(1-p) \geq 5$, проверочная статистика рассчитывается следующим образом:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}.$$

Ситуация усложняется, когда мы хотим испытать две независимые выборки.

Для испытания двух выборочных дисперсий (s_1^2 и s_2^2) используем:

$$F_{(n_1 - 1), (n_2 - 1)} = \frac{n_1 s_1^2}{(n_1 - 1)} \Big/ \frac{n_2 s_2^2}{(n_2 - 1)}$$

с большей дисперсией в числителе.

Для испытания выборочных средних ($\bar{x}_1$ и $\bar{x}_2$), если мы знаем генеральные отклонения (σ_1^2 и σ_2^2), статистическим критерием является:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Чтобы проверить две выборочные средние, если нам не известны генеральные дисперсии, мы должны сначала проверить гипотезу о равенстве генеральных дисперсий, используя F-критерий. Если F-критерий показывает, что мы можем предположить $\sigma_1^2 = \sigma_2^2$, то для проверки гипотезы о равенстве средних мы используем проверочную статистику:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

с $(n_1 + n_2 - 2)$ степенями свободы.

Если F-критерий показывает, что мы должны принять $\sigma_1^2 \neq \sigma_2^2$, то поскольку объемы выборок по крайней мере 30, то статистикой является:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{(n_1 - 1)} + \frac{s_2^2}{(n_2 - 1)}}}.$$

Для испытания выборочных долей $\hat{p}_1$ и $\hat{p}_2$, если обе выборки не менее 30, статистикой является:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

где $\bar{p}$ — доля из объединенных выборочных совокупностей. (Выше мы охватывали две ситуации, когда объемы выборок меньше, чем 30).

Если мы имеем две зависимые выборки, то берем разницы между зависимыми парами и обрабатываем их как единую выборку. Критериальной статистикой является:

$$t = \frac{\bar{x}_d - 0}{s_d / \sqrt{(n-1)}}$$

с $(n-1)$ степенями свободы.

Если мы хотим сравнить частоты, то используем непараметрический χ^2 критерий. Проверочная статистика всегда рассчитывается тем же путем, используя наблюдаемые и ожидаемые частоты, то есть:

$$\chi^2 = \sum \left[\frac{(f_o - f_E)^2}{f_E}\right].$$

Если мы имеем два свойства, то используем испытание для того, чтобы увидеть, имеется ли связь между ними. Строится таблица сопряженности с наблюдаемыми частотами. Таблица и нулевая гипотеза используются для расчета ожидаемых частот. Число степеней свободы равняется:

(число строк − 1) × (число столбцов − 1).

УПРАЖНЕНИЯ

Упражнение 6.1

Частная компания "Betta Bolt" (товарищество с ограниченной ответственностью) производит различные промышленные крепежные детали. Одним из видов выпускаемой продукции является противомагнитный болт. Болты производятся со средней длиной 5 см с известным стандартным отклонением в 0,05 см. Из последнего выпуска продукции была взята выборка объемом 25 единиц. Средняя длина болтов в выборке оказалась равной 5,025 см. Какие выводы могут быть сделаны о средней длине болтов в этом производственном выпуске?

Упражнение 6.2

Частная компания "Shush" выпускает прохладительные напитки. На данный момент она производит поллитровые бутылки лимонада. Наполняющая машина установлена так, что наполняемость бутылки равна 500 мл. Каждый час отбираются наугад 30 бутылок и проверяются для того, чтобы определить, соответствует ли содержимое, необходимому объему. Средний объем в последней выборке равняется 495, 6 мл и стандартное отклонение составляет 8,3 мл. Имеется ли основание предполагать, что требуется переналадка машины?

Упражнение 6.3

Компания упаковывает продукт своего производства — специально просушенное кофе — в пакетики весом 500 г. Руководство компании заявляет, что в среднем пакетики содержат по крайней мере 500 г. Служащий из Министерства торговых стандартов посетил компанию, ее завод, с намерением проверить это утверждение. Служащий случайно выбрал 10 пакетиков и взвесил их содержимое. Было установлено следующее содержание пакетиков:

Номер пакетика:	1	2	3	4	5	6	7	8	9	10
Вес, г	497	498	502	503	495	496	497	500	501	496

Требуется:

1. Установить нулевую гипотезу и альтернативную гипотезу для проведения соответствующего испытания.
2. Провести испытание H_0 используя вышеприведенные данные на 1% уровне значимости. Сделать выводы.

Замечание. Можно предположить, что вес пакетиков нормально распределен.

Упражнение 6.4

Компания по производству глиняной посуды предполагает, что 10% выпуска кофейных чашек имеют дефекты и должны продаваться по сниженным ценам, как товар второго сорта. Недавно компания ввела новые процедуры контроля качества. Для оценки эффективности таких процедур менеджер отобрал случайно 100 чашек и исследовал их. Он нашел 7 бракованных изделий. Можно ли заключить на основе только этого факта, что наблюдается значительное улучшение качества продукта?

Упражнение 6.5

Большой супермаркет имеет в продаже кукурузные хлопья (сухой завтрак). Эти хлопья продаются в коробках с маркой изготовителя, а также в простых коробках с маркой супермаркета. Производитель хлопьев наполняет оба типа коробок. Однако есть некоторое подозрение, что поставщик в среднем наполняет коробки с маркой супермаркета меньшим количеством хлопьев. Из выпуска двух типов упаковок были взяты случайные выборки и рассчитана следующая статистика:

Упаковки с маркой производителя — 25 упаковок — $\bar{x}_1 = 503{,}6$г, $s_1 = 5{,}03$ г.

Упаковки с маркой супермаркета — 30 упаковок — $\bar{x}_2 = 498{,}2$г, $s_2 = 4{,}39$ г.

Есть ли основание для оправдания такого подозрения?

Упражнение 6.6

Выработка является главным фактором, влияющим на общие затраты связанные с химическим процессом. Частная компания с ограниченной ответственностью "СОМ" постоянно пытается повысить выработку своего завода, который производит химическое вещество АВС1. Завод проработал 8 недель до того, как была проведена последняя регулировка, и были получены следующие недельные выработки:

Неделя	1	2	3	4	5	6	7	8
Процент выработки	91,0	90,4	90,8	91,4	90,0	89,8	91,6	90,6

Установки завода затем были отрегулированы и завод проработал следующие 6 недель. Были получены следующие данные выработки:

Неделя	1	10	11	12	13	14
Процент выработки	91,0	90,8	91,2	91,7	90,9	91,6

Есть ли основание предполагать, что регулировки были удачны и выработка АВС1 возросла?

Упражнение 6.7

Частная компания с ограниченной ответственностью "СОМ" также производит химическое вещество АВС2, используя два разных процесса. Компания хочет сравнить выработку этих процессов. Известно, что на выработку может серьезно влиять чистота сырья. Управляющий производством выбирает восемь пачек сырья и использует их в двух процессах. Получены следующие данные:

Номер пачки	1	2	3	4	5	6	7	8
Выработка процесса 1	70,1	67,0	68,6	68,8	70,2	68,0	69,5	68,4
Выработка процесса 2	73,9	70,3	68,5	71,3	69,7	71,0	69,8	68,2

Какие выводы вы можете сделать о выработке химического вещества ABC2 процессом 1 и процессом 2?

Упражнение 6.8

Внешний аудитор проверяет систему учета малой компании. Первая проверка 100 сделок показала, что 56 сделок ошибочны. Компании предоставили месяц для исправления системы. Когда аудитор провел вторую проверку, он обнаружил, что 28 сделок ошибочны из 78 проведенных сделок. Есть ли какое-нибудь основание предпологать, что доля ошибок уменьшилась между проверками?

Упражнение 6.9

Частная отечественная химическая корпорация была уполномочена провести опрос потребительского рынка для выявления потребительских предпочтений пяти марок моющих средств, которые она производит. Было опрошено 200 человек и получены следующие ответы:

Марка	A	B	C	D	E
Процент предпочитаемой марки	19	20	15	25	21

Показывают ли данные какую-либо предпочтительность отдельной марки?

Упражнение 6.10

Относится к упражнению 6.9. Отечественная химическая корпорация решила начать новую рекламную компанию для марки E. Было опрошено 100 человек. Были получены следующие результаты:

Марка	A	B	C	D	E
Процент предпочитаемой марки	20	19	14	25	22

1. Показывают ли данные предпочтительность какой-либо марки над остальными?
2. Изменилась ли та часть рынка, которая предпочитает покупать товар марки E в результате проведения рекламной компании?
3. Изменилась ли структура предпочтений в результате рекламной компании?

Упражнение 6.11

Опрос касался ссуд банка или строительного общества. Была проведена случайная выборка и опрошено 250 человек, которые впервые брали деньги под залог в течение последнего года. Результаты опроса сведены в таблицу:

Источник ссуды	*Меньше 10000 ф. ст.*	*10000 ф. ст. и более*
Банк	25	27
Строительное общество	125	75

Проведите соответствующее статистическое испытание для того, чтобы выяснить, есть ли какие-либо основания для обнаружения связи между источником и размером ссуды.

Упражнение 6.12

Лондонский Мидлэнд-Банк и Шотландский Банк в настоящее время пересматривают систему отчетности для владельцев текущих счетов. В соответствии с настоящей системой клиенты получают отчетность каждый квартал (каждые три месяца), и имеют право получения отчетности в любое другое время. Однако предполагалось, что владельцы текущих счетов должны определить, как часто они хотели бы получать отчетность, выбрав один из следующих четырех вариантов:

1) Безрегулярной отчетности.
2) Каждый месяц.
3) Каждые 3 месяца.
4) Каждые 6 месяцев.

К тому же для клиентов сохранялась бы свобода запроса отчетности. По оценкам ожидается более 600000 нерегулярных запросов отчетности в год. В настоящее время банк имеет приблизительно 1,2 млн текущих счетов.

Для оценки приемлемости этой новой системы была взята случайная выборка владельцев текущих счетов объемом 150. Были установлены предпочтения по поводу выбора одного из четырех вариантов получения отчетности. Результаты оказались следующими:

Предпочитаемая частота	*Без отчетности*	*Каждый месяц*	*Каждые 3 месяца*	*Каждые 6 месяцев*
Число клиентов	11	65	61	13

Требуется:

1. Построить частотное распределение регулярной отчетности в год для выборки объемом 150 клиентов и определить среднюю и стандартное отклонение вашего распределения.
2. Определить доверительный интервал с вероятностью 95% для истинной средней количества регулярных отчетностей в год и объяснить значение этого интервала.
3. Вся отчетность поставляется ежедневно из банковского компьютерного центра. В соответствии с настоящими условиями возможен выпуск 33000 отчетов каждый день. Предполагая, что компьютерный центр работает 300 дней в году, прокомментируйте, достаточной ли этих возможностей для вероятного числа отчетов, которые должны быть выпущены. Ваш ответ должен включать соответствующее испытание значимости, и должны быть четко установлены предположения, принятые вами.

Упражнение 6.13

Компания с ограниченной ответственностью "Martin Electronic Controls" — мелкий производитель электрических компонентов — специализируется на промышленном контрольном оборудовании.

1. Главный бухгалтер провел проверку 300 последних накладных по продаже для того, чтобы быть уверенным, что число незначительных погрешностей — таких, как неправильная дата, неправильное суммирование, не четкая печать или отсутствие подписи ответственного лица — лежит внутри принятой компанией нормы ошибки в 2%. Если 9 из 300 выборочных единиц содержат незначительную неисправность, то имеется ли статистическое основание того, что обусловленная норма ошибки нарушена?
2. Сколько из 300 выбранных единиц, описанных в п.1 потребуется для определения того, что доля накладных, содержащих незначительные погрешности, превысит установленную долю ошибки на 5%-ном уровне значимости?

Упражнение 6.14

При анализе портфеля инвестиций двух схожих компаний, которые оперируют на одном рынке, проведено сравнение прибыли на акцию. Результаты деятельности компаний за последние шесть лет приведены в таблице:

Год	*Прибыль на акцию (пенсов)*	
	Компания A	*Компания B*
19X0	14,3	13,8
19X1	15,6	14,6
19X2	17,2	16,4
19X3	16,4	16,8
19X4	14,9	15,0
19X5	17,6	16,4

Были рассчитаны средние и стандартные отклонения результатов каждой компании и затем определена t-статистика.

Компания	*Средняя (пенсов)*	*Стандартное отклонение (пенсов)*
A	16,0	1,2977
B	15,5	1,2050

$$t = \frac{16{,}0 - 15{,}5}{\sqrt{\frac{1{,}2977^2}{6} + \frac{1{,}2050^2}{6}}} = 0{,}69$$

(10 степеней свободы).

Аналитик заключил, что результат незначимый при испытаниях с одной и с двумя границами, и поэтому он решил, что эти две компании имеют в среднем одинаковую прибыль на акцию.

Требуется:

1. Оценить критически анализ, который был проведен, и указать три причины при которых использованная процедура и сделанное заключение являются неподходящими и неправильными.
2. Провести соответствующий анализ данных и установить свои заключения.

Упражнение 6.15

а) Укажите обстоятельства, при которых объединенная оценка дисперсии должна быть использована для испытания разницы между двумя средними.

б) Главная нефтяная компания рассматривает добавку к их бензину для увеличения количества миль пробега на литр машиной соответствующей марки. Имеются две добавки Апэкс и Бимакс от разных поставщиков. Для испытания действия этих добавок независимое агенство испытало 8 машин с Апэкс-добавкой и 10 машин с Бимакс-добавкой. Все машины одной марки, производства и размера. Достигнутые результаты приведены ниже:

Добавка	*Расстояние на литр, миль*									
Апэкс	10,4	10,7	10,2	10,5	9,8	10,0	10,6	10,8		
Бимакс	10,7	11,3	9,9	10,8	10,1	10,4	11,1	10,2	10,5	10,6

По очень большой выборке машин такого типа среднее расстояние — 10 милей на литр бензина без добавок.

Требуется:

1) Оценить увеличение пробега в милях на литр для каждой из двух добавок.

2) Испытать является ли значимым увеличение миль на литр в результате использования Бимакс-добавки при сравнении с установленным потреблением бензина.

3) Провести испытание для того, чтобы определить, значима ли разница между средним потреблением бензина при Апэкс-добавке и Бимакс-добавке.

4) Предполагая, что возрастание средней в милях на литр, данное в п. 1) может считаться типичным при использовании добавок, и, что цена на бензин без добавок составляет 40 пенсов за литр, рассчитайте для каждой добавки максимум повышения цены за литр бензина, чтобы это было экономически выгодно для автомобилиста.

Упражнение 6.16

1. Объясните кратко значение термина "значимое различие" в статистическом испытании гипотез.
2. Четыре производственных завода компании с ограниченной ответственностью "Zeus" находятся в Аубридже, Бидоне, Крамбурне и Дипуле. Выбранные случайно рабочие этих четырех заводов были опрошены для выяснения их

мнения, должна ли заработная плата быть основана, как предлагает компания, на производительности труда.

Данные опроса приведены ниже:

Взгляд	*Заводы компании*			
	Аубридж	*Бидон*	*Крамбурн*	*Дипул*
За	80	40	50	25
Против	35	30	40	25

Требуется:

а) Анализ χ^2 таблицы сопряженности дал величину χ^2, равную 7,34. Испытайте гипотезу, что нет значимой разницы во взглядах рабочих разных заводов. Объяснить, что нужно сделать для достижения ясного решения.

б) Рабочих в Аутбридже, Бидоне, Крамборне и Дипуле попросили также указать, кому из них меньше 40 лет. Число тех, кому меньше 40 лет для всех четырех заводов было: 75, 48, 54 и 57 соответственно, из которых 60, 28, 30 и 45 собственно были за предложение компании относительно заработной платы. Проанализируйте две результирующие 2×4 таблицы сопряженности отдельно для двух возрастных групп для испытания гипотезы, что нет значимой разницы во взглядах между рабочими этих заводов.

в) Оцените общий процент рабочих, которые выступают за сдельную оплату труда для каждой возрастной группы. Значимо ли различие процентов по этим группам?

Глава 7. СТАТИСТИЧЕСКИЙ КОНТРОЛЬ КАЧЕСТВА

7.1. ВВЕДЕНИЕ

Качество — наиболее важная характеристика любого бизнеса, которая должна быть в центре внимания всех работников безотносительно к их служебному положению. Контроль качества — наиболее важный элемент достижения качества и может рассматриваться на двух уровнях. На более низком уровне для управления качеством в целях уменьшения числа бракованных изделий используются статистические методы. На более высоком уровне качество можно рассматривать как стиль работы всей компании по отношению к поставщикам, производимой продукции и к заказчикам. Один из основных тезисов лидеров движения за качество состоит в том, что применение статистических методов на начальном уровне имеет весьма небольшой эффект, если не подкрепляется соответствующим стилем работы организации в целом. Качество должно стать неотъемлемой частью любой организации.

Поэтому абсолютно необходимо понять, что статистические методы контроля качества должны применяться при наличии соответствующих общих условий. Эти условия были описаны в начале 50-х годов доктором У. Е. Демингом (W. E. Deming), одним из первых специалистов в области контроля качества.

Г-н Деминг утверждал, что заказчики могут получить более, чем просто удовлетворение от поставки продукции; они должны получить такое удовольствие от нее, что вопрос о следующем заказе у них просто не будет вызывать сомнений. Для того, чтобы этого достичь, компания должна предложить заказчику товар высочайшего качества по разумной цене. Г-н Деминг подчеркивал, что указанные цели не являются несовместимыми. Улучшение качества совсем не означает увеличение цены товара. На самом деле улучшение качества приводит к снижению издержек, поскольку число товаров, которые придется отнести к браку или подвергнуть переработке, заметно снижается. Кроме того, ввиду повышения производительности уменьшается число сбоев в производственном процессе. (рис. 7.1).

Для повышения качества продукции необходимо знать, каковы требования заказчика. Следовательно, составной частью цикла качества должны стать обследования потребителей (рис. 7.2).

При разработке способов повышения качества продукции важно, чтобы администрация предприятия принимала во внимание, что большую часть проблем можно решить, изменив систему управления в целом. И только решение небольшого числа проблем зависит от усилий отдельных рабочих или от оборудования на некоторых производственных участках. Решающая роль стиля управления в создании общих условий для производства высококачественной продукции неоднократно подчеркивалась ведущими экспертами в области контроля качества.

Улучшение качества продукции

⇩

Снижение издержек

⇩

Повышение производительности

⇩

Увеличение объемов продаж посредством реализации продукции более высокого качества по более низким ценам

⇩

Расширение бизнеса и увеличение числа рабочих мест

Рис. 7.1. Результаты улучшения качества продукции

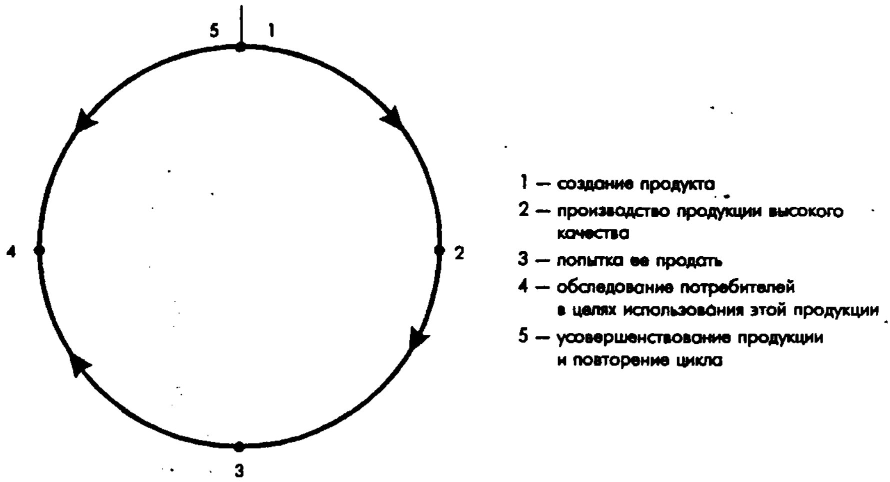


Рис. 7.2. Цикл качества Деминга

Статистические методы, которые будут описаны ниже, способствуют достижению компанией целей повышения качества, однако при отсутствии общего настроя и соответствующей системы управления эти методы окажутся неэффективными. В данной главе мы остановимся на двух основных аспектах проблемы контроля качества. В первой части главы мы рассмотрим способы определения степени соответствия некоторого технологического процесса спецификации качества;

иными словами, каковы возможности контролирования технологического процесса. Затем мы обсудим методы управления технологическим процессом, которые позволят сохранить достигнутый уровень качества и контролировать сам процесс.

7.2. ИЗМЕНЧИВОСТЬ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Любой технологический процесс подвержен изменчивости, даже если он функционирует в полном соответствии с установленными нормативами. Рассмотрим, например, следующие технологические процессы:

- розлив фруктового сока по картонным упаковкам определенного объема с помощью специального станка;
- проверка диаметра отверстий, просверленных в металлическом листе;
- проверка веса упаковки сахара;
- определение длины металлических стержней, нарезаемых металлорежущим станком.

В каждом из этих процессов станок работает в соответствии с некоторым заданным средним значением, однако возможно отклонение параметров отдельных изделий в ту или иную сторону. Обычно значения параметров, меньшие или большие среднего значения, находятся в некотором балансе и имеют нормальное распределение значений переменной. Дисперсия, или разброс, распределения также изменяется в зависимости от вида станка (рис. 7.3).

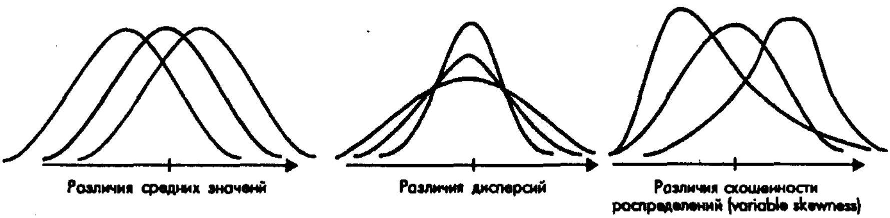


Рис. 7.3. Изменчивость технологического процесса

Эта изменчивость содержит в себе две компоненты:

- изменчивость под воздействием простых или неустранимых причин;
- изменчивость под воздействием неслучайных, или специальных, причин.

Простые причины изменчивости существуют всегда, и фактически их нельзя устранить до тех пор, пока сам процесс не будет изменен. Появление такой изменчивости связано с типом применяемых станков и общими условиями функционирования процесса. Ее значение зависит от конкретного станка или конкретных условий. Для преобразования общей изменчивости необходимо либо использовать новый, или модифицированный, станок, либо осуществлять контроль за условиями функционирования процесса.

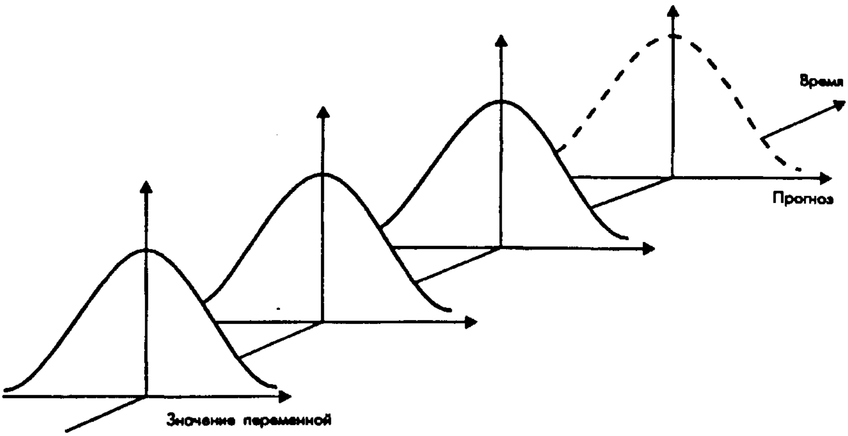


Рис. 7.4. Изменчивость, вызванная простыми причинами

Если технологический процесс находится под контролем, т.е. осуществляется правильно, то общие причины изменчивости приводят к распределению, которое устойчиво во времени и, следовательно, поддается прогнозу.

Вариация, которая появляется под действием общих причин, определяет границы функционирования технологического процесса при наличии контроля за соблюдением определенных условий, т.е. при наличии правильных стартовых параметров, правильного материально-технического обслуживания, управления со стороны специалиста соответствующей классификации, использовании соответствующего сырья.

Неслучайные, или специальные, причины колеблемости появляются ввиду возникновения особых изменений в самом технологическом процессе или окружающей среде, которые можно выявить. Например:

- ошибка оператора при наладке станка;
- частичная поломка или замедление работы станка;
- поломка заводских кондиционеров и неожиданное увеличение температуры воздуха;
- несоблюдение пропорций при смешивании различных ингредиентов сырья.

Неслучайные причины изменчивости приводят к нестабильности распределения значений переменной. Предсказать вид распределения больше невозможно. Технологический процесс выходит из-под контроля.

Статистический контроль за технологическим процессом используется для определения условий, при которых этот процесс можно контролировать, или условий, при которых возникают неполадки и процесс выходит из-под контроля. Если технологический процесс невозможно контролировать ввиду появления неслучайных причин изменчивости, то с помощью статистического метода выявить эти причины нельзя. Этот метод лишь позволяет оператору установить факт наличия *вероятных* неполадок. Оператор должен первым узнать, что технологический

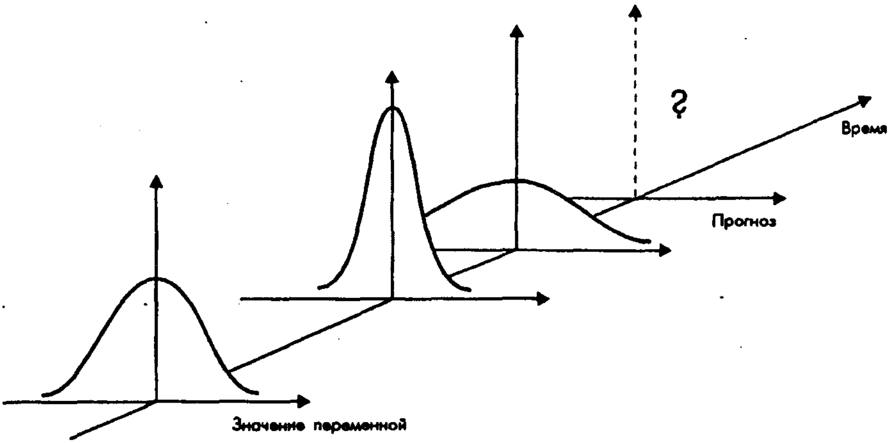


Рис. 7.5. Изменчивость, вызванная неслучайными причинами

процесс нарушен. Если он не в состоянии установить причину непредвиденных изменений, то должен сообщить об этом определенному должностному лицу своей организации. Именно быстрота и результативность ответной реакции на сигнал, полученный благодаря применению статистического метода, определяют сознательное отношение к проблеме качества во всей организации.

7.3. ПРОИЗВОДСТВЕННЫЕ ВОЗМОЖНОСТИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Необходимость расчета производственных возможностей процесса определяется двумя причинами. Во-первых, можно определить, какие виды работ могут быть хорошо выполнены на данном станке. Если возможности станка таковы, что он может работать в соответствии с допустимым отклонением ±1 мм, то было бы ошибкой предполагать, что он удовлетворяет более узкому промежутку допустимых отклонений в ±0,5 мм. Измерение производственных возможностей процесса позволяет определить, какие виды работ можно выполнить на данном станке, а какие — нет.

Во-вторых, для того, чтобы проверить, находится технологический процесс под контролем или нет, необходимо определить уровень изменчивости, при котором технологический процесс находится под контролем. Изменчивость процесса в целом нужно определить таким образом, чтобы можно было выделить изменчивость, вызванную простыми и особыми причинами. Необходимо определить производственные возможности процесса при правильных стартовых параметрах, соответствующем сырье и управлении со стороны профессионального оператора. После того, как определены и измерены общие причины изменчивости, появляется возможность определить вариацию, возникающую вследствие неслучайных причин.

Производственные возможности процесса определяются как допустимый интервал неустранимой или простой изменчивости технологического процесса в нормальных условиях. Этот интервал позволяет судить о возможности изготовления конкретного продукта в рамках данного технологического процесса.

При расчете производственных возможностей процесса, как правило, предполагается, что рассматриваемая переменная нормально распределена. Если мы предположим, что в некотором технологическом процессе переменная имеет нормальное распределение со средним значением μ и стандартным отклонением σ, то производственные возможности рассчитываются обычно как $\mu + 3\sigma$, т.е. величина интервала производственных возможностей составляет 6σ. На практике оценку производственных возможностей технологического процесса рассчитывают с помощью выборки, т.е. с использованием не 6σ, а $6s$, где s — стандартное отклонение выборки. Еще раз отметим, что технологический процесс должен функционировать в нормальных условиях по тем параметрам, которые определяют его производственные возможности.

Из всей продукции, полученной с помощью данного технологического процесса, случайным образом осуществляется выборка не менее, чем в 50 единиц, в условиях, когда процесс является контролируемым. Каждое изделие измеряется соответствующей переменной, на основе которой рассчитывается стандартное отклонение.

❑ **Пример 7.1.** На некотором станке выполняется резка стальной проволоки определенной длины. Наладка станка осуществлялась из расчета, что длина каждого отрезка составляет 100 мм. Контроль за работой ведет оператор. Из всей партии случайным образом было отобрано 60 отрезков проволоки и замерена их длина. Было установлено, что средняя длина одного куска составляет 100,1 мм, а стандартное отклонение равно 0,2 мм. Каковы производственные возможности этого станка?

Решение

Производственные возможности составляют: 6 × стандартное отклонение = 6 × 0,2 = 1,2 мм. Идеальной была бы ситуация, когда осуществлялась бы проверка полученного значения производственных возможностей для отрезков проволоки другой длины, что позволило бы убедиться в нормальном функционировании станка независимо от того, на какое значение параметра осуществлялась его наладка.

После того, как измерены производственные возможности, появляется база для принятия решения о выполнении конкретных задач при наличии спецификации работ и заданных допустимых отклонениях. Если данный технологический процесс не удовлетворяет требованиям спецификации, процесс производства нельзя начинать до тех пор, пока не будет найдено решение проблемы (в данном случае использование другого станка или изменение существующей технологии). Следствием применения неадекватных технологических процессов является производство изделий низкого качества, которые приходится признать бракованными или подвергнуть переработке, что, в свою очередь, приведет к увеличению затрат и нарушению привычного ритма работы.

❑ **Пример 7.2.** Обратимся к ситуации, которая была описана в примере 7.1. Компания получила четыре заказа на стальные прутья. Спецификации и допустимые отклонения, соответствующие каждому заказу:

Заказ	Длина	мм
1	100	+0,5
2	95	0,0 +1,0
3	105	+1,0
4	110	+0,7

Станок можно наладить на производство стальных прутьев необходимой длины, обычно такая переналадка осуществляется в расчете на середину интервала допустимых отклонений. Какой из этих заказов может быть выполнен на данном станке?

Решение

Возможности данного технологического процесса равны 1,2 мм. Продукция, которая будет выпущена с использованием данного станка, удовлетворяет спецификации, если допустимое отклонение больше, чем производственная возможность.

Заказ 1. Допустимое отклонение составляет +0,5 мм, длина интервала — 1,0 мм. Станок, имеющий возможность производства прутьев в 1,2 мм, не удовлетворяет этой спецификации. Необходимо использовать другой станок, который может производить прутья 1,0 мм или менее, либо пересмотреть спецификацию по договоренности с заказчиком.

Заказ 2. Осуществляется наладка станка на производство прутьев размером 95,5 мм. Допустимые отклонения: от 0,0 до +1,0 = 1,0 мм. Станок, который производит прутья 1,2 мм, не удовлетворяет этой спецификации. Таким образом, этот станок не может использоваться и в данном случае.

Заказ 3. Допустимое отклонение составляет +1,0 мм, длина интервала равна 2 мм. Станок, который может производить прутья 1,2 мм, удовлетворяет этой спецификации.

Заказ 4. Допустимое отклонение составляет +0,7 мм, длина интервала равна 1,4 мм. Производительность станка удовлетворяет и этой спецификации.

Идея о существовании простых или неустранимых или неизбежных причин изменчивости технологического процесса является основой для всей теории контроля качества. Она затрагивает всех сотрудников организации — разработчиков, сбытовиков, инженеров, бухгалтеров, но в первую очередь — заказчика. В примере **7.2** специалист из отдела продаж не должен принимать заказы 1 и 2, если рассматриваемый станок — единственный вид оборудования, которым располагает фирма. В этом случае ему следует обсудить спецификацию с заказчиком. Если же стальные прутья предназначены для использования внутри компании, разработчики должны быть ознакомлены с производственными возможностями оборудования, имеющегося на предприятии. Им следует обсудить этот вопрос с инженерами в связи с возможностью внесения некоторых изменений. Нередко должностные лица этого уровня предполагают, что изделия нестандартной длины можно отсортировать в процессе контроля за ходом технологического процесса. Однако главный бухгалтер может возражать по поводу проведения этого мероприятия в связи с его высокой стоимостью. Этот пример иллюстрирует необходимость создания продуманной системы управления и объединения усилий всех сотрудников, направленных на повышение качества.

7.4. КОНТРОЛЬНЫЕ КАРТЫ

Контрольные карты используются для оценки "контролируемости" или "неконтролируемости" процесса. Эту оценку можно получить:

- осуществляя проверку замеров важнейших параметров изделия, например, веса сахара в одной упаковке, диаметра отверстия, просверленного в листе металла, длины стального прута;
- осуществляя проверку отдельных качественных характеристик изделия, например, прочно ли упакован пакет, правильно ли закрыта крышкой бутылка, не повреждено ли фарфоровое изделие и т.д.

В первом случае используются контрольные карты количественного признака, а во втором — контрольные карты качественного признака. Поскольку каждый из указанных видов контрольных карт имеет свою область применения, мы рассмотрим их в отдельности.

7.4.1. Контрольные карты количественных признаков при известных μ и σ

Эти контрольные карты были предложены Шухартом (Shewhart) еще в 20-е годы, в период его работы в телефонной компании Bell Laboratories (США), поэтому их иногда называют карты Шухарта. С помощью этих карт выявляется различие между изменчивостью технологического процесса, вызванной простыми причинами, и изменчивостью, появившейся под действием неслучайных причин. Существует несколько видов карт Шухарта. Мы остановимся более подробно на двух типах: контрольной карте средних арифметических и контрольной карте изменчивости технологического процесса.

КОНТРОЛЬНАЯ КАРТА СРЕДНИХ АРИФМЕТИЧЕСКИХ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Если генеральная совокупность имеет нормальное (или близкое к нормальному) распределение со средним значением μ и стандартным отклонением σ, выборочное распределение выборочного среднего также является нормальным и имеет такое же среднее значение и стандартную ошибку, равную $\sigma/\sqrt{n}$, где n — объем выборки. Для любого нормального распределения между граничными значениями, равными

$$\mu \pm 2 \times \text{стандартное отклонение},$$

заключено примерно 95% распределения. Вероятность того, что полученное значение окажется больше, чем $\mu + 2 \times$ стандартное отклонение, составляет 2,5%, или один случай из 40, вероятность получения значения, меньшего $\mu - 2 \times$ стандартное отклонение, также составляет 2,5%. Аналогично интервал

$$\mu \pm 3 \times \text{стандартное отклонение}$$

охватывает около 99,8% распределения. Вероятность того, что полученное значение превысит $\mu + 3 \times$ стандартное отклонение или окажется меньше, чем $\mu - 3 \times$ стандартное отклонение, составляет 0,1%, т.е. это событие будет иметь место в одном случае из 1000.

Графическая иллюстрация этих крайних значений приведена на рис. 7.6 для выборочного распределения выборочного среднего. Эта диаграмма является основой для составления контрольной карты среднего арифметического технологического процесса.

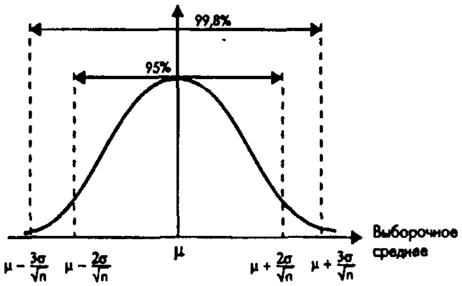


Рис. 7.6. Выборочное распределение выборочного среднего

Для построения графика, приведенного на рис. 7.6, необходимо, чтобы значения μ и σ были известны. Их оценки получают по результатам расчетов среднего значения и стандартного отклонения соответствующих параметров технологического процесса на протяжении длительного промежутка времени.

95%-ные границы распределения называются верхней и нижней **предупреждающими границами**. 98%-ные границы распределения называются верхней и нижней **границами регулирования**.

Построение контрольной карты состоит в нанесении на график выборочных средних в соответствии с номером выборки (рис. 7.7).

Стандартная процедура использования этих контрольных карт состоит из следующих шагов:

1. Через равные промежутки времени проводится выборка объемом n и рассчитывается выборочное среднее.

2. Полученное значение выборочного среднего наносится на контрольную карту в соответствии с номером выборки.

3. Если выборочное среднее лежит за пределами границы регулирования, производится остановка технологического процесса в целях выявления неслучайных причин вариации.

4. Если два последовательно полученных значения выборочных средних находятся в промежутке между предупреждающей границей и границей регулирования,

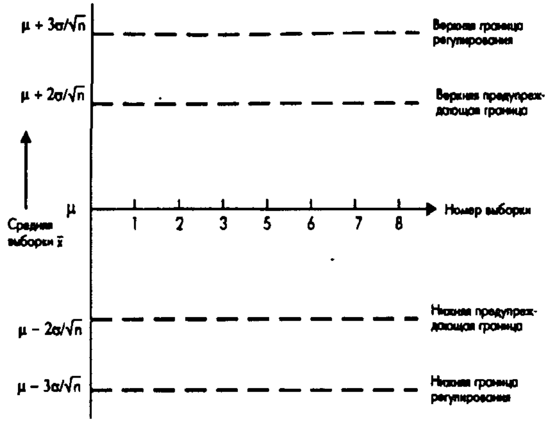


Рис. 7.7. Контрольная карта Шухарта выборочного среднего

предпринимаются немедленные действия по остановке процесса производства и выявлению неисправностей. Если некоторое среднее значение лежит за пределами предупреждающих границ, следующая выборка производится сразу же, до момента проведения очередной выборки.

5. Если точки на графике образуют явный возрастающий или убывающий тренд, предпринимаются определенные меры даже в случаях, когда эти точки находятся в пределах предупреждающих границ. Этот тренд может оказаться индикатором наличия неслучайных причин, например, снижения параметров наладки станка.

Применение этой процедуры иногда приводит и к необоснованным остановкам технологического процесса, однако это случается крайне редко. Любые издержки, связанные с остановками процесса производства, будут больше, чем та экономия, которая может быть получена вследствие улучшения качества продукции.

Данная процедура в сущности представляет собой не что иное, как проверку гипотез, о которой шла речь в гл. 6. Нулевая гипотеза заключается в том, что технологический процесс находится под контролем, причем все технологические параметры соответствуют установленным производственным возможностям. Альтернативная гипотеза утверждает, что процесс не является контролируемым. Каждый раз, когда мы проводим выборку, осуществляется процедура проверки гипотез. Если выборочное среднее лежит за предупреждающими границами, H_0 отклоняется при 5%-ном уровне значимости. Если оно находится за пределами границ регулирования, мы отклоняем H_0 при 2%-ном уровне значимости. С помощью контрольных карт можно без труда показать результаты периодически повторяющейся проверки гипотез.

❒ **Пример 7.3.** Производится расфасовка чая в упаковки объемом по 125 г. Известно, что фасовочный станок работает со стандартным отклонением в 0,15 г. Для обеспечения необходимого веса достаточно наладить станок на среднее значение в 125 г. Через каждые полчаса производится случайная выборка объемом в 5 упаковок. Каждую упаковку взвешивают. Ниже приведены результаты шести последовательных выборок.

Номер выборки	*1*	*2*	*3*	*4*	*5*	*6*
Вес упаковки, г	125,1	124,9	125,2	125,0	124,8	124,9
	125,3	125,0	125,1	125,0	124,8	125,1
	125,1	125,1	125,3	124,7	125,2	125,0
	124,8	124,9	125,0	125,2	125,1	124,9
	125,1	124,7	125,1	125,1	124,9	125,2

Построим по этим данным контрольную карту арифметического среднего и опишем функционирование процесса расфасовки.

Решение

Центральная ось контрольной карты соответствует уровню $\mu = 125$ г.

Предупреждающие границы строятся на уровнях:

$$\mu \pm 2 \times \frac{\sigma}{\sqrt{n}} = 125 \pm 2 \times \frac{0{,}15}{\sqrt{5}},$$

т.е. для 124,866 г и 125,134 г.

Границы регулирования строятся на уровнях:

$$\mu \pm 3 \times \frac{\sigma}{\sqrt{n}} = 125 \pm 3 \times \frac{0{,}15}{\sqrt{5}},$$

т.е. для 124,80 г и 125,20 г.

Вычислим среднее значение каждой из выборок.

Номер выборки	*1*	*2*	*3*	*4*	*5*	*6*
Среднее значение, $\bar{x}$, г	125,08	124,92	125,14	125,0	124,96	125,02

Нанесем средние значения на контрольную карту.

Среднее значение выборки *3* находится выше верхней предупреждающей границы, однако, среднее значение следующей выборки находится внутри контрольных границ и, следовательно, можно предположить, что поводов для беспокойства нет. Предполагается, что выборка *4* производится сразу же после выборки *3*, в которой были обнаружены некоторые отклонения параметра.

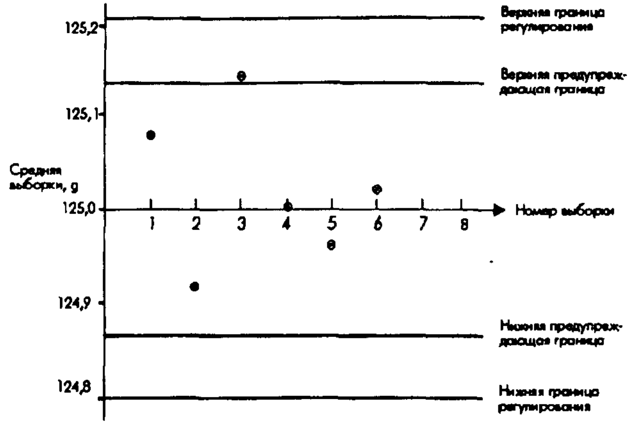


Рис. 7.8. Контрольная карта выборочного среднего

КОНТРОЛЬНЫЕ КАРТЫ ИЗМЕНЧИВОСТИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Результатом процесса производства может быть выпуск изделий, параметры которых в среднем удовлетворительны, однако возрастание значений относительно средней должно являться причиной для беспокойства. Производственные возможности станка изменяются, и его работа становится неконтролируемой. Размеры значительной доли изделий могут оказаться слишком большими или слишком маленькими даже в случае, когда среднее значение является допустимым. Показатель **размаха** значений каждой выборки может использоваться при построении контрольной карты размаха, характеризующей изменения стандартного отклонения. Эту карту строят строго с учетом данных контрольной картой среднего, а затем обе карты используются одновременно.

Размах = Максимальное значение Минимальное значение.

В основу контрольной карты размаха положено достаточно сложное распределение, поэтому при определении положения линии центра предупреждающих границ и границ регулирования, используются специальные таблицы. Как правило, определяют только верхнюю предупреждающую границу и границу регулирования, поскольку в основном только большой размах приводит к осложнениям в технологическом процессе. Однако иногда для выявления чрезвычайно низких размахов, которые могут означать либо улучшение процесса производства, либо наличие измерительной ошибки, используются и нижние границы контрольной карты.

Положение каждой из прямых определяется следующим образом:

Центральная линия	$d_n \times \sigma$,	где n — объем выборки, σ — стандартное отклонение технологического процесса;
Верхняя предупреждающая граница	$r_w \times \sigma$	отсекает 2,5% значений в верхней части распределения, т.е. при условии нормального протекания процесса размах выборки превысит данное значение в одном случае из 40;
Верхняя граница регулирования	$r_A \times \sigma$	отсекает 0,1% значений в верхней части распределения, т.е. при условии нормального протекания процесса размах выборки превысит это значение в одном случае из 1000.

d_n, r_w, r_A — параметры распределения вероятностей размахов выборок, полученных из нормального распределения. Их значения зависят от объема выборки n и определяются по статистическим таблицам. Ниже приводится выдержка из соответствующей таблицы.

Таблица 7.1. **Параметры, используемые для построения контрольной карты размахов**

n	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
d_n	1,128	1,693	2,059	2,326	2,534	2,704	2,847	2,970	3,078	3,173
r_w	3,17	3,68	3,98	4,20	4,35	4,49	4,61	4,70	4,79	4,86
r_A	4,65	5,05	5,30	5,45	5,60	5,70	5,80	5,90	5,95	6,05

❑ **Пример 7.4**. По данным примера **7.3**, относящихся к весу упаковок с чаем, построим контрольную карту размахов.

Решение.

Размер выборки равен 5, стандартное отклонение процесса составляет 0,15 г. Используя данные табл. 7.1, находим:

$$d_5 = 2{,}326\,,\quad r_w = 4{,}20\,,\quad r_A = 5{,}45,\ \text{следовательно}$$

Центральная линия	$d_n \times \sigma = 2{,}326 \times 0{,}15 = 0{,}35$ г.;
Верхняя предупреждающая граница	$r_w \times \sigma = 4{,}20 \times 0{,}15 = 0{,}63$ г.;
Верхняя граница регулирования	$r_A \times \sigma = 5{,}45 \times 0{,}15 = 0{,}82$ г.

Соответствующий график приведен на рис. 7.9.

Используя данные, приведенные в примере 7.3, вычислим размах каждой выборки.

Номер выборки	*1*	*2*	*3*	*4*	*5*	*6*
Размах выборки, г	0,5	0,4	0,3	0,5	0,4	0,3

Нанесем полученные значения размахов на контрольную карту (рис. 7.9).

В данном примере и контрольная карта средних значений, и контрольная карта размахов говорят о том, что технологический процесс находится под контролем.

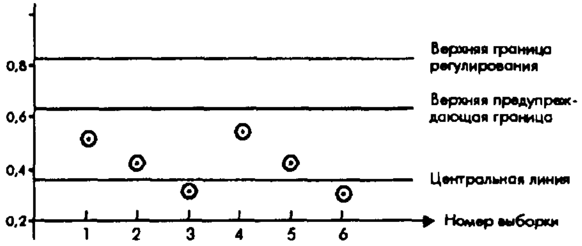


Рис. 7.9. Контрольная карта размахов

Обе карты — примеры наиболее часто используемых типов контрольных карт. Необходимо, чтобы они находились вблизи от оператора станка, что обеспечит возможность быстро и без труда их заполнить.

Для того, чтобы облегчить поиск неисправностей в случае их появления, на контрольной карте необходимо отмечать такие события, как начало новой смены, переналадка станка, получение новой партии сырья. Цель построения контрольных карт — предотвратить технологический процесс при наличии неисправностей, выявляя появившиеся неслучайные причины.

7.4.2. Контрольные карты количественных признаков при неизвестных μ и σ

В разделе 7.4.1 перед тем, как начать построение контрольных карт, мы предполагали, что значения μ и σ известны. Однако это не всегда так, и в этих случаях приходится осуществлять оценку значений μ и σ. Один из методов получения оценок основан на использовании данных предыдущих выборок.

Оценка μ вычисляется как среднее всех индивидуальных значений. Это равносильно нахождению среднего значения выборочных средних. Для каждой выборки рассчитывается $\bar{x}$, а затем — среднее из всех полученных значений $\bar{x}$. Это и есть $\bar{\bar{x}}$.

Если значение σ неизвестно, оценку разброса значений в генеральной совокупности получают с использованием среднего значения размахов выборки $\bar{R}$.. Для того, чтобы оценить σ, мы должны знать характер взаимосвязи между $\bar{R}$ и σ.

В табл. 7.1 приводится коэффициент, применяемый при расчете центральной линии контрольной карты размахов. Эта центральная линия и есть $\bar{R}$. Следовательно, искомая взаимосвязь достаточно проста:

$$\bar{R} = d_n \sigma ,$$

где n — размер выборки, а d_n находится из табл. 7.1. Таким образом, вместо используемого во всех предшествующих расчетах σ мы воспользуемся

$$\sigma = \bar{R} / d_n .$$

Применение этого метода проиллюстрируем на примере расчета границ регулирования. Отметим, что эта же процедура может использоваться и при нахождении предупреждающих границ. Границы регулирования контрольной карты среднего рассчитывались как $\mu \pm \frac{3\sigma}{\sqrt{n}}$. В данном случае эта формула примет следующий вид:

$$\bar{\bar{x}} \pm 3 \frac{\bar{R} / d_n}{\sqrt{n}} .$$

Верхняя граница регулирования контрольной карты размахов рассчитывалась как $r_A \sigma$. В данном случае имеем:

$$r_A \frac{\bar{R}}{d_n} ,$$

где r_A и d_n приведены в табл. 7.1.

Отметим, что при изложении вопроса о вероятности статистического превышения границ регулирования в различных учебниках, к сожалению, используются различные обозначения и исходные предпосылки. Это приводит к появлению таблиц и формул, которые дают несколько отличные друг от друга значения границ. Американские учебники отличаются в этом отношении от английских, и все это сопровождается изменениями во времени.

Несмотря на то, что описанные выше методы определения положения границ регулирования и предупреждающих границ сами по себе являются сложными, они тем не менее требуют проведения некоторых дополнительных расчетов. Если контрольные карты будет использовать работник, не имеющий достаточного уровня статистических знаний, то очевидное преимущество будет иметь подход, позволяющий предельно упростить все вычисления.

Вернемся вновь к границам регулирования контрольной карты среднего. Эти границы рассчитываются по следующей формуле:

$$\bar{\bar{x}} \pm \frac{3\bar{R}/d_n}{\sqrt{n}}.$$

В данном случае $\frac{3/d_n}{\sqrt{n}}$ является константой. Обозначим ее через А, тогда положение границ регулирования будет определяться как

$$\bar{\bar{x}} \pm A\bar{R}.$$

Существует специальная таблица значений А для различных значений n.

Верхняя граница регулирования контрольной карты размахов определяется как:

$$\frac{r_A \bar{R}}{d_n}.$$

В данном случае $\frac{r_A}{d_n}$ вновь представляет собой константу. Назовем ее, например С, тогда уровень верхней границы регулирования будет равен $C\bar{R}$. Значения С определяются в зависимости от значений n по специальным таблицам.

Наконец, нижняя граница регулирования контрольной карты размахов определяется как $B\bar{R}$. Значения В находятся из соответствующих таблиц в зависимости от значений А и С (см. статистические таблицы).

❑ **Пример 7.5**. В примере 7.3 приводятся данные о 125-граммовых упаковках с чаем. Через равные интервалы в 30 мин проводится случайная выборка, в примере приводятся результаты измерений по шести выборкам. Используем эти данные для того, чтобы проиллюстрировать получение оценок μ и σ.

Выборочное среднее и размах выборки рассчитаны в разделе 7.4.1 и будут использоваться при нахождении $\bar{\bar{x}}$ и $\bar{R}$.

Таблица 7.2

Номер выборки	*Среднее значение $\bar{\bar{x}}$, г*	*Размах $\bar{R}$, г*
1	125,08	0,5
2	124,92	0,4
3	125,14	0,3
4	125,00	0,5
5	124,96	0,4
6	125,02	0,3
Итого	750,12	2,4

Среднее значение $\bar{\bar{x}} = 125{,}02$ г, $\bar{R} = 0{,}4$ г.

Оценка границы регулирования для контрольной карты среднего имеет следующий вид:

$$\bar{\bar{x}} \pm A\bar{R}.$$

Находим из таблиц, что при n = 5, A = 0,58, следовательно, граница регулирования равна 125,02 ± 0,58 × 0,4 = (125,02 ± 0,232) г. Полученные границы регулирования, равные 125,25 г и 124,79 г, можно сравнить с теми значениями, которые были получены в разделе **7.4.1** с использованием μ и σ, и равны 125,2 г и 124,8 г соответственно.

Контрольная карта размахов

Центральная линия $\bar{R}$ = 0,4 г

Верхняя граница регулирования $C\bar{R}$ = 2,11 × 0,4 = 0,844 г
Сравните со значением 0,82 г, полученным в разделе 7.4.1.

Нижняя граница регулирования $B\bar{R}$ = 0 × 0,4 = 0.
Значения C и B определяются из соответствующих таблиц при n = 5.

7.4.3. Контрольные карты качественных признаков

Оценка качества продукции через измерение количественных параметров, например, длины изделия, не всегда целесообразна. В некоторых случаях наличие или отсутствие у изделия дефектов зависит от определенных качественных признаков. Например, керамическая фабрика производит китайскую керамику. По окончании процесса обжига каждое изделие подвергается проверке. Контролер выявляет изделия с дефектами лакового покрытия: трещинами, отколотыми кусочками и т.д. Если некоторое изделие имеет хотя бы один из этих признаков, оно относится к браку. В каждой партии продукции обязательно найдется несколько бракованных изделий. Конечно, ошибки всегда имеют место, но их появление должно быть редким и носить случайный характер. Вопрос, который должен задать себе производитель: соответствует ли доля бракованных изделий той доле, которая бывает при нормальных условиях, или имеются какие-то неполадки. В случае, если удельный вес брака слишком высок, каковы причины такой ситуации? К примеру, температура в печи для обжига может изменяться в процессе обжига и оказывать воздействие на качество лакового покрытия.

Различают два типа контрольных карт качественных признаков. В p-картах используется удельный вес бракованных изделий, а в c-картах — число бракованных изделий, приходящихся на одну выборку. В данном параграфе мы остановимся более подробно на p-картах.

Долю бракованных изделий в генеральной совокупности p в условиях контролируемого технологического процесса оценивают на основе большого числа выборок:

$$\text{Наилучшая оценка } p = \frac{\text{Общее число бракованных изделий во всех выборках}}{\text{Общее число изделий, подвергшихся проверке}}$$

В p-картах долю бракованных изделий в выборке размера n принято обозначать через $\hat{p}$.

$$\hat{p} = \frac{\text{Число бракованных изделий в выборке}}{\text{Число изделий в выборке}} = \frac{r}{n}.$$

Число бракованных изделий r в выборке размером в n изделий, полученной случайным образом из достаточно большой генеральной совокупности, имеет биномиальное распределение, если вероятность того, что изделие окажется бракованным, p является константой.

$$P\text{ (r дефектов в выборке размера n)} = {}^nC_r \times p^r \times q^{n-r}, \text{ где } r = 0, 1, 2, \ldots, n.$$

Так как $\hat{p} = \frac{r}{n}$, $\hat{p}$ также имеет биномиальное распределение.

Стандартная ошибка распределения $\hat{p}$ равна:

$$SE_{\hat{p}} = \sqrt{\frac{pq}{n}}.$$

Расчет параметров биномиального распределения достаточно трудоемок, поэтому для упрощения расчетов его можно аппроксимировать либо распределением Пуассона, либо нормальным распределением. Мы не будем приводить точные формулировки правил такой аппроксимации, а ограничимся лишь ее общими принципами, которые состоят в следующем:

В процессе аппроксимации используется распределение Пуассона, если $n \geq 30$, $p < 0{,}1$ и $np < 5$.

В процессе аппроксимации используется нормальное распределение, если $n \geq 30$, $0{,}1 < p < 0{,}9$, $np > 5$ и $n(1\ p) > 5$.

При использовании любого из указанных распределений в процессе аппроксимации построение контрольной карты типа p аналогично построению контрольной карты среднего. Однако при аппроксимации нормальным распределением процедура значительно упрощается, что показано ниже.

Центральная линия:	она строится на уровне доли бракованных изделий в условиях контролируемого технологического процесса $\hat{p}$, оцененной по выборочным значениям в течение достаточно длительного промежутка времени.	
Предупреждающие границы:	$p \pm 2 \times \sqrt{\frac{pq}{n}}$	в условиях контролируемого технологического процесса значение $\hat{p}$ окажется за пределами этих границ примерно в одном случае из 40.
Границы регулирования:	$p \pm 3 \times \sqrt{\frac{pq}{n}}$	в условиях контролируемого технологического процесса значение $\hat{p}$ окажется за пределами этих границ примерно в одном случае из 1000.

Построенную в результате этого алгоритма контрольную карту можно интерпретировать точно так же, как и контрольные карты среднего и размахов. Если для аппроксимации использовалось нормальное распределение, значения нижней предупреждающей границы и нижней границы регулирования могут оказаться отрицательными. Поскольку в данном случае отрицательные значения недопустимы, можно либо не принимать во внимание нижние границы карты, либо провести аппроксимацию заново с использованием распределения Пуассона.

❒ **Пример 7.6.** Компания производит пластмассовые походные чашки. В течение времени, когда было точно известно, что технологический процесс находится под контролем, было проведено 25 выборок, каждая объемом в 100 единиц. Оборудование было соответствующим образом налажено, использовалось сырье допустимого качества, наблюдение за ходом процесса осуществлял оператор соответствующей квалификации. Был произведен контроль изделий в каждой выборке. В табл. 7.3 приведены данные об обнаруженных бракованных изделиях.

Таблица 7.3. **Число бракованных изделий в каждой из 25 выборок размером в 100 единиц**

Номер выборки	*Число бракованных изделий*	*Номер выборки*	*Число бракованных изделий*
1	8	14	4
2	6	15	5
3	1	16	4
4	6	17	2
5	3	18	8
6	2	19	4
7	7	20	3
8	6	21	9
9	7	22	7
10	3	23	5
11	5	24	3
12	4	25	5
13	6		

Нужно построить контрольную карту качественного признака.

Решение

Общее число бракованных изделий в 25 выборках равно 123, следовательно, оценка доли бракованных изделий в генеральной совокупности составит:

$$\hat{p} = \frac{123}{25 \times 100} = 0{,}049 .$$

Хотя значение p достаточно мало и np = 4,9, т.е. меньше 5, прибегнем сначала к аппроксимации нормальным распределением и определим положение границ на контрольной карте.

Центральная линия: 0,049.

Предупреждающие границы: $0{,}049 \pm 2 \times \sqrt{\dfrac{0{,}049 \times (1 - 0{,}049)}{100}} = 0{,}049 \pm 0{,}043$, т.е. 0,006 и 0,092.

Границы регулирования: $0{,}049 \pm 3 \times \sqrt{\dfrac{0{,}049 \times (1 - 0{,}049)}{100}} = 0{,}049 \pm 0{,}065$, т.е. −0,016 и 0,114.

Нижняя граница регулирования отрицательная, поэтому ее либо не наносят на контрольную карту, либо полагают равной нулю.

Так как $p < 0{,}1$ и $np < 5$, в данном случае аппроксимация с использованием распределения Пуассона, вероятно, позволила бы получить лучшие результаты. Среднее число дефектов в выборке $np = 4{,}9$. Следовательно, распределение вероятностей Пуассона имеет вид:

$$P\,(r \text{ дефектов в выборке}) = \frac{4{,}9^r\, e^{-4{,}9}}{r!}, \quad r = 0,\ 1,\ 2,\ \ldots$$

Значения предупреждающих границ и границ регулирования, равные 0,001, 0,025, 0,975 и 0,999, определяются в процессе расчета кумулятивных вероятностей, что показано в табл. 7.4.

Положим, что нижняя граница регулирования равна нулю, поскольку она не может иметь отрицательные значения. Положение оставшихся трех границ определяется как середина соответствующего интервала. Например, верхняя предупреждающая граница лежит между $p = 0{,}09$ и $p = 0{,}10$ и, следовательно, находится на уровне $p = 0{,}095$.

Таблица 7.4. **Использование аппроксимации распределением Пуассона для определения положения границ на контрольной карте**

Число бракованных изделий, r	*Доля бракованных изделий, $\hat{p}$*	*Вероятность, P(r)*	*Кумулятивная вероятность,*
1	2	3	4
			Нижняя граница регулирования 0,001
0	0,00	0,0074	0,0074
			Нижняя предупреждающая граница 0,025
1	0,01	0,0365	0,0439
2	0,02	0,0894	0,1333

Продолжение табл. 7.4

1	2	3	4	
3	0,03	0,1460	0,2793	
4	0,04	0,1789	0,4582	
5	0,05	0,1753	0,6335	
6	0,06	0,1432	0,7767	
7	0,07	0,1002	0,8769	
8	0,08	0,0614	0,9382	
9	0,09	0,0334	0,9717	
				Верхняя предупреждающая граница 0,975
10	0,10	0,0164	0,9880	
11	0,11	0,0073	0,9953	
12	0,12	0,0030	0,9983	
				Верхняя граница регулирования 0,999
13	0,13	0,0011	0,9994	
14	0,14	0,0004	0,9998	

Ниже приведены результаты аппроксимации двумя указанными распределениями.

	Положение контрольной границы по результатам аппроксимации	
	нормальным распределением	*распределением Пуассона*
Нижняя граница регулирования	(–0,016)	0
Нижняя предупреждающая граница	0,006	0,005
Верхняя предупреждающая граница	0,092	0,095
Верхняя граница регулирования	0,114	0,125

На рис. 7.10 построена контрольная карта по результатам аппроксимации распределением Пуассона.

❑ **Пример 7.7.** Та же компания продолжает осуществлять случайную выборку из готовой продукции, полученной в ходе данного технологического процесса, объемом в 100 чашек. Табл. 7.5 содержит информацию о числе бракованных изделий в 15-ти следующих друг за другом выборках. Используя эти данные, оценим, является ли технологический процесс контролируемым.

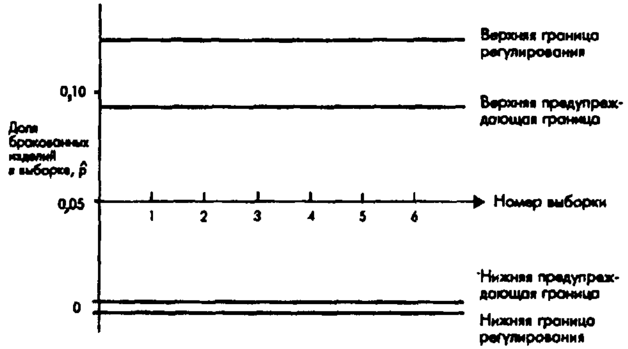


Рис. 7.10. Контрольная карта качественного признака

Таблица 7.5. **Число бракованных изделий в каждой из 15 выборок объемом в 100 штук**

Номер выборки	*Число бракованных изделий*	*Доля брака*	*Номер выборки*	*Число бракованных изделий*	*Доля брака*
1	5	0,05	9	4	0,04
2	4	0,04	10	7	0,07
3	2	0,02	11	11	0,11
4	4	0,04	12	4	0,04
5	6	0,06	13	3	0,03
6	4	0,04	14	4	0,04
7	5	0,05	15	2	0,02
8	3	0,03			

Решение

Нанесем выборочные доли на контрольную карту, построенную на рис. 7.10 (рис. 7.11).

Как следует из рис. 7.11, процесс находился под контролем до момента проведения выборки № 11. Доля бракованных изделий в выборке № 11 выходит за пределы верхней предупреждающей границы. Предполагается, что выборка № 12 производится не через установленный интервал времени, а сразу после выборки № 11. Поскольку значение выборки № 12 вновь попадает в промежуток между двумя предупреждающими границами, мы можем сделать вывод о том, что выборка № 11 является исключением, а не индикатором того, что процесс вышел из-под контроля. Кроме того, следует отметить, что большая часть значений лежит ниже уровня центральной линии, равного 0,049. Этот факт можно интерпретировать как улучшение технологии.

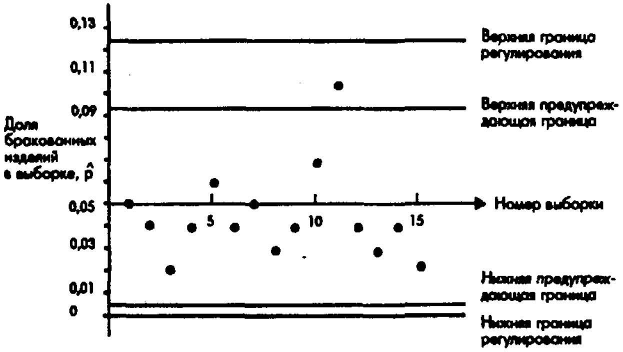


Рис. 7.11. Контрольная карта качественного признака для процесса производства пластмассовых чашек

7.5. СТАТИСТИЧЕСКИЙ ПРИЕМОЧНЫЙ КОНТРОЛЬ КАЧЕСТВА НЕКОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ

В предыдущем параграфе мы предполагали, что продукция производится с использованием станков и технологий внутри компании, которые также находятся под её контролем. Между тем компания может иметь намерение проверять качество продукции, закупаемой у внешних поставщиков, чтобы определить, удовлетворяет ли она стандартам, оговоренным в соглашении. Аналогичным образом клиенты компании могут потребовать, чтобы перед отправкой им продукции была произведена ее последняя проверка. Такие проверки качества входящих и исходящих потоков продукции и составляют две области применения выборки при приемочном контроле качества. Эта процедура представляет собой итоговую проверку результатов всех мероприятий по контролю качества, которые имели место в ходе технологического процесса компании.

Когда на одном из этих этапов осуществляется проверка продукции, нужно сформулировать правила, которыми следует руководствоваться в случае обнаружения бракованных изделий. Количество бракованных изделий в выборке позволяет сделать вывод о доле брака в партии продукции в целом. Таким образом соответствующие правила касаются решений, которые следует принять по поводу партии продукции, если в выборке было обнаружено определенное число бракованных изделий. Приведем пример такой схемы выборки:

1. Производится случайная выборка из партии продукции объемом в 15 единиц, затем проверяется каждое изделие в выборке.

2. Если в выборке не было ни одного или обнаружено одно бракованное изделие, предполагается, что все изделия в партии имеют приемлемый уровень качества. Партия продукции принимается.

3. Если в выборке было обнаружено два или более бракованных изделия, предполагается, что качество всей партии продукции ниже допустимого уровня. Партия продукции отклоняется.

Это — пример схемы одноэтапной выборки. При необходимости можно пользоваться схемой, предполагающей двухэтапную выборку. Например:

1. Производится случайная выборка из партии продукции объемом в 15 единиц, затем проверяется каждое изделие в выборке.

2. Если в выборке не было ни одного бракованного изделия, предполагается, что все изделия имеют приемлемый уровень качества, и партия продукции принимается.

3. Если в выборке было обнаружено одно или два бракованных изделия, производится вторая выборка объемом в 20 единиц, затем проверяется каждое изделие в этой выборке:

а) если во второй выборке обнаружено не более одного бракованного изделия, осуществляется приемка партии продукции;

б) если во второй выборке обнаружено более одного бракованного изделия, то отказываются от приемки всей партии продукции.

4. Если в первой выборке обнаружено три или более бракованных изделия, предполагается, что качество всей партии продукции ниже допустимого уровня, и принимается решение об отказе от приемки партии продукции.

Этот вид схемы выборочного приемочного контроля качества основан на биномиальном распределении вероятностей, что означает, что большинство расчетов сложны и требуют длительных временных затрат. В результате этого при вычислении параметров схемы выборки обычно используют специальные таблицы и карты. Так как основная идея этого метода достаточно проста, в данном параграфе мы остановимся на принципах работы алгоритма. Детали расчетов оговариваются в приведенном ниже примере.

❑ **Пример.** Осуществляется поставка партии продукции от внешнего поставщика. Случайным образом производится выборка, и попавшие в нее изделия подвергаются проверке. Система проверки характеризуется тремя параметрами — удельным весом бракованных изделий p, объемом выборки n и максимально допустимым числом бракованных изделий в выборке c. Схема выборки определяется с помощью n и c. Если, например, $n = 12$, а $c = 1$, из партии продукции производится выборка 12 изделий. Акт приемки партии имеет место, если в выборке не было ни одного или было обнаружено одно бракованное изделие. Если же число бракованных изделий равно двум или более, от партии продукции отказываются.

Совершенно необязательно, чтобы объем выборки представлял собой определенный процент от размера всей партии продукции, однако важнейшим условием является принцип случайности выборки, обеспечивающий репрезентативность партии продукции.

Выбор численных значений n и c зависит от доли бракованных изделий в партии продукции, которую допускает клиент. Например, заказчик намерен осуществлять приемку любой партии продукции, число бракованных изделий в которой не превышает 4%, и отказывается от партий продукции с большей долей бракованных изделий. Максимально допустимая доля бракованных изделий называется **допустимым уровнем качества** (AQL). В идеале необходима такая схема выборки, с помощью которой осуществлялась бы приемка всех партий продукции, доля бракованных изделий в которых не превышает AQL, и отказ от тех партий продукции, доля бракованных изделий в которых превосходит AQL (рис. 7.12).

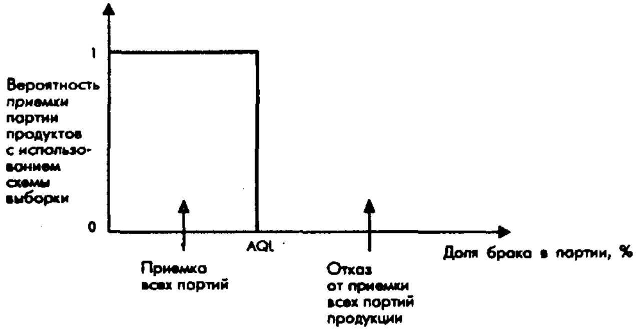


Рис. 7.12. Идеальная кривая оперативной характеристики

Приведенный на рис. 7.12 график – пример построения **кривой оперативной характеристики (O-C Curve — кривая O-C)**. На практике таких идеальных ситуаций не существует, поэтому обычно заказчика просят установить еще один измеритель. Он получил название **допустимого процента бракованных изделий в партии (LTPD)**. LTPD определяет максимальную долю бракованных изделий, которую заказчик допускает в партии продукции. Клиент готов осуществить приемку некоторых партий продукции, доля бракованных изделий в которых лежит в промежутке между AQL и LTPD, но он отказывается принимать те партии продукции, в которых этот показатель выше LTPD. Производитель не должен поставлять партии продукции, в которых удельный вес брака выше, чем LTPD. Следовательно, схема выборки должна быть продумана таким образом, чтобы выявить такие партии продукции как можно более точно. Кривая O-C для этого случая приведена на рис. 7.13.

К сожалению, пока еще не разработаны методы практического построения схем выборки, которые работали бы в точном соответствии со схемой, приведенной на рис. 7.13. На практике кривая О-С выглядит чаще всего так, как показано на рис. 7.14.

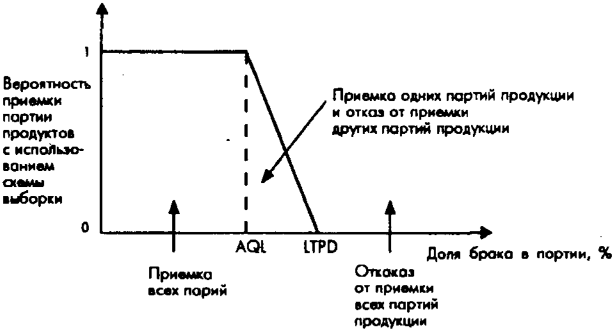


Рис. 7.13. Почти идеальная кривая оперативной характеристики

Обратимся теперь к проблеме вероятности принятия неверных решений. Риск производителя α — это вероятность того, что применение схемы выборки приведет к отказу от приемки партии продукции, которая оказалась бы приемлемой с точки зрения потребителя. Риск потребителя β — это вероятность того, что применение схемы выборки приведет к приемке партии продукции, неприемлемой для потребителя. Схема выборки, применяемая на практике, должна быть направлена на сведение каждой из вероятностей к минимуму.

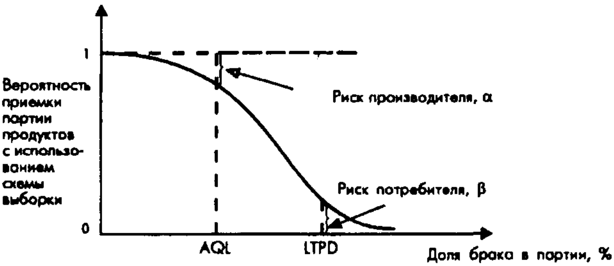


Рис. 7.14. Реальная кривая оперативной характеристики

Для того чтобы выработать подходящую схему выборки, производитель и потребитель должны заключить соглашение по следующим вопросам:

1. Допустимый уровень качества.

2. Риск производителя α , т.е. вероятность того, что применение данной схемы ошибочно приведет к отказу от приемки партии продукции, в которой удельный вес бракованных изделий равен AQL — партии, которую потребитель мог бы принять.

3. Допустимый процент бракованных изделий в партии.

4. Риск потребителя β, т.е. вероятность того, что применение данной схемы ошибочно приведет к приемке партии продукции, удельный вес брака в которой равен LTPD — партии, от приемки которой потребитель бы отказался.

Для любых заданных объема выборки n и максимально допустимого числа отказов от приемки c легко можно построить кривую оперативной характеристики, используя биномиальное распределение. Для любых заданных значений AQL и LTPD можно найти соответствующие вероятности α и β. Гораздо сложнее построить соответствующую кривую, если значения α и β заданы априорно. В данном случае нужно обратиться к литературе, содержащей готовые таблицы систем контроля.

❑ **Пример 7.8.** Построим кривую оперативной характеристики:

1. Построим кривые оперативной характеристики для следующих двух схем выборочного контроля.

Схема А: объем выборки $n = 8$, $c = 1$, т.е. осуществляется приемка партий продукции, если число бракованных изделий в выборке не больше одного включительно.

Схема В: объем выборки $n = 16$, $c = 2$, т.е. осуществляется приемка партий продукции, если число бракованных изделий в выборке не больше двух включительно.

Рассмотрим значения доли бракованных изделий в партии p от 0,05 до 0,40, двигаясь с шагом 0,05.

2. Если производитель и потребитель приходят к соглашению о том, что AQL равен 5%, риск производителя равен 0,05, если LTPD составляет 25%, а риск потребителя — 0,05, какая из схем будет предпочтительнее в данной ситуации?

3. Разработаем более подходящую схему выборки и построим кривую О-С.

Решение

1. По формулам для биномиального распределения находим вероятность наличия r бракованных изделий в выборке размером n:

$$P(r) = {}^{n}C_{r} \times p^{r} \times q^{n-r}, \quad \text{где } r = 0, 1, 2, \ldots, n.$$

Вероятность приемки партии продукции определяется как $P(r \le c)$.

Схема А:

$P(r) = {}^8C_r \times p^r \times q^{8-r}$, где $r = 0, 1, 2, \ldots, 8$,

т.е. $P(0) = q^8$ и $P(1) = 8 \times p \times q^7$,

$P(\text{приемки партии}) = P(r \leq 1) = P(r = 0) + P(r = 1)$.

Таблица 7.6. **Вероятность приемки партии продукции с заданной долей бракованных изделий (p) — применение схемы А**

p	*q*	*P(r=0)*	*P(r=1)*	*P(r ≤ 1)*
0,05	0,95	0,66342	0,27933	0,94275
0,10	0,90	0,43047	0,38264	0,81311
0,15	0,85	0,27249	0,38469	0,65718
0,20	0,80	0,16777	0,33554	0,50331
0,25	0,75	0,10011	0,26697	0,36708
0,30	0,70	0,05765	0,19765	0,25530
0,35	0,65	0,03186	0,13726	0,16912
0,40	0,60	0,01680	0,08958	0,10638

Примечание. Эти расчеты можно значительно облегчить путем применения таблиц рассеяния.

Значения $P(r \leq 1)$, т.е. P(приемки партии продукции при заданном p) нанесены на график (рис. 7.15) в зависимости от значений p.

Схема В:

$P(r) = {}^{16}C_r \times p^r \times q^{16-r}$, где $r = 0, 1, 2, \ldots, 16$.

$P(\text{приемки партии}) = P(r \leq 2) = P(r = 0) + P(r = 1) + P(r = 2)$.

Таблица 7.7. **Вероятность приемки партии продукции с заданной долей бракованных изделий (p) — применение схемы В**

p	*P(r =0)*	*P(r =1)*	*P(r =2)*	*P(r ≤ 2).*
0,05	0,44013	0,37063	0,14630	0,95706
0,10	0,18530	0,32943	0,27452	0,78925
0,15	0,07425	0,20965	0,27748	0,56138
0,20	0,02815	0,11259	0,21111	0,35185
0,25	0,01002	0,05345	0,13363	0,19710
0,30	0,00332	0,02279	0,07325	0,09936
0,35	0,00102	0,00875	0,03533	0,04510
0,40	0,00028	0,00301	0,01505	0,01834

Значения $P(r \leq 2)$, т.е. P(приемки партии продукции с использованием схемы В) также нанесены на график (рис. 7.15).

2. Из графика, приведенного на рис. 7.15, следует, что вероятность отказа от приемки удовлетворительной партии продукции для AQL, равного 0,05, составляет:

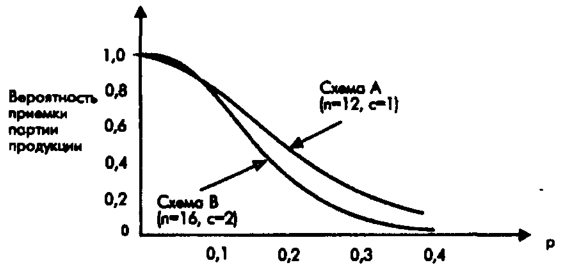


Рис. 7.15. Характеристики кривых по схемам А и В

Схема А:

Р(ошибочного отказа от приемки пригодной партии продукции) = 1 – 0,943 = 0,057.

Схема В:

Р(ошибочного отказа от приемки пригодной партии продукции) = 1 – 0,957 = 0,043.

В обеих схемах риск производителя составляет приблизительно 0,05.

Если LTPD равно 0,25 (достаточно высокое значение), вероятность приемки партии продукции составит:

Схема А:

Р(ошибочной приемки непригодной партии продукции)= 0,367.

Схема В:

Р(ошибочной приемки непригодной партии продукции)= 0,197.

В обеих схемах риск потребителя очень велик и значительно превышает 0,05.

Обе схемы непригодны для потребителя, мнение которого в процессе выработки соглашения является более весомым. Ведь именно потребитель оплачивает заказ на партию продукции.

3. Для ответа на этот вопрос должны использоваться понятия "испытания" и ошибки при определении соответствующей комбинации n и c. Это достаточно легко сделать с помощью составления таблиц рассеяния или использования готовых таблиц без них трудоемкость расчетов заметно возрастает.

Удовлетворительной является комбинация n = 28 и c = 3.

Таблица 7.8. **Вероятность приемки партии продукции при заданной доле бракованных изделий**

p	$P(r = 0)$	$P(r = 1)$	$P(r = 2)$	$P(r = 3)$	$P(r \le 3)$
0,05	0,23783	0,35048	0,24903	0,11359	0,95093
0,10	0,05233	0,16282	0,24423	0,23518	0,69456
0,15	0,01056	0,05219	0,12433	0,19015	0,37723
0,20	0,00193	0,01354	0,04570	0,09901	0,16018
0,25	0,00032	0,00296	0,01333	0,03852	0,05513
0,30	0,00005	0,00055	0,00319	0,01186	0,01565
0,35	0,00001	0,00009	0,00063	0,00295	0,00368
0,40	0,00000	0,00001	0,00010	0,00060	0,00071

Если AQL=0,05, то P(отказа от приемки пригодной партии продукции) = 1 − 0,951 = 0,049. Полученное значение достаточно близко к заданному риску производителя, равному 0,05.

Если LTPD=0,25, то P(ошибочной приемки непригодной партии продукции)=0,055. Это значение также близко к заданному риску потребителя, равному 0,05.

РЕЗЮМЕ

Статистический контроль технологических процесов — лишь часть общего подхода к обеспечению качества. Следует иметь в виду, что существует также множество организационных проблем, которые не рассматривались в данной главе, но являются не менее важными, чем сами статистические методы. Статистический контроль технологических процессов основывается на выборках из готовой продукции, полученной в ходе данного технологического процесса. Характеристики этих выборок позволяют сделать выводы о функционировании процесса производства. Первый шаг состоит в определении производственных возможностей технологического процесса в условиях, когда он находится под контролем, т.е. когда существует соответствующая система управления, операторы соответствующей квалификации и соответствующее сырье. Неустранимая или простая изменчивость характеристик выпускаемой продукции измеряется с помощью данных о выборках из готовой продукции. Обычно производственные возможности процесса определяются как промежуток в ±3 стандартных отклонения.

Показатель производственных возможностей используется в процессе принятия решения о том, какие виды операций можно выполнять с помощью данного технологического процесса и для того, чтобы выявить, когда технологический процесс становится неконтролируемым под воздействием неслучайных причин.

Одним из методов проверки того, находится ли технологический процесс под контролем и функционирует ли он в пределах заданных производственных возможностей, являются контрольные карты Шухарта. Если измеряются количественные признаки, например, длина изделия, с помощью контрольных карт обычно проводится проверка изменений среднего значения и разброса или размаха результатов измерений. В случаях, когда осуществляется проверка качественных характеристик продукции, контрольные карты используются для регулирования доли бракованных изделий в выборке. В предположении, что процесс является контролируемым, определяются предупреждающие границы и границы регулирования. Их уровень рассчитывают, как правило, с использованием нормального распределения. Положение предупреждающих границ устанавливается на уровне ±2 стандартных ошибки, а положение границ регулирования — на уровне ±3 стандартных ошибки. При нахождении этих границ используются также биномиальное распределение и распределение Пуассона. Вне пределов каждой из предупреждающих границ находятся значения 2,5% выборок для процесса, который является контролируемым. Аналогично вне каждой из границ регулирования находится 0,1% выборок для процесса под контролем. Если выборочная характеристика лежит вне предупреждающих границ, то это может означать, что технологический процесс вышел из-под контроля. В таком случае сразу же осуществляют следующую выборку. Если результаты выборки оказываются за пределами границ регулирования, немедленно останавливают процесс производства и выявляют причины такой ситуации.

Схемы выборочного контроля при приемке товара используются для определения доли брака в партии продукции. Производитель и потребитель товара оговаривают допустимый (допустимый уровень качества — AQL) и максимальный (допустимый процент бракованных изделий в партии — LTPD) уровни брака. Разрабатывается схема выборки, по которой рассчитывается риск производителя, заключающийся в ошибочном отказе потребителя от приемки пригодной партии продукции, и риск потребителя, заключающийся в ошибочной приемке партии продукции, в которой число бракованных изделий оказалось максимальным. Схемы выборки основаны на биномиальном распределении, при их построении используются специальные таблицы.

УПРАЖНЕНИЯ

Упражнение 7.1

В некоторой компании планируется использовать для производства медных дисков диаметром 25 ± 0,2 мм штамповочный пресс. Производственная возможность пресса по данным измерений равна 0,28 мм.

Требуется:

1. Определить, можно ли использовать данный пресс для производства дисков, удовлетворяющих спецификации.
2. Если ответ в п.1 положительный, определить границы изменения среднего значения параметра наладки пресс таким образом, чтобы не производить диски, диаметр которых выходит за пределы, оговоренные в спецификации.

Упражнение 7.2

Для производства стержней из нержавеющей стали диаметром 60,30 ± 0,25 мм используется токарный станок. Производственная возможность станка по данным измерений, произведенных в нормальных условиях, составляет 0,40 мм.

Требуется:

1. Определить, может ли токарный станок удовлетворить требуемой спецификации.
2. Имеется второй токарный станок, производственная возможность которого равна 0,60 мм. Может ли этот токарный станок удовлетворить спецификации на производство стальных стержней?
3. Предположим, что для выполнения этой операции должен использоваться токарный станок, о котором шла речь в п. 2. Стержни с диаметром, меньшим оговоренного в спецификации, придется выбросить, а стержни с диаметром, большим оговоренного в спецификации, можно после предварительной сортировки пустить в дальнейшую переработку. Каково среднее значение диаметра, на которое должен быть настроен токарный станок, если стержни с диаметром, меньшим оговоренного в спецификации, должны появляться не чаще одного раза из 1000? Какую долю стержней при данном значении параметра нагрузки придется пустить в дальнейшую переработку?

Упражнение 7.3

Компания "Small plc" занимается производством столовых приборов. Она располагает режущими станками для нарезки заготовок из нержавеющей стали на отрезки определенной длины. Производственные возможности трех станков по данным измерений равны:

Режущий станок А: 0,16 см;
Режущий станок В: 0,29 см;
Режущий станок С: 0,10 см.

Заготовки для производства ножей, вилок и ложек различны и, следовательно, их нарезка осуществляется в соответствии с различными спецификациями.

Нож — длина равна $21{,}25 \pm 0{,}06$ см;
Вилка — длина равна $20{,}00 \pm 0{,}09$ см;
Ложка — длина равна $21{,}20 \pm 0{,}16$ см.

Требуется:

1. Определить при условии, что каждый из имеющихся режущих станков налажен на изготовление определенного вида заготовок в определенный момент времени, можно ли одновременно выпускать все три типа заготовок? Как в этом случае следует распределить заготовки по станкам?
2. Найти значения верхней и нижней границы параметра наладки каждого из станков.

Упражнение 7.4

Компания "Greens Food Products" выпускает консервированные супы нескольких видов. В настоящее время фасовочный станок налажен на расфасовку банок с овощным супом весом в 425 г. По данным измерений, производимых в течение длительного периода времени, было установлено, что стандартное отклонение станка равно 10 г.

Контроль за работой фасовочного станка осуществляется с помощью выборок объемом в 5 банок, которые производятся через каждые 20 мин. Содержимое каждой банки тщательно взвешивается.

Требуется:

1. Построить для фасовочного станка контрольные карты средних арифметических и изменчивости технологического процесса.
2. Ниже приведены результаты измерений для восьми выборок. Нанести эти показатели на контрольные карты средних арифметических и изменчивости технологического процесса и размаха. Описать работу фасовочного станка за период времени, в течение которого производились выборки.

Номер выборки	*Вес банки с супом*				
1	453,3	417,5	423,2	424,3	407,3
2	393,8	411,2	433,5	426,5	428,8
3	374,0	419,3	422,2	427,8	428,8
4	447,7	416,5	431,8	425,5	407,3
5	461,8	413,7	428,2	423,2	428,0
6	379,7	405,2	427,8	433,7	424,7
7	439,2	423,7	406,8	407,2	426,3
8	447,7	431,8	422,2	426,3	429,2

Упражнение 7.5

Компания "HVJ Engineering plc" выпускает детали для производства автомобилей. На одной из производственных линий осуществляются нарезка и обработка стальных осей. Длина каждой из заготовок, нарезаемых в настоящее время, равна 6,65 см. По результатам наблюдения за режущим станком в условиях, когда его работа была контролируемой, было установлено, что стандартное отклонение технологического процесса составляет 0,07 см. Контроль за ходом процесса осуществляется с помощью случайных выборок в количестве 4 оси, производимых через каждые 30 мин, и точных замеров осей, попавших в выборку.

Требуется:

1. Построить для данного режущего станка вышеуказанные контрольные карты средних арифметических и изменчивости технологического процесса и размаха
2. Ниже в таблице приведены результаты измерений по последним десяти выборкам из деталей, изготовленных на данном станке:

Номер выборки	*Длина стальной оси, см*			
1	6,79	6,68	6,60	6,61
2	6,46	6,56	6,75	6,59
3	6,39	6,60	6,61	6,63
4	6,65	6,72	6,73	6,56
5	6,62	6,67	6,70	6,70
6	6,65	6,58	6,75	6,79
7	6,69	6,73	6,61	6,61
8	6,65	6,51	6,72	6,73
9	6,55	6,66	6,71	6,58
10	6,61	6,70	6,59	6,74

Нанести эти данные на контрольные карты средних арифметических и изменчивости технологического процесса и размаха, построенные в п.1. Описать работу режущего станка за период времени, в течение которого производились выборки.

3. При условии, что режущий станок был ошибочно налажен таким образом, что средняя длина заготовки оказалась равной 6,78 см, определить вероятность того, что ближайшая следующая выборка не обнаружит эту ошибку.
4. Повторить расчеты границ регулирования в контрольных картах средних арифметических и изменчивости технологического процесса в предположении, что стандартное отклонение технологического процесса неизвестно.

Упражнение 7.6

В компании по производству одежды осуществляется контроль за выпуском рубашек типа Т два раза в день путем проверки изделий, попавших в случайную выборку объемом в 150 единиц. Каждый экземпляр считается либо прошедшим приемку, либо бракованным. Результаты последних 20 выборок изделий, изготовленных работником соответствующей квалификации на машине, тщательно подготовленной и отлаженной, с использованием хорошего сырья, использовались для построения контрольной карты по доле брака. Ниже приводятся соответствующие результаты:

Номер выборки	*Число бракованных изделий на каждые 150 рубашек*	*Номер выборки*	*Число бракованных изделий на каждые 150 рубашек*
1	4	11	2
2	1	12	4
3	6	13	8
4	3	14	3
5	4	15	5
6	7	16	4
7	3	17	6
8	9	18	5
9	6	19	3
10	5	20	2

Требуется:

1. Используя приведенные выше данные, построить контрольные карты для доли брака p. Определить контрольные границы, используя для аппроксимации биномиального распределения нормальное распределение и распределение Пуассона.
2. По контрольной карте, построенной для аппроксимации распределением Пуассона, проверить результаты следующих 10 выборок.

Номер выборки	*Число бракованных изделий на каждые 150 рубашек*	*Номер выборки*	*Число бракованных изделий на каждые 150 рубашек*
21	5	26	2
22	6	27	3
23	12	28	8
24	10	29	10
25	6	30	3

Прокомментируйте свои действия.

Упражнение 7.7

Компания "Taylors plc" закупает электронные детали у "JC Jones Ltd". Taylors собирается разработать схему выборочного контроля приемки продукции, чтобы проверить качество закупаемых деталей. В настоящее время рассматриваются три варианта схем выборки.

Ниже приведены кривые оперативной характеристики, соответствующие каждому из трех вариантов:

Схема А

Производится случайная выборка 10 деталей из каждой партии. Приемка партии деталей принимаются, если в выборке обнаружено не более одной бракованной детали.

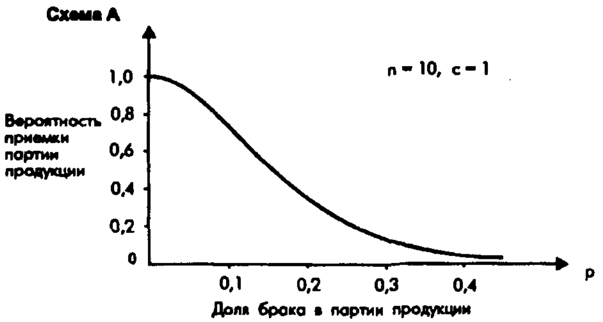


Схема В

Производится случайная выборка 20 деталей из каждой партии. Партия деталей принимается, если в выборке обнаружено не более двух бракованных деталей.

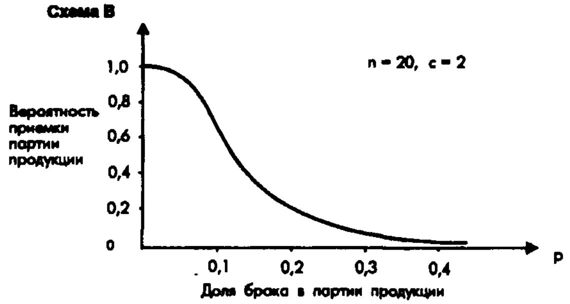
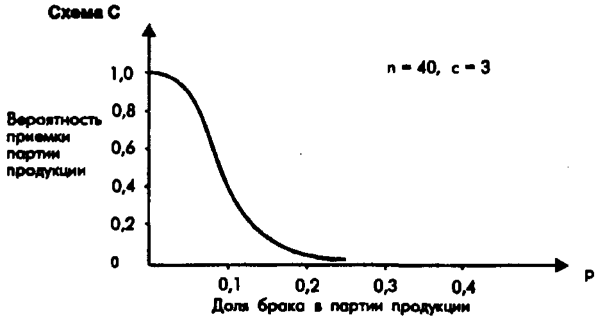


Схема С

Производится случайная выборка 40 деталей из каждой партии. Партия деталей принимается, если в выборке обнаружено не более трех бракованных деталей.



Требуется:

1. Определить, для какой из двух схем — А или В — приемка партии деталей, содержащей более 5% брака, является наиболее вероятной?
2. Определить, для какой из двух схем — А или В — отказ от приемки партии деталей, содержащей более 30% брака, наиболее вероятен?
3. Какая из схем в наибольшей степени отвечает интересам Taylor?
4. Какая из схем в наибольшей степени отвечает интересам Jones?
5. При условии, что допустимый уровень качества составляет 10%, какая из трех схем — А, В или С — является наиболее жесткой?
6. Если Taylor устанавливает, что допустимый уровень качества (AQL) равен 5%, а допустимый процент брака в партии (LTPD) — 20%, каковы значения риска производителя и риска потребителя в каждой из трех схем? Какая из этих трех схем является наилучшей с точки зрения Taylor?

7. Если с использованием схемы С осуществляется проверка 100 партий деталей, каждая из которых содержит 8% брака, каким приблизительно будет число партий деталей, которые не прошли приемку?
8. Если с использованием схемы С осуществляется проверка 1000 партий деталей, каждая из которых содержит 3% брака, каким приблизительно будет число партий деталей, которые прошли приемку?

Упражнение 7.8

Компания "Denton plc" выпускает пружины. По окончании процесса производства осуществляется контроль готовой продукции с использованием следующей схемы выборки: из каждой партии продукции случайным образом отбирается 35 пружин. Если в выборке обнаружено не более одного бракованного изделия, то осуществляется приемка всей партии продукции.

Заказчик "Denton plc" установил, что AQL равен 1%, а LTPD — 10%. Определите риск производителя и риск потребителя.

Упражнение 7.9

Администрация некоторого промышленного предприятия собирается сравнить две следующие схемы выборочного контроля качества поступающих деталей:

Схема А: $n = 12$, $c = 1$;
Схема В: $n = 25$, $c = 2$.

Администрация считает приемлемыми AQL, равный 4%, и LTPD, равный 25%.

Требуется:

1. Построить кривую оперативной характеристики для каждой из схем.
2. Сравнить обе схемы и определить, какая из них лучше.

Глава 8. ЛИНЕЙНАЯ РЕГРЕССИЯ

8.1. ВВЕДЕНИЕ

Статистический анализ в предыдущих главах касался поведения отдельных переменных. Теперь перейдем к анализу поведения двух или более переменных и связи между ними.

Например, рассмотрим компанию, которая регулярно помещает рекламу на один из своих товаров в местную газету. Компания ведет записи ежемесячно о суммах денег, затраченных на рекламу и поступивших от продажи этого товара.

Если реклама эффективна, то можно предположить, что вероятно существует какая-то связь между затратами на рекламу и соответствующими ежемесячными объемами продаж. Предположим, что чем больше сумма затрат на рекламу, тем больше объем продаж (по крайней мере, в определенных пределах). Не существует теоретической основы, исходя из которой мы могли бы написать уравнение, которое точно показало бы связь продаж с расходами на рекламу. Имеется ряд факторов, неразрывно связанных между собой, которые точно определяют ежемесячный объем реализаций. Это такие факторы, как цена товара, цена товара-конкурента, период времени, погодные условия. Тем не менее, если расходы на рекламу являлись бы главным фактором, определяющим продажу, то знание связи между этими двумя переменными было бы очень полезным для оценки объема продаж и соответствующего планирования финансовой политики компании.

Обычно для определения связи между переменными используется термин **«ассоциация»**. Термин «регрессия» используется для описания природы связи, термин «корреляция» — для измерения тесноты связи.

Нам необходимо знать, например, сильная ли связь между ежемесячными расходами на рекламу и ежемесячным объемом продаж. Знание этого фактора может обеспечить надежную оценку продаж. Если связь слабая, то ее изучение обеспечивает только описание продаж при весьма низкой надежности этого описания.

Процедура анализа связи между переменными необходима для установления природы любой связи. Теперь мы может разработать математическое уравнение или модель для описания этой связи с математической точки зрения. Линейные уравнения — простейшие для анализа, поэтому мы постараемся описать связь между переменными посредством линейной модели. Этот процесс носит название построения **линейной регрессии**. Степень пригодности линейной модели к исходным данным является индикатором силы линейной связи между переменными, а следовательно, и надежности любых оценок, производимых при помощи этой модели. На этой стадии полезно графическое представление данных.

Рис. 8.1 показывает, что линейная модель может быть применена при описании связи между продажей и расходами на рекламу.

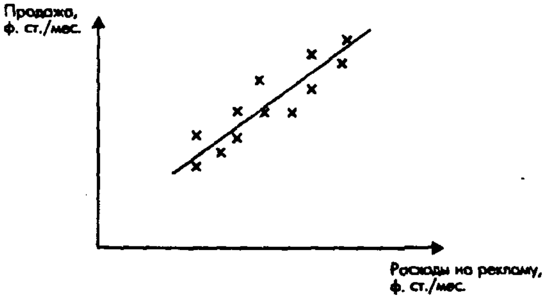


Рис. 8.1. Пример линейной связи

Если бы мы получили другой график (см. рис. 8.2), то можно было бы сделать вывод, что линейная модель не применима при описании связи между объемом продаж и расходами на рекламу.

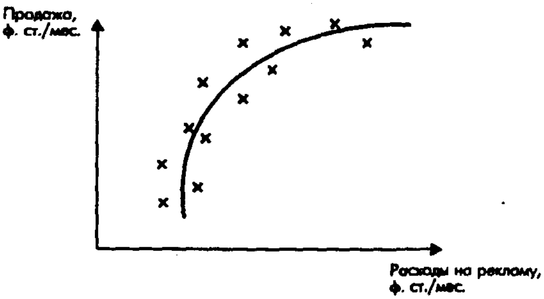


Рис. 8.2. Пример нелинейной связи

Линейная регрессия — первый пример использования математических моделей. Цель любой модели — помочь понять конкретную ситуацию, а, возможно, и объяснить ее путем последующего анализа. Мы можем использовать модель для того, чтобы делать какие-либо прогнозы или предсказания. Модель обычно является упрощением реальной ситуации. Мы должны сделать простейшие предположения, чтобы суметь сконструировать модель, которая давала бы возможность управления, но сама по себе модель должна быть все-таки достаточно реалистичной, чтобы заслуживать внимания. Модели линейной регрессии используются наиболее часто.

Они включают в себя как простые модели для двух переменных, с которыми мы главным образом столкнемся, так и более совершенные модели для многих переменных, которых мы лишь коснемся. Эти модели широко используются потому, что существуют пакеты прикладных программ (ППП), которые осуществляют требуемые расчеты. Нужно быть предельно внимательным при использовании ППП для того, чтобы окончательно убедиться, что мы досконально понимаем результаты и правильно их оцениваем.

Эта глава охватывает анализ простой модели линейной регрессии, построенной на конкретных данных (*парной линейной регрессии*). В конце главы рассматриваются множественные регрессионные модели, несколько моделей нелинейной связи и, наконец, измерение корреляции с использованием коэффициента ранговой корреляции Спирмена.

8.2. ПРОСТАЯ МОДЕЛЬ ЛИНЕЙНОЙ РЕГРЕССИИ

Простая *линейная регрессия* связана с тем, что мы называем двумерным распределением, т.е. распределением двух переменных. Существует ли линейная связь между двумя переменными или нет? Всегда лучше использовать две переменные, нежели одну. Например, нас интересует соотношение между ростом и весом у определенной группы людей; между ценой и количеством проданного товара; возрастом служащих и их заработной платой; возрастом и весом кур; еженедельными издержками и отработанным временем в отделах; пройденной дистанцией и затраченным временем.

Первым шагом в анализе является изучение переменных: какие из них относятся к факторам, каково их влияние друг на друга. Предположим, что фермер хочет предсказать вес кур, которых он выращивает. Вес — это переменная, которую он желает предсказать, поэтому это будет **зависимая переменная**. Отмечать значения зависимой переменной будем на **оси OY**. Пусть вес курицы зависит от ее возраста. Тогда возраст — это **независимая переменная**, значение которой нам известно по предположению и которое мы можем использовать при оценке ее веса. Независимая переменная будет отмечаться нами на **оси OX**. Если мы установим природу связи между возрастом и весом курицы, то сможем предсказать вес курицы в данном возрасте. Любая курица, для которой реальный вес значительно отличается от прогнозируемого, может быть подвергнута обследованию.

Теперь мы должны ответить на вопрос: как изменяется вес в зависимости от изменения возраста. Во-первых, можно предположить, что вес увеличивается с возрастом. Когда курица совсем взрослая, мы можем предположить, что ее вес с небольшими отклонениями зависит от пищи и погодных условий. Прибавка в весе и ее вес в зрелом возрасте также будет зависеть от породы и способа ее выращивания и кормления. Существует также множество других факторов, помимо возраста, влияющих на вес. Процесс исследования возможной связи переменных — зависит ли зависимая переменная у от независимой переменной х и от других факторов, которые также могут повлиять на связь, — очень важная часть статистического моделирования. Наша цель — не просто построить какую-то любую линейную регрессию, а постараться выяснить, чем объясняется вариация веса курицы с помощью моделирования, и решить, можно ли определить вес курицы, зная только ее возраст.

Вероятно, выводом из вышеприведенной задачи будет то, что существуют несколько взаимосвязанных между собой факторов для определения точного веса конкретно взятой курицы. Общая картина связи представлена на рис. 8.3.

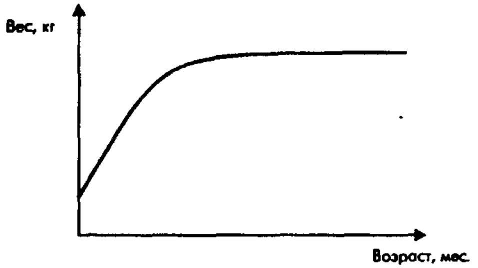


Рис. 8.3. Возможная связь между весом и возрастом курицы определенной породы

Теперь мы должны собрать данные для того, чтобы проверить правильность наших предположений о наличии и характере связи между переменными.

❒ **Пример 8.1.** Пример касается времени, которое занимают поставки. Мы займемся специальными услугами — поставками на короткие расстояния внутри города. Оценим стоимость услуги, определив время поставки при любом расстоянии.

Факторы, помимо пройденного расстояния, которые повлияют на затраченное время: пробки на дорогах, время суток, дорожные работы, погода, дорожная система, водитель, вид транспорта. Однако первоначальное исследование будет предельно простым, насколько это возможно: будем рассматривать связь только между расстоянием, измеряемым кратчайшим маршрутом на линиях, и затраченным временем в минутах. Рассмотрим всевозможные поездки за определенный период, которые могут быть совершены в городе. Измерим время и расстояние каждой десятой поездки, начиная с произвольно выбранного часа и дня недели. Фирма работает шесть дней в неделю, кроме воскресенья. Случайное число, выброшенное игральной костью — 2, таким образом, следующий вторник — выбранный нами день. Услуги оказываются с 8 ч утра до 6 ч вечера. Случайное число от 0 до 9, полученное из таблицы случайных чисел для выбора времени, оказалось числом 6. Таким образом, первая поездка после часа дня (т. е. шестой час, начиная с восьми утра). Затем мы отберем каждую десятую поставку. Выборочные данные первых десяти поставок будут использованы для анализа.

Таблица 8.1. **Исходные данные о расстоянии и времени поставок**

Расстояние, миль	*Время, мин*
3,5	16
2,4	13
4,9	19
4,2	18
3,0	12
1,3	11
1,0	8
3,0	14
1,5	9
4,1	16

Нам нужно объяснить изменения времени (переменная y), принимая расстояние в качестве независимой переменной x. Предположим, что затраченное время растет по мере увеличения расстояния. Представим данные на графике, чтобы определить связь, которая существует между переменными (рис. 8.4).

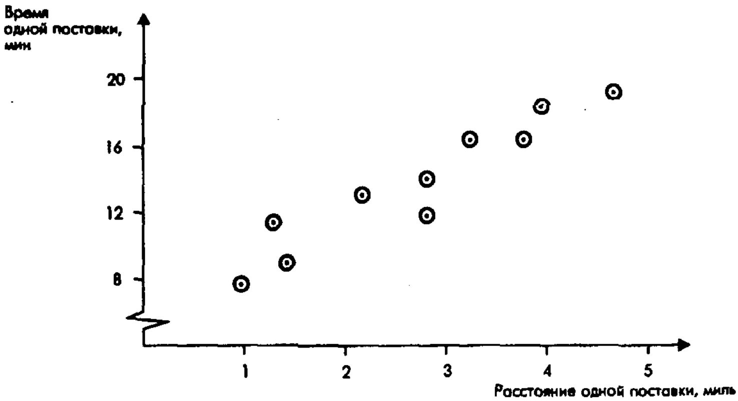


Рис. 8.4. Зависимость времени поставок от расстояния по совокупности случайных данных о поставках

Рис. 8.4 свидетельствует об общем увеличении времени с увеличением расстояния. Точки на графике собраны в пучки вокруг прямой линии. Это означает, что мы можем использовать линейную модель для описания связи между двумя переменными. Точки не находятся точно на линии. Но было бы удивительно, если бы это было так, с точки зрения остальных факторов, которые могут повлиять на время поездки. Линейная модель, описывающая связь между двумя переменными, будет приближением к действительности — к истинному времени и расстоянию. Рис. 8.6 показывает наилучший вариант.

Для совокупности, данные из которой мы используем, существует множество различных расстояний при различном времени. Фактически, для любого расстояния существует распределение возможного времени поставок. Наш пример включает десять поездок. Их можно сгруппировать по дальности поставки. Например, поставки на расстояние 1,0 мили 1,3, 1,5, 2,4, 3,0 миль и т.д.

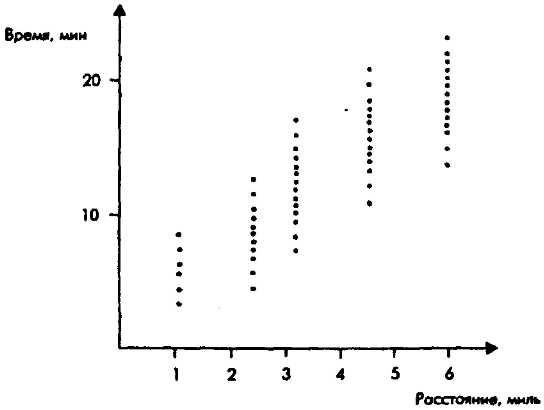


Рис. 8.5. Распределение времени поставок при определенном расстоянии

Эта идея важна для последующего анализа. Вернемся к нашим предположениям: наилучшей моделью для описания связи между временем поездки и расстоянием будет линейная модель. Теперь нам необходимо найти способ для нахождения приемлемой схемы определения точек этой линии по исходным данным. Эта линия называется **линией наилучшего подбора**.

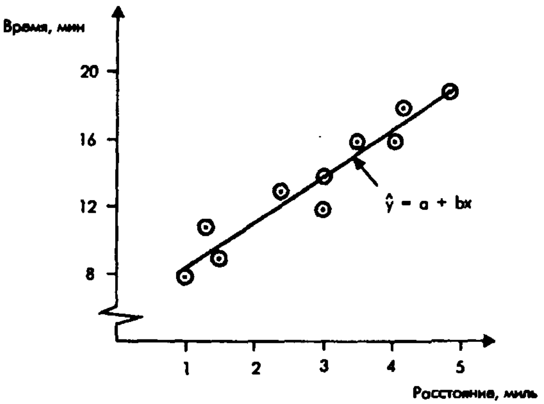


Рис. 8.6. Зависимость времени поставки от расстояния поставки

Показанная на рис. 8.6 линия — это возможная линейная модель для описания связи между переменными. Уравнение этой линии может быть записано следующим образом:

$$\hat{y} = a + bx,$$

где a — определяется как пересечение линии регрессии с осью y; b — угол наклона линии регрессии, называется **коэффициентом регрессии**. Рассмотрим конкретное значение пройденного пути, которое мы обозначим как x_i. Для x_i фактическое время будет y_i, тогда как время, прогнозируемое линейной моделью (теоретическое) определяется из уравнения:

$$\hat{y}_i = a + bx_i.$$

Разница между этими двумя значениями:

$$e_i = y_i - \hat{y}_i,$$

называется **ошибкой**, или **отклонением**, или **остатком**. Мы можем определить величину ошибки для всех отмеченных точек. Линейная модель, которая наилучшим образом аппроксимирует данные — одна из тех, для которой общая ошибка выборки имеет наименьшее значение. Чтобы рассчитать ее, нужно избежать позитивных и негативных значений. Это можно сделать, возведя все ошибки в квадрат и делая их положительными величинами. Линия наилучшего подбора — та, которая минимизирует квадраты разниц между рассматриваемыми значениями y и соответствующими значениями x, рассчитанными с помощью линии наилучшего подбора. Эта линия называется линией регрессии, полученной методом наименьших квадратов. Может быть избран и другой критерий подбора наилучшей линии.

Используя различные расчеты, можно определить наклон и пересечение линии регрессии с осью OY методом наименьших квадратов.

Формулы для определения угла наклона линии регрессии и ее пересечения с осью OY следующие:

$$\text{наклон } b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2},$$

где n — размер выборки;

$$\text{пересечение } a = \frac{\sum y}{n} - \frac{b\sum x}{n}.$$

Соответствующие расчеты для выборки n=10 даны ниже. Линейная модель:

$$\hat{y} = a + bx.$$

Таблица 8.2. **Расчет линии регрессии**

Расстояние, x, миль	*Время, y, мин*	*xy*	x^2	y^2
3,5	16	56,0	12,25	256
2,4	13	31,2	5,76	169
4,9	19	93,1	24,01	361
4,2	18	75,6	17,64	324
3,0	12	36,0	9,0	144
1,3	11	14,3	1,69	121
1,0	8	8,0	1,0	64
3,0	14	42,0	9,0	196
1,5	9	13,5	2,25	81
4,1	16	65,6	16,81	256
Итого: 28,9	136	435,3	99,41	1972

Последние три графы используются в последующих расчетах.

$$\text{Показатель наклона } b = \frac{10 \cdot 435{,}3 - 28{,}9 \cdot 136}{10 \cdot 99{,}41 - 28{,}92} = \frac{422{,}6}{158{,}9} = 2{,}66;$$

$$\text{Показатель пересечения } a = \frac{136}{10} - \frac{2{,}66 \cdot 29{,}9}{10} = 5{,}91.$$

Подставим эти значения в линейную модель:

$$\hat{y} = 5{,}91 + 2{,}66x.$$

Время поставки (мин.) = 5,91 + 2,66 расстояние (миль).

Наклон линии регрессии (2,66 минут на милю) — это рассчитанное количество минут, приходящиеся на одну милю расстояния поставки. Пересечение (5,91 минут) — это рассчитанное время для подготовки к поездке и доставки товаров, т.е. необходимое для каждой поездки время в сравнении с реально затраченным временем. Пересечение дает средний эффект всех влияющих факторов за исключением расстояния. Напомним, что в нашем примере эти значения основаны на малом объеме данных. Мы должны оценить реальность полученных оценок, т.е. подсчитать доверительные интервалы для параметров a и b.

8.3. ТЕСНОТА ЛИНЕЙНОЙ СВЯЗИ — КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ r

В приведенном выше примере данные подтвердили обоснованность линейной модели. Однако мы не имеем объективного представления о том, насколько хорошо аппроксимирует данные линейная модель. Подбор на основе графика в данном случае оказался точным, но он может быть обманчивым, так как распределение точек на графике зависит от выбора масштаба. Необходимо объективное измерение тесноты линейной связи.

Мы полагаем, что связь между переменными существует. Рассмотрим две переменные x и y. Поле точек представлено на диаграмме рассеяния (рис. 8.7), на которой показана и линия регрессии, полученная методом наименьших квадратов. На этом графике добавлена линия $y = \bar{y}$.

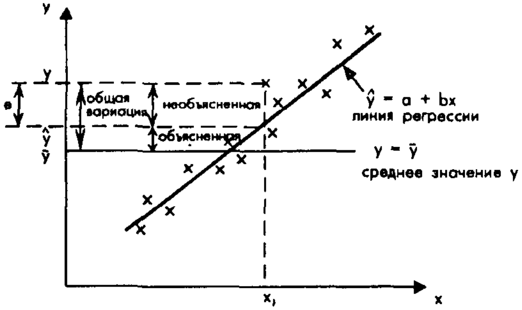


Рис. 8.7. Структура дисперсии зависимой переменной y

Если мы возьмем конкретное значение x, допустим x_1, то в любой точке выборки значению x будет соответствовать значение y. Фактически это могут быть несколько точек с одним и тем же значением x и разными значениями y, но в каждом случае фактическое значение y может быть разбито на два компонента. Это можно записать как: действительное значение y равно значению исходя из линейной связи между y и x плюс значение y, обусловленное другими факторами:

$$y = \bar{y} + e,$$

где e — остаток, разница между фактическим значением y и значением y на прямой.

Линейная связь только частично объясняет вариации значений y. Необъясненная часть является остатком, e. Если бы связь между x и y была абсолютно линейной, то все e были бы равными 0. По мере того, как сила линейной связи уменьшается, остаток увеличивается. Это соотношение формирует основу, на которой мы можем рассчитать силу линейной связи. Мы должны рассмотреть все точки, а не только одну–две. Общая вариация значения y может быть записана как:

$$\sum (y - \bar{y})^2,$$

Общая вариация значений y не зависит от значения x. Общее изменение y с учетом линейной связи между x и y может быть записано:

$$\sum (\hat{y} - \bar{y})^2.$$

Это выражение соответствует той части вариации y, которая объясняется регрессией, т.е. введением независимой переменной x, поскольку вариация x и y связывается уравнением y = a + bx. Вариация y, которая не объясняется линейной связью, записывается как:

$$\sum(y-\hat{y})^2.$$

Эта вариация возникает из-за других факторов, не включенных в линейную модель, т.е. эта вариация не объясняется данной регрессией.

Отношение объясненной вариации к общей вариации используется как мера линейности связи. Чем теснее связь, тем ближе это отношение к 1. Это отношение называется **коэффициентом детерминации**, обозначается r^2 и имеет вид:

$$r^2=\frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}.$$

Коэффициент детерминации часто выражается в процентах и показывает величину дисперсии y, которая объясняется независимой переменной x, включенной в модель в случае полной линейной связи между x и y $r^2=1$, или 100%. Если связь отсутствует, то r^2 равно 0. Коэффициент детерминации не определяет, увеличивается ли или уменьшается y с ростом x. Эта информация может быть получена с помощью коэффициента корреляции Пирсона, который включает произведение переменных x и y; он обозначается r. Этот коэффициент может быть получен как квадратный корень из коэффициента детерминации:

$$r=\sqrt{\frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}}.$$

Для вычислений полезно алгебраически преобразовать это выражение и воспользоваться следующей формулой:

$$r=\frac{n\sum xy-\sum x\cdot\sum y}{\sqrt{(n\sum x^2-(\sum x)^2)\cdot(n\sum y^2-(\sum y)^2)}}.$$

Это и есть выборочный коэффициент корреляции. Значение r всегда лежит между −1 и +1. Знак r такой же, как и знак коэффициента регрессии b. Если b — положителен, показывая положительную связь между переменными, то коэффициент корреляции r будет также положительным. Если коэффициент регрессии b меньше нуля, то и коэффициент корреляции r также отрицательный.

По мере того, как возрастает сила линейной связи, точки на графике будут лежать более близко к прямой линии, а величина r будет ближе к 1. По мере уменьшения силы связи значение r будет ближе к 0, а точки будут более рассеяны. При r=0 линейной связи не существует. Но это не значит, что не существует вообще никакой связи. На рис. 8.8 и 8.9 отражены случаи, когда значения коэффициента корреляции приближаются к 0.

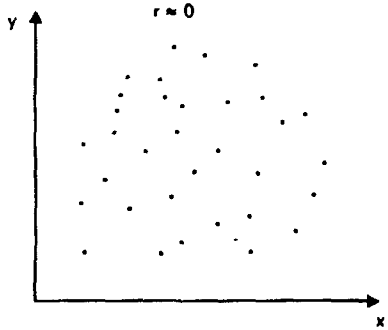


Рис. 8.8. Случай отсутствия связи между переменными

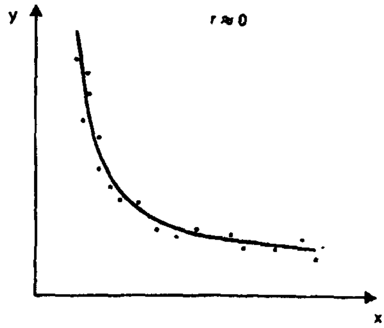


Рис. 8.9. Сильная нелинейная связь между переменными

Вернемся к примеру 8.1, в котором рассматривается модель прогноза времени поставки в зависимости от расстояния внутри города. Коэффициент корреляции рассчитывается следующим образом:

$$r = \frac{n\sum xy - \sum x \cdot \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2) \cdot (n\sum y^2 - (\sum y)^2)}} .$$

По нашим данным коэффициент корреляции равен:

$$r = \frac{10 \cdot 435{,}2 - 28{,}9 \cdot 136}{\sqrt{(10 \cdot 99{,}41 - 28{,}9^2) \cdot (10 \cdot 1972 - 136^2)}} = \frac{422{,}6}{\sqrt{158{,}9 \cdot 1224}} = 0{,}958.$$

Это значение коэффициента корреляции очень близко к единице, что свидетельствует об очень тесной линейной связи между расстоянием и временем поставки. Этот вывод подтверждает первоначальное предположение, сделанное исходя из диаграммы.

Коэффициент детерминации ($r^2 \cdot 100\%$) показывает процент общей вариации времени поставки, который зависит от расстояния. В нашем случае коэффициент детерминации высок:

$$r^2 = 0{,}958^2 \cdot 100 = 91{,}8\% .$$

По выборочной модели можно вычислить ожидаемое время при заданном расстоянии поставки:

время (мин.) = 5,91 + 2,66 (расстояние в милях).

Выборочная модель объясняет 91,8% вариации времени доставки. Не объясняется 8,2% вариации времени поездки. Эта часть вариации обусловлена всеми остальными факторами, влияющими на время поездки, но не включенными в модель.

8.4. ПРЕДСКАЗАНИЯ И ПРОГНОЗЫ НА ОСНОВЕ ЛИНЕЙНОЙ МОДЕЛИ РЕГРЕССИИ

8.4.1. Прогнозы с упорядоченными данными

Мы можем использовать модель для прогноза времени поездки на любые расстояния. Если расстояние равно 4,0 мили, то среднее время поставки:

$$\hat{y} = 5{,}91 + 2{,}66 \cdot 4{,}0 = 16{,}6 \text{ мин.}$$

В расчетах такого рода требуется осторожность: не рекомендуется использовать модель для прогноза при тех значениях независимой переменной, которые не входят в исходные данные. В нашем случае расстояние изменяется от 1,0 мили до 4,9 мили. Не очевидно, что модель подойдет для данных, не входящих в этот интервал.

Связь между временем и расстоянием может изменяться по мере увеличения расстояния. Например, дальняя поездка может включать использование скоростных шоссе, тогда как наша модель описывала связь с учетом медленных городских поездок. Дальние перевозки должны включать остановки на отдых или перекус, которые безусловно изменяют затраченное время.

Если бы нам нужно было экстраполировать расчеты для расстояния, выходящего за указанные пределы, мы должны были бы собрать больше данных. Если бы мы решили не делать этого, то должны быть предельно осторожны при использовании прогнозных значений времени поездок. Но эти прогнозы были бы, вероятно, ненадежны.

8.4.2. Оценки, ошибки и остатки

Насколько точными должны быть наши прогнозы? В следующей части мы рассмотрим вопросы, связанные с доверительными интервалами. Однако также полезно оценить надежность, сравнив значения зависимой переменной y и предсказанные значения $\bar{y}$ для каждого значения независимой переменной x. Эти ошибки, или остатки e — необъясненная часть каждого наблюдаемого значения y, являются чрезвычайно важными по двум причинам. Во-первых, они позволяют проверить, применима ли данная модель и те предположения, на которых она основана. Во-вторых, мы можем использовать их для того, чтобы дать грубую оценку вероятных ошибок прогнозов, сделанных на основе линейной модели.

Табл. 8.3 содержит значения остатков для примера 8.1.

Таблица 8.3. **Расчет остатков $(y-\hat{y})$**

Расстояние, миль x	*Фактическое время, мин* y	*Вычисленное время* $y = 5{,}91+2{,}66x$, *мин*, $\hat{y}$	*Остаток* $e=(y-\hat{y})$, *мин*
3,5	16	15,22	+0,78
2,4	13	12,29	+0,71
4,9	19	18,94	+0,06
4,2	18	17,08	+0,92
3,0	12	13,89	−1,89
1,3	11	9,37	+1,63
1,0	8	8,57	−0,57
3,0	14	13,89	+0,11
1,5	9	9,90	−0,90
4,1	16	16,82	−0,82

Мы можем проверить удовлетворительность модели, нанося остатки $(y-\hat{y})$ на ось ординат против вычисленных значений y, с учетом значений x. Эта процедура очень важна при построении множественной регрессии, когда исходные данные не могут быть нанесены на исходную диаграмму, т.е. линейность предположенной связи может быть оценена в полной мере только через анализ остатков. Если линейная модель является точной, разности, или остатки, будут носить случайный характер и их сумма будет близка к нулю. Изображение разностей, или остатков, для данного примера дано на рис. 8.10.

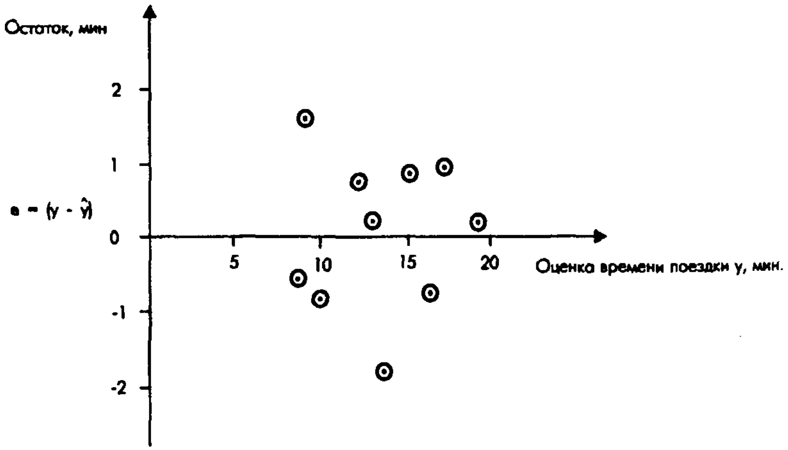


Рис. 8.10. График остатков $(y-\hat{y})$ против вычисленных значений $\hat{y}$

Если бы связь была нелинейной, то рисунок показал бы это очень четко. Пример эффекта линейной модели показан на рис. 8.11 и 8.12.

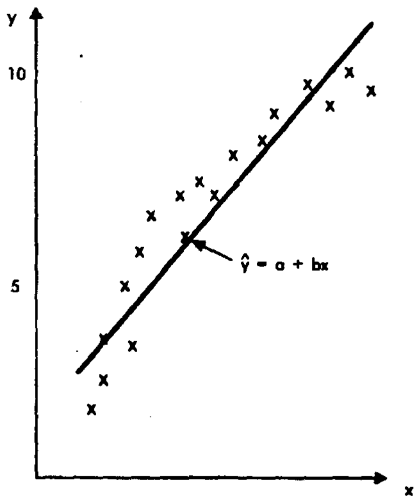


Рис. 8.11. Исходные данные и вычисленная линия регрессии

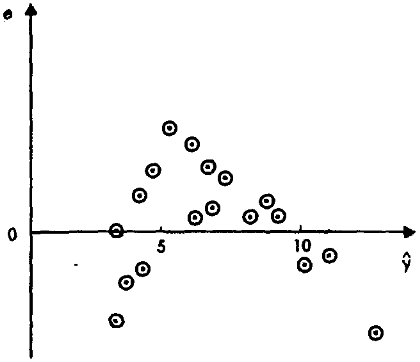


Рис. 8.12. Остатки, свидетельствующие о криволинейности связи

Остатки позволяют оценить рассеяние ошибок. Одним из основных предположений в методе наименьших квадратов является то, что рассеяние данных возле линии регрессии одинаково при всех значениях x (см. рис. 8.13 и рис. 8.14):

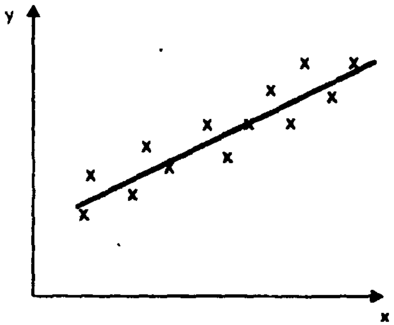


Рис. 8.13. Постоянная вариация для всех значений x

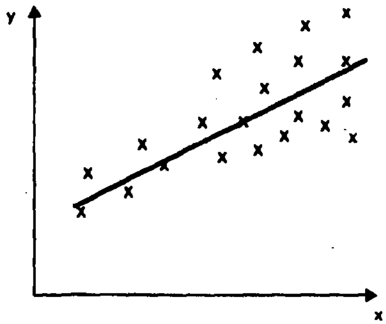


Рис. 8.14. Вариация, изменяющаяся со значениями x

Варьирующие данные постоянно пересекают линию, следовательно, для подбора наилучшей линии может быть использован метод наименьших квадратов (рис. 8.13).

На рис. 8.14 показан пример данных, которые распределяются вдоль линии регрессии неравномерно, в этом случае метод наименьших квадратов непригоден для подбора "наилучшей" линии.

Если остатки распределяются так, как на рис. 8.15, то мы делаем вывод, что вариация y изменяется с изменениями x.

В нашем примере выделяются два значения остатка (−1,89 и 1,63). Они свидетельствуют, что данные не соответствуют предположениям о едином характере вариации. Следовательно, доверительный интервал, описанный в следующей части, не будет применим.

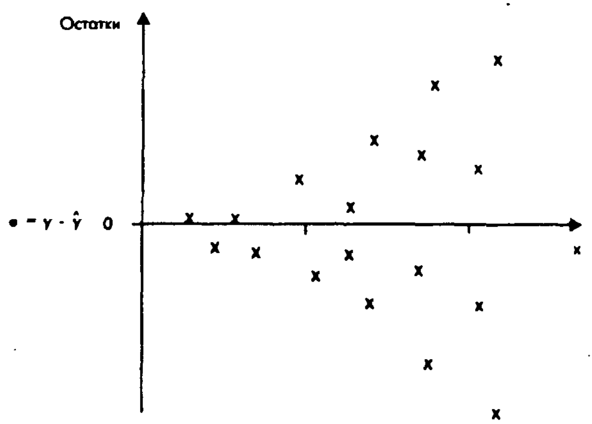


Рис. 8.15. График остатков для данных, вариация которых не является постоянной для всех значений x

Единственным способом продолжать статистический анализ доверительных интервалов и испытание гипотез в таком случае является трансформация данных (часто используются логарифмы значений x) до тех пор, пока график остатков не покажет случайное рассеяние точек относительно $e = 0$ при небольших значениях остатков.

Оценка устойчивости и колеблемости линейной модели регрессии может быть длительной процедурой, особенно если ряд данных увеличивается не постоянно, а с замедлением. Существует множество ППП для статистической обработки. Может показаться, что дело обстоит просто: нужно собрать данные и ввести их в компьютер. Программа рассчитывает линейную модель, не обращая внимание на то, годится она или нет. В следующем разделе мы рассмотрим, как установить и оценить линейную регрессионную модель.

8.5. СТАТИСТИЧЕСКИЙ ВЫВОД В АНАЛИЗЕ ЛИНЕЙНОЙ РЕГРЕССИИ

8.5.1. Основные предпосылки

В этом разделе мы обсудим одно из обязательных предположений, лежащих в основе дальнейшего анализа линейной регрессионной модели. Данные, по которым построена модель линейной регрессии, являются выборкой пар значений x и y. В сущности, мы используем выборку для построения модели, которая в общем виде представит связь. Связь между зависимой переменной y и независимой переменной x описывается как:

$$y = \alpha + \beta x + \varepsilon,$$

где ε — отклонение от значения y на линии:

$$y = \alpha + \beta x.$$

Для данного значения x: $y = \alpha + \beta x$ — линейная модель, которую можно построить, имея все необходимые данные.

Для любого данного x генеральная совокупность значений y должна быть нормально распределена вдоль линии регрессии, при постоянной вариации для всех x (см. рис. 8.16), y — среднее значение всех значений y для данного x.

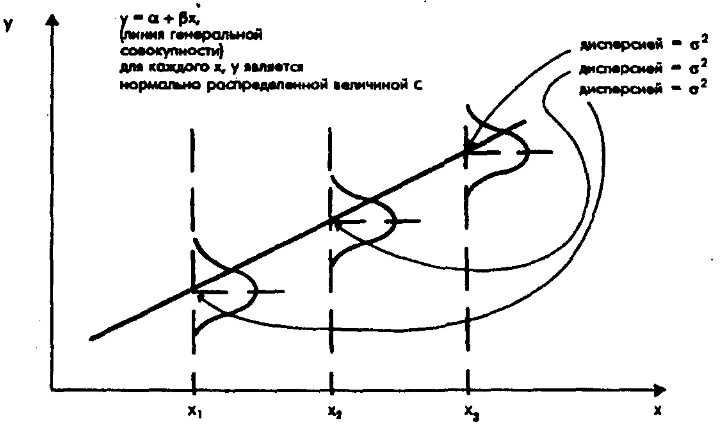


Рис. 8.16. Распределение значений y в генеральной совокупности

Как и прежде, греческие буквы обозначают параметры совокупности, такие как μ и σ. ε — ошибки или остатки, разность между фактическими значениями y и средней величиной y на линии. Если метод наименьших квадратов применен, чтобы получить линию, которая была наиболее подходящей к данным, то это может быть достигнуто, если мы минимизируем $\sum e^2$. Линейная модель, которую мы вычисляем по выборке, имеет вид:

$$\hat{y} = a + bx,$$

где $\hat{y}$ – оценка генеральной средней y для данного значения x, a и b — выборочные оценки параметров генеральной совокупности α и β.

Как в любом случае, если мы произведем вторую выборку, значения a и b будут другими. Существует аналогия между использованием $\bar{x}$ для оценки μ и использованием a для оценки α. Делая предположение относительно выборочного распределении $\bar{x}$, мы находим доверительный интервал для величины генеральной средней μ.

Точно такая же процедура может быть использована для получения α и β путем вывода выборочных величин a и b. Нашей основной моделью является:

$$y = \alpha + \beta x + \varepsilon .$$

Предположения:

1. Связь является линейной;
2. Независимая переменная x предполагается известной и может быть использована для прогноза y;
3. Ошибки, или остатки ε, нормально распределены;
4. Для любых данных x ожидаемое значение e равно 0, т.е. $E(\varepsilon)=0$;
5. Дисперсия постоянна для всех значений x, т.е. дисперсия $\varepsilon=\sigma^2$;
6. Ошибки независимы.

Если придерживаться этих предположений, то распределение значений y в генеральной совокупности для данного x является нормальным со средней:

$$\mu_{y/x} = \alpha + \beta x,$$

где $\mu_{y/x}$ обозначает среднее y для данного x при дисперсии, равной σ^2

Линия регрессии, построенная по выборочным данным, является лучшей оценкой линии генеральной совокупности, с a — лучшей оценкой α и b — лучшей оценкой β. Так как существует множество всевозможных выборок, которые могут быть произведены из данной генеральной совокупности, нельзя быть уверенным, что эта выборка произведена именно из данной генеральной совокупности. Должно быть проведено испытание гипотез по данным выборки для того, чтобы установить соответствие выборки генеральной совокупности. Прежде всего, насколько уверенно мы можем говорить о линейной связи в исходной совокупности. Если в совокупности линейная связь отсутствует, то коэффициент корреляции генеральной совокупности ρ будет равен нулю и β — показатель наклона линии регрессии также будет равен нулю. До проверки линейности нам необходимо вычислить доверительные интервалы для показателя наклона β, точки пересечения α, при среднем значении y для данного x или при индивидуальном значении y для данного x. Как и в предыдущих главах, будем использовать случайную выборку для расчета выборочных статистик и для оценки соответствующих параметров совокупности.

8.5.2. Испытание гипотезы для оценки линейности связи

Воспользуемся данными случайной выборки из генеральной совокупности для измерения линейной связи для совокупности. Мы не знаем, является ли связь в генеральной совокупности линейной. Случайная выборка может свидетельствовать о линейности связи переменных, тогда как в действительности в генеральной совокупности связь может быть нелинейной. Такого рода возможности показаны на рис. 8.17 и 8.18.

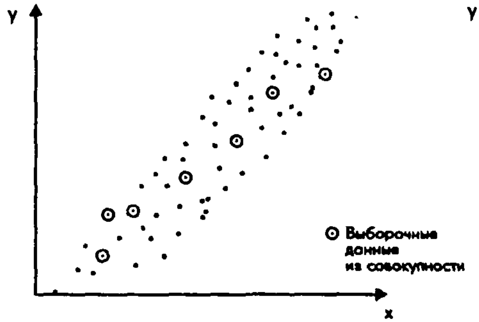


Рис. 8.17. Случайная выборка из генеральной совокупности, с линейной связью

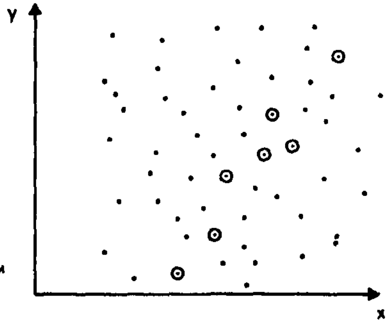


Рис. 8.18. Случайная выборка с нелинейной связью

Необходимо установить вероятность того, что линейная связь в выборочной совокупности свидетельствует о линейной связи в генеральной совокупности. В решении этой задачи нам поможет испытание гипотезы. Как в любой ситуации, где используются гипотезы, мы не можем без сомнения утверждать, что связь в генеральной совокупности совместима со связью в выборочной совокупности. Определим совместимость через испытание нулевой гипотезы. Линейная регрессия отображается в нескольких статистиках и можно провести проверку гипотезы для каждой из них, а потом сделать совокупный вывод. Нулевые гипотезы при этом формируются аналогично вышеуказанным. В данном случае нулевая гипотеза означает отсутствие линейной связи между зависимой и независимой переменными в генеральной совокупности.

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ρ

Вычисление коэффициента корреляции Пирсона зависит от размера выборки. Если мы интерпретируем величину r с точки зрения выборки, то не принимаем во внимание ее размер. Вывод о связи в генеральной совокупности зависит от размера выборочной совокупности. Так, если мы получили коэффициент корреляции, например 0,90, который рассчитан для выборочной совокупности из шести единиц, и сравниваем его с таким же значением, которое было рассчитано для совокупности из двадцати единиц, то во втором случае мы более уверены, что связь в генеральной совокупности — линейная. Шанс получения выборочной совокупности, в которой связь линейна, из генеральной совокупности, в которой связь не линейна, уменьшается по мере увеличения размера выборки. Коэффициент корреляции оценивается с помощью t-критерия:

H_0: Между переменными x и y не существует линейной связи, иначе говоря, независимая переменная x не помогает в предсказании значений y, т.е. $\rho=0$.

H_1: $\rho \neq 0$, т.е. между переменными x и y существует некая линейная связь, x помогает в прогнозировании y.

Используя эти альтернативные гипотезы, мы получим двусторонний критерий. Если бы мы решили, что ρ должно быть только положительным, то H_1: $\rho>0$ и мы использовали бы односторонний критерий:

$$t = \sqrt{\frac{r^2(n-2)}{(1-r^2)}}.$$

Количество степеней свободы равно $(n-2)$, так как мы рассчитали $\bar{x}$ и $\bar{y}$ для нахождения r, используя две степени свободы, n — число пар значений выборки. Если бы нам понадобилось провести испытание при 5%-ном уровне значимости, используя двусторонний критерий, полученное значение критерия нужно сравнить с $t_{0,025,(n-2)}$ из Приложения 2.

Для того чтобы проиллюстрировать наши действия, вернемся к примеру 8.1. Мы получили значение коэффициента корреляции $r=0,958$. Тогда значение критерия:

$$t = \sqrt{\frac{0,958^2 \cdot 8}{(1-0,958^2)}} = \sqrt{\frac{7,342}{0,082}} = 9,45.$$

Количество степеней свободы: $(10-2) = 8$

По таблицам Приложения 2 находим: $t_{0,025,8} = 2,306$.

Рассчитанное значение критерия (9,45) больше, чем 2,306. Поэтому мы отвергнем H_0 на 5%-ном уровне значимости и выберем гипотезу H_1, т.е. мы в праве предположить, что коэффициент корреляции в генеральной совокупности не равен нулю, и что между временем и расстоянием существует линейная связь. Этот результат можно было предвидеть, учитывая высокое значение коэффициента корреляции r.

КРИТЕРИЙ ПРОВЕРКИ ГИПОТЕЗЫ О СВЯЗИ НА ОСНОВЕ ПОКАЗАТЕЛЯ НАКЛОНА ЛИНИИ РЕГРЕССИИ

В простой линейной регрессии критерий показателя наклона — коэффициента регрессии, выполняет те же функции, что и критерий коэффициента корреляции. Поэтому мы проводим либо испытание r, либо b, но не оба сразу. В уравнении множественной регрессии, где имеется коэффициент регрессии для каждой независимой переменной, необходимы оба критерия, и они выполняют различные функции.

H_0: Между переменными нет линейной связи и x не помогает в прогнозе y, т.е. $\beta=0$.

H_1: $\beta \neq 0$, т.е. существует линейная связь, и x помогает в прогнозе y.

В этом случае используют двусторонний критерий. Однако как и при испытании r, мы можем заменить этот критерий на односторонний, если предполагаем, что $\beta>0$ или $\beta<0$ — более значимые гипотезы. Формула критерия похожа на ту, что мы использовали для μ и p в гл.6. Когда дисперсия генеральной совокупности неизвестна, тестовая статистика для выборочной средней определяется как:

$$t = \frac{\text{(выборочная статистика – параметр, предполагаемый в гипотезе } H_0)}{\text{наилучшая оценка стандартной ошибки статистики}} = \frac{(\bar{x} - \mu)}{s/\sqrt{n-1}}.$$

Тестовая статистика для коэффициента регрессии b:

$$t = \frac{b - 0}{\text{оцененная стандартная ошибка } b}.$$

Оцененная стандартная ошибка b:

$$se_b = \frac{\hat{\sigma}_e}{\sqrt{\sum (x - \bar{x})^2}},$$

где $\hat{\sigma}_e$ — дисперсия распределения остатков вдоль линии регрессии генеральной совокупности. Предположим, что дисперсия одинакова для всех значений x. Лучшей оценкой генеральной дисперсии σ_e^2 является:

$$\hat{\sigma}_e^2 = \frac{\sum e^2}{(n-2)} = \frac{\sum (y - \hat{y})^2}{(n-2)}.$$

Алгебраически это можно выразить как:

$$\hat{\sigma}_e^2 = \frac{(\sum y^2 - a \sum y - b \sum xy)}{(n-2)}.$$

Чтобы проиллюстрировать наши действия, вернемся к примеру 8.1 о времени и расстоянии. Используем первое выражение для $\hat{\sigma}_e^2$:

$$\hat{\sigma}_e^2 = \frac{0{,}78^2 + 0{,}71^2 \ldots (-0{,}82)^2}{8} = \frac{10{,}01}{8} = 1{,}25,$$

Поэтому

$$\hat{\sigma}_e^2 = 1{,}12 \text{ мин.}$$

$$\text{и } \sqrt{\sum (x - \bar{x})^2} = \sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)} = \sqrt{15{,}889} = 3{,}99,$$

Отсюда: $$se_b = \frac{1,12}{3,99} = 0,281.$$

Значение критерия для β:

$$t = \frac{2,66}{0,281} = 9,47.$$

Если принять допустимые погрешности, то с учетом округления значение t т
же, как и значение t для коэффициента корреляции: 9,47 по сравнению с 9,4

Чтобы вычислить двусторонний критерий на 5%-ном уровне, сравним знач
данного критерия со значением из Приложения 2:

$$t_{0,025,8} = 2,306.$$

Так как 9,47 > 2,306, отвергнем гипотезу H_0 и принимаем гипотезу H_1
5%-ном уровне принятия решения с ошибкой доказательство будет непосле,
тельным при наличии только нулевых гипотез. Таким образом, мы приш
такому же выводу, что и ранее. Мы выбрали предположение, что сущес
линейная связь между временем и расстоянием, т.е. $\beta \neq 0$, x помогает объяс
варьируемость признака y.

8.5.3. Доверительный интервал в линейном регрессионном анали

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ НАКЛОНА ЛИНИИ РЕГРЕССИИ

Доверительный интервал (1–p) 100% для показателя наклона β опред
пределы значений b с ошибкой выборки β, в которых с вероятностью (1–p) · 100 н
дится β. Другими словами, для (1–p) · 100% выборок истинное значение β б
лежать в данном доверительном интервале. Доверительный интервал имеет то
вид, что и т е , которые рассматривались в гл. 5:

$$b + t_{(p/2)\,(n-2)} \cdot se_b.$$

Из вышеизложенного мы знаем, что:

$$se_b = \frac{\hat{\sigma}_e}{\sqrt{\sum (x - \bar{x})^2}}.$$

Рассчитаем 95%-ный доверительный интервал для наклона линии регрес
примере 8.1 о времени и расстоянии поставок:

$$b \pm t_{0,025,8}\, se_b = 2,66 \pm 2,31 \cdot 0,281 = 2,66 + 0,65.$$

С вероятностью 95% мы можем сказать, что значения β лежат между 2,
3,31 мин на милю; и 5% — вероятность того, что значение β лежит вне дан
интервала.

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ СРЕДНЕГО ЗНАЧЕНИЯ y ПРИ ДАННОМ ЗНАЧЕНИИ x

Теперь вернемся к основному предположению о том, что для данного значения x, обозначим его x_0, возможные значения y нормально распределены. Среднее значение этих нормальных распределений — значение y на линии регрессии генеральной совокупности. Обозначим среднее значение $\mu_{y|x}$. Доверительный интервал для $\mu_{y|x}$: с вероятностью $(1-p) \cdot 100$ имеет вид:

$$\hat{y} \pm t_{(p/2),\,(n-2)}\,\hat{\sigma}_e^2 \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2}}\,,$$

где $\hat{y}$ — оцененное значение y, вычисленное из выборочной регрессии $\hat{y} = a + bx$.

Заметим, что доверительный интервал зависит от значения x_0. Поэтому ширина интервала варьирует по мере изменения x. Интервал является наименьшим, когда $x_0 = \bar{x}$. В этом случае интервал примет более знакомый нам вид:

$$\hat{y} \pm t_{(p/2),\,(n-2)}\,\hat{\sigma}_e^2/\sqrt{n}\,.$$

Ширина интервала растет по мере того, как увеличивается отличие x_0 от $\bar{x}$ в любом направлении.

В примере 8.1 при вероятности 95% доверительный интервал $\mu_{y|x}$ равен:

$$\hat{y} \pm 2{,}31 \cdot 1{,}12 \sqrt{\frac{1}{10} + \frac{(x_0 - 2{,}89)^2}{15{,}89}}\,,$$

где: $\hat{y} = 5{,}91 + 2{,}66\,x_0$.

Значения для этого интервала будут вычислены в следующем разделе.

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ИНДИВИДУАЛЬНЫХ ЗНАЧЕНИЙ y ПРИ ДАННОМ ЗНАЧЕНИЯ x

Остановимся на следующем предположении для данной модели, что значения y распределены вдоль линии регрессии с вариацией σ_e^2, которая одинакова для всех значений x. Так как мы используем выборку, то существуют два элемента изменяемости признака y. Один исходит из оцененной позиции математического ожидания (среднего) $\mu_{y|x}$, а другой — из отклонений индивидуальных значений от своего среднего значения.

Эти два элемента различны: во-первых, благодаря колебаниям внутри выборки, которые могут быть сокращены, если увеличится размер выборки; во-вторых, благодаря природе переменных эти колебания неизбежны. Поэтому утверждать доверительный интервал для индивидуальных значений y не похож на другие доверительные интервалы, которые полностью подвержены эффекту выборочных колебаний. Некоторые исследователи считают их интервалами «прогноза», а не доверительными интервалами. Но как бы их ни называли, важно понять различие между (1−p) 100%, интервалом для $\mu_{y/x}$ и интервалом для индивидуальных значений y при данном x.

Выражения доверительных интервалов очень похожи между собой. Единственным значимым различием является то, что вариация для индивидуальных y при данном значении x увеличивается на величину $\hat{\sigma}_e^2 \cdot (1-p)$. 100%-ный доверительный интервал для индивидуальных значений y при $x = x_0$ имеет вид:

$$\hat{y} \pm t_{(p/2),\,(n-2)}\,\hat{\sigma}_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2}}\,,$$

где: $\hat{y} = a + bx_0$.

С вероятностью 95% доверительный интервал в примере 8.1 составит:

$$\hat{y} \pm 2{,}31 \cdot 1{,}12 \sqrt{1 + \frac{1}{10} + \frac{(x_0 - 2{,}89)^2}{15{,}89}}\,.$$

Табл. 8.4 и рис. 8.19 показывают изменения двух доверительных интервалов по мере того, как меняется x_0.

Таблица 8.4. **Вычисление доверительных интервалов для $\mu_{y|x}$ и для y при данном значении x_0, по данным примера 8.1**

Расстояние, x_0, миль	*Время, мин $\hat{y}=5,91+2,66x_0$*	*95%-ные доверительные интервалы для*	
		$\mu_{y/x}\pm$ мин	*y при данном x_0 $\pm$ мин*
1,0	8,57	1,47	2,98
2,0	11,23	1,00	2,77
2,89 ($\bar{x}$)	13,60	0,82	2,71
3,0	13,89	0,82	2,71
4,0	16,55	1,09	2,81
4,9	18,94	1,54	3,01

Даже всего лишь при 20-ти значениях приходится делать долгие и сложные вычисления. Используя ППП, большую часть работы можно сделать очень быстро. Важно понять, что делает та или иная программа и как интерпретировать ее результаты. К сожалению, различные программы имеют небольшое количество символов и знаков. Если вы имеете пакет программ по регрессионному анализу, то разберитесь сначала с простым примером, как в этой книге, а затем используйте ППП. Сравните производительность компьютера и вычислений вручную и тогда вы все поймете. Доскональное понимание двух различных линейных моделей также поможет, когда вы приступите к множественной регрессии, расчеты по которой всегда должны производиться на компьютере.

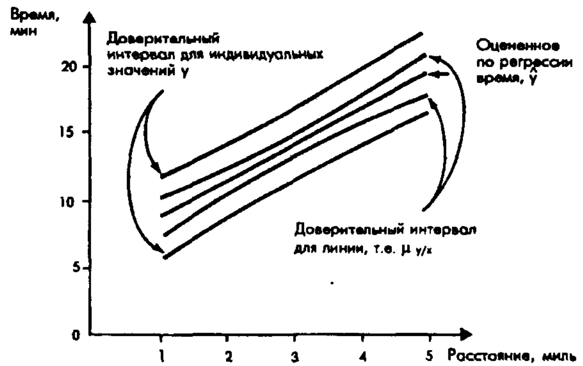


Рис. 8.19. Доверительные интервалы для $\mu_{y/x}$ для y при данном x_0 для выборки, n=10

8.6. МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

В предыдущих разделах было упомянуто о том,что вряд ли выбранная независимая переменная является единственным фактором, который повлияет на зависимую переменную. В большинстве случаев мы можем идентифицировать более одного фактора, способного влиять каким-то образом на зависимую переменную. Так, например, разумно предположить, что расходы цеха будут определяться количеством отработанных часов, использованного сырья, количеством произведенной продукции. По видимому, нужно использовать все факторы, которые мы перечислили для того, чтобы предсказать расходы цеха. Мы можем собрать данные об издержках, отработанном времени, использованном сырье и т.д. за неделю или за месяц Но мы не сможем исследовать природу связи между издержками и всеми другими переменными посредством корреляционной диаграммы. Начнем с предположений о линейной связи, и только если это предположение будет неприемлимо, попробуем использовать нелинейную модель. Линейная модель для множественной регрессии:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon.$$

Вариация y объясняется вариацией всех независимых переменных, которые в идеале должны быть независимы друг от друга. Например, если мы решим использовать пять независимых переменных, то модель будет следующей:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon.$$

Как и в случае простой линейной регрессии мы получаем по выборке оценки β_1, β_2, α, и т.д. Наилучшая линия для выборки:

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n.$$

Коэффициент a и коэффициенты регрессии b вычисляются с помощью рия минимальности суммы квадратов ошибок $\sum (y - \hat{y})^2$. Для дальнейшего ан регрессионной модели используют следующие предположения об ошибка любого данного x:

1. $E(\varepsilon)=0$.
2. Дисперсия e равна σ_e^2 и одинакова для всех x.
3. Ошибки независимы друг от друга.

Эти предположения те же, что и в случае простой регрессии. Однако в д случае они ведут к очень сложным вычислениям. К счастью, ППП выполня вычисления, позволяя нам сосредоточиться на интерпретации и оценке мно торной модели. В следующем разделе мы определим шаги, которые необх предпринять в случае множественной регрессии, но в любом случае мы полагаться на компьютер.

ШАГ 1. ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ

Первый шаг обычно предполагает обдумать, как зависимая переменная быть связана с каждой из независимых переменных. Нет смысла включать д нительные переменные x, если они не дают возможность объяснения вариа Вспомним, что наша задача состоит в том,чтобы объяснить вариацию y по изменения независимой переменной x. Нам необходимо рассчитать коэффиц корреляции r для всех пар переменных при условии независимости наблю друг от друга. Это даст нам возможность определить, связаны x с y линейн же нет, независимы ли x и x_i между собой. Это важно в множественной регр Мы можем вычислить каждый из коэффициентов корреляции, как пока разделе 8.5, чтобы посмотреть, насколько их значения отличны от нуля нужно выяснить, нет ли высокой корреляции между значениями незави переменных. Если мы обнаружим высокую корреляцию, например, между x то маловероятно, что обе эти переменные должны быть включены в оконч ную модель.

ШАГ 2. ОПРЕДЕЛЕНИЕ ВСЕХ СТАТИСТИЧЕСКИ ЗНАЧИМЫХ МОДЕЛ

Мы можем исследовать линейную связь между y и любой комбинацией неза мых переменных. Но модель имеет силу только в том случае, если суще значимая линейная связь между y и всеми x и если каждый коэффи регрессии b значимо отличен от нуля.

Мы можем оценить значимость модели в целом, используя F-критерий. I того, мы должны использовать t-критерий для каждого коэффициента регр b, чтобы определить, значимо ли он отличен от нуля. Если коэффициент р сии не значимо отличается от нуля, то соответствующая независимая перем не помогает в прогнозе значения y и модель не имеет силы.

Полная процедура заключается в том, чтобы установить множествену нейную регрессионную модель для всех комбинаций независимых переме Оценим каждую модель, используя F-критерий для модели в целом и t-кри для каждого коэффициента регрессии. Если F-критерий или любой из t-крит незначимы, то эта модель не имеет силы и не может быть использована.

модели исключаются из рассмотрения. Этот процесс занимает очень много времени. Например, если у нас имеются пять независимых переменных, то возможно построение 31 модели: одна модель со всеми пятью переменными, пять моделей, включающие четыре из пяти переменных, десять — с тремя переменными, десять — с двумя переменными и пять моделей с одной.

Можно получить множественную регрессию не исключая последовательно независимые переменные, а расширяя их круг. В в этом случае мы начинаем с построения простых регрессий для каждой из независимых переменных поочередно. Мы выбираем лучшую из этих регрессий, т.е. с наивысшим коэффициентом корреляции, затем добавляем к этому, наиболее приемлемому значению переменной y вторую переменную. Этот метод построения множественной регрессии называется прямым.

Обратный метод начинается с исследования модели, включающей все независимые переменные; в нижеприведенном примере их пять. Переменная, которая дает наименьший вклад в общую модель, исключается из рассмотрения, остается только четыре переменных. Для этих четырех переменных определяется линейная модель. Если же эта модель не верна, исключается еще одна переменная, дающая наименьший вклад, остается три переменных. И этот процесс повторяется со следующими переменными. Каждый раз, когда исключается новая переменная, нужно проверять, чтобы значимая переменная не была удалена. Все эти действия нужно производить с большим вниманием, так как можно неосторожно исключить нужную, значимую модель из рассмотрения.

Не важно, какой именно метод используется, может быть несколько значимых моделей и каждая из них может иметь огромное значение.

ШАГ 3. ВЫБОР ЛУЧШЕЙ МОДЕЛИ ИЗ ВСЕХ ЗНАЧИМЫХ МОДЕЛЕЙ

Эта процедура может быть рассмотрена с помощью примера, в котором определились три важнейших модели. Первоначально было пять независимых переменных x: x_1, x_2, x_3, x_4, x_5, но три из них — x_2, x_4 и x_5 — исключены из всех моделей. Эти переменные не помогают в прогнозировании y.

Поэтому значимыми моделями оказались:

Модель 1: y прогнозируется только x_1.
Модель 2: y прогнозируется только x_3.
Модель 3: y прогнозируется x_1 и x_3 вместе.

Для того, чтобы сделать выбор из этих моделей, проверим значения коэффициента корреляции и стандартного отклонения остатков $\hat{\sigma}_e$. Коэффициент множественной корреляции — есть отношение "объясненной" вариации y к общей вариации y и вычисляется так же, как и коэффициент парной корреляции для простой регрессии при двух переменных. Модель, которая описывает связь между y и несколькими значениями x, имеет множественный коэффициент корреляции R, который близок к $\pm$ 1 и значение $\hat{\sigma}_e$ очень мало. Коэффициент детерминации r^2, который часто предлагается в ППП, описывает процент изменяемости y, которая объясняется моделью. Модель имеет значение в том случае, когда r^2 близко к 100%.

В данном примере мы просто выбираем модель с наибольшим значением r^2 и наименьшим значением $\hat{\sigma}_e$. Предпочтительной моделью оказалась модель 1. На следующем шаге необходимо сравнить модели 1 и 3. Различие между этими моделями состоит во включении переменной x_3 в модель 3. Вопрос в том, повышает ли значительно x_3 точность предсказания значения y или же нет. Следующий критерий поможет ответить нам на этот вопрос — это *частный F-критерий*. Рассмотрим пример, иллюстрирующий всю процедуру построения множественной регрессии.

❐ **Пример 8.2.** Руководство большой шоколадной фабрики заинтересовано в построении модели для того, чтобы предсказать реализацию одной из своих уже долго существующих торговых марок. Были собраны следующие данные.

Таблица 8.5. **Построение модели для прогноза объема реализации**

Дата	*Реализация за 6 мес., млн. ф.ст.*	*Реклама, млн. ф.ст.*	*Цена, пенсы за ед.*	*Цена конкурента, пенсы за ед.*	*Индекс потребительских расходов*
19X0					
январь-июнь	126	4,0	15,0	17,0	100,0
июль-декабрь	137	4,8	14,8	17,3	98,4
19X1					
январь-июнь	148	3,8	15,2	16,8	101,2
июль-декабрь	191	8,7	15,5	16,2	103,5
19X2					
январь-июнь	274	8,2	15,5	16,0	104,1
июль-декабрь	370	9,7	16,0	18,0	107,0
19X3					
январь-июнь	432	14,7	18,1	20,2	107,4
июль-декабрь	445	18,7	13,0	15,8	108,5
19X4					
январь-июнь	367	19,8	15,8	18,2	108,3
июль-декабрь	367	10,6	16,9	16,8	109,2
19X5					
январь-июнь	321	8,6	16,3	17,0	110,1
июль-декабрь	307	6,5	16,1	18,3	110,7
19X6					
январь-июнь	331	12,6	15,4	16,4	110,3
июль-декабрь	345	6,5	15,7	16,2	111,8
19X7					
январь-июнь	364	5,8	16,0	17,7	112,3
июль-декабрь	384	5,7	15,1	16,2	112,9

Определим "лучшую" модель для прогноза объема реализации.

Решение.

Шаг 1. Просмотр данных. Реализация за шесть месяцев — зависимая переменная у. У нас пять независимых переменных х, четыре из них — расходы на рекламу, цена товара, конкурентная цена и индекс потребительских затрат. Пятая переменная — время, которое может быть обозначено для первого периода — Январь-июнь 19Х0– период 1, следующий период–2 и т.д., до 16 — последнего периода, июль-декабрь 19Х7. Вычислим коэффициенты корреляции, r, для всех шести переменных.

Воспользуемся процедурой проверки гипотез для определения значимости этих коэффициентов.

H_0: $\rho=0$ – коэффициент корреляции в генеральной совокупности равен нулю. Между парой переменных не существует никакой линейной связи.

В идеале это должно выполняться для всех пар независимых переменных.

H_1: $\rho\neq0$ коэффициент корреляции не равен нулю. Между парой переменных существует линейная связь.

Это должно выполняться для пар, образованных зависимой переменной с каждой независимымой переменной.

Проверим эти гипотезы на 5%-ном и 1%-ном уровнях значимости, используя двусторонний критерий. Из таблиц t-распределения значение t на 5%-ном уровне значимости составляет:

$$t_{0,025} = 2,145,$$

а на 1%-ном уровне:

$$t_{0,005} = 2,997.$$

Формула критерия:

$$t = \sqrt{\frac{r^2(n-2)}{1-r^2}}\,.$$

с (n–2) степенями свободы.

Коэффициенты корреляции и соответствующий уровень значимости приведены ниже:

Таблица 8.6. **Коэффициенты корреляции r (в скобках указан уровень значимости)**

	Зависимая переменная Реализация	*Время*	*Независимые переменные*		
			реклама	*цена*	*конкурентная цена*
Время	0,68(1%)	—			
Реклама	0,64(1%)	0,10	—		
Цена	0,23	0,17	–0,01	—	
Конкурентная цена	0,23	–0,05	0,21	0,70(1%)	—
Индекс	0,82(1%)	0,96(1%)	0,27	0,23	0,03

Зависимая переменная, т.е. объем реализации, имеет невероятно сильную линейную связь со временем, расходами на рекламу товара и индексом потребительских расходов. К сожалению, независимые переменные, время и индекс потребительских расходов, очень высоко коррелированы. Маловероятно, что обе переменные должны быть включены в окончательную модель. Это же верно и для двух ценовых переменных с коэффициентом корреляции 0,70. Будем иметь это в виду в ходе выполнения шага 2.

Шаг 2. Нахождение всех статистически значимых моделей. Будем использовать обратный метод для нахождения значимых моделей. Начнем с рассмотрения всех переменных в модели и затем придем к четырем переменным вместо пяти и так далее, пока не будут определены значимые модели. Модель для пяти переменных имеет вид:

$$\text{Реализация} = a + b_1x(\text{время}) + b_2x(\text{реклама}) + b_3x(\text{цена}) + \\ + b_4x(\text{конкурентная цена}) + b_5x(\text{индекс}).$$

Установим сначала общую значимость модели, используя F-критерий. Компьютер производит обычно табличный анализ дисперсии, в котором общая вариация реализации разделена на две части: часть, которая объясняется моделью, и часть, которая не объясняется моделью, т.е. на вариацию, объясненную регрессией и необъясненную, или остаточную вариацию. Компьютер рассчитывает два показателя:

$$\text{Средняя из квадратов отклонений, обусловленных регрессией} = \frac{\sum(\hat{y}-\bar{y})^2}{df_{\text{регрес.}}},$$

эта величина измеряет вариацию, объясненную регрессионной моделью.

$$\text{Средняя из квадратов остатков} = \frac{\sum(y-\hat{y})^2}{df_{\text{остат.}}},$$

которая измеряет вариацию, не объясненную регрессией.

Замечание: Общее число степеней свободы равно $n-1$, где n — число данных в совокупности, в данном примере $n=16$, $df_{\text{регрес.}}$ — число степеней свободы для регрессии, которая задана числом независимых переменных k. В данной модели $k=5=df_{\text{остат.}}$ — число степеней свободы для остатков может быть найдено как:

$$df_{\text{остат.}} = df_{\text{общ.}} - df_{\text{регрес.}} = (n-1) - \text{число независимых переменных}.$$

В данном примере $df_{\text{остат.}} = 16-1-5=10$.

Если модель описывает связь между y и всеми независимыми переменными x, то величина остаточной вариации будет очень малой. Для всей модели в целом:

H_0: нет линейной связи между какими-либо независимыми переменными и продажей, т.е. $\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=0$.

H_1: *существует линейная связь между одной или большим числом независимых переменных*, т.е. по крайней мере одна величина $\beta_i \neq 0$.

Для того чтобы модель была полезной и имела силу, мы должны отвергнуть H_0 и принять H_1. Значение F-критерия есть соотношение двух величин, описанных выше:

$$F = \frac{\text{Средний квадрат отклонениеи, обусловленных регрессией}}{\text{Средний квадрат отклонений, обусловленных остатками}}.$$

Этот критерий с одним «хвостом» (односторонний), потому, что средний квадрат, обусловленный регрессией, должен быть больше, чтобы мы могли принять H_1. В предыдущих разделах, когда мы использовали F-критерий, критерии были двусторонние, так как во главу угла ставилось большее значение вариации, каким бы оно ни было. В регрессионном анализе нет выбора — наверху (в числителе) всегда вариация y по регрессии. Если она меньше, чем вариация по остаточной величине, мы принимает H_0, так как модель не объясняет изменений y. Это значение F-критерия сравнивается с табличным:

$$F_{0,05,\ k,(n-1-k)}.$$

Из таблиц стандартного распределения F-критерия:

$$F_{0,05,\ 1,10} = 3,326.$$

В нашем примере значение критерия:

$$F = 28271\ /\ 1736 = 16,3.$$

16,3 > F, поэтому мы получили результат с высокой достоверностью.

Проверим каждое из значений коэффициентов регрессии. Предположим, что компьютер сосчитал все необходимые t-критерии. Для первого коэффициента гипотезы формулируются так:

H_0: время не помогает объяснить изменение продаж при условии, что остальные переменные присутствуют в модели, т.е. $\beta_1=0$.

H_1: время дает существенный вклад и должно быть включено в модель, т.е. $\beta_1\neq0$.

Проведем испытание гипотезы на 5%-ном уровне, пользуясь двусторонним t-критерием при:

$$(n-1-k) = 10 \text{ степенях свободы}.$$

Граничные значения на данном уровне:

$$t_{0,025,10} = \pm2,228.$$

Значение критерия:

$$t = \frac{b_1 - 0}{s_{b_1}}.$$

Рассчитанные значения t-критерия должны лежать вне указанных границ для того, чтобы мы смогли отвергнуть гипотезу H_0.

Так как все независимые переменные подчиняются этому правилу, то результаты проверки значимости для пяти переменных в данной модели обобщены в следующей таблице:

Таблица 8.7. **Проверка значимости коэффициентов регрессии для пяти переменных**

Независимая переменная	*Коэффициент регрессии b*	*Значение t-критерия*	*Достоверность на 5%-ном уровне*
Время	−13,4	−1,29	не достоверно
Реклама	6,6	2,21	почти достоверно
Цена	−6,4	−0,41	не достоверно
Конкурентная цена	12,1	0,84	не достоверно
Индекс	30,5	2,65	достоверно

Сейчас мы видим, что наша модель не достоверна, потому что четыре коэффициента регрессии не значимо отличны от нуля. Нам нужно решить, какую переменную следует исключить из модели.

В табл. 8.8 представлены шаги, предпринятые по мере того, как мы сокращаем число переменных в модели от 5 до 4, затем до 3, до 2 и, наконец, до 1 независимой переменной. Прочерки показывают, что переменная не включена в модель. Пользуясь результатами наших исследований, можно решить, какая переменная должна быть исключена из рассмотрения. Для каждой модели мы испытываем всю регрессию и отдельные коэффициенты регрессии. Если модель подходит по всем критериям, то $\hat{\sigma}_\varepsilon$ должно быть малым, а r — как можно ближе к 1.

Таблица 8.8. **Исследование различных моделей регрессии**

Количество переменных в модели	*Значимость всей модели на 5%-ном уровне по F-критерию*	*Значимость коэффициентов регрессии по t-критерию*					*Имеет ли значимость модель*	$\hat{\sigma}_e$	*r*
		Время	*Реклама*	*Цена*	*Конкурентная цена*	*Индекс*			
5	да	нет	?	нет	нет	да	нет	42	0,94
4	да	–	да	нет	нет	да	нет	43	0,93
3	да	–	да	–	нет	да	нет	41	0,93
2	да	–	да	–	–	да	да	41	0,93
2	да	да	да	–	–	–	да	50	0,89
1	да	–	–	–	–	да	да	62	0,82

Шаг 3. Какую из значимых моделей нужно использовать?

В нашем примере значение модели появились лишь тогда, когда количество переменных сократилось до двух. Сравнение моделей необходимо проводить через сопоставление стандартных отклонений остатков. Отклонение должно быть предельно малым числом. Первая модель с расходами на рекламу и индексом потребительских расходов является наилучшей, так как $\hat{\sigma}_\varepsilon = 41$ по сравнению с $\hat{\sigma}_\varepsilon = 50$ для модели с расходами на рекламу и временем как независимыми переменными. *Последний шаг* — сравнение лучшей модели с двумя переменными с лучшей моделью с одной переменной. По величине корреляции выбираем лучшую модель с одной переменной при r=0,82. Если бы добавление еще одной независимой переменной значительно улучшило модель, то мы смогли бы применить частный F-критерий для проверки. Этот критерий показывает, что введение величины расходов на рекламу значительно улучшает модель и нам нужно использовать две переменные: индекс потребительских расходов и расходы на рекламу. Окончательная модель:

Объем реализации = −1476 + 9,54х (реклама) + 15,8х (индекс) (млн. ф. ст./6 мес.).

Коэффициенты регрессии для расходов на рекламу и индекса потребительских расходов положительны, как мы и предполагали. Постоянная − 1476 (ф. ст.) выглядит абсурдной, но вспомним, что модель имеет силу только для значений, входящих в выборочную совокупность. Расходы на рекламу изменяются от 3,8 млн. ф. ст. до 19,8 млн. ф. ст., а индекс — от 98,4 до 112,9.

Этот пример хорошо показывает все сложности объяснения и расчета каждой отдельной величины многофакторной модели. Цель статистической модели — объяснить вариацию продаж, а не предоставить особую информацию по изолированному влиянию рекламы или индекса потребительских цен на реализацию. По данным выборки модель дает некоторое представление о таких эффектах. В выборочной совокупности всегда возникает противоречие между теми или иными переменными. Поэтому коэффициенты регрессии при отдельных переменных должны использоваться с особым вниманием.

Наконец, при анализе мы должны проверить структуру и размер ошибок, а потому мы должны заранее предполагать большие ошибки. Ошибки рассчитываются следующим образом:

Ошибка = фактический объем реализации — ожидаемая реализация.

Таблица 8.9. **Размер ошибок (млн. ф. ст.)**

Фактическая реализация	*Ожидаемая реализация*	*Ошибка*		*Фактическая реализация*	*Ожидаемая реализация*	*Ошибка*
126	142	-16		367	424	-57
137	125	12		367	350	17
148	159	-11		321	346	-25
191	242	-51		307	335	-28
274	247	27		331	387	-56
370	307	63		345	350	- 5
432	361	71		364	354	10
445	417	28		384	362	22

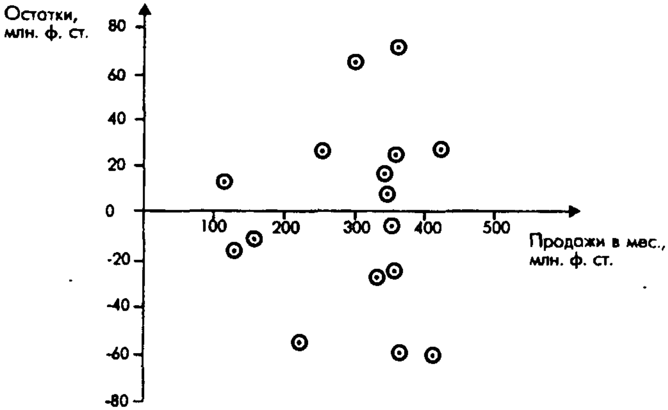


Рис. 8.20. Распределение остатков для модели с двумя переменными

Оказалось восемь ошибок с отклонениями 10% или более от фактического объема продаж. Наибольшая из них — 27%. Будет ли размер ошибки принят компанией при планировании деятельности? Ответ на этот вопрос будет зависеть от степени надежности других методов.

8.7. НЕЛИНЕЙНЫЕ СВЯЗИ

Вернемся к ситуации, когда у нас всего две переменные, но связь между ними нелинейная. На практике многие связи между переменными являются криволинейными. Например, связь может быть выражена уравнением:

$$y = ax^2 + bx + c;$$

или:

$$y = ax^3 + bx^2 + cx + d;$$

или:

$$y = a + b/x;$$

или:

$$y = ae^{bx}.$$

Если связь между переменными сильная, т.е. отклонение от криволинейной модели относительно небольшое, то мы сможем догадаться о природе наилучшей модели по диаграмме (полю корреляции). Однако трудно применить нелинейную модель к выборочной совокупности. Было бы легче, если бы мы могли манипулировать нелинейной моделью в линейной форме. В первых двух записанных моделях функциям x^2 и x^3 могут быть присвоены разные имена, и тогда будет использоваться множественная модель регрессии. Например, если модель:

$$y = ax^3 + bx^2 + cx + d$$

лучше всего описывает связь между y и x, то перепишем нашу модель, используя независимые переменные x, x^2, x^3:

$$y = aZ + bX + cx + d,$$

где $Z = x^3$ и $X = x^2$.

Эти переменные рассматриваются как обыкновенные независимые переменные, даже если мы знаем, что Z, X и x не могут быть независимы друг от друга. Лучшая модель выбирается так же, как и в предыдущем разделе.

Третья и четвертая модели рассматриваются по-другому. Тут мы уже встречаемся с необходимостью так называемой линейной трансформации. Например, если связь

$$y = 1 + 4 / x,$$

то на графике это будет изображено кривой линией. Все необходимые действия могут быть представлены следующим образом:

Таблица 8.10. **Расчет $y = 1 + 4/x$ и $x = 1/x$**

x	$y = 1 + 4/x$	$x = 1/x$
0	бесконечность	бесконечность
1	5	1
2	3	0,5
3	2,33	0,33
4	2	0,25

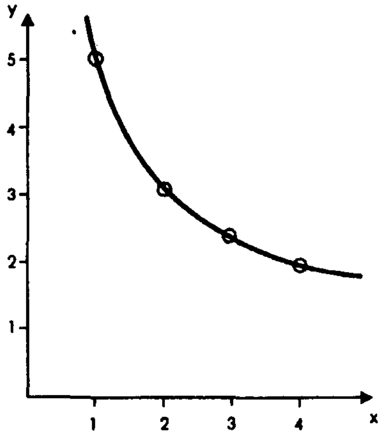


Рис. 8.21. Нелинейная связь

Линейная модель, при трансформированной связи:

$$y = 1 + 4/x,$$

где $X = 1/x$.

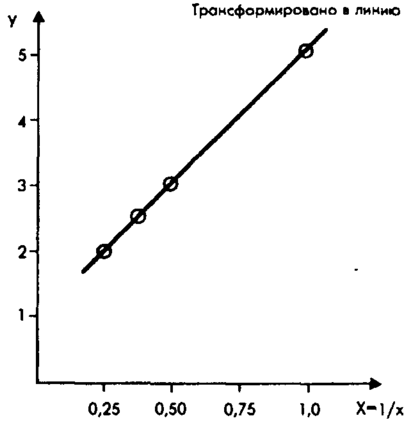


Рис. 8.22. Линейная трансформация связи

В общем, если исходная диаграмма показывает, что связь может быть изображена в форме: $y = \alpha + \beta/x$, то представление y против X, где $X=1/x$ определит прямую линию. Воспользуемся простой линейной регрессией для установления модели: $\hat{y} = a + bX$. Рассчитанные значения a и b — лучшие значения α и β.

Четвертая модель, приведенная выше, включает трансформацию y с использованием натурального логарифма:

$$y = ae^{bx}.$$

Взяв логарифмы по e обеих сторон уравнения, получим:

$$\ln(y) = \ln(a) + \ln(e^{bx}),$$

поэтому: $\ln(y) = \ln(a) + bx$, где $\ln = \log_e$ и $\ln(e) = 1$.

Если $Y=\ln(y)$ и $A=\ln(a)$, то $Y=A+b$ – уравнение линейной связи между Y и x. Пусть $y = ae^{bx}$ — связь между y и x, тогда мы должны трансформировать каждое значение y взятием логарифма по e. Определяем простую линейную регрессию Y ($=\ln(y)$) по x для того, чтобы найти значения A и b. Антилогарифм записан ниже.

Таким образом, метод линейной регрессии может быть применен к нелинейным связям. Однако в этом случае требуется алгебраическое преобразование при записи исходной модели.

❒ **Пример 8.3.** Следующая таблица содержит данные об общем годовом объеме производства промышленной продукции в определенной стране за период 19X0–X7:

Таблица 8.11. **Годовой объем продукции 19X0–X7**

Год	*Годовая продукция, тыс. т*
19X0	740
19X1	804
19X2	879
19X3	961
19X4	1042
19X5	1137
19X6	1242
19X7	1357

Требуется:

1. Нарисовать диаграмму, прокомментировать ее.
2. Нарисовать новую диаграмму за следующий год.
3. Предположим, что связь между общей годовой продукцией и временем может быть описана как:

$$y = ab^x,$$

где y — общий объем годовой продукции и x — число лет с 19X0.

Используйте выборочную совокупность для оценки a и b. Объяснить получение b.

4. Используйте модель, описанную в п.3, для прогнозирования общего объема продукции. Прокомментируйте ваш прогноз.

Решение

1. Диаграмма по указанным данным.

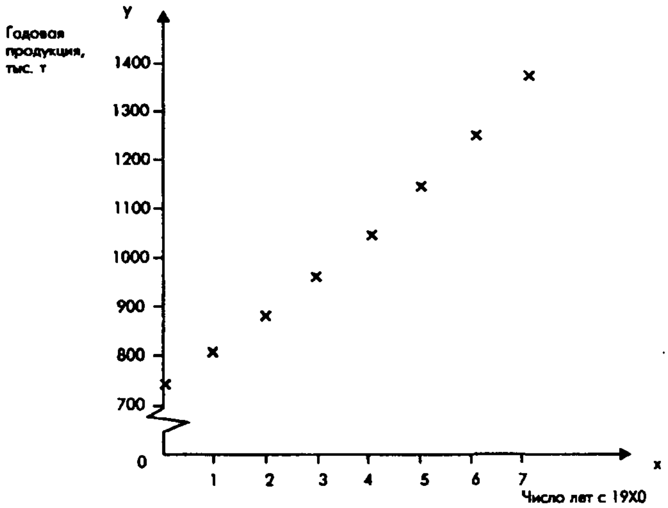


Рис. 8.23. Диаграмма общего объема производства

Рисунок показывает, что между y и x существует некая связь, но она может быть криволинейной.

2. Трансформируем для начала значения объема производства в логарифмической форме.

Таблица 8.12. **Логарифмы значений объема производства**

Год	*Значение логарифма* Y = log (y)
19X0	2,869
19X1	2,905
19X2	2,944
19X3	2,983
19X4	3,018
19X5	3,056
19X6	3,094
19X7	3,133

Линейная связь между log (y) и x четко видна из диаграммы. Нелинейная же связь между y и x должна быть трансформирована в:

$$Y = \log(y) = A + Bx,$$

где: $Y = \log(y)$.

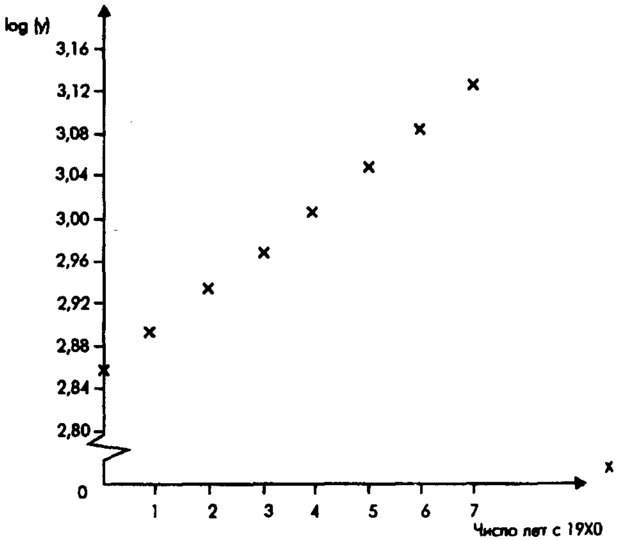


Рис. 8.24. Изображение log (y), соответствующих значениям x

3. Предполагаемая модель:

$$y = ab^x.$$

Пролагорифмируем обе части:

$$\log(y) = \log(ab^x) - \log(a) + \log(b^x).$$

Отсюда:

$$\log(y) = \log(a) + x\log(b).$$

Пусть $A = \log(a)$ и $B = \log(b)$. Тогда ожидаемая линейная модель:

$$Y = \log(y) = A + Bx.$$

Для оценки A и B используем технику простой линейной регрессии:

Таблица 8.13. **Расчеты для модели регрессии**

Год	Log *(годовая продукция)* $Y = \log(y)$	*Число лет с 19X0*	x^2	$x\log(y)$
19X0	2,869	0	0	0
19X1	2,905	1	1	2,905
19X2	2,944	2	4	5,888
19X3	2,983	3	9	8,949
19X4	3,018	4	16	12,072
19X5	3,056	5	25	15,280
19X6	3,094	6	36	18,564
19X7	3,133	7	49	21,931
Всего	24,002	28	140	85,589

Для выявления наилучшей модели применяется метод наименьших квадратов в выборочной совокупности:

$$B = \frac{n\sum(x\log(y)) - \sum x \sum \log(y)}{n\sum x^2 - (\sum x)^2};$$

$$B = \frac{8 \cdot 85{,}589 - 28 \cdot 24{,}002}{8 \cdot 140 - (28)^2} = 0{,}03767.$$

Коэффициент A находится как:

$$nA = \sum \log(y) - B\sum x,$$

т.е. $8A = 24{,}002 - 0{,}03767 \cdot 28,$

$A = 2{,}8684.$

Поэтому линейная модель:

$$\log (y) = 2{,}8684 + 0{,}03767x,$$

так как A = log (a), то a = 10^A = 738,6. Так как B = log (b), то b = 10^B =

Связь между общим объемом годовой продукции и числом лет с 19X0 быть описана:

$$y = 739 \cdot (1{,}091)^x.$$

Интерпретация b: если мы перепишем эту модель таким образом: $y = 739\ (1+9{,}1/100)^x$, то природа связи станет более очевидна. Годовая про 739 (тыс. т) в 19X0 году при x=0. Затем мы видим, что продукция растет н в год. b — отношение объема производства в текущем году к объему произв в предыдущем году. Прогноз будущего объема продукции:

$$19X8:\ x = 8,\ y = 739\ (1{,}091)^8 = 1483 \text{ тыс. т};$$

$$19X9:\ x = 9,\ y = 748\ (1{,}091)^9 = 1618 \text{ тыс. т}.$$

Нужно быть *предельно внимательными*, когда мы расширяем рамки м построенной по выборочной совокупности. Предположим, что условия с 1! 19X7 годы остаются неизменными. Это предложение может быть оправд для прогноза на 19X8 год, но по мере того как мы будем двигаться далее, п станет все менее надежен.

8.8. РАНГОВЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ СПИРМЕНА r

В предыдущих разделах предполагалось, что данные включают в себя ка измерения. Если же возможны какие-либо процедуры, связанные с постро последовательности, то действия без параметров неуместны. Типичным при являются маркетинговые исследования, например, исследование вкуса пищи нам понадобится испробовать четыре вида супа, то мы можем располож в последовательности в порядке предпочтения от 1 до 4, но мы не сможем у точный принцип предпочтения. Такой вид данных называется **порядк** Данные, которыми мы пользовались до настоящего времени, называются **и вальными**.

Положим, двух людей попросили попробовать четыре вида супа. Вер что будет иметь место два вида вариаций принципов предпочтения. Две по вательности составляют две порядковые переменные. Мы можем сравни последовательности по степени согласованности, используя **ранговый ко: циент корреляции Спирмена**.

Коэффициент Спирмена r_s не является параметрическим критерием в чии от коэффициента Пирсона:

$$r_s = 1 - \frac{6 \sum d^2}{n\ (n^2 - 1)},$$

где d — разница между значениями рангов для каждой единицы. Так же как и коэффициент Пирсона, значение r_s изменяется от -1 до $+1$. Значимость коэффициента Спирмена может быть проверена таким же образом, что и значение коэффициента Пирсона.

□ **Пример 8.4.** Два человека дегустируют 10 супов. Нам нужно узнать, какой уровень согласованности достигнут между этими двумя дегустаторами.

Таблица 8.14. **Определение уровня согласованности между двумя дегустаторами**

Суп / *Дегустатор*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*Всего*
Первый	1	3	4	9	8	10	2	7	6	5	
Второй	3	2	5	7	9	10	1	6	4	8	
Разница (d)	-2	1	-1	2	-1	0	1	1	2	-3	0
d^2	4	1	1	4	1	0	1	1	4	9	26

Отсюда:

$$r_s = 1 - \frac{6 \cdot 26}{10 \cdot 99} = 1 - 0{,}158 = 0{,}842 .$$

Для того чтобы оценить достоверность этого результата, укажем гипотезы, которыми мы руководствуемся.

H_0: между двумя дегустаторами либо нет связи, либо они просто не согласованы друг с другом, т. е. $\rho \le 0$.

H_1: существует согласованность (положительная связь), т.е. $\rho_s > 0$.

Это односторонний критерий. Двусторонний критерий мы бы использовали, если бы нас интересовала либо положительная, либо отрицательная связь.

Значение критерия для $n \ge 10$ (приближенно) имеет вид:

$$z = \frac{(r_s - 0)}{1/\sqrt{(n-1)}} = \frac{0{,}842}{1/3} = 2{,}53.$$

Из стандартных таблиц:

$$z_{0,05} = 1{,}645.$$

Полученное значение критерия больше, чем 1,645. Результат достоверен на 5%-ном уровне. Отвергаем H_0 и становится очевидной положительная связь между двумя дегустаторами. Нормальному приближение распределения верно только для $n \ge 10$. Для $n < 10$ существуют специальные таблицы.

РЕЗЮМЕ

Построение линейной регрессии — это не что иное, как построение модели. Нам нужно объяснить изменяемость зависимой переменной в результате изменения независимой переменной.

Построение модели всегда начинается с выбора факторов, которые, вероятно, могут повлиять на зависимую переменную. Необходимо выбрать основные из них. В простой линейной регрессии (парной) имеется только одна независимая переменная. Модель выражается как:

$$y = \alpha + \beta x + \varepsilon,$$

где ε – остаток, часть y, которая не объясняется моделью, рассчитываемой обычно по выборочной совокупности при помощи метода наименьших квадратов. Для прямой линии показатель наклона рассчитывается по формуле:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2},$$

а пересечение с осью OY:

$$a = \frac{\sum y}{n} - \frac{b\sum x}{n}.$$

Теснота линейной связи измеряется коэффициентом корреляции Пирсона r:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}.$$

Чем ближе r к –1 или к +1, тем точнее линейная модель описывает связь между x и y. Коэффициент детерминации, r^2, измеряет долю вариации y, объяснимой x.

Кроме того, предполагается, что остатки нормально распределены с нулевым средним значением и стандартным отклонением σ_e, которое постоянно для всех значений x. Доверительные интервалы могут быть установлены для параметров ρ, α и β с использованием выборочных оценок r, a и b. Для оценки достоверности используют гипотезы о значениях параметров. Для всех критериев нулевые гипотезы являются предположением о том, что в генеральной совокупности между переменными в генеральной совокупности не существует линейной связи. Значение критерия для ρ будет следующим:

$$t = \sqrt{\frac{r^2(n-2)}{(1-r^2)}},$$

при (n–2) степенях свободы;

для β:

$$t = \frac{b - 0}{\hat{\sigma}_e / \sqrt{\sum (x - \bar{x})^2}},$$

при (n−2) степенях свободы.

Для данного x, скажем x_0, рассчитывается доверительный интервал для среднего значения y и для индивидуальных значений y, близких к среднему значению. Ширина таких интервалов изменяется, и они могут быть изображены с помощью кривых линий с каждой стороны линии регрессии.

Множественная регрессия имеет более чем одну независимую переменную в модели. Обычно в этом случае все расчеты производятся с помощью компьютера. Для проверки достоверности связи также используют гипотезы. Если мы исследуем большое количество независимых переменных, применяется быстрый метод исключения недостоверных и не имеющих силу моделей и построения лучших моделей.

Нелинейные модели алгебраически могут быть превращены в линейные.

Коэффициент ранговой корреляции Спирмена рассчитывается аналогично, как и коэффициент Пирсона, когда нам важно прежде всего упорядочивание данных, нежели прямое измерение.

УПРАЖНЕНИЯ

Упражнение 8.1

1. Торговцу нужно выяснить, как изменяется количество пучков салата, продаваемого ежедневно в розницу.

   Требуется определить факторы, которые могут повлиять на количество.

2. Имеются следующие сведения о количестве и цене:

Количество, тыс./день	28	29	34	35	37	37	41	46
Цена, ф. ст. за единицу	30	31	25	26	22	24	16	12

Требуется:

а) изобразить данные на графике, вычислить коэффициент корреляции r, найти достоверность на уровне 1%;

б) построить модель линейной регрессии и объяснить значения коэффициентов;

в) если бы цена салата была равна 45 ф. ст. за каждый пучок, то сколько было бы продано? Прокомментируйте ваши предположения.

Упражнение 8.2

Провели исследование, сколько сберегает население и сколько оно зарабатывает за год. Были получены следующие данные для случайно отобранных 9 чел.:

Доход, тыс. ф. ст.	15	6	9	3	20	11	14	10	12
Сбережения, ф. ст.	2000	200	500	500	2500	1800	1500	1500	1600

Требуется:
1. Изобразить данные на графике. Измерить тесноту линейной связи.
2. Построить модель регрессии и вычислить коэффициенты.
3. Какие еще факторы вы предлагаете рассмотреть?

Упражнение 8.3

Туристическую фирму крупного курортного города интересует связь между числом отпускников, останавливающихся в отелях и расходами на рекламу отелей. Взято случайное число отелей — 6, сходных по размеру. Была собрана следующая информация за текущий сезон:

Отель	1	2	3	4	5	6
Реклама, ф. ст.	9000	6000	10000	8000	7000	4000
Число гостей	1100	1200	1600	1300	1100	800

Требуется:
1. Построить модель для объяснения изменения числа гостей и проверить достоверность этой модели.
2. Прокомментировать точность прогнозов, производимых по этой модели.

Упражнение 8.4

Для установления стандартных издержек бухгалтер фирмы собрал следующие данные:

Выпуск, тыс. ед. в день	12	19	12	17	15	15	17	16	18	19
Издержки, тыс. ф. ст. за день	3,2	4,1	2,9	3,8	3,6	3,5	3,9	3,7	4,0	4,2

Требуется:
1. Установить предварительные стандарты.
2. Прокомментировать, насколько точна оценка стандартных издержек. Какие факторы могут быть дополнительно включены в стандарты для увеличения точности?

Упражнение 8.5

Компания "Garden Groceries" владеет 12 магазинами. Размер размещенных магазинов велик. Финансовый директор группы магазинов рассматривает возможность слияния числа мелких магазинов для увеличения прибыльности компании. Он предположил, что оборот магазинов вследствие слияния останется прежним. Ему необходимо установить связь между прибылью и оборотом. Данные для каждого магазина в отдельности за последний финансовый год приведены ниже:

Магазин	Годовая прибыль, тыс. ф. ст.	Оборот, тыс. ф. ст.
1	2	50
2	4	60
3	11	85
4	17	85
5	18	100
6	28	120
7	34	140
8	36	155
9	48	180
10	55	210
11	71	250
12	85	300

Требуется:

1. Построить модель для описания связи между прибылью и оборотом. Интерпретировать каждую постоянную данной модели. Прокомментировать применимость модели.
2. Дать советы финансовому директору по поводу слияния магазинов.

Упражнение 8.6

15 студентов сдали экзамены по бухгалтерскому учету и математике. Их ранжирование по итогам экзаменов дано ниже. Существует ли какая-нибудь связь между этими двумя результатами?

Студент	Бухгалтерский учет	Математика
1	10	13
2	5	4
3	12	10
4	1	1
5	6	11
6	2	2
7	7	8
8	11	9
9	15	14
10	3	5
11	9	7
12	14	12
13	13	15
14	4	3
15	8	6

Упражнение 8.7

Компания имеет большое количество продавцов, которые предъявили отчеты о своих поездках. Десять из них предъявили требования на оплату этих поездок. За два квартала компания выплатила 17,5 пенсов за каждую милю проезда.

Продавец	*Пройдено миль в I квартале*	*Пройдено миль во II квартале*
А	28	45
Б	26	54
В	35	49
Г	20	36
Д	30	50
Е	16	32
Ж	40	64
З	36	56
И	24	44
К	25	50

Требуется:

1. Не используя данные за I квартал, найти: вычисленное значение, доверительный интервал при 95% вероятности для среднего уровня расходов на поездку каждого продавца, который должен быть выплачен во II квартале.
2. Определить линию регрессии методом наименьших квадратов, с помощью которой можно предсказать пройденные мили во II квартале, за базис принять количество миль, пройденных в I квартале.

Среднее расстояние, пройденное продавцом за I квартал, равно 3000 миль. С помощью полученного уравнения регрессии рассчитайте значение средних расходов каждого продавца за II квартал.

Какое из двух полученных значений вы бы рекомендовали использовать? Почему?

Упражнение 8.8

Главный бухгалтер "Pen Prodakts" проанализировал время, затраченное на производство основных продуктов компании. Он получил следующие данные для одного конкретного продукта, который производился на серийной основе:

Серия продуктов	*Размер серии продуктов*	*Время, ч*
1	32	21,4
2	24	17,0
3	30	20,4
4	45	29,6
5	15	12,6
6	26	19,1
7	50	34,2
8	18	15,2
9	20	16,3
10	40	29,2

Данные расположены в хронологическом порядке по времени производства серии каждого типа предметов за прошедший год. Получив эти данные, бухгалтер построил уравнение регрессии для связи между временем y и размером серии x:

$$y = 3{,}5 + 0{,}6x$$

с коэффициентом корреляции 0,99. Учитывая высокую корреляцию, бухгалтер заключил, что размер серии является "предсказателем" величины затраченного времени.

Требуется:

1. Объяснить полученные значения коэффициентов регрессии в контексте данного примера.
2. Вычислить отклонения рассматриваемого производственного времени от линии регрессии и показать, что эти отклонения в сумме равны нулю. Объяснить, почему они всегда будут в сумме стремиться к нулю.
3. Вычислить сумму квадратов отклонений и объяснить важность полученных значений.
4. Изобразить на графике отклонения в связи с размером серии продуктов и номером серий. Какой вывод вы можете сделать из полученного графика? Согласны ли вы с мнением, что размер партии товара является величиной, с помощью которой можно точно предсказать время?

Упражнение 8.9

"Pandora" — включает региональною распределенную систему с большим числом товарных вагонов, прикрепленных к определенным семи складам. Когда на каждую поставку был определен свой склад и конкретный день поставки, каждому поставщику соответствовала определенная группа потребителей. Поставщик поставляет со склада картон каждому потребителю, следующий определенным маршрутом. Определите эффективность системы распределения, используя следующие данные:

Группа поставщиков	*Затраченное время, ч*	*Число потребителей*	*Общий вес груза (картон), т*
1	6,3	6	24
2	4,7	3	15
3	5,0	4	17
4	8,5	7	30
5	3,8	4	14
6	4,1	6	20

Требуется:

1. Изобразить данные в удобной форме и прокомментировать появившиеся связи.
2. Определить коэффициент корреляции Пирсона между величинами времени и каждой из оставшихся переменных. Прокомментируйте эти две переменные как величины, помогающие предсказать общее время поставки.
3. Если регрессия для общего время поставки включает число потребителей и величину общего веса груза, то получен коэффициент множественной корреляции, равный 0,99. Объясните значение этой величины и сравните ее с двумя парными коэффициентами корреляции, которые вы определили в п. 2.

Упражнение 8.10

Компания "Themis Processing Ltd", полностью подчиненная компании "Tantalus Products", была основана в 19X6 г. для добычи железной руды. С той поры "Themis" построил 10 новых фирм, строящих заводы различных размеров, которые успешно функционируют и получают немалую прибыль. В настоящее время эта компания рассматривает перспективу построения двух дополнительных заводов. Так как средние производственные издержки — это критический фактор в общей прибыльности предприятия, важно оценить будущие издержки на ранних стадиях производства.

Главный бухгалтер полагает, что некоторые из ключевых факторов влияют на производственные издержки и это такие факторы как: экономия масштаба, они присутствуют на заводах различных размеров. Чтобы доказать свою точку зрения, бухгалтер получил следующие данные об издержках для десяти заводов:

Завод	*Первые операции*	*Выработка, т в мес.*	*Средние издержки, ф. ст. за 1 т руды*
1	19X6	900	51,95
2	19X6	500	57,18
3	19X8	1750	46,90
4	19X9	2000	45,37
5	19X0	1400	46,03
6	19X0	1500	48,15
7	19X1	3000	44,22
8	19X2	1100	48,72
9	19X2	2600	45,40
10	19X3	1900	44,69

Средние издержки за 1 т для каждого завода были получены суммированием всех прямых и косвенных расходов, а затем делением общих издержек на количество тонн, фактически произведенных в 19X5 г. рассматриваемым заводом.

Требуется:

1. Если выработка x, а средние издержки за 1 т y, то как на графике будет выглядеть y по x и y против $1/x$. Прокомментируйте форму полученных графиков.
2. Результаты линейной регрессии y по $1/x$ следующие:

Коэффициент	*Значение*	*Стандартная ошибка*
Пересечение (a)	41,75	0,602
Градиент (b)	7925,0	667,2

Остаточная стандартная ошибка (s) = 0,984.

Доля объясненной вариации (r^2) = 0,946.

Изобразите на графике регрессию для двух случаев для п.1.

Объясните значение этих данных.

3. На диаграмме для у по 1/х укажите 95%-ые пределы на основе остаточной стандартной ошибки. Объясните, почему эти пределы нахождения величины ошибки не обеспечивают прогноз доверительных интервалов при прогнозировании средних издержек на 1 т для данного значения 1/х.

Упражнение 8.11

Применение внутренней электрической системы было вынесено на внешний рынок в 1982 г. Покупателям была предоставлена возможность купить страховой полис для покрытия затрат на ремонт на 5 лет. Таблица внизу показывает общий объем продаж лицензии за 1982–1988 гг. вместе с продажей страховых полисов и с учетом общего индекса цен на товары электрической промышленности.

Год	*Продажа лицензий, тыс. ф. ст.*	*Продажа полисов, ед*	*Индекс цен*
1982	3600	400	120
1983	6250	300	126
1984	9170	600	131
1985	14000	1200	140
1986	21600	1700	144
1987	27000	2200	150
1988	41600	2000	160

1. Объясните, почему важно установить первоначальные объемы продаж лицензии при инфляции, когда проявляется связь между объемом продаж лицензий на применение и числом проданных страховых полисов. Сопоставьте с ценами 1980 г.
2. Вычислите коэффициент детерминации между дефлированным уровнем продаж лицензий и числом страховых полисов, объяснить полученный результат.
3. Вычислите уравнение линейной регрессии с помощью метода наименьших квадратов для прогноза количества проданных полисов при данном дефлированном уровне продаж лицензий.
4. Общие продажи лицензии в 1989 г. оценивались в 51 млн. ф. ст. в ценах 1989 г.; в 1989 г. индекс цен на электротехнические товары к ценам 1980 г. должен равняться 170. Предскажите продажи страховых полисов в 1989 г. с помощью регрессионного анализа.
5. Прокомментируйте недостатки использования данного типа исследования для прогноза продаж полисов в 1989 г.

Упражнение 8.12

Компания "Venus Tableware Ltd" является основным производителем керамической посуды в Блюленде. Сумма импорта и экспорта керамических изделий незначима по сравнению с местным объемом производства. Таким образом можно сделать вывод о том, что общий объем производства равен общему объему продаж. Рассмотрим следующую таблицу:

Год	*Общий объем производства страны, тыс. т*	*Производство керамической посуды, тыс. т*	*Доля фирмы, %*
19X0	744	113	15,2
19X1	733	108	14,0
19X2	828	131	15,8
19X3	900	144	16,0
19X4	936	146	15,6
19X5	977	157	16,1
19X6	1007	163	16,2
19X7	1066	175	16,4

Требуется:

1. Изобразить на графике логарифм общего объема производства по каждому году. Объяснить полученный график.
2. Выразить общий объем производства в форме:

$$Y = ab^x,$$

где Y — общий объем производства, тыс. т, X– 19X0 г.

Вычислить a и b и объяснить полученное значение b.

3. Используйте модель, описанную в п. 2 для прогноза общего объема производства за следующие три года: 19X8, 19X9, 19X0.
4. Главный бухгалтер "Venus Tableware" хочет использовать общий объем производства для прогноза объема реализации компании. Так как фирма увеличила долю на рынке с 15,2% до 16,4%, главный бухгалтер считает уместным произвести два типа прогнозов:

а) пессимистичный прогноз — доля компании останется 16,4%;

б) оптимистичный прогноз — доля компании будет увеличиваться на 0,2% в каждом следующем году, достигая к 19X0 году 17,0% .

Получить набор двух типов прогнозов для ежегодного объема продаж в течение следующих трех лет.

5. Обрисовать в общих чертах основные выводы из прогнозов, основанных на регрессионном анализе.

Глава 9. ВРЕМЕННЫЕ РЯДЫ И ПРОГНОЗИРОВАНИЕ

9.1. ВВЕДЕНИЕ

Каким бы видом бизнеса вы ни занимались, вам приходится планировать предпринимательскую деятельность на будущий период. При составлении как краткосрочных, так и долгосрочных планов менеджеры вынуждены прогнозировать будущие значения таких важнейших показателей, как, например, объем продаж, ставки процента, издержки и т.д. В этой главе мы рассмотрим возможности применения в целях прогнозирования фактических данных за прошлые промежутки времени.

В предыдущей главе при характеристике регрессионных методов колебания зависимой переменной объяснялись на основе изучения соответствующих значений независимой переменной. В данной главе мы будем использовать аналогичный подход, причем в качестве независимой будет выступать переменная времени. К примеру, мы хотим объяснить колебания объемов продаж только через изменение значений этого показателя во времени, без учета каких-либо других факторов. Если удается выявить определенную тенденцию изменения фактических значений, то ее можно использовать для прогнозирования будущих значений данного показателя. Множество данных, в которых время является независимой переменной, называется **временным рядом.**

Модель, построенную по ретроспективным данным, не всегда можно использовать в прогнозировании отдельных показателей. Например, план некоторой компании может коренным образом измениться, если эта компания несет убытки. Кроме того, существует множество внешних факторов, которые могут полностью изменить тенденцию, существовавшую ранее. К таким факторам можно отнести существенные изменения цен на сырье, резкое увеличение уровня инфляции в мире в целом или стихийные бедствия, которые непредсказуемым образом могут повлиять на предпринимательскую деятельность.

В разделе 9.2 мы рассмотрим временные ряды, которые содержат такие элементы, как собственно тренд, сезонная вариация и циклическая вариация. Эти элементы можно объединять с помощью нескольких способов. Остановимся на двух типах моделей: **модели с аддитивной компонентой** и **модели с мультипликативной компонентой**. Как следует из их названий, элементы в этих моделях либо складываются друг с другом, либо перемножаются. Каждой из моделей соответствуют различные методы расчета компоненты тренда. Мы будем использовать сочетание методов скользящего среднего и линейной регрессии.

Следует иметь в виду, что описанные выше методы — это далеко не весь, а иногда и не лучший инструментарий для составления прогнозов. Существует множество других, более изощренных статистических методов. Помимо количественных, существуют также качественные методы, которые используются в условиях недостаточного количества или отсутствия фактических данных. Среди них можно назвать, например, **метод Дельфи**, который используется экспертами для прогнозирования возможных будущих последствий, и **метод написания сценария**.

9.2. ЭЛЕМЕНТЫ ВРЕМЕННОГО РЯДА

Значения некоторой переменной (например, объемы продаж) изменяются во времени под воздействием целого ряда факторов. Если, к примеру, некоторая компания предлагает на рынке новый вид продукции, то с течением времени объемы продаж этой продукции возрастают. Общее изменение значений переменной во времени называется **трендом** и обозначается через **T**. В примерах, которые будут рассмотрены ниже, тренд является линейным. Это означает, что модель тренда легко построить, используя для расчета параметров прямой, наилучшим образом аппроксимирующий данный тренд, метод регрессии. Затем данная модель может использоваться для прогнозирования будущих значений тренда. В действительности тренд в чистом виде либо не существует, например, при колебании значений спроса вокруг некоторой фиксированной величины, либо в большинстве случаев он является нелинейным. На приведенных ниже рис. 9.1 и 9.2 проиллюстрирован тренд значений спроса в соответствии с различными стадиями жизненного цикла продукта. Новым видам продукции соответствует возрастающий тренд, тогда как устаревшим продуктам на заключительной стадии их жизненного цикла — убывающий.

Метод скользящего среднего, изложенный ниже, можно использовать для выделения тренда из модели, содержащей сезонную компоненту. Этот метод позволяет выравнивать тренд фактических значений через сглаживание сезонных колебаний. Однако тренды, полученные с использованием метода скользящего среднего, как правило, не используются для прогнозирования будущих значений, поскольку процесс их получения предполагает высокий уровень неопределенности.

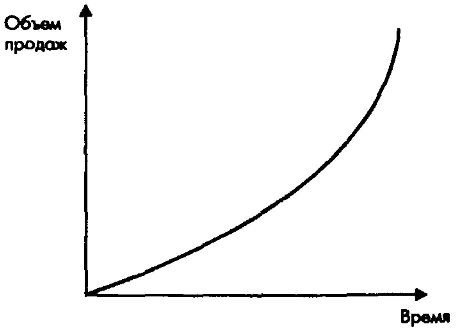


Рис. 9.1. Объемы продаж новой продукции, пользующейся спросом

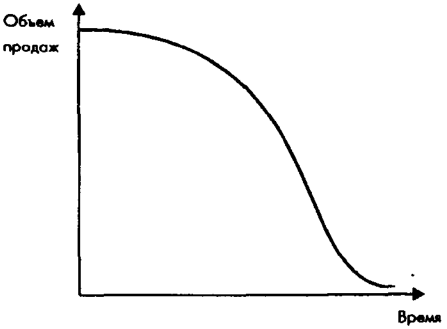


Рис. 9.2. Объемы продаж устаревшей продукции в конце жизненного цикла

В большинстве случаев значения переменных характеризуют не только тренд. Часто они подвержены циклическим колебаниям. Если эти колебания повторяются в течение небольшого промежутка времени, то они называются **сезонной вариацией.** Колебания, повторяющиеся в течение более длительного промежутка времени, называются **циклической вариацией**. Модели, содержащие сезонную компоненту, которые будут рассмотрены в данной главе, основаны на традиционном понятии сезона, однако, в более широком смысле термин «сезон» в прогнозировании применим к любым систематическим колебаниям. Например, при изучении товарооборота в течение недели под термином «сезон» подразумевается 1 день. При исследовании транспортных потоков дня или в течение недели также может использоваться модель с сезонной компонентой. Любые колебания относительно тренда, построенного по годовым значениям некоторого показателя, можно описать в виде модели с циклической компонентой. Не будем рассматривать примеры с циклическим фактором. Этот фактор можно выявить только по данным за длительные промежутки времени в 10, 15 или 20 лет, однако в данном случае колебания значений тренда могут быть вызваны воздействием общеэкономических факторов.

Наличие подобных циклических факторов можно легко обнаружить в данных за 1960–75 гг. В этот период было разработано множество методов прогнозирования, однако впоследствии тенденции общеэкономического развития претерпели значительные изменения. Остановимся подробнее на моделировании более коротких промежутков времени и не будем учитывать воздействие циклической компоненты.

Последняя предпосылка нашей модели также следует из метода линейной регрессии. Она связана со значением **ошибки**, или **остатка**, т.е. той части значения наблюдения, которую нельзя объяснить с помощью построенной модели. Величину ошибок можно использовать в качестве меры степени соответствия модели исходным данным. Обычно применяют два вида таких мер. Это **среднее абсолютное отклонение** (mean absolute deviation — *MAD*):

$$MAD = \frac{\sum |\text{Фактические значения} - \text{Прогнозные значения}|}{n} = \frac{\sum |E_t|}{n},$$

равное отношению суммы величин всех ошибок без учета их знака к общему числу наблюдений, и **среднеквадратическая ошибка** (mean square error — *MCE*):

$$MSE = \frac{\sum (E_t)^2}{n},$$

которая представляет собой отношение суммы квадратов ошибок к общему числу наблюдений. Последняя из указанных мер резко возрастает при наличии высоких ошибок.

В процессе анализа временного ряда мы стараемся определить все имеющиеся факторы и построить модель, которая соответствующим образом отражала бы их.

❑ **Пример 9.1.** Представленные ниже данные — это количество продукции, проданной компанией "Lewplan plc" в течение последних 13 кварталов.

Таблица 9.1 **Количество продукции, проданной в течение последних 13 кварталов**

Дата	*Количество проданной продукции, тыс. шт.*
Январь–март 19X6	239
Апрель–июнь	201
Июль–сентябрь	182
Октябрь–декабрь	297
Январь–март 19X7	324
Апрель–июнь	278
Июль–сентябрь	257
Октябрь–декабрь	384
Январь–март 19X8	401
Апрель–июнь	360
Июль–сентябрь	335
Октябрь–декабрь	462
Январь–март 19X9	481

Необходимо проанализировать указанное множество данных и установить, можно ли обнаружить тенденцию. Если устойчивая тенденция действительно

Рис. 9.3. Объемы продаж компании Lewplan plc по кварталам в натуральном выражении

существует, данная модель будет использоваться нами для прогнозирования количества проданной продукции в следующие кварталы.

Решение

На рис. 9.3 нанесены соответствующие значения. При построении диаграммы временного ряда полезно последовательно соединить точки отрезками, чтобы более четко увидеть любую тенденцию.

Как следует из диаграммы, возможен возрастающий тренд, содержащий сезонные колебания. Объемы продаж в зимний период (1 и 4) значительно выше, чем в летний (2 и 3). Сезонная компонента практически не изменится в течение трех лет. Тренд показывает, что в целом объем продаж возрос примерно с 230 тыс. шт. в 19X6 г. до 390 тыс. шт. в 19X8 г., однако увеличения сезонных колебаний не произошло. Этот факт свидетельствует в пользу модели с аддитивной компонентой (см. 9.3).

9.3. АНАЛИЗ МОДЕЛИ С АДДИТИВНОЙ КОМПОНЕНТОЙ: A = T + S + E

Моделью с аддитивной компонентой называется такая модель, в которой вариация значений переменной во времени наилучшим образом описывается через сложение отдельных компонент. Предположив, что циклическая вариация не учитывается, модель фактических значений переменной A можно представить следующим образом:

Фактическое значение = Трендовое значение + Сезонная вариация + Ошибка, т.е.

$$A = T + S + E.$$

В моделях как с аддитивной, так и с мультипликативной компонентой общая процедура анализа примерно одинакова:

Шаг 1. Расчет значений сезонной компоненты.

Шаг 2. Вычитание сезонной компоненты из фактических значений. Этот процесс называется десезонализацией данных. Расчет тренда на основе полученных десезонализированных данных.

Шаг 3. Расчет ошибок как разности между фактическими и трендовыми значениями.

Шаг 4. Расчет среднего отклонения (MAD) или среднеквадратической ошибки (MSE) для обоснования соответствия модели исходным данным или для выбора из множества моделей наилучшей.

9.3.1. Расчет сезонной компоненты в аддитивных моделях

❐ **Пример 9.2.** Вернемся к примеру 9.1 предыдущего параграфа, в котором рассматриваются квартальные объемы продаж компании Lewplan plc. Мы уже *выяснили, что этим данным отвечает аддитивная модель*, т.е. фактически объемы продаж можно выразить следующим образом:

$$A = T + S + E.$$

Для того чтобы элиминировать влияние сезонной компоненты, воспользуемся методом скользящей средней. Просуммировав первые четыре значения, получим общий объем продаж в 19X6 г. Если поделить эту сумму на четыре, можно найти средний объем продаж в каждом квартале 19X6 года, т. е.

$$(239 + 201 + 182 + 297)/4 = 229{,}75.$$

Полученное значение уже не содержит сезонной компоненты, поскольку представляет собой среднюю величину за год. У нас появилась оценка значения тренда для середины года, т.е. для точки, лежащей в середине между кварталами II и III. Если последовательно передвигаться вперед с интервалом в три месяца, можно рассчитать средние квартальные значения на промежутке: апрель 19X6 — март 19X7 (251), июль 19X6 — июнь 19X7 (270,25) и т.д. Данная процедура позволяет генерировать скользящие средние по четырем точкам для исходного множества данных. Получаемое таким образом множество скользящих средних представляет наилучшую оценку искомого тренда.

Теперь полученные значения тренда можно использовать для нахождения оценок сезонной компоненты. Мы рассчитываем:

$$A - T = S + E.$$

К сожалению, оценки значений тренда, полученные в результате расчета скользящих средних по четырем точкам, относятся к несколько иным моментам времени, чем фактические данные. Первая оценка, равная 229,75, представляет собой точку, совпадающую с серединой 19Х6 г., т.е. лежит в центре промежутка фактических значений объемов продаж во II и III кварталах. Вторая оценка, равная 251, лежит между фактическими значениями в III и IV кварталах. Нам же требуются десезонализированные средние значения, соответствующие тем же интервалам времени, что и фактические значения за квартал. Положение десезонализированных средних во времени сдвигается путем дальнейшего расчета средних для каждой пары значений. Найдем среднюю из первой и второй оценок, центрируя их на июль-сентябрь 19Х6 г., т. е.

$$(229{,}75 + 251)/2 = 240{,}4.$$

Это и есть десезонализированная средняя за июль-сентябрь 19Х6 г. Эту десезонализированную величину, которая называется **центрированной скользящей средней**, можно непосредственно сравнивать с фактическим значением за июль-сентябрь 19Х6 г., равным 182. Отметим, что это означает отсутствие оценок тренда за первые два или последние два квартала временного ряда. Результаты этих расчетов приведены в табл. 9.2.

Таблица 9.2. **Расчет по 4 точкам центрированных скользящих средних значений тренда для модели A — T = S + E**

Дата	*Объем продаж, тыс. шт.*	*Итого за четыре квартала*	*Скользящая средняя за четыре квартала*	*Центрированная скользящая средняя*	*Оценка сезонной компоненты A −T=S + E*
1	2	3	4	5	6
Январь–март 19Х6	239		–		
Апрель–июнь	201		–		
		919	229,75		
Июль–сентябрь	182			240,4	− 58,4
		1004	251		
Октябрь–декабрь	297			260,6	+ 36,4
		1081	270,25		
Январь–март 19Х7	324			279,6	+ 44,4
		1156	289		
Апрель–июнь	278			299,9	− 21,9
		1243	310,75		
Июль–сентябрь	257			320,4	− 63,4
		1320	330		
Октябрь–декабрь	384			340,3	+ 43,8
		1402	350,5		

Продолжение

1	2	3	4	5	6
Январь–март 19Х8	401			360,2	+ 40,8
		1480	370		
Апрель–июнь	360			379,8	− 19,8
		1558	389,5		
Июль–сентябрь	335			399,5	− 64,5
		1638	409,5		
Октябрь–декабрь	462		—		
Январь–март 19Х9	481		—		

Для каждого квартала мы имеем оценки сезонной компоненты, которые включают в себя ошибку или остаток. Прежде чем мы сможем использовать сезонную компоненту, нужно пройти два следующих этапа. Найдем средние значения сезонных оценок для каждого сезона года. Эта процедура позволит уменьшить некоторые значения ошибок. Наконец, скорректируем средние значения, увеличивая или уменьшая их на одно и то же число таким образом, чтобы общая их сумма была равна нулю. Это необходимо, чтобы усреднить значения сезонной компоненты в целом за год. Корректирующий фактор рассчитывается следующим образом: сумма оценок сезонных компонент делится на 4. В последнем столбце табл. 9.2 эти оценки записаны под соответствующими квартальными значениями. Сама процедура приведена в табл. 9.3.

Таблица 9.3. **Расчет средних значений сезонной компоненты**

	Год	*Номер квартала*				
		1	*2*	*3*	*4*	
	19Х6	–	–	−58,4	+ 36,4	
	19Х7	+ 44,4	− 21,9	− 63,4	+ 43,8	
	19Х8	+ 40,8	− 19,8	− 64,5	–	
Итого		+ 85,2	− 41,7	−186,3	+ 80,2	
Среднее значение		85,2 ÷ 2	− 41,7 ÷ 2	− 186,3 ÷ 2	80,2 ÷ 3	
Оценка сезонной компоненты		+ 42,6	− 20,8	− 62,1	+ 40,1	Сумма = −0,2
Скорректированная сезонная компонента*		+ 42,6	− 20,7	− 62,0	+ 40,1	Сумма = 0

* В данном случае производилось округление двух значений сезонной компоненты до ближайшего большего числа, а двух значений — до ближайшего меньшего числа таким образом, чтобы общая сумма была равна нулю.

Значения сезонной компоненты еще раз подтверждают наши выводы, сделанные в **9.2** на основе диаграммы. Объемы продаж за два зимних квартала превышают среднее трендовое значение приблизительно на 40 тыс. шт., а объемы продаж за два летних периода ниже средних на 21 и 62 тыс. шт. соответственно.

Аналогичная процедура применима при определении сезонной вариации за любой промежуток времени. Если, например, в качестве сезонов выступают дни недели, для элиминирования влияния ежедневной «сезонной компоненты» также рассчитывают скользящую среднюю, но уже не по четырем, а по семи точкам. Эта скользящая средняя представляет собой значение тренда в середине недели, т.е. в четверг; таким образом, необходимость в процедуре центрирования отпадает.

9.3.2. Десезонализация данных при расчете тренда

Шаг 2 состоит в десезонализации исходных данных. Она заключается в вычитании соответствующих значений сезонной компоненты из фактических значений данных за каждый квартал, т.е. A − S = T + E, что показано ниже.

Таблица 9.4 **Расчет десезонализированных данных**

Дата	*Номер квартала*	*Объем продаж, тыс. шт.* A	*Сезонная компонента* S	*Десезонализированный объем продаж, тыс. шт.* A − S = T + E
Январь–март 19X6	1	239	(+42,6)	196,4
Апрель–июнь	2	201	(−20,7)	221,7
Июль–сентябрь	3	182	(−62,0)	244,0
Октябрь–декабрь	4	297	(+40,1)	256,9
Январь–март 19X7	5	324	(+42,6)	281,4
Апрель–июнь	6	278	(−20,7)	298,7
Июль–сентябрь	7	257	(−62,0)	319,0
Октябрь–декабрь	8	384	(+40,1)	343,9
Январь–март 19X8	9	401	(+42,6)	358,6
Апрель–июнь	10	360	(−20,7)	380,7
Июль–сентябрь	11	335	(−62,0)	397,1
Октябрь–декабрь	12	462	(+40,1)	421,9
Январь–март 19X9	13	481	(+42,6)	438,4

Новые оценки значений тренда, которые еще содержат ошибку, можно использовать для построения модели основного тренда. Если нанести эти значения на исходную диаграмму, можно сделать вывод о существовании явного линейного тренда.

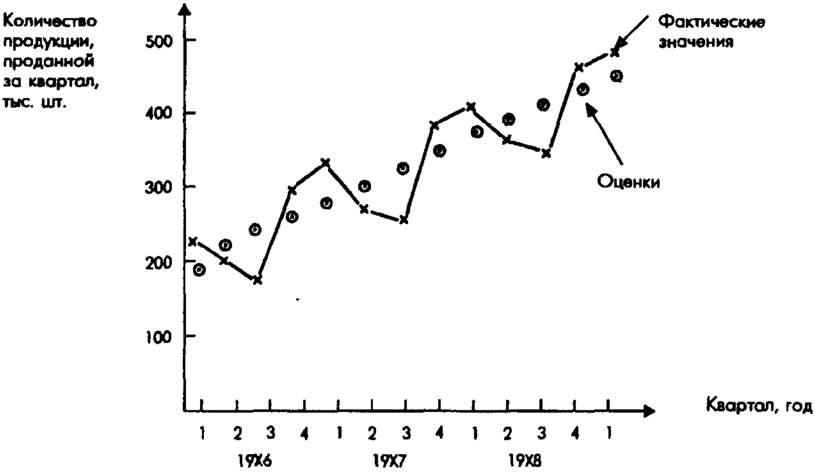


Рис. 9.4. Фактические и десезонализированные квартальные объемы продаж компании Lewplan plc

Уравнение линии тренда имеет вид:

$$T = a + b \times \text{ — номер квартала},$$

где a и b характеризуют точку пересечения с осью ординат и наклон линии тренда. Для определения параметров прямой, наилучшим образом аппроксимирующей тренд, можно использовать метод наименьших квадратов. Таким образом, как мы знаем из предыдущей главы о линейной регрессии, уравнения для расчета параметров a и b будут иметь вид:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2},$$

$$a = \frac{\sum y}{n} - \frac{b\sum x}{n},$$

где x — порядковый номер квартала, y — значение (T + E) в предыдущей таблице. С помощью калькулятора подсчитаем:

$$\sum x = 91, \quad \sum x^2 = 819, \quad \sum y = 4158{,}7, \quad \sum xy = 32747{,}1, \quad n = 13.$$

Подставив найденные значения в соответствующие формулы, получим:

$$b = 19{,}978, \qquad a = 180{,}046.$$

Следовательно, уравнение модели тренда имеет следующий вид:

Трендовое значение объема продаж, тыс. шт. = 180,0 + 20,0 × номер квартала.

9.3.3. Расчет ошибок

Шаг 3 нашего алгоритма, предшествующий составлению прогнозов, состоит в расчете ошибок или остатка. Наша модель имеет следующий вид:

$$A = T + S + E.$$

Значение *S* было найдено в разделе **9.3.1**, а значение T — в разделе **9.3.2.** Вычитая каждое это значение из фактических объемов продаж, получим значения ошибок.

Таблица 9.5. **Расчет ошибок для модели с аддитивной компонентой**

Дата	*Номер квартала*	*Объем продаж, тыс. шт.* *A*	*Сезонная компонента* *S*	*Трендовое значение, тыс. шт.* *T*	*Ошибка, тыс. шт.* *A − S−T=E*
Январь–март 19X6	1	239	(+42,6)	200	− 3,6
Апрель–июнь	2	201	(−20,7)	220	+ 1,7
Июль–сентябрь	3	182	(−62,0)	240	+4,0
Октябрь–декабрь	4	297	(+40,1)	260	− 3,1
Январь–март 19X7	5	324	(+42,6)	280	+ 1,4
Апрель–июнь	6	278	(−20,7)	300	− 1,3
Июль–сентябрь	7	257	(−62,0)	320	− 1,0
Октябрь–декабрь	8	384	(+40,1)	340	+ 3,9
Январь–март 19X8	9	401	(+42,6)	360	− 1,6
Апрель–июнь	10	360	(−20,7)	380	+ 0,7
Июль–сентябрь	11	335	(−62,0)	400	− 3,0
Октябрь–декабрь	12	462	(+40,1)	420	+ 1,9
Январь–март 19X9	13	481	(+42,0)	440	− 1,6

Последний столбец этой таблицы можно использовать в шаге 4 при расчете среднего абсолютного отклонения (**MAD**) или средней квадратической ошибки (**MSE**):

$$MAD = \frac{\sum |E_t|}{n} = \frac{28,7}{13} = 2,2\ ,$$

$$MSE = \frac{\sum (E_t)^2}{n} = \frac{78{,}85}{13} = 6{,}1 .$$

В нашем случае ошибки достаточно малы и составляют от 1 до 2%. Тенденция, выявленная по фактическим данным, достаточно устойчива и позволяет получить хорошие краткосрочные прогнозы.

9.3.4. Прогнозирование по аддитивной модели

Прогнозные значения по модели с аддитивной компонентой рассчитываются как

$$F = T + S \text{ (тыс. шт. за квартал)},$$

где трендовое значение T = 180 + 20 × номер квартала, а сезонная компонента S составляет +42,6 в январе–марте, – 20,7 в апреле–июне, 62,0 в июле–сентябре и +40,1 в октябре–декабре.

Порядковый номер квартала, охватывающего ближайшие три месяца с апреля по июль 19X9 г., равен 14, таким образом прогнозное трендовое значение составит:

$$T_{14} = 180 + 20 \times 14 = 460 \text{ (тыс. шт. за квартал)} .$$

Соответствующая сезонная компонента равна – 20,7 тыс. шт. Следовательно, прогноз на этот квартал определяется как:

$$F \text{ (апрель–июнь 19X9 г.)} = 460 - 20{,}7 = 439{,}3 \text{ тыс. шт.}$$

Не следует забывать: чем более отдаленным является период упреждения, тем меньшей оказывается обоснованность прогноза. В данном случае мы предполагаем, что тенденция, обнаруженная по ретроспективным данным, распространяется и на будущий период. Для сравнительно небольших периодов упреждения такая предпосылка может действительно иметь место, однако ее выполнение становится менее вероятным по мере составления прогнозов на более отдаленную перспективу.

9.4. АНАЛИЗ МОДЕЛИ С МУЛЬТИПЛИКАТИВНОЙ КОМПОНЕНТОЙ: A = T × S × E

В некоторых временных рядах значение сезонной компоненты не является константой, а представляет собой определенную долю трендового значения. Таким образом, значения сезонной компоненты увеличиваются с возрастанием значений тренда.

❒ **Пример 9.3.** Компания CD plc осуществляет реализацию нескольких видов продукции. Объемы продаж одного из продуктов за последние 13 кварталов представлены в таблице 9.6.

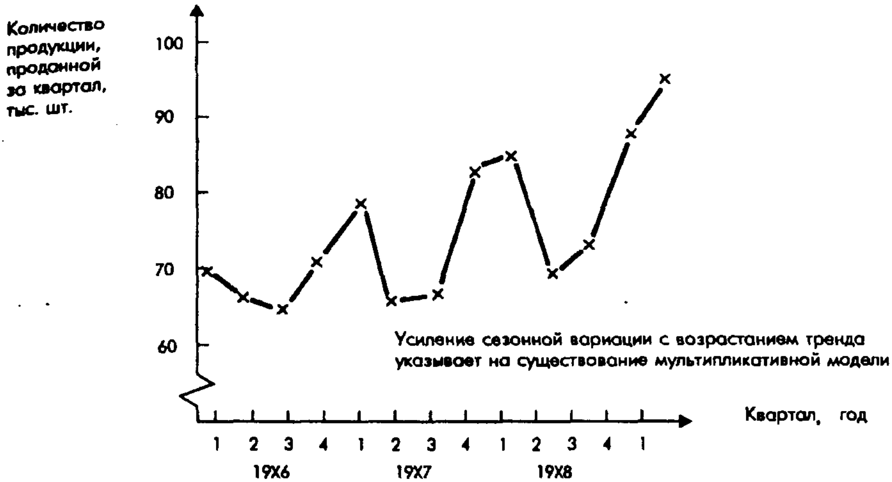


Рис. 9.5. Квартальные объемы продаж компании CD plc

Таблица 9.6. **Квартальные объемы продаж компании CD plc**

Дата	*Номер квартала*	*Количество проданной продукции, тыс. шт. A*
Январь–март 19X6	1	70
Апрель–июнь	2	66
Июль–сентябрь	3	65
Октябрь–декабрь	4	71
Январь–март 19X7	5	79
Апрель–июнь	6	66
Июль–сентябрь	7	67
Октябрь–декабрь	8	82
Январь–март 19X8	9	84
Апрель–июнь	10	69
Июль–сентябрь	11	72
Октябрь–декабрь	12	87
Январь–март 19X9	13	94

Построим по этим данным точечную диаграмму:

Объем продаж этого продукта так же, как и в предыдущем примере, подвержен сезонным колебаниям, и значения его в зимний период выше, чем в летний. Однако размах вариации фактических значений относительно линии тренда постоянно возрастает. К таким данным следует применять модель с мультипликативной компонентой:

Фактическое значение = Трендовое значение × Сезонная вариация × Ошибка,

т. е.

$$A = T \times S \times E.$$

В нашем примере есть все основания предположить существование линейного тренда, но чтобы полностью в этом убедиться, проведем процедуру сглаживания временного ряда.

9.4.1. Расчет значений сезонной компоненты

В сущности эта процедура ничем не отличается от той, которая применялась для аддитивной модели. Так же вычисляются центрированные скользящие средние для трендовых значений, однако оценки сезонной компоненты представляют собой коэффициенты, полученные по формуле $A/T = S \times E$. Результаты расчетов приведены в табл. 9.7.

Таблица 9.7. **Расчет значений сезонной компоненты для CD plc**

Дата	*Номер квартала*	*Объем продаж, тыс. шт. A*	*Скользящая средняя за четыре квартала*	*Центрированная скользящая средняя*	*Коэффициент сезонности* $A/T = S \times E$
1	2	3	4	5	6
Январь–март 19Х6	1	70			
Апрель–июнь	2	66			
			68		
Июль–сентябрь	3	65		69,13	0,940
			70,25		
Октябрь–декабрь	4	71		70,25	1,011
			70,25		
Январь–март 19Х7	5	79		70,50	1,121
			70,75		
Апрель–июнь	6	66		72,13	0,915
			73,50		
Июль–сентябрь	7	67		74,13	0,904
			74,75		

Продолжение

1	2	3	4	5	6
Октябрь–декабрь	8	82		75,13	1,092
			75,50		
Январь–март 19X8	9	84		76,13	1,103
			76,75		
Апрель–июнь	10	69		77,38	0,892
			78		
Июль–сентябрь	11	72		79,25	0,909
			80,50		
Октябрь–декабрь	12	87		–	–
Январь–март 19X9	13	94		–	–

Значения сезонных коэффициентов получены на основе квартальных оценок по аналогии с алгоритмом, который применялся для аддитивной модели. Так как значения сезонной компоненты — это доли, а число сезонов равно четырем, необходимо, чтобы их сумма была равна четырем, а не нулю, как в предыдущем случае. (Если бы в исходных данных предполагалось семь сезонов в течение недели по одному дню каждый, то общая сумма значений сезонной компоненты должна была бы равняться семи). Если эта сумма не равна четырем, производится корректировка значений сезонной компоненты точно таким же образом, как это уже делалось ранее. В таблице оценки, рассчитанные в последнем столбце предшествующей табл. 9.8, расположены под соответствующим номером квартала.

Таблица 9.8. **Расчет значений сезонной компоненты для CD plc**

	Год	*Номер квартала*				
		1	2	3	4	
	19X6	–	–	0,940	1,011	
	19X7	1,121	0,915	0,904	1,092	
	19X8	1,103	0,892	0,909	–	
Итого		2,224	1,807	2,753	2,103	
Среднее значение		2,224 + 2	1,807 + 2	2,753 + 2	2,103 +2	
Оценка сезонной компоненты		1,112	0,903	0,918	1,051	Сумма = 3,984
Скорректированная сезонная компонента*		1,116	0,907	0,922	1,055	Сумма = 0

* Скорректированная оценка сезонной компоненты получена в результате умножения соответствующей доли на (4/3,984).

Как показывают оценки, в результате сезонных воздействий объемы продаж в январе—марте увеличиваются на 11,6% соответствующего значения тренда (1,116). Аналогично сезонные воздействия в октябре-декабре приводят к увеличению объема продаж на 5,5% от соответствующего значения тренда. В двух других кварталах сезонные воздействия состоят в снижении объемов продаж, которое составляет 90,7 и 92,2% от соответствующих трендовых значений.

9.4.2. Десезонализация данных и расчет уравнения тренда

После того как оценки сезонной компоненты определены, можем приступить к процедуре десезонализации данных по формуле $A/S = T \times E$. Результаты расчетов этих оценок значений тренда приведены в табл. 9.9.

Таблица 9.9. **Расчет уравнения тренда для компании CD plc**

Дата	*Номер квартала*	*Объем продаж, тыс. шт.* A	*Коэффициент сезонности* S	*Десезонализированный объем продаж, тыс. шт.* $A/T = S \times E$
Январь-март 19X6	1	70	1,116	62,7
Апрель-июнь	2	66	0,907	72,8
Июль-сентябрь	3	65	0,922	70,6
Октябрь-декабрь	4	71	1,055	67,3
Январь-март 19X7	5	79	1,116	70,8
Апрель-июнь	6	66	0,907	72,8
Июль-сентябрь	7	67	0,922	72,7
Октябрь-декабрь	8	82	1,055	77,7
Январь-март 19X8	9	84	1,116	75,2
Апрель-июнь	10	69	0,907	76,1
Июль-сентябрь	11	72	0,922	78,2
Октябрь-декабрь	12	87	1,055	82,4
Январь-март 19X9	13	94	1,116	84,2

Полученные трендовые значения наносятся на исходную точечную диаграмму.

Точки, образующие представленный на графике тренд, достаточно сильно разбросаны. Объемы продаж в данном случае не образуют такой строгой последовательности, как в предыдущем примере с компанией Lewplan plc. Скорее всего, пример с CD plc более близок к реальной действительности.

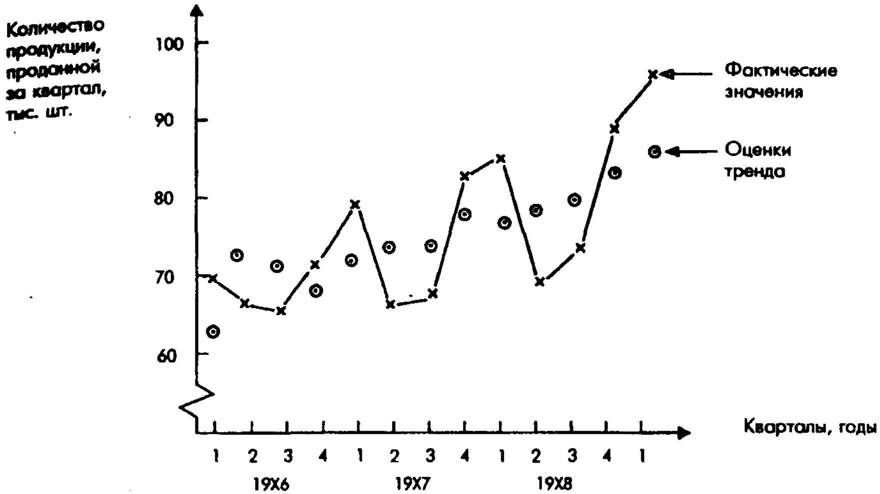


Рис. 9.6. Фактический и десезонализированный объем продаж по 3-месячной средней

Теперь нужно принять решение о том, какой вид будет иметь уравнение тренда. Очевидно, что линия тренда — не кривая, наоборот, она несколько больше напоминает прямую, хотя отдельные точки, особенно значения за 19X6 г, расположены хаотически. Предположим для простоты, что тренд линейный, и для расчета параметров прямой, наилучшим образом его аппроксимирующей, будем применять метод наименьших квадратов. Воспользовавшись той же процедурой, что и в разделе **9.3.2**, находим, что

$$T = 64{,}6 + 1{,}36 \times \text{номер квартала (тыс. шт. в квартал)} .$$

Это уравнение будем использовать в дальнейшем для расчета оценок трендовых объемов продаж на каждый момент времени.

9.4.3. Расчет ошибок: A/(T × S) = E или A — (T × S) = E

Итак, мы нашли значения тренда и сезонной компоненты. Теперь мы можем использовать их для того, чтобы рассчитать ошибки в прогнозируемых по модели объемах продаж T × S по сравнению с фактическими значениями A. В табл. 9.10 эти ошибки рассчитаны как отношение E = A/(T × S).

Для каждого рода ошибки достаточно велики, что видно из графика десезонализированных значений. Однако начиная с первого квартала 19X7 г. величина ошибки составляет в среднем 2–3% от фактического значения, и можно сделать вывод о соответствии построенной модели фактическим данным.

Таблица 9.10. **Расчет ошибок для компании CD plc**

Дата	Номер квартала	Объем продаж, тыс. шт. A	Сезонная компонента S	Трендовое значение, тыс. шт. T	Ошибка		
					$T \times S$	$A/(T\times S)$	$A-(T\times S)$
Январь–март 19X6	1	70	1,116	66,0	73,7	0,95	− 3,7
Апрель–июнь	2	66	0,907	67,3	61,0	1,08	+ 5,0
Июль–сентябрь	3	65	0,922	68,7	63,3	1,03	+ 1,7
Октябрь–декабрь	4	71	1,055	70,0	73,9	0,96	− 2,9
Январь–март 19X7	5	79	1,116	71,4	79,7	0,99	− 0,7
Апрель–июнь	6	66	0,907	72,8	66,0	1,00	0
Июль–сентябрь	7	67	0,922	74,1	68,3	0,98	− 1,3
Октябрь–декабрь	8	82	1,055	75,5	79,7	1,03	+ 2,3
Январь–март 19X8	9	84	1,116	76,8	85,7	0,98	− 1,7
Апрель–июнь	10	69	0,907	78,2	70,9	0,97	− 1,9
Июль–сентябрь	11	72	0,922	79,6	73,3	0,98	− 1,3
Октябрь–декабрь	12	87	1,055	80,9	85,4	1,02	+ 1,6
Январь–март 19X9	13	94	1,116	82,3	91,9	1,02	+ 2,1

9.4.4. Прогнозирование по модели с мультипликативной компонентой

При составлении прогнозов по любой модели предполагается, что можно найти уравнение, удовлетворительно описывающее значения тренда. В обоих изложенных выше примерах эта предпосылка была успешно выполнена. Тренд, который нами рассматривался, был очевидно линейным. Если бы исследуемый тренд представлял собой кривую, мы были бы вынуждены моделировать эту связь с помощью одного из методов формализации нелинейных взаимосвязей, рассмотренных в предыдущей главе. После того как параметры уравнения тренда определены, процедура составления прогнозов становится совершенно очевидной. Прогнозные значения определяются по формуле:

$$F = T \times S,$$

где

$$T = 64{,}6 + 1{,}36 \times \text{номер квартала (тыс. шт. за квартал)},$$

а сезонные компоненты составляют 1,116 в первом квартале, 1,097 — во втором, 0,922 — в третьем и 1,055 в четвертом квартале. Ближайший следующий квартал — это второй квартал 19X9 г., охватывающий период с апреля по июнь и имеющий во временном ряду порядковый номер 14. Прогноз объема продаж в этом квартале составляет:

$$F = T \times S = (64{,}6 + 1{,}36 \times 14) \times 0{,}907 = 83{,}64 \times 0{,}907 = 75{,}9 \text{ (тыс. шт. за квартал)}.$$

С учетом величины ошибки прогноза мы можем сделать вывод, что данная оценка будет отклоняться от фактического значения не более чем на 2–3%. Аналогично, прогноз на октябрь–декабрь 19X9 г., рассчитывается для квартала с порядковым номером 16 с использованием значения сезонной компоненты для IV квартала года:

$$F = T \times S = (64{,}6 + 1{,}36 \times 16) \times 1{,}055 = 83{,}36 \times 1{,}055 = 91{,}1 \text{ (тыс. шт. за квартал)}.$$

Разумно предположить, что величина ошибки данного прогноза будет несколько выше, чем предыдущего, поскольку этот прогноз рассчитан на более длительную перспективу.

РЕЗЮМЕ

Под временным рядом понимается любое множество данных, относящихся к определенным моментам времени. Это могут быть, скажем, годы, кварталы, месяцы или недели. В моделях временного ряда ретроспективная тенденция используется для прогнозирования поведения переменной в будущем. Краткосрочные прогнозы являются более точными, чем долгосрочные. Если прогноз составлялся на более длительный период времени при условии, что существующая тенденция сохранится в будущем, то тем больше величина ошибки.

Для моделирования временных рядов используются два типа моделей — аддитивная и мультипликативная. В обоих случаях предполагается, что значение переменной включает в себя ряд компонент. Временной ряд может состоять из собственно тренда — общей тенденции изменения значений переменной; сезонной вариации — краткосрочных периодических колебаний значений переменной; циклической вариации — долгосрочных периодических колебаний значений переменной; ошибки или остатка. В данном учебном пособии не рассматривались массивы данных за длительные промежутки времени, содержащие циклическую вариацию.

Рассмотренные нами модели имеют следующий вид:

$$\text{Аддитивная } A = T + S + E,$$

$$\text{Мультипликативная } A = T \times S \times E.$$

В обоих видах моделей для десезонализации данных применяется метод скользящего среднего. Затем десезонализированные данные используются при построении модели тренда. По этой модели составляют прогнозы будущих значений тренда. В случае линейной модели для нахождения параметров прямой, наилучшим образом аппроксимирующей фактические значения, используется метод наименьших квадратов. Процесс построения нелинейных моделей гораздо более сложен.

В отличие от линейных регрессионных моделей для оценки обоснованности или точности прогнозных моделей статистические методы, как правило, не используются. Наилучшую среди нескольких моделей выбирает специалист, составляющий прогноз. Чтобы определить, насколько точно рассматриваемая модель аппроксимирует прошлые данные, применяются два показателя:

$$\text{Среднее абсолютное отклонение (MAD)} = \frac{\sum |E_t|}{n}.$$

$$\text{Среднеквадратическая ошибка (MSE)} = \frac{\sum (E_t)^2}{n}.$$

УПРАЖНЕНИЯ

Упражнение 9.1

За последние 11 кварталов товарооборот компании "Amada plc", скорректированный на инфляцию, составил:

Год	1				2				3		
Квартал	1	2	3	4	1	2	3	4	1	2	3
Товарооборот	22	28	34	27	31	43	43	41	46	53	56

Требуется:

1. В предположении существования линейного тренда построить модель с аддитивной компонентой.
2. Сделать прогноз на ближайшие три квартала. Прокомментировать вопрос о вероятной точности ваших прогнозов.

Упражнение 9.2

Спрос на стулья, которые продает компания "Peace Retailers", составил:

Год	1		2				3			
Квартал	3	4	1	2	3	4	1	2	3	4
Спрос	157	137	156	151	153	141	154	152	154	142

Требуется:

1. Построить соответствующую модель с аддитивной компонентой.
2. Дать прогноз на первые два квартала четвертого года.

Упражнение 9.3

Ниже приведены квартальные объемы выпуска продукции компании "Cobournes plc":

Год	1				2				3		
Квартал	1	2	3	4	1	2	3	4	1	2	3
Объем выпуска	24	50	56	63	79	89	79	80	93	100	88

Требуется:

1. Проанализировать значения квартальных объемов выпуска на основе модели с аддитивной компонентой.
2. Прокомментировать поведение тренда.

Упражнение 9.4

Положение дел в компании "Doble-Flood" достаточно тяжелое. Ниже приводятся данные о прибыли компании за последние 10 кварталов (скорректированные на инфляцию).

Год	1				2				3	
Квартал	1	2	3	4	1	2	3	4	1	2
Прибыль	146	106	123	89	97	74	80	53	56	35

Требуется:

1. Построить модель временного ряда с мультипликативной компонентой.
2. Дать прогноз на 2 следующих квартала года 3.

Упражнение 9.5

Объемы выпуска компании "Banham and Barsey" возрастают из года в год. Ниже приведены значения этого показателя.

Год	1			2				3		
Квартал	2	3	4	1	2	3	4	1	2	3
Объем выпуска	400	715	600	585	560	975	800	765	720	1235

Требуется:

1. Построить соответствующую модель временного ряда с мультипликативной компонентой.
2. Дать прогноз на два следующие квартала.

Упражнение 9.6

Используя данные для "Peace Retailers" из упражнения 9.2.

Требуется:

1. Рассчитать ошибку, среднее абсолютное отклонение (MAD) и среднеквадратическую ошибку (MSE) для модели с аддитивной компонентой, построенной в упражнении **9.2**.
2. Используя простую экспоненциальную модель скользящего среднего при F_0 = 150 и a = 0,2, вычислить прогнозные значения, ошибку и соответствующие MAD и MSE.
3. Сравнить результаты, полученные в п.п. **1** и **2**, и сделать вывод о том, какая из моделей лучше.

Упражнение 9.7

В июне 1990 г. директор-распорядитель крупного мебельного магазина "Cushair Designs" пригласил консультанта по менеджменту с целью разработать простой и практичный метод прогнозирования уровней квартальных объемов продаж магазина на шесть месяцев вперед. В процессе решения этой задачи консультант пришел к выводу, что метод прогнозирования, отвечающий этой цели, в первую очередь предполагает десезонализацию валового объема продаж за последние 30 месяцев. Затем полученные значения временного ряда можно нанести на график, построить прямую, наилучшим образом их аппроксимирующую, и экстраполировать полученную тенденцию на ближайшие два квартала. Применение соответствующего индекса сезонности к этим значениям позволит получить оценки объемов продаж на два следующих квартала.

Значения валового объема продаж Cushair Designs

Период продаж	*Объем розничной продажи, тыс. ф. ст.*
Январь–март 1988	285
Апрель–июнь 1988	310
Июль–сентябрь 1988	315
Октябрь–декабрь 1988	385
Январь–март 1989	340
Апрель–июнь 1989	370
Июль–сентябрь 1989	375
Октябрь–декабрь 1989	460
Январь–март 1990	395
Апрель–июнь 1990	425

Консультант по менеджменту дал также некоторые рекомендации по поводу того, как можно избежать расчета индексов сезонности для универмага. Он счел эту продедуру нецелесообразной, поскольку в его распоряжении было слишком мало информации за прошлые периоды. Он решил использовать национальный поквартальный индекс сезонности, значения которого публикуются в прессе. Он предполагал, что ассортиментный набор мебельного магазина его клиента незначительно отличается от того ассортиментного набора, на основе которого строится национальный индекс.

Национальный поквартальный индекс сезонности продаж для мебели

	Январь–март	*Апрель–июнь*	*Июль–сентябрь*	*Октябрь–декабрь*
Мультипликативный индекс сезонности	94	98	96	112

Требуется:

а) Нанести на график фактические значения квартальных объемов продаж и объяснить, почему мультипликативная модель соответствует этим данным в большей степени, чем аддитивная.

б) Найти значения десезонализированных данных и нанести их на график.

в) Используя метод наименьших квадратов, определить параметры уравнения прямой, проходящей через десезонализированные данные.

г) Рассчитать оценки валового объема продаж в третьем и четвертом кварталах 1990 г.

д) Определить вероятную точность полученных оценок.

Часть 3

ПЛАНИРОВАНИЕ В БИЗНЕСЕ

В третьей и четвертой частях данной книги рассматриваются модели исследования операций.

Единого общепринятого определения исследования операций никогда не существовало. Однако в рассматриваемом нами контексте утверждение, что исследованием операций называется применение научных методов в решении проблем бизнеса, можно принять в качестве рабочего определения. Исследование операций часто, хотя и не во всех случаях, включает в себя математические и статистические модели. Обычно эти модели применяются в сложных ситуациях, в которых фигурирует множество взаимодействующих друг с другом целей и переменных.

В главах, посвященных регрессии и анализу временных рядов, нами были разработаны модели, позволяющие объяснить изменчивость и выявить ее причины. Тем не менее для применения решений в различных типах сложных ситуаций были созданы модели исследования операций. Они представляют собой стандартный набор методов, которые, как предполагают специалисты в области исследования операций, должны получить широкое распространение. Между тем на практике возникающие проблемы достаточно редко укладываются в рамки этих моделей. Гораздо чаще эти методы позволяют исследователю выработать возможные направления вложения инноваций. Необходимо помнить, что построенная модель должна как можно точнее воспроизводить моделируемую ситуацию.

В третьей части рассматривается применение методов сетевого анализа календарного планирования проектов и управления запасами в планировании бизнеса.

Глава 10. СЕТЕВОЙ АНАЛИЗ И КАЛЕНДАРНОЕ ПЛАНИРОВАНИЕ ПРОЕКТОВ

10.1. ВВЕДЕНИЕ

Сетевой анализ — это метод планирования работ проектного характера, т.е. работ, операции в которых, как правило, не повторяются. Этот метод применим, например, при составлении календарного плана выполнения операций, входящих в программу инсталлирования компьютерной системы в некоторой компании, или операций, являющихся составными частями улучшения обстановки оффиса. Процессы инсталлирования компьютерных систем или улучшения обстановки оффиса в данной компании могут протекать непрерывно, однако, вряд ли два любых проекта окажутся совершенно одинаковыми.

Методы сетевого анализа позволяют осуществить анализ проекта, который включает в себя большое число взаимосвязанных операций. Мы можем определить вероятную продолжительность выполнения работ, их стоимость, возможные размеры экономии времени или денежных средств, а также то, выполнение каких операций нельзя отсрочить, не задержав при этом срок выполнения проекта в целом. Немаловажной является и проблема обеспечения ресурсами. Методы сетевого анализа могут быть использованы при составлении календарного плана выполнения операций, удовлетворяющего существующим ограничениям на обеспечение ресурсами.

Анализ любого проекта осуществляется в три этапа:

1. Расчленение проекта на ряд отдельных работ(или операций), из которых затем составляется логическая схема. Под операцией понимается деятельность или процесс, выполнение которых требует затрат временных и/или иных ресурсов.
2. Оценка продолжительности выполнения каждой операции; составление календарного плана выполнения проекта и выделение работ, которые определяют завершение выполнения проекта в целом.
3. Оценка потребностей каждой операции в ресурсах; пересмотр плана выполнения операций с учетом обеспечения ресурсами либо перераспределение денежных или других ресурсов, которое улучшит план.

Рассмотрим каждый из этих этапов в отдельности.

10.2. СЕТЕВЫЕ ГРАФЫ

Первым шагом в анализе любого проекта является составление списка входящих в него операций. Детали такого списка зависят от специфики конкретного проекта. Тем не менее во всех случаях необходимо выделить непосредственно предшествующую операцию или операции. Непосредственно предшествующими называются операции, выполнение которых должно быть закончено прежде, чем может начаться данная операция. Например, при постройке дома крыша не может быть построена до того момента, пока не закончится возведение стен.

После того как составлен список, логическая последовательность выполнения операций может быть проиллюстрирована с помощью графа. Существуют различные типы графов, но наиболее широкое применение получили так называемые **вершинные** и **стрелочные графы**. Однако каждый из них имеет свои преимущества и недостатки, и выбор того или иного графа является вопросом личных предпочтений или же определяется целью создания и использования данного графа.

10.2.1. Стрелочные графы

В этом типе графов (рис. 10.1) каждая операция представлена стрелкой. Длина стрелок значения не имеет. Направление стрелки отражает ход времени и обычно указывается слева направо. Начало и окончание каждой операции называются событиями и изображаются на графе кружочками или узлом.

Рис. 10.1. Изображение операции на стрелочном графе

Операции обозначают буквой или словом, а события — числом. Поскольку любая операция характеризуется парой событий, ее можно также обозначать с помощью чисел, соответствующих этим событиям. Например, на рис. 10.1 операция A означает то же самое, что и операция (1, 2). Одному узлу может соответствовать (входить или выходить из него) несколько операций. Событие, изображаемое на графе с помощью узла, не считается свершившимся до тех пор, пока не окончены все входящие в него операции. Операция, выходящая из некоторого узла, не может начаться до тех пор, пока не будет достигнуто **начальное событие**, т.е. пока не будут завершены все операции, входящие в **узловое начальное событие**.

Если операция C не может быть начата до момента окончания работ A и B, логическую схему данной ситуации можно представить графически следующим образом (см. рис. 10.2).

Начальным событием для C является конечное событие для A и B. Существенно, что в стрелочном графе сохраняется логическая зависимость операций. Иногда, чтобы достичь этого, необходимо включить в граф одну или более **фиктивных логических операций**.

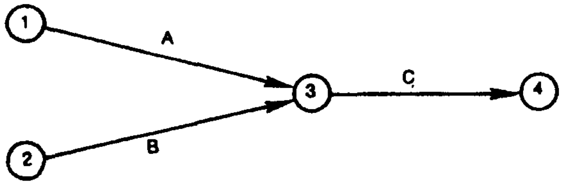


Рис. 10.2. Логические взаимосвязи в стрелочном графе

Фиктивная логическая стрелка вводится в граф, если необходимо отразить, что некоторое событие не может появиться раньше другого события, а с помощью обычных стрелок, соответствующих операциям, этого сделать нельзя. Функция фиктивной логической операции состоит в том, чтобы показать последовательность появления событий.

Фиктивным логическим операциям ставится в соответствие нулевая продолжительность выполнения, а изображаются они обычно пунктиром. Например, если работу C нельзя начать прежде, чем завершится операция A, а работу D нельзя начать до тех пор, пока не завершатся работы A и B, соответствующий стрелочный граф будет выглядеть следующим образом:

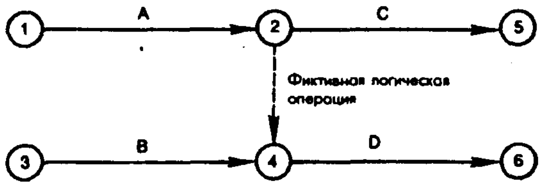


Рис. 10.3. Использование в стрелочном графе фиктивной логической операции

Кроме того, в стрелочных графах для избежания неоднозначности используются **фиктивные операции идентификации**. В некоторых пакетах прикладных программ, используемых в сетевом анализе, операции обозначаются не с помощью букв или слов, а числами, обозначающими соответствующие им события. Если же две или более операций выполняются одновременно и имеют одни и те же начальное и конечное события, то компьютер не сможет отличить их друг от друга и не воспримет вводимую исходную информацию. Как показано на рис. 10.4, включение фиктивной операции идентификации позволяет решить данную проблему. На практике принято нумеровать события таким образом, чтобы номер конечного события был больше, чем номер начального события.

Первый шаг после составления списка операций, входящих в проект, состоит в том, чтобы создать таблицу операций, в которой отражаются все операции, а также операции, непосредственно им предшествующие.

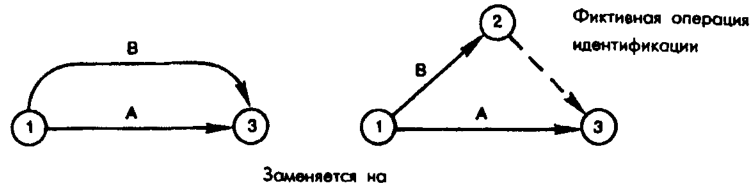


Рис. 10.4. Использование в стрелочном графе фиктивной операции идентификации

В данный список не включаются фиктивные логические операции или операции идентификации. На основе полученного списка строится стрелочный сетевой граф, включающий действительные и фиктивные операции и отражающий установленные взаимосвязи между ними. После того, как закончено построение исходного графа, можно выявить и исключить из рассмотрения ненужные фиктивные операции. Затем для улучшения логической схемы исходный граф можно модифицировать и перекомпоновать.

Ненужные фиктивные логические операции можно выявить с помощью простого практического правила. Если единственной операцией, выходящей из некоторого узла, является фиктивная логическая операция, то по всей вероятности без нее можно обойтись.

❏ **Пример 10.1.** Компания "Delco plc" — это промышленная фирма, которая заключила контракт о производстве партии станков, предназначеных к использованию крупным предприятием обувной промышленности для массового производства обуви. *Ниже перечислены операции, которые необходимо выполнить в* процессе разработки и производства этих станков (табл. 10.1):

Таблица 10.1. **Таблица операций для задачи из примера 10.1**

Операции		*Непосредственно предшествующая операция*
A	Составление сметы затрат	–
B	Согласованные оценки	A
C	Покупка собственного оборудования	B
D	Подготовка конструкторских проектов	B
E	Строительство основного цеха	D
F	Монтаж оборудования	C, E
G	Испытания оборудования	F
H	Определение типа модели	D
I	Проектирование внешнего корпуса	D
J	Создание внешнего корпуса	H, I
K	Конечная сборка	G, J
L	Контрольная проверка	K

Нужно изобразить операции с помощью стрелочного графа.

Решение.

Сетевой граф должен начинаться с единственного начального события, которое показано на рис. 10.5 кружочком, и заканчиваться единственным конечным событием. Построение графа мы начали с первого события. С этого события начинаются все операции, которым не предшествуют никакие виды работ. Начинать построение полезно с примерного эскиза будущего графа:

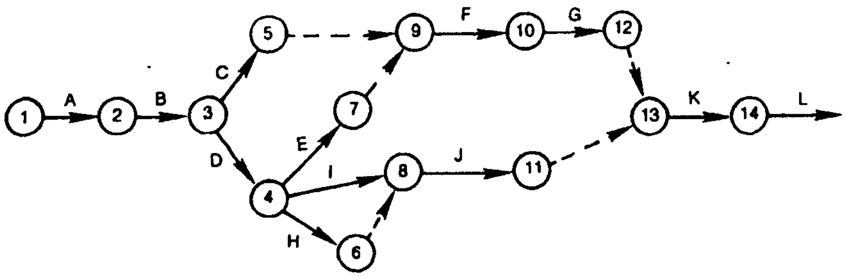


Рис. 10.5. Примерный эскиз графа для примера 10.1

В соответствии с приведенной выше таблицей необходимо тщательно, переходя от одной операции к другой, проверить построенный в первом приближении граф. В случае необходимости следует провести его корректировку, а затем для совершенствования схемы построить новый. В данном случае можно исключить все фиктивные логические операции и оставить одну фиктивную операцию идентификации (рис. 10.6).

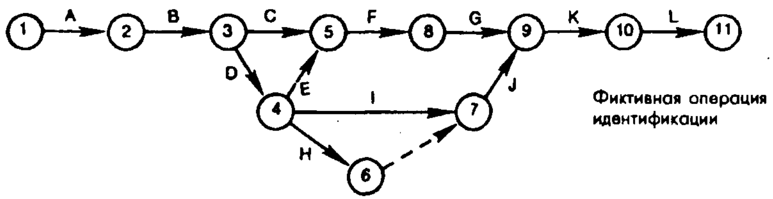


Рис. 10.6. Новый чертеж стрелочного графа для примера 10.1

❐ **Пример 10.2.** Компания "Delco plc" является участником другого проекта, детали которого приведены ниже:

Таблица 10.2. **Таблица операций для примера 10.2**

Операция	*Непосредственно предшествующая операция*	*Операция*	*Непосредственно предшествующая операция*
A	–	E	B, C
B	–	F	C
C	–	G	D, E
D	A, B	H	F, G

Изобразим данный проект при помощи стрелочного графа.

Решение

Построение начинаем с начального события, обозначенного кружком 1. Из таблицы следует, что существуют три операции — A, B и C, которым не предшествует ни одна из операций. Поэтому из начального события выходят три стрелки. На первый взгляд таблица операций выглядит чрезвычайно простой, однако отразить присущую ей логику с помощью сетевого графа достаточно трудно, вследствие чего мы вынуждены использовать три фиктивные логические операции (см. рис. 10.7).

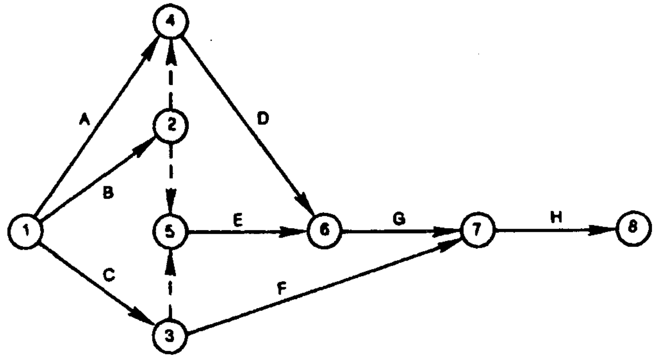


Рис.10.7. Стрелочный граф для примера 10.2

10.2.2. Вершинные графы

В этом типе сетевых графов операции представлены узлами графа, а стрелками изображаются их взаимосвязи. В таких графах не возникает необходимости вводить фиктивные операции. Как и в предыдущем случае, течение времени следует изображать в направлении слева направо.

❑ **Пример 10.3.** Обратившись к данным из примера **10.2**, модифицируем полученную в этом примере схему, поставив в соответствие операциям узлы графа.

Решение

Логическую схему, приведенную в данном примере, гораздо проще проиллюстрировать используя метод построения сетевых графов по схеме "операция–узел", однако с его помощью труднее получить общую картину переходов от одной операции к другой. Построение вершинного графа начинают с начального узла, за которым следуют первые три операции — А, В и С. Построить такой граф достаточно просто.

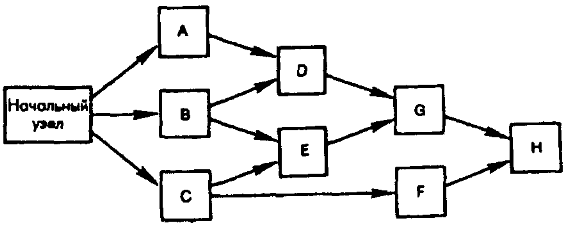


Рис. 10.8. Вершинный граф

Каждый из описанных типов графов имеет свои преимущества и недостатки. Обычно не имеет принципиального значения, какая из систем используется. Если в стрелочные графы приходится вводить достаточно большое число фиктивных операций, то гораздо более предпочтительным является выбор вершинного графа. Ниже приведено сравнение двух видов изображения операций и их основных особенностей (см. рис. 10.9).

10.3. АНАЛИЗ КРИТИЧЕСКОГО ПУТИ

После того как проведена идентификация операций, можно оценить их продолжительность. На основе продолжительности выполнения каждой операции и руководствуясь логической схемой, можно найти время выполнения проекта в целом. На данном этапе предполагается, что продолжительность выполнения каждой операции является фиксированной величиной, не испытывающей влияния неопределенности. В последнем разделе главы мы рассмотрим вопрос о том, какие поправки следует внести в этот анализ, чтобы учесть неопределенность времени выполнения операций. В каждом графе существует несколько возможных путей. Общее время, необходимое для того, чтобы пройти какой-либо путь, есть сумма времени выполнения всех операций, принадлежащих данному пути. Продолжительность выполнения всего проекта занимает наибольшее время. Более длительные операции называются **критическими**. Любая задержка срока начала или окончания выполнения этих работ повлечет за собой задержку срока выполнения

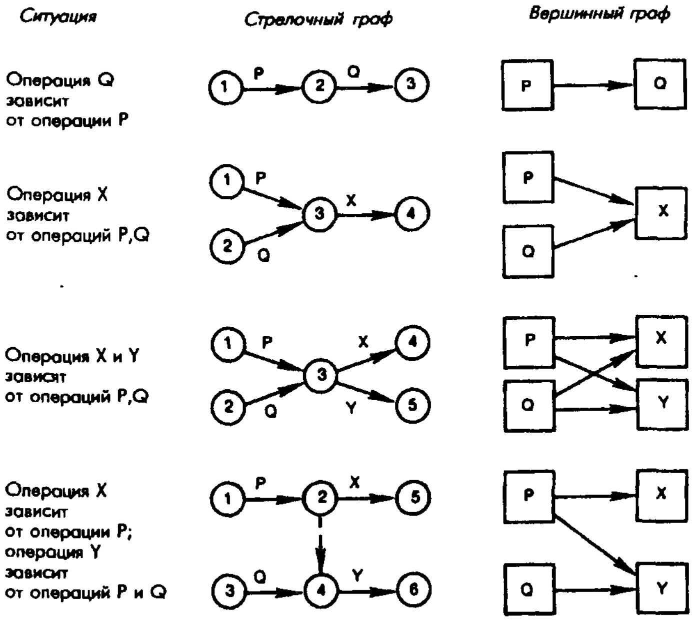


Рис. 10.9. Сравнение сетевых стрелочного и вершинного графов

проекта в целом. Критические операции образуют непрерывную цепь, проходящую через весь граф. Эта цепь критических операций называется **критическим путем**. В каждом графе найдется по крайней мере один критический путь.

Для того чтобы найти общую продолжительность выполнения проекта, нужно определить продолжительность критического пути. В большинстве графов идентифицировать все идущие сквозь граф пути, чтобы выявить среди них тот, который занимает наибольшее время, достаточно трудно. Существуют два возможных метода, позволяющих отследить движение времени в графе:

1. Определение для каждой операции наиболее ранних сроков начала и окончания ее выполнения.
2. Определение для каждого события наиболее раннего срока его наступления.

Следует отметить, что второй метод может использоваться только в стрелочных графах.

10.3.1. Анализ критического пути с применением вершинных графов

❒ **Пример 10.4.** В табл. 10.3 указана продолжительность выполнения каждой операции проекта, о котором шла речь в примерах **10.2** и **10.3**. Определим общую продолжительность выполнения проекта.

Таблица 10.3. **Операции и их продолжительность для примера 10.4**

Операция	*Непосредственно предшествующая операция*	*Время, дней*
A	–	8
B	–	10
C	–	6
D	A, B	8
E	B, C	9
F	C	14
G	D, E	14
H	F, G	6

Вершинный граф, соответствующий данному проекту, был построен в примере **10.3.**

Решение

Предположим, что каждая из исходных операций A, B и C начинается в нулевой момент времени. Это наиболее ранний срок начала этих ES операций. Наиболее ранний срок, к которому их выполнение может быть завершено, определяется следующим образом:

Наиболее ранний срок окончания EF=ES+ Продолжительность операции.

Обычно найденные значения этих сроков наносятся непосредственно на граф, однако, мы занесем их сначала в таблицу, чтобы продемонстрировать методику проведения расчетов.

Таблица 10.4. **Расчет наиболее ранних сроков начала и окончания операций для примера 10.4**

Операция	*Продолжительность, дней*	*Наиболее ранний срок начала*	*Наиболее ранний срок окончания*	*Комментарий*
A	8	0	0 + 8 = 8	
B	10	0	0 + 10 = 10	
C	6	0	0 + 6 = 6	
D	8	10	10 + 8 = 18	Нельзя начать, пока не завершены A и B
E	9	10	10 + 9 = 19	Нельзя начать, пока не завершены B и C
F	14	6	14 + 6 = 20	Нельзя начать, пока не завершена C
G	14	19	14 + 19 = 33	Нельзя начать, пока не завершены D и E
H	6	33	33 + 6 = 39	Нельзя начать, пока не завершены F и G

Наиболее ранние сроки начала и окончания операций занесены в вершинный граф, изображенный на рис. 10.10. Нетрудно заметить, что операция H завершится на 39-й день, следовательно, это значение дает нам искомую продолжительность выполнения проекта в целом.

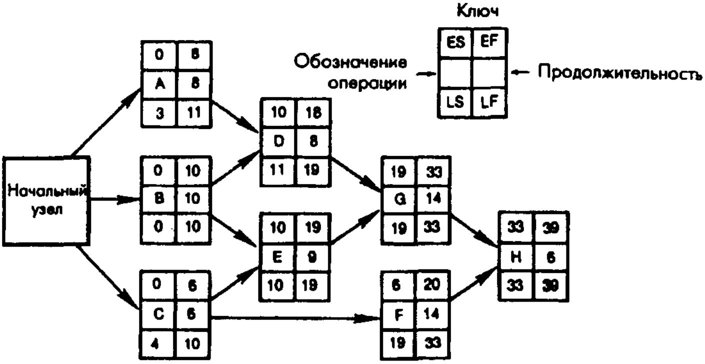


Рис.10.10 Вершинный граф для примера 10.4

На данном этапе мы еще не можем определить критические операции. Чтобы это осуществить, необходимо для каждой операции рассчитать два срока, ей соответствующие, а именно **наиболее поздний срок начала** LS и **наиболее поздний срок окончания** LF операции. В данном случае процедуру расчетов мы начнем с последней операции в графе и предположим, что наиболее поздний и наиболее ранний сроки ее окончания совпадают. Затем вычитанием из этой величины продолжительности выполнения операций находим наиболее поздний срок ее начала. Ход выполнения расчетов показан в табл. 10.5.

Таблица 10.5. **Расчет наиболее поздних сроков начала и окончания операций для примера 10.4**

Операция	*Продолжительность, дней*	*Наиболее поздний срок окончания*	*Наиболее поздний срок начала*	*Комментарии*
1	2	3	4	5
H	6	39	39 − 6 = 33	
G	14	33	33 − 14 = 19	G нужно завершить до наступления наиболее позднего срока начала H
F	14	33	33 − 14 = 19	F нужно завершить до наступления наиболее позднего срока начала H
E	9	19	19 − 9 = 10	E нужно завершить до наступления наиболее позднего срока начала G

Продолжение

1	2	3	4	5
D	8	19	19 − 8 = 11	D нужно завершить до наступления наиболее позднего срока начала G
C	6	10	10 − 6 = 4	C нужно завершить до наступления наиболее позднего срока начала E и F. Нужно использовать наименьший из этих сроков, равный 10 дням.
B	10	10	10 − 10 = 0	B нужно завершить до наступления наиболее позднего срока начала D и E. Нужно использовать наименьший из этих сроков, равный 10 дням.
A	8	11	11 − 8 = 3	A нужно завершить до наступления наиболее позднего срока начала D.

Критической является операция, для которой справедливы следующие соотношения:

$$ES = LS \quad \text{и} \quad EF = LF,$$

т. е. операция, для которой не существует резерва времени между наиболее ранним сроком ее начала и наиболее поздним сроком ее окончания. Нетрудно заметить, что в нашем примере критическими являются операции B, E, G и H. Путь в вершинном графе, соединяющий эти операции, называется критическим путем. В нашем примере критическим является путь B—E—G—H.

10.3.2. Анализ критического пути с применением стрелочных графов

Приведенная выше методика анализа аналогичным образом может использоваться и для стрелочных графов. Значения сроков ES, EF, LS и LF записываются в графе вдоль стрелок, соответствующих операциям:

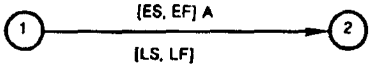


Рис. 10.11. Нанесение на стрелочный граф сроков, соответствующих операциям

Можно провести подобный анализ в терминах сроков наступления каждого события. Производится расчет наиболее раннего срока, к которому может завершиться каждое событие. Этот срок называется **наиболее ранним сроком события** (earliest event time – EET). Общая продолжительность выполнения проекта определяется EET конечного узла графа. EET исходного события равен нулю.

Для того чтобы выявить критические операции, необходимо, начиная с конца графа, вычислить **наиболее поздние сроки событий** (latest event time – LET), к которым события могут закончиться. События, для которых выполняются соотношения

$$\text{LET}_{\text{начала}} - \text{EET}_{\text{окончания}} + \text{продолжительность} = 0$$

или

$$\text{EET}_{\text{начала}} - \text{LET}_{\text{окончания}} + \text{продолжительность} = 0,$$

являются критическими.

❐ **Пример 10.5.** Применив EET и LET, повторим задачу из примера **10.4** при условии, что продолжительность выполнения фиктивных операций равна нулю.

Решение

В первую очередь для каждого события вычислим значение наиболее раннего срока. Если некоторому событию соответствует более одной операции, появляется проблема выбора соответствующего значения. Поскольку событие считается незавершенным до тех пор, пока не будет завершено выполнение всех составляющих его операций, следует выбрать наибольшее из значений.

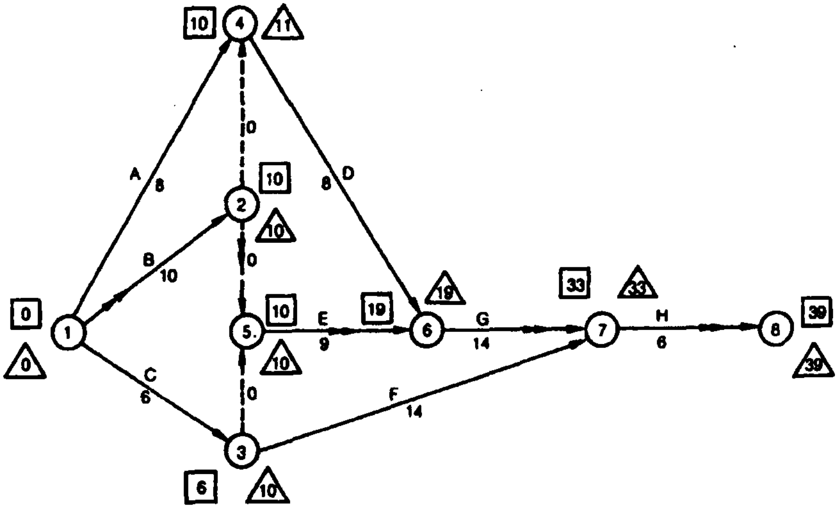


Рис. 10.12. Стрелочный граф для примера 10.5 с указанием EET и событий

☐ – наиболее ранний срок события, △ – наиболее поздний срок события (стандартный срок, дней)

Полученные значения сроков наносятся на стрелочный граф, как это показано на рис. 10.12.

EET последнего события равно 39 дням, которые также определяют общую продолжительность выполнения проекта.

Таблица 10.6. **Расчет значений EET для примера 10.5**

Узел	*EET, дней*	*Комментарии*
1	0	Начальное событие
2	0 + 10 = 10	EET узла 1 + продолжительность операции B
3	0 + 6 = 6	EET узла 1 + продолжительность операции C
4	0 + 8 = 8 или 10 + 0 = 10*	EET узла 1 + продолжительность операции A EET узла 2 + продолжительность фиктивной операции Выбирается максимальный срок, т. е. 10 дней
5	10 + 0 = 10* или 6 + 0 = 6	EET узла 2 + продолжительность фиктивной операции EET узла 3 + продолжительность фиктивной операции Выбирается максимальный срок, т. е. 10 дней
6	10 + 8 = 18 или 10 + 9 = 19*	EET узла 4 + продолжительность операции D EET узла 5 + продолжительность операции E Выбирается максимальный срок, т. е. 19 дней
7	19 + 14 = 33* или 6 + 14 = 20	EET узла 6 + продолжительность операции G EET узла 3 + продолжительность операции F Выбирается максимальный срок, т. е. 33 дня
8	33 + 6 = 39	EET узла 7 + продолжительность операции H

* Выбранное значение EET

Чтобы определить критические операции, будем двигаться по графу начиная с конечного узла и вычисляя LET каждого события. Предположим, что для конечного события EET = LET. Если в некоторый узел входит более одной стрелки, то возникает проблема выбора значения LET. Так как событие должно завершиться к сроку, удовлетворяющему всем наиболее поздним срокам начала событий, которые выходят из данного узла для LET, следует выбрать наименьшее значение.

Найденные значения сроков наносятся на стрелочный граф, изображенный на рис. 10.12.

Операция является критической, если для нее справедливы следующие соотношения:

$$EET_{\text{начала}} = LET_{\text{начала}}$$

и

$$EET_{\text{окончания}} = LET_{\text{окончания}}$$

$$LET_{\text{окончания}} - EET_{\text{начала}} - \text{Продолжительность} = 0.$$

Из рисунка 10.12 видно, что критическими, как и ранее, являются операции B, E, G и H.

Таблица 10.7. **Расчет значений LET для примера 10.5**

Узел	*LET, дней*	*Комментарий*
8	39	Конечный узел LET = EET
7	39 − 6 = 33	LET узла 8 − продолжительность операции H
6	33 − 14 = 19	LET узла 7 − продолжительность операции G
5	19 − 9 = 10	LET узла 6 − продолжительность операции E
4	19 − 8 = 11	LET узла 6 − продолжительность операции D
3	10 − 0 = 10* или 33 − 14 = 19	LET узла 5 − продолжительность фиктивной операции или LET узла 7 − продолжительность операции F Выбирается минимальный срок, т.е. 10 дней
2	10 − 0 = 10* или 11 − 0 = 11	LET узла 5 − продолжительность фиктивной операции или LET узла 4 − продолжительность фиктивной операции Выбирается минимальный срок, т.е. 10 дней
1	11 − 8 = 3 или 10 − 10 = 0* или 10 − 6 = 4	LET узла 4 − продолжительность операции A или LET узла 2 − продолжительность операции B или LET узла 3 − продолжительность операции C Выбирается минимальный срок, т.е. 0 дней

* Выбранное значение LET.

Любые замедления на критическом пути приведут к задержке срока выполнения всего проекта. Между тем для некритических путей можно допустить некоторые задержки при выполнении составляющих их операций или пересмотреть график их выполнения. Запас времени, который существует в схеме проекта, называется **резервом времени**. Различают несколько видов резерва времени, возникающих под влиянием различных воздействий, которые оказывает запас времени на схему выполнения проекта. **Общим резервом** называется количество времени, на которое можно увеличить продолжительность операции в результате продления срока ее выполнения или пересмотра плана, не влияющего на продолжительность выполнения проекта в целом. **Свободным резервом** называется количество времени, на которое можно увеличить продолжительность операции в результате продления срока ее выполнения или пересмотра плана, не оказывающего воздействия на наиболее ранний срок выполнения любой последующей операции. Иногда используют третий вид, так называемый **независимый резерв времени**. Он не оказывает никакого влияния на предшествующие или последующие операции. Для любой операции

$$\text{Общий резерв времени} = LET_{\text{окончания}} - EET_{\text{начала}} - \text{Продолжительность},$$

а также

$$\text{Свободный резерв времени} = EET_{\text{окончания}} - EET_{\text{начала}} - \text{Продолжительность}$$

и

$$\text{Независимый резерв} = EET_{\text{окончания}} - LET_{\text{начала}} - \text{Продолжительность}.$$

Иногда бывает полезно изобразить на графе имеющийся в наличии резерв времени, особенно если план выполнения операций необходимо пересмотреть. В этом случае одним из возможных методов является **график Ганта.**

❐ **Пример 10.6.** По данным примера **10.5** для каждой операции найдем общий резерв времени.

Таблица 10.8. **Расчет резерва времени операций для примера 10.5 (дней)**

Операция	*LET окончания*	*EET начала*	*Продолжительность*	*Общий резерв времени*
A	11	0	8	3
B	10	0	10	0
C	10	0	6	4
D	19	10	8	1
E	19	10	9	0
F	33	6	14	13
G	33	19	14	0
H	39	33	6	0

Операции, общий резерв времени которых равен нулю, являются критическими. На рис. 10.13 построен график Ганта, и отмечены возможно наиболее ранние сроки начала операций.

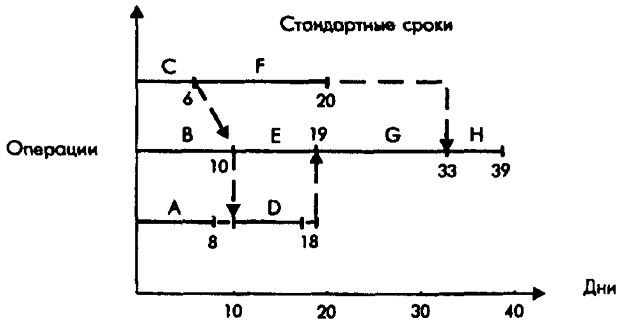


Рис. 10.13. График Ганта для примера 10.5

10.4. СТОИМОСТЬ ПРОЕКТА

Общая стоимость проекта зависит от стоимости выполнения каждой операции, а также от любых дополнительных переменных или постоянных расходов. Так как необходимо завершить все операции, независимо от того, являются они критическими или нет, общая стоимость выполнения операций представляет собой арифметическую сумму отдельных значений стоимости каждой операции.

Можно снижать продолжительность выполнения некоторых операций с помощью дополнительных ресурсов. Косвенным последствием такой меры является увеличение стоимости данных операций. Однако если операция критическая, то экономия времени ее выполнения может привести к общей экономии времени выполнения проекта в целом, а следовательно, и к снижению общей стоимости проекта.

Наименьший возможный срок, к которому можно завершить операцию, получил название **критического срока.** В некоторых случаях завершить операцию можно только либо к стандартному, либо к критическому сроку их выполнения, но не между ними. Иногда, напротив, существует возможность постепенно уменьшать время выполнения операции до того момента, пока не будет достигнут критический срок ее выполнения. Рассмотрим, как действует уменьшение времени выполнения операций на календарный план, и стоимость выполнения проекта, принимая во внимание две различные цели:

1. Минимизацию общего времени выполнения проекта;
2. Минимизацию общей стоимости проекта.

10.4.1. Минимизация общей продолжительности проекта с минимальными дополнительными расходами

Для этой цели необходимо обладать информацией о стоимости каждой операции, любом возможном уменьшении времени ее выполнения и о дополнительных издержках, связанных со снижением времени выполнения операции.

❒ **Пример 10.7.** Обратимся к данным примера **10.2.** Ниже приводится дополнительная информация о стоимости операций и возможном уменьшении времени их выполнения.

Таблица 10.9. **Значения стандартных и критических сроков и соответствующих издержек выполнения операций для примера 10.7**

Операция	*Непосредственно предшествующие операции*	*Стандартное значение*		*Критическое значение*	
		времени, дней	*стоимости, ф. ст.*	*времени, дней*	*стоимости, ф. ст.*
A	–	8	7500	4	9000
B	–	10	8500	8	11000
C	–	6	6000	5	7000
D	A, B	8	13000	5	16000
E	B, C	9	14000	6	16500
F	C	14	14500	11	18000
G	D, E	14	13500	10	18750
H	F, G	6	5500	4	6500
Общие издержки выполнения операций			82500		102750

Показатели критических значений отражают минимальное время, за которое можно выполнить операцию, и общую стоимость выполнения операции в течение этого времени. Необходимо сделать выбор между стандартными значениями времени и издержек и их критическими значениями. Практически невозможно получить экономию времени выполнения операции в один день при пропорциональном возрастании ее стоимости. Помимо стоимости каждой операции необходимо учесть стоимость строительной площадки, составляющую 1000 ф. ст. в день.

1. Каково минимальное время, в течение которого можно завершить проект?
2. Какова соответствующая минимальная дополнительная стоимость?

Решение

Минимальное время можно найти, рассчитав для всех, как критических, так и некритических операций, критическое время их выполнения. Ниже изображен стрелочный граф, построенный в примере **10.2**. На граф нанесены значения EET и LET, найденные на основе критических значений времени выполнения операций.

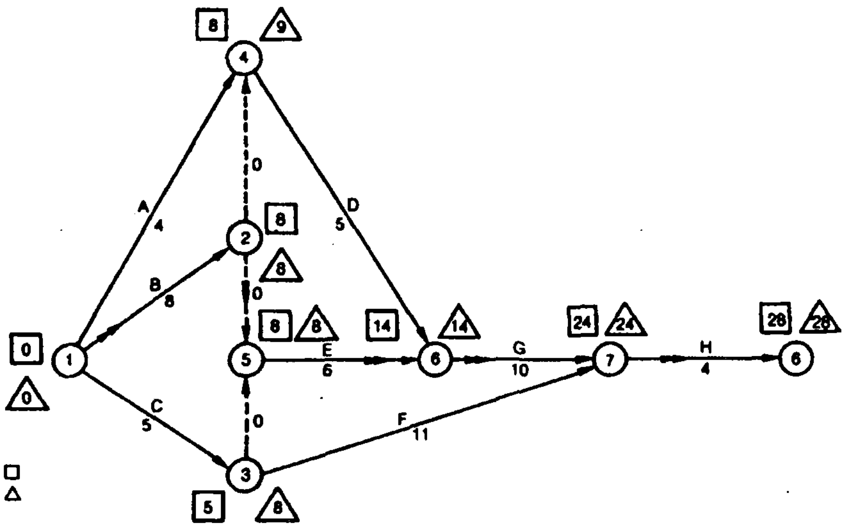


Рис. 10.14. Стрелочный граф для примера 10.7 с указанием критического времени

□ – наиболее ранний срок события, дней

△ – наиболее поздний срок события, дней

Нетрудно заметить, что EET узла 8 равно 28 дням, поэтому минимальное время выполнения проекта также составляет 28 дней. Критический путь остается неизменным: B – E – G – H.

Общую стоимость можно найти из следующего уравнения:

Общая стоимость = Критическая стоимость операций + 28 ×
× Стоимость строительной площадки в день = 102 750 ф.ст. + 28 × 1000 ф. ст. =
= 130750 ф.ст.

Между тем найденное значение стоимости выполнения проекта в указанное время не является минимальным, поскольку необходимости использовать критические значения для некритических операций нет. Некритическими являются операции A, C, D и F. Поэтому необходимо найти эффект от использования соответствующих этим операциям некритических значений показателей. В случае, если существует возможность восстановить их стандартную продолжительность, не увеличивая при этом общую продолжительность выполнения проекта, можно будет одновременно достичь и экономию стоимости в результате использования ее некритических значений.

Критические значения для операции A можно не использовать, поскольку увеличение продолжительности ее выполнения до 8 дней не меняет EET узла 4, и, следовательно, не оказывает воздействия на выполнение остальных операций календарного плана. Использование некритических значений для операции A позволяет достичь экономии, составляющей 1500 ф.ст.

Увеличение продолжительности операции C с 5 до 6 дней приведет к увеличению значения EET узла 3 до 6, однако не окажет воздействия на EET узлов 5 и 7. Вследствие этого продолжительность выполнения проекта останется неизменной. Использование некритических значений, соответствующих операции C, позволит получить экономию 1000 ф.ст.

Если использовать некритические значения показателей операции D, то EET узла 6 возрастет до 16 дней. Узел 6 принадлежит критическому пути, поэтому для того, чтобы достичь минимального общего времени выполнения проекта, составляющего 28 дней, необходимо применять критические значения времени и стоимости операции D.

Использование некритических значений для операции F не изменит EET узла 7 и не приведет к увеличению продолжительности проекта в целом. Используя для F некритические значения, мы сможем достичь экономии, составляющей 3500 ф. ст.

Минимальная стоимость выполнения проекта за 28 дней составит:

$$130750 - 1500\ (A) - 1000\ (C) - 3500\ (F) = 124750 \text{ ф.ст.}.$$

Стоимость выполнения проекта в стандартные сроки равна:

82500 (стоимость операций) + 39000 (стоимость строительной площадки) = 121500 ф.ст.

Следовательно, дополнительная стоимость, связанная с завершением выполнения проекта на 11 дней раньше, будет равна:

$$124750 - 121500 = 3250 \text{ ф. ст.}$$

❑ **Пример 10.8.** *Обратимся к данным примера* **10.1**. В табл. 10.10 приводится дополнительная информация о стоимости операций и возможном сокращении времени их выполнения.

Таблица 10.10. **Стандартные и критические значения сроков выполнения и стоимости операций для примера 10.8**

Операция	*Стандартное значение*		*Возможное сокращение времени, недель*	*Критическое время, недель*	*Дополнительные издержки сокращения времени на неделю, ф. ст.*
	времени, недель	*стоимость, ф. ст.*			
A	2	400	1	1	400
B	1	0	0	1	0
C	4	200	2	2	125
D	6	450	4	2	175
E	3	700	2	1	250
F	3	200	2	1	200
G	4	600	3	1	125
H	2	0	0	2	0
I	3	250	1	2	200
J	8	600	4	4	100
K	2	450	1	1	250
L	2	200	1	1	150
Стоимость операций		4050			

Переменные накладные расходы составляют 300 ф. ст. в неделю в течение всего времени выполнения проекта.

1. Определить стандартные значения общего времени выполнения и общей стоимости проекта.
2. Найти минимальное время, за которое можно выполнить данный проект, и соответствующее ему минимальное значение стоимости.

Решение

На рис. 10.15 воспроизведен стрелочный граф, построенный в примере **10.1.** Для каждой операции на графе указаны значения EET и LET. Стандартный срок выполнения проекта составляет 24 недели, а соответствующий ему критический путь имеет следующий вид:

A – B – D – I – J – K – L.

Общая стоимость проекта составляет:

4050 (стоимость операций) + 24×300 (переменные накладные расходы) =
= 11250 ф. ст.

Чтобы определить минимальное время, требующееся для выполнения проекта в целом, каждой операции поставим в соответствие минимальный срок ее завершения. На рис.10.16 показаны значения этих сроков и итоговые значения EET и LET.

Минимальная продолжительность проекта составляет 12 недель. В данном случае критическими оказываются следующие пути:

A – B – D – I – J – K – L.

и

A – B – D – H – J – K – L.

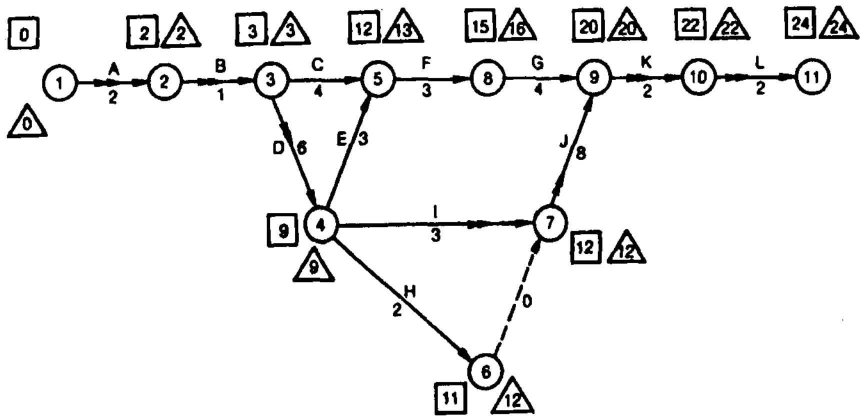


Рис. 10.15. Стрелочный граф для примера 10.8 с указанием стандартных сроков

☐ – наиболее ранний срок события, △ – наиболее поздний срок события (стандартный срок, дней)

Проверим, можно ли, используя некритические значения для некоторых некритических операций, получить экономию денежных средств.

Некритическими являются операции C, E, F и G. Продолжительность операций в данном примере можно изменять по интервалам в одну неделю, так как единицей измерения продолжительности является неделя. В первую очередь рассмотрим операции, которые, если использовать их некритические значения, могут принести наибольшую экономию денежных средств. Операции будем рассматривать в следующем порядке: E (250 ф.ст.), F (200 ф.ст.), и C (125 ф.ст.) или G (125 ф.ст.).

Минимальная стоимость выполнения проекта за 12 недель составила:
(4050 (*стандартная стоимость операций*) + 1 · 400 (A) + 4 · 175 (D) + 1 · 200 (F) + 3 · 125 (G) + 1 · 200 (I) + 4 · 100 (J) + 1 · 250 (K) + 1 · 150 (L) (предельные издержки) + 12 · 300 (переменные накладные расходы) = 4050 + 2675 + 3600 = 10325 ф.ст.

Таблица 10.11. **Использование некритических значений показателей для некритических операций из примера 10.8**

Операция	*Изменение продолжительности*	*Эффект*
E	Увеличение на 2 недели	EET узла 5 становится равным 7 неделям; EET узла 8 становится равным 8 неделям, не влияя при этом на EET узла 9, принадлежащего критическому пути; других воздействий нет. E выполняется в стандартный срок
F	Увеличение на 1 неделю	EET узла 8 становится равным 9 неделям. Операции E, F и G становится критическими.
G	Увеличение невозможно	
C	Увеличение на 2 недели	На узел 5 не оказывается никакого воздействия. C выполняется в стандартный срок.

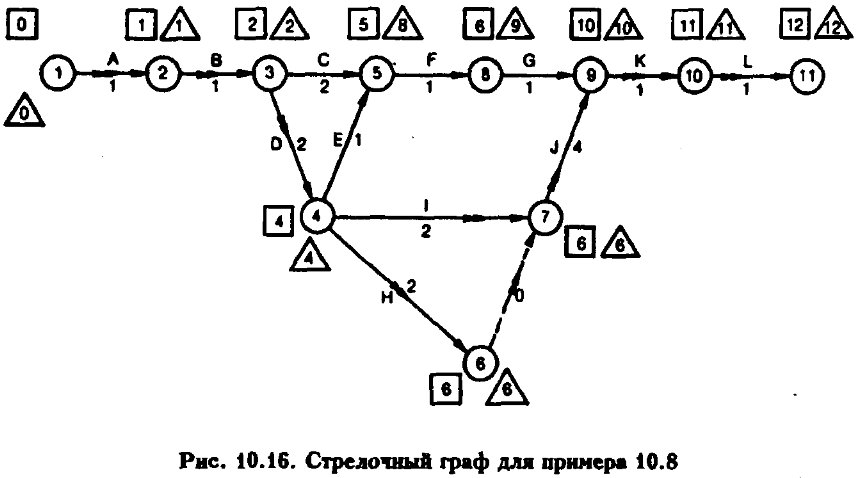


Рис. 10.16. Стрелочный граф для примера 10.8 с указанием критических сроков

□ – наиболее ранний срок события, △ – наиболее поздний срок события (стандартный срок, дней)

10.4.2. Выполнение проекта с минимальными издержками

Если выполнение проекта требует оплаты переменных накладных расходов, таких, например, как расходы, связанные с оборудованием строительной площадки, то может оказаться выгодным снижение продолжительности выполнения проекта. Поскольку сами эти сокращения влекут за собой определенные издержки, необходимо подвести баланс. Экономия времени может быть достигнута только в том случае, если сократить продолжительность критических операций. Критические значения должны использоваться только по тем критическим операциям, по которым величина экономии накладных расходов превышает стоимость выполнения операции за критическое время.

❑ **Пример 10.9.** Обратившись к данным примера **10.7**, определим минимальную стоимость проекта и соответствующее время его выполнения. Предполагается, что операции можно выполнять либо в стандартные, либо в критические сроки, но не в промежутке между ними.

Решение

Используя граф, построенный в примере **10.5** для стандартных сроков выполнения операций, перечислим все критические операции и соответствующие им показатели максимально возможной экономии времени и чистой экономии стоимости.

Таблица 10.12. **Расчет минимальной стоимости проекта для примера 10.9**

Операция	*Число дней экономии для критического времени*	*Дополнительная стоимость критического времени, ф. ст.*	*Экономия, ф. ст.*	*Чистая экономия, ф. ст.*	*Комментарии*
B	2	2500	2 · 1000	−500	Критические значения не используются
Либо E	1*	2500	1 · 1000	−1500	Критические значения не используются
Либо E и D	3	5500	3 · 1000	−2500	Критические значения не используются
G	4	5250	4 · 1000	−1250	Критические значения не используются
H	2	1000	2 · 1000	+ 1000	Используются критические значения. Снижение продолжительности проекта с 39 до 37 дней

* Достичь экономии, равной 3 дням, нельзя, поскольку в этом случае путь A–D–G–H становится критическим. Поэтому общая продолжительность снижается только на один день. Если же использовать критические значения одновременно для E и D, достигается экономия времени, равная 3 дням. Однако соответствующая стоимость становится равной: 2500 ф.ст. + 3000 ф.ст., т.е. такая экономия времени не целесообразна.

Минимальная стоимость проекта равна: 121500 − 1000 = 120500 ф.ст. Соответствующее время его выполнения составляет 37 дней.

10.5. НЕОПРЕДЕЛЕННОСТЬ ВРЕМЕНИ ВЫПОЛНЕНИЯ ОПЕРАЦИЙ

В приведенных выше методах анализа предполагалось, что время выполнения операций точно известно. Однако на практике сроки выполнения операций обычно являются довольно неопределенными. Управляющий производством может выдвинуть некоторые предположения о том, сколько времени потребуется для выполнения каждой работы, но не может предусмотреть возможные трудности или задержки выполнения. Неопределенность сроков выполнения операций означает, что общая продолжительность проекта также подвержена неопределенности.

Выбор метода, позволяющего учесть эту неопределенность, зависит от типа проекта и природы неопределенности. Если можно определить минимальную и максимальную продолжительности каждой операции, то их рассчитывают с помощью показателей ожидаемой (средней) продолжительности и ожидаемого времени выполнения проекта. Алгоритм, получивший наиболее широкое применение, называется **методом оценки и пересмотра проектов (Project Evaluation and Reiew Technique – PERT)**. При вычислении ожидаемого времени выполнения проекта методом PERT используются показатели ожидаемого времени выполнения операций. Оставшаяся часть алгоритма аналогична описанным выше алгоритмам, применяемым в случаях, когда время выполнения операций является фиксированной величиной.

Если время выполнения операций подвержено влиянию неопределенности, то большое значение приобретают некритические пути в графе, когда могут изменяться сроки выполнения всех операций. На практике может оказаться, что путь, который на основе ожидаемых значений сроков считался некритическим, становится критическим в соответствии с результатами метода определения критического пути.

В основу метода PERT положена предпосылка о проведении продолжительности операции. Предполагается, что время выполнения каждой отдельно взятой операции аппроксимируется β-распределением. Если это верно, то распределение времени выполнения проекта в целом является нормальным. Метод PERT может применяться при анализе конкретного проекта только в случае выполнения данной предпосылки. График β-распределения изображен на рис. 10.17. Наименьшее возможное время выполнения операции называют **оптимистическим сроком** (a), а наибольшее возможное время ее выполнения – **пессимистическим сроком** (b).

Рис. 10.17. Стандартное β-распределение для времени выполнения операций

Пику распределения соответствует наиболее вероятное время выполнения операции (m). Необходимо произвести оценку каждого из этих трех сроков для всех операций, входящих в граф.

Исходя из этих трех значений можно найти ожидаемую продолжительность операции (t) и ее дисперсию. Ожидаемая продолжительность операции определяется следующим образом:

$$t = \frac{a + 4m + b}{6}.$$

Соответствующая дисперсия ожидаемой продолжительности определяется по формуле:

$$\sigma_t^2 = \left(\frac{b - a}{6}\right)^2.$$

Время выполнения проекта можно найти непосредственно из графа, используя для этого ожидаемые значения продолжительности операций. Предполагается, что время выполнения проекта в целом распределено по нормальному закону.

В предположении, что сроки выполнения операций не зависят друг от друга, среднее значение нормального распределения определяется как сумма математических ожиданий продолжительности критических операций, а дисперсия — как сумма их дисперсий. Полученное нормальное распределение можно использовать для оценки вероятности завершения проекта к заранее установленной дате.

Алгоритм метода PERT аналогичен анализу сетевого графа с фиксированными значениями продолжительности операций.

1. Составить список всех операций, входящих в проект, с указанием непосредственно предшествующих операций, а также оптимистического, наиболее вероятного и пессимистического сроков их выполнения.
2. Построить сетевой граф.
3. В предположении, что время выполнения любой операции аппроксимируется β-распределением, оценить для каждой операции ожидаемое время ее выполнения и его дисперсию.
4. Используя ожидаемые значения сроков выполнения операций, найти продолжительность проекта в целом.
5. Определить критические операции и критический путь.
6. С помощью значений дисперсии для критических операций оценить дисперсию ожидаемой продолжительности всего проекта.

❑ **Пример 10.10.** Процесс создания и серийного производства нового вида продукта компаний "ABC" включает в себя следующие операции (см. *табл.* 10.13).

1. Определим ожидаемое число недель, необходимое для выполнения проекта. Какие операции являются критическими?
2. Какова вероятность того, что выполнение проекта займет *более* 16 недель?

Таблица 10.13. **Таблица операций и сроков их выполнения для примера 10.10**

Операция	Непосредственно, предшествующие операции	Сроки выполнения операций, недель		
		оптимистический, a	наиболее вероятный, m	пессимистический, b
A	—	1,5	2	2,5
B	A	2	2,5	6
C	—	1	2	3
D	C	1,5	2	2,5
E	B, D	0,5	1	1,5
F	E	1	2	3
G	B, D	3	3,5	7
H	G	3	4	5
I	F, H	1,5	2	2,5

Решение

Ожидаемые сроки выполнения операций и соответствующие дисперсии имеют следующие значения:

Таблица 10.14. **Расчет ожидаемых сроков выполнения операций и их дисперсий по данным примера 10.10**

Операция	Ожидаемый срок выполнения, недель	Дисперсия, недель^2
A	$\frac{1,5+4\cdot 2+2,5}{6}=2$	$\left(\frac{2,5-1,5}{6}\right)^2=1/36$
B	$\frac{2+4\cdot 2,5+6}{6}=3$	$\left(\frac{6-2}{6}\right)^2=16/36$
C	$\frac{1+4\cdot 2+3}{6}=2$	$\left(\frac{3-1}{6}\right)^2=4/36$
D	2	$=1/36$
E	1	$=1/36$
F	2	$=4/36$
G	4	$=16/36$
H	4	$=4/36$
I	2	$=1/36$

Ниже приведен сетевой граф с указанием ожидаемой продолжительности каждой операции (см. рис. 10.18).

Расчет ожидаемого срока выполнения проекта в целом производится обычным способом. Как показано на рис. 10.18, выполнение проекта предполагается осуществить за 15 недель. Критическими являются операции A, B,G, H и I. Приведем для сравнения другие возможные пути в графе:

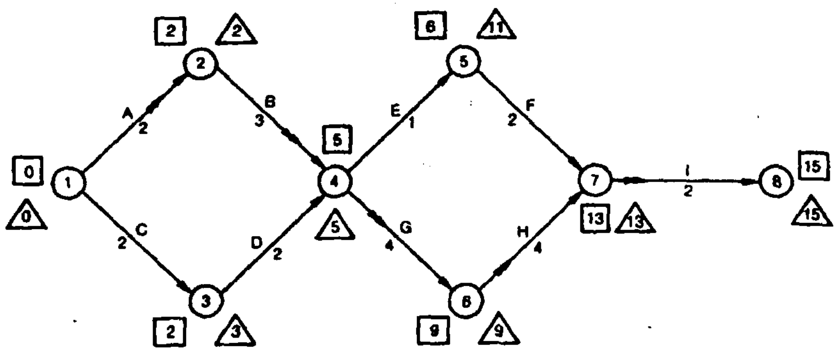


Рис. 10.18. Стрелочный граф с указанием ожидаемых сроков выполнения операций для примера 10.10

□ – наиболее ранний срок события

△ – наиболее поздний срок события (ожидаемые сроки, недель)

A, B, E, F, I	— занимает 10 недель,
C, D, E, F, I	— занимает 9 недель,
C, D, G, H, I	— занимает 14 недель.

Следует отметить, что путь — C,D,G,H, I — занимает время, которое меньше выполнения критического пути всего на одну неделю. Поэтому небольшие изменения времени выполнения некоторых операций могут привести к изменению критического пути.

Дисперсия ожидаемого времени выполнения всего проекта определяется как сумма дисперсий критических операций:

$$\sigma^2 = \sigma_A^2 + \sigma_B^2 + \sigma_G^2 + \sigma_H^2 + \sigma_I^2 ,$$

следовательно,

$$\sigma^2 = 1/36 + 16/36 + 16/36 + 4/36 + 1/36 = 38/36 = 1{,}11 \text{ недель}^2.$$

Стандартное отклонение времени выполнения проекта составит:

$$\sqrt{1{,}11} = 1{,}03 \text{ недель}.$$

Вероятность того, что выполнение проекта займет более 16 недель, можно найти следующим образом:

Шестнадцать недель составляют z стандартных отклонений от среднего, где:

$$z = \frac{16 - 15}{1{,}03} = 0{,}97.$$

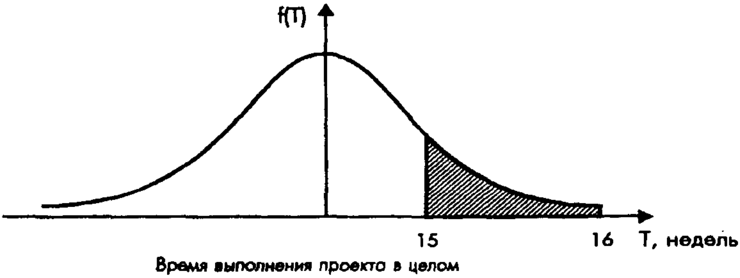


Рис. 10.19. Распределение времени выполнения проекта для примера 10.10

По таблице стандартного нормального распределения находим:

$$P\ (z \geq 0{,}97) = 0{,}166.$$

Следовательно, вероятность того, что выполнение проекта займет более 16 недель, равна 16,6%.

10.6. РАСПРЕДЕЛЕНИЕ РЕСУРСОВ

Сетевой граф отражает логическую последовательность выполнения операций, входящих в проект. Ни в одном из видов анализа, рассмотренных нами выше, не принимались во внимание какие-либо ограничения на обеспечение ресурсами. Исходный календарный план выполнения операций составлялся при условии, что все необходимые ресурсы имеются в достаточном количестве. Однако такая ситуация имеет место далеко не всегда, а если это так, то использование ресурсов в соответствии с потребностями, указанными в исходном календарном плане, может оказаться неэкономичным.

Метод составления календарного плана с учетом обеспечения ресурсами зависит от конкретных целей лиц, осуществляющих контроль за ходом выполнения проекта. Например, вопросом первостепенной важности может оказаться завершение проекта к определенному сроку безотносительно к затратам ресурсов — такие планы ограничены по времени. И наоборот, в условиях ограниченности в денежных средствах на выполнение проекта отводится определенное количество ресурсов, тогда как срок выполнения не принимается в расчет — такие планы ограничены по ресурсам. В данном контексте к ресурсам можно отнести рабочую силу, оборудование, сырье, денежные средства, производственные площади и т.д.

Перед тем как приступить к выполнение проекта, управляющий производством должен четко сформулировать критерий, в соответствии с которым будет осуществляться распределение ресурсов. В качестве такого критерия можно выбрать:

1. Максимальное использование ресурсов. Оценить использование ресурсов можно через соответствующий коэффициент:

$$\text{Коэффициент использования} = \frac{\text{Общее количество используемых ресурсов}}{\text{Общее количество наличных ресурсов}}.$$

2. Минимизацию максимальных потребностей в ресурсах.
3. Минимизацию максимальных изменений потребностей в ресурсах.

Кроме названных, существует множество других критериев.

Существует также множество возможных методов решения проблемы распределения ресурсов, таких, как, например, эвристические методы, методы линейного и других видов математического программирования. Рассмотрим один из простейших алгоритмов, в котором используются графики ресурсов и "метод проб и ошибок".

10.6.1. Графики ресурсов

Если общая потребность в некотором ресурсе определяется на основе постоянных интервалов, например, за один день или за одну неделю, то можно построить **график ресурса**. Ресурсы, требуемые для осуществления каждой работы, складываются по всем работам, выполняемым одновременно, в предположении, что каждая работа начинается в наиболее ранний срок ее выполнения. Необходимо построить отдельные графики по каждому виду ресурса. На рис.10.20 схематично изображен график ресурса "рабочая сила". Как следует из приведенного графика, иногда потребности в рабочей силе превышают ее наличие, но в то же время общее число требуемых человеко-часов не превосходит их наличного количества.

Если потребность в ресурсе превысила его лимит, необходимо либо вложить в проект дополнительное количество ресурса, либо пересмотреть календарный план выполнения операций. Иногда в таких ситуациях необходимо задержать срок

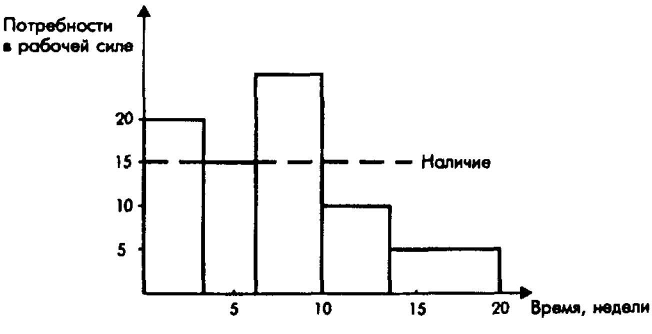


Рис. 10.20. График ресурса "рабочая сила"

выполнения проекта. Несмотря на то, что некоторые операции проекта не имеют явной логически последовательной взаимосвязи, одновременное их выполнение часто оказывается невозможным вследствие ограничений на ресурсы. Это ограничение можно отразить на графике ресурса, если провести линию, соответствующую наличному количеству данного ресурса. Такой прием позволит не планировать выполнение определенных операций на один и тот же период.

☐ **Пример 10.11.** Компания с ограниченной ответственностью "XYZ" заключила контракт на проведение работ по асфальтированию стоянки автомобилей. Менеджер проекта установил, что данная работа состоит из восьми основных операций. Приведем детальное описание этих операций:

Таблица 10.15. **Операции для примера 10.11, с указанием сроков выполнения и потребностей в рабочей силе**

Операция	*Предшествующие операции*	*Время, дней*	*Число человек, требуемое для выполнения операции*
A	–	3	1
B	–	6	1
C	–	7	2
D	A	8	2
E	C	4	1
F	B, E	3	2
G	C	10	2
H	F, G	3	1

Ввиду необходимости срочного выполнения работ на других участках, "XYZ" может выделить только четырех человек для проведения работ на автомобильной стоянке. Определим, сколько времени займет проведение работ и как следует распределить рабочих. Предположим, что каждый из рабочих может выполнять любую операцию.

Решение

Предположив, что все операции начинаются в наиболее ранний срок, построим соответствующий график "рабочей силы". После этого можно составить календарный план выполнения операций, удовлетворяющий ограничению на количество работников. Сначала построим сетевой граф и определим критический путь.

Время выполнения проекта в целом, если не принимать во внимание обеспечение ресурсами, составляет 20 дней. Критический путь выглядит следующим образом: C – G – H.

В предположении, что выполнение всех операций начинается в наиболее ранние сроки, посмотрим график Ганта (см. **10.3.2.**) и соответствующий график ресурса. График Ганта отражает распределение резерва времени на момент окончания каждой операции. С его помощью мы можем определить, какие операции выполняются одновременно и по каким операциям можно изменить календарный план их выполнения таким образом, чтобы эти изменения не привели к задержке выполнения проекта в целом.

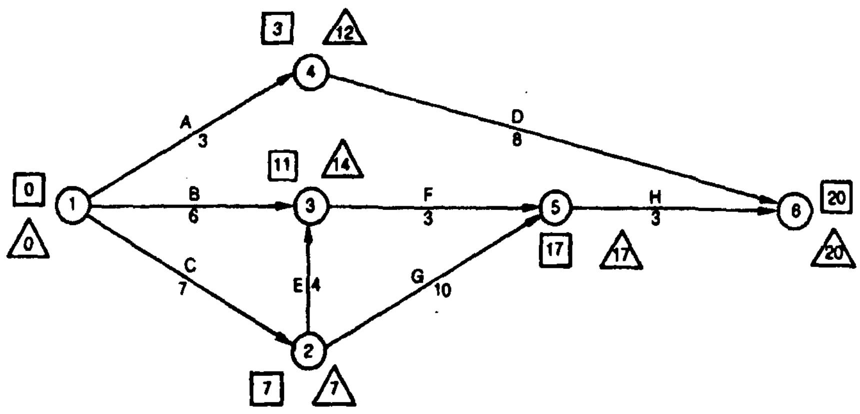


Рис. 10.21. Стрелочный граф для примера 10.11

□ – наиболее ранний срок события

△ – наиболее поздний срок события (стандартные сроки, дней)

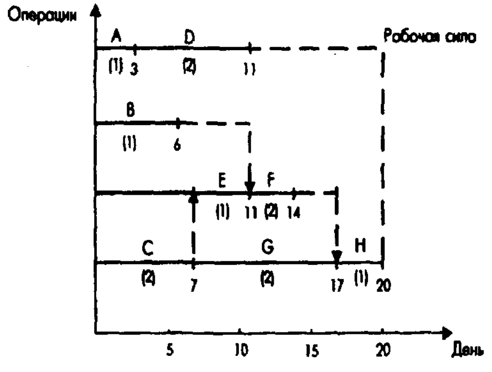


Рис. 10.22. График Ганта для примера 10.11

Из графика ресурса следует, что лимит, равный четырем рабочим, превышается, когда выполнение операции D попадает в промежуток между 3 и 11 днями осуществления проекта. Пересмотреть календарный план и полностью удовлетворить *потребности в рабочей силе, соответствующие* операции D; нельзя. Для выполнения критических операций C и G требуются два человека, поэтому операция D не может быть начата в течение 17 дней, т.е. до тех пор, пока не закончится выполнение остальных некритических операций. Если операцию D

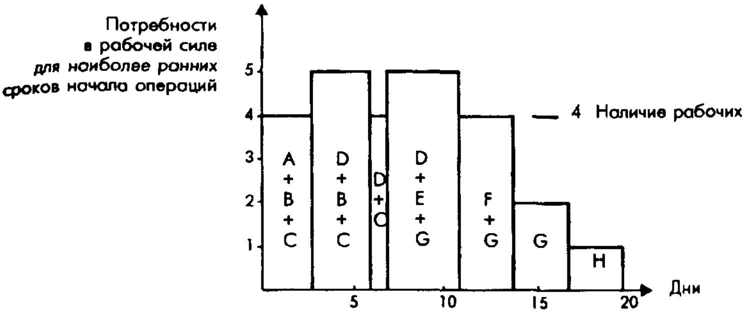


Рис. 10.23. График ресурса для примера 10.11, соответствующий наиболее ранним срокам начала выполнения операций

отложить на 12 дней, то в дни с 12 по 14 потребность в рабочих все еще будет превышать их наличие: в эти дни будут выполняться операции G (2 человека), F(2 человека) и D (2 человека). В этом случае придется либо привлечь к работе одного рабочего дополнительно на указанный период, либо отложить операцию D до момента, когда будет завершена операция F, т.е. до 14-го дня. При последнем варианте будет иметь место задержка в выполнении проекта, равная двум дням. Таким образом, его продолжительность возрастает с 20 до 22 дней.

РЕЗЮМЕ

Сетевой анализ используется при разработке и планировании проектов. Он предполагает разбиение проекта на отдельные виды работ или операции. Логическая взаимосвязь между операциями изображается с помощью сетевого графа. На основе значений сроков выполнения операций производится расчет общей продолжительности проекта, возможных сроков начала и окончания каждого вида работ и определяются операции, принадлежащие критическому пути.

Операции в сетевых графах можно изображать либо с помощью стрелок, либо с помощью узлов. Альтернативным методом изображения сети операций является график Ганта, в котором используется шкала времени.

В большинстве проектов определенные виды работ могут быть выполнены в более сжатые сроки, однако, это требует дополнительных затрат. Соответствующие показатели называются критическими сроками и критическими затратами. Они могут использоваться при составлении календарных планов реализации проектов за "минимальное время" или с "минимальной стоимостью".

Сроки выполнения операций могут быть подвержены влиянию неопределенности. В этом случае для анализа проекта можно использовать метод оценки и пересмотра проектов (PERT), основанный на предпосылке об аппроксимации сроков выполнения операций β-распределением с минимальным значением a, наиболее вероятным значением m и максимальным значением b. Ожидаемая продолжительность операции в методе PERT рассчитывается по следующей формуле:

$$\frac{a + 4m + b}{6},$$

а соответствующая дисперсия равна

$$\frac{(b - a)^2}{6^2}.$$

Продолжительность выполнения проекта имеет нормальное распределение, среднее значение которого равно сумме значений ожидаемых сроков выполнения операций, принадлежащих критическому пути. Дисперсия данного нормального распределения есть сумма дисперсий критических операций. Распределение времени выполнения проекта в целом используют в расчетах вероятности завершения проекта к заранее заданному сроку.

В анализ проектов можно включить также вопросы, связанные с наличием и распределением ресурсов. Для определения направлений пересмотра календарного плана выполнения операций и достижения конкретной цели применяются графики Ганта и графики ресурсов.

УПРАЖНЕНИЯ

Упражнение 10.1

Компания с ограниченной ответственностью "MR" разрабатывает строительный проект небольшого масштаба. Основные операции проекта, соответствующие им непосредственно предшествующие операции и время их выполнения приведены в таблице:

Операция	*Непосредственно предшествующая операция*	*Продолжительность, недель*
A	–	4
B	–	6
C	A, B	7
D	B	3
E	C	4
F	D	5
G	E, F	3

Требуется:

1. Дать иллюстрацию проекта с помощью стрелочного сетевого графа.

2. Определить критические операции и общую продолжительность выполнения проекта.

Упражнение 10.2

Используя данные упражнения 10.1:

1. Дать иллюстрацию проекта с помощью вершинного графа;
2. На основе графа, построенного в п.1, определить влияние на ход выполнения проекта задержки операции D на четыре недели.

Упражнение 10.3

В Чартерском Институте подготовки специалистов по принятию количественных решений (CIQDM) действует ежегодная программа чтения лекций сотрудникам института. Подготовка программы на следующий год ведется сотрудниками ректората института, начиная с осени предыдущего года. Эта программа содержит детальные сведения о лекторах и их лекциях, а также список членов института. Ниже перечислены операции, входящие в процесс подготовки программы, с указанием соответствующих непосредственно предшествующих операций:

Операция		*Непосредственно предшествующие операции*	*Стандартное время, дней*	*Критическое время, дней*	*Дополнительные издержки, ф. ст.*
A	Выбор дат проведения лекций	–	5	5	–
B	Назначение лекторов и согласование лекционных тем	A	20	10	100
C	Подготовка для программы рекламных материалов	–	15	10	150
D	Обновление списка студентов, обучающихся заочно	–	15	5	200
E	Подготовка списка оплачиваемых сотрудников	D	30	25	50
F	Распечатка программы и списка членов на принтере	B, C, E	10	5	100
G	Корректировка напечатанных программы и списка членов	F	10	5	50
H	Печать и раскладка программы по экземплярам	G	15	10	75
I	Получение распечатанного на компьютере списка адресов членов института	E	5	2	50
J	Рассылка программы	H,I	5	2	50

Если в процессе подготовки программы будет занято стандартное число сотрудников ректората, соответствующее штатному расписанию, то, как было оценено, каждая операция будет выполнена в указанные выше стандартные сроки. При этом предполагается, что управленческий персонал работает 5 дней в неделю. Однако существует возможность принять на работу несколько временных работников дополнительно в помощь основному персоналу на этот период. Продолжительность выполнения операций в этих условиях определяется критическими сроками, значения которых, а также соответствующие значения дополнительных издержек, связанных с выполнением операций в критические сроки, указаны выше. Для простоты расчетов предполагается, что все операции могут быть выполнены только либо в стандартные, либо в критические сроки.

Требуется:

1. Изобразить данный проект с помощью сетевого графа.
2. Определить общее время, требуемое для подготовки и рассылки программы, при условии, что временные работники не будут приняты на работу в этот период. Какие операции являются критическими?
3. Как повлияет на общую продолжительность проекта тот факт, что время, необходимое для получения рекламных материалов, было оценено неправильно, и на самом деле данная операция занимает 30 дней?
4. Каково значение наименьшего возможного срока, к которому можно закончить подготовку и рассылку программы? Какова минимальная дополнительная стоимость завершения проекта к этому сроку?

Упражнение 10.4

Компания с ограниченной ответственностью "Jubilee Computer Systems Ltd" выполняет заказ, полученный от ее потребителя. Необходимая информация приведена ниже.

Операция	*Непосредственно предшествующие операции*	*Срок, дней*			*Стоимость для ожидаемой продолжительности, ф. ст.*
		оптимистический	*наиболее вероятный*	*пессимистический*	
A	–	3	4	5	1000
B	–	4	7	10	1400
C	–	4	5	6	2000
D	A	5	6	7	1200
E	B	2	2,5	6	900
F	C	10	10,5	14	2500
G	D, E	3	4	5	800
H	G, F	1	2	9	300

Косвенные издержки, связанные с выполнением проекта, составляют 300 ф. ст. в день. В контракте, заключенном с потребителем, оговорено, что если заказ не будет выполнен в течение 15 дней, сумма штрафа составит 100 ф. ст. за каждый последующий день.

Требуется:

1. Построить сетевой граф. Каково ожидаемое значение времени выполнения всего проекта? Каково значение соответствующей стоимости?
2. Какой путь в графе является критическим? Прокомментируйте продолжительности некритических путей.
3. Какова вероятность того, что проект будет завершен без выплаты штрафов?

Упражнение 10.5

Компания "Rogers plc" намерена учредить дочернюю издательскую компанию. В нижеследующей таблице приведены необходимые операции, их взаимозависимости и продолжительность.

Операция	*Непосредственно предшествующие операции*	*Продолжительность, недель*
A	–	3
B	A	4
C	A	2
D	A	6
E	B	3
F	D	2
G	D	4
H	G	7
I	C, E, F	5
J	G, I	3

Требуется:

1. Определить ожидаемое время выполнения проекта в целом.
2. В предположении, что для выполнения каждой операции в установленные сроки требуется один человек, определить скорректированную ожидаемую продолжительность проекта при условии, что в распоряжении компании для выполнения данной работы имеются только два человека, каждый из которых может выполнять любую из операций.

Упражнение 10.6

Администрация компании "Saturnite plc" собирается реализовать исследовательский проект по изучению характеристик нового продукта. Итогом выполнения проекта должен быть отчет, содержащий рекомендации по выпуску нового продукта. Ниже приведены операции, которые необходимо осуществить в процессе выполнения исследовательского проекта:

Операция	*Описание*	*Непосредственно предшествующие операции*	*Ожидаемое время выполнения, недель*	*Потребности в персонале*
A	Первичные разработки	–	5	3
B	Исследование рынка	–	3	2
C	Получение технических стандартов	A	2	2
D	Создание образца	A	5	5
E	Подготовка рыночной базы	A	3	3
F	Расчет стоимости	C	2	2
G	Испытание продукта	D	4	5
H	Выборочный контроль	B, E	6	4
I	Оценки цены	H	2	1
J	Итоговый отчет	F, G, I	6	2

Требуется:

1. Построить сетевой граф, отражающий приведенные выше операции и их взаимосвязи. Определить критический путь и наименьшую продолжительность выполнения проекта.
2. В предположении, что началом выполнения проекта служит нулевой момент времени, а каждая операция начинается с наиболее раннего срока, построить график, изображающий потребности в персонале на любой момент времени.
3. Администрация компании приняла решение, что на выполнение изложенного проекта в любой момент времени будет выделено не более 9 человек персонала.

Опишите, как следует выполнять проект в данных условиях за наименьшее время, найденное в п 1. В течение какого количества недель в выполнении проекта будут участвовать все 9 человек персонала?

Упражнение 10.7

1 сентября каждого года администрация компании с ограниченной ответственностью "Salemis Ltd" составляет бюджет на следующий год. Было установлено, что процесс составления бюджета включает в себя следующие этапы:

Этап		*Предшествующие этапы*	*Время, недель*
1		2	3
A	Оценка ставок заработной платы	–	2
B	Разработка прогнозов рынка	–	4
C	Определение цен продаж	–	3
D	Составление бюджета для объемов продаж	B	3

Продолжение

1		2	3
E	Составление бюджета доходов от продажи	C, D	1
F	Составление бюджета расходов по продаже	A, D	3
G	Составление бюджета объемов производства	D	6
H	Составление бюджета накладных расходов	A	4
I	Составление бюджета трудовых ресурсов	A, G	2
J	Составление бюджета сырья	G	3
K	Составление бюджета производственных площадей и оборудования	G	5
L	Выработка прогноза общей прибыли	E, F, H, I, J, K	1

Составление бюджета необходимо закончить к концу декабря, таким образом администрация располагает периодом в 17 рабочих недель.

Требуется:

1. Построить сетевой граф, отражающий последовательность выполнения этапов, включенных в подготовку бюджетов. Можно ли закончить данный процесс в течение 17 недель?
2. Если бы потребовалось сократить время, отведенное на составление бюджетов, на какие этапы следовало бы обратить внимание и почему?
3. Объясните различие между понятиями общего, свободного и независимого резерва времени. Докажите, что свободный резерв времени этапа 1 равен трем неделям, причем две из них — это независимый резерв времени.

Упражнение 10.8

Некоторый проект включает в себя 10 операций, продолжительность и взаимосвязи которых указаны ниже.

Продолжительность операций F и H является неопределенной, поскольку на данной стадии ее оценка вызывает некоторые затруднения.

Операция	*Продолжительность, дней*	*Непосредственно предшествующие операции*
A	6	–
B	1	A
C	2	A
D	1	B
E	1	D
F		B
G	1	C
H		F, G
I	4	E, H
J	5	I

Требуется:

1. Построить соответствующий сетевой граф, отражающий взаимосвязи между 10 операциями.
2. Определить минимальное время, требующееся для реализации проекта, не учитывая при этом влияние операций F и H.
3. Если бы возникла необходимость закончить выполнение проекта за 19 дней, какие ограничения накладывало бы это условие на продолжительность операций F и H?
4. После проведения дальнейших исследований было установлено, что ожидаемыми сроками выполнения операций F и H являются два дня и один день соответственно. Кроме того, можно предположить, что неопределенность в продолжительности этих двух операций можно аппроксимировать с помощью распределения Пуассона. Исходя из этой информации найдите вероятность того, что выполнение проекта займет не более 19 дней. В нижеследующей таблице приведены некоторые значения вероятностей, соответствующие распределению Пуассона:

Среднее значение μ	*Вероятность*				
	0	*1*	*2*	*3*	*4 и более*
1	0,368	0,368	0,184	0,061	0,019
2	0,135	0,271	0,271	0,180	0,143

Упражнение 10.9

Фирма "Hydra Company" выпускает ряд средств для ухода за волосами и для бритья, включая опасные бритвы. Ее конкурент организовал недавно производство нового вида опасных бритв, которые за последние шесть месяцев приобрели большую популярность на потребительском рынке, что оказало обратное воздействие на объемы продаж фирмы "Hydra Company". Администрация приняла решение о скорейшем внедрении в производство конкурентоспособной продукции и поручила главному бухгалтеру составить план разработки нового продукта и внедрения его на потребительский рынок.

Первый шаг, предпринятый бухгалтером при разработке этого проекта, состоял в определении основных задач, которые необходимо решить в процессе создания нового продукта. Эти задачи перечислены ниже. Он произвел также оценку времени, которое займет решение каждой задачи, и выявил задачи, которые ей предшествуют.

Задача		*Время, недель*	*Предшествующие задачи*
1		2	3
A	Создание новой продукции	8	–
B	Создание упаковки	4	–
C	Подготовка производственных мощностей	4	A
D	Получение сырья и материалов	2	A

Продолжение

1		2	3
E	Выпуск опытной партии продукции	3	C, D
F	Упаковка	2	B
G	Принятие решения о выборе пробного рынка сбыта	1	—
H	Упаковка опытной партии	2	E, F
I	Поставка продукции на пробный рынок сбыта	3	H, G
J	Продажа продукции на пробном рынке сбыта	4	I
K	Оценка результатов внедрения продукции на рынок	3	J
L	Планирование выпуска продукции на национальном уровне	4	K

1. Постройте сетевой граф, отражающий логическую последовательность решения указанных задач, и определите, какой период времени пройдет с момента разработки плана до налаживания серийного выпуска новой продукции (можно предположить, что выпуск продукции на национальном уровне будет иметь место сразу же после составления его плана).
2. Рассчитайте значения резерва времени, соответствующие каждой из некритических операций.
3. Время, которое потребуется для выполнения задач A, B, D, K и L, подвержено влиянию неопределенности, поэтому для получения наиболее вероятных значений сроков выполнения этих операций, которые приведены выше, были разработаны следующие оценки оптимистических и пессимистических сроков:

Задача	*Оптимистический срок, недель*	*Пессимистический срок, недель*
A	5	13
B	2	6
D	1	4
K	2	6
L	2	8

С учетом приведенной выше информации определите ожидаемое время, которое пройдет до момента серийного выпуска продукции, и вероятность того, что этот период превысит 35 недель (следует ввести предпосылку о том, что продолжительность проекта в целом аппроксимируется нормальным распределением).

Глава 11. ПЛАНИРОВАНИЕ И УПРАВЛЕНИЕ ЗАПАСАМИ

11.1. ВВЕДЕНИЕ

Одно из последствий изменения экономической ситуации состоит в том, что промышленные компании вынуждены пересматривать свою политику хранения и управления запасами, включая и сырье, и конечную продукцию.

Если некоторая компания имеет товарные запасы, то капитал, овеществленный в этих товарах, замораживается. Этот капитал, который нельзя использовать, представляет для компании потерянную стоимость в форме невыплаченных процентов или неиспользуемых возможностей инвестирования. Кроме того, наличие запасов влечет за собой определенные издержки, поскольку для их хранения необходимо создать определенные условия и выделить определенные площади; необходимо оплачивать работу персонала, осуществляющего управление запасами; запасы должны быть застрахованы и т.п. В этой связи разумно предположить, что целью любой компании является хранение по возможности наименьшего запаса. Однако, следует принять во внимание и другие соображения. Спрос на продукцию чаще всего содержит долю неопределенности. Поэтому чем меньше уровень запаса, тем больше вероятность того, что возникнет дефицит продукции. Наличие дефицита тех или иных товаров уже само по себе является для компании источником определенных убытков либо в сфере производства, либо в связи с потерей клиентов.

Если компания создает товарный запас исходя из размера производственной партии деталей, то вероятнее всего экономически выгодным окажется производство крупных партий деталей, однако, такая политика подразумевает высокий уровень исходного запаса. В данной области существует множество проблем, которые предстоит решить. На сегодняшний день все более важной функцией администрации большинства компаний становится анализ эффективности и оценка политики управления запасами. В отдельных компаниях теория управления запасами находит все более широкое применение. Если в некоторой компании используются современные методы управления, такие, например, как метод "точно в срок", то особую важность приобретает знание менеджерами компании механизма функционирования системы управления запасами. Методы моделирования систем управления запасами были разработаны еще несколько лет назад. Диапазон этих моделей достаточно обширен – от базисных моделей простых детерминированных систем до более сложных моделей, учитывающих неопределенность спроса или сроков поставки заказа. Если система управления запасами является достаточно сложной, то для ее моделирования можно использовать имитационные методы (см. гл. 14). В любом случае цель процесса моделирования состоит в том, чтобы помочь лицу, принимающему решение, определить с учетом некоторого критерия принятия решения уровень заказа и срок его подачи.

Целью практически любого решения является минимизация общих издержек, связанных с хранением запасов. Не менее важен анализ последствий применения неоптимальной схемы запаса, что предполагает анализ модели на чувствительность.

В данной главе мы рассмотрим основную модель управления запасами, а также возможные способы ее адаптации к различным проблемам, возникающим в связи с наличием запасов, включая и проблему неопределенности. Базисные модели управления запасами часто подвергают критике, основанной на утверждении об отсутствии их связи с реальностью. Хотя такая критика отчасти оправдана, эти модели могут служить полезным введением в существо предмета и проблем, которые необходимо решить.

11.2. ОСНОВНАЯ МОДЕЛЬ УПРАВЛЕНИЯ ЗАПАСАМИ

Рассмотрим для начала проблемы управления запасами, связанные либо с заказом на партию деталей внешнему поставщику, либо с выпуском партии деталей. Политика организации производства или подачи заказов в этой ситуации должна быть такой, чтобы общие издержки были минимальными.

В любой системе управления запасами уровень последних изменяется в соответствии с циклической моделью. Процесс снижения уровня запасов определяется соответствующей моделью спроса. В некоторой точке для пополнения запаса будет сделан новый заказ. По прошествии некоторого времени, называемого **временем поставки**, заказ будет получен, и уровень запасов возрастает. После этого начинается новый цикл запасов (см. рис. 11.1.).

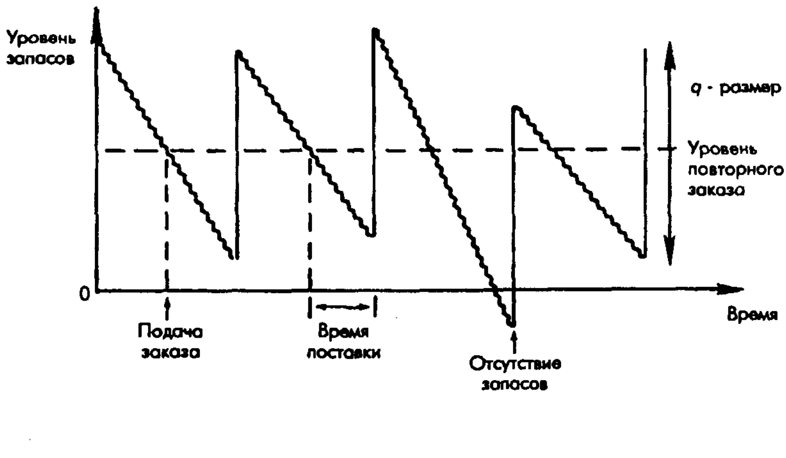


Рис. 11.1. Стандартная модель хранения запасов

11.2.1. Система предпосылок основной модели управления запасами

Для упрощения процесса моделирования в модель вводится ряд предпосылок:

1. Спрос на продукцию является постоянным, или приблизительно постоянным. Если коэффициент использования запасов является постоянным, то уровень запасов также будет уменьшаться с постоянным коэффициентом.
2. Предполагается, что время поставки известно и является постоянной величиной. Это означает, что заказ можно осуществлять в точке с определенными значениями временного параметра и размера запаса (**уровень повторного заказа**), которые обеспечивают получение заказа в тот момент, когда уровень запасов равен нулю.
3. Отсутствие запасов является недопустимым.
4. В течение каждого цикла запасов делается заказ на постоянное количество продукции (q).

Окончательный вид модели управления запасами является следующим:

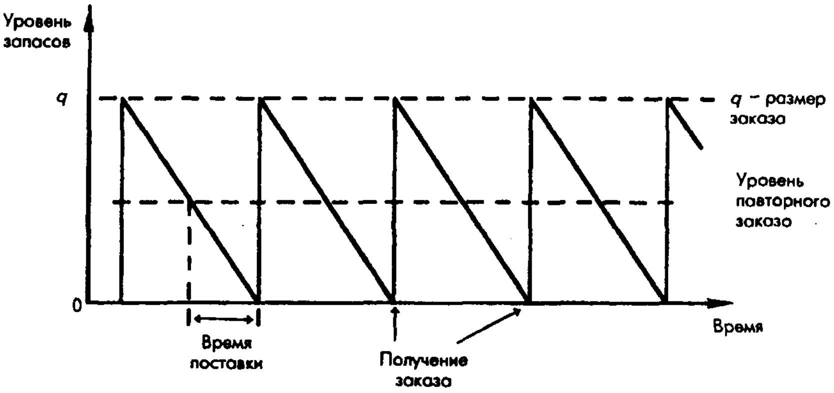


Рис. 11.2. Схема управления запасами для основной модели

Все циклы запасов являются одинаковыми. Максимальное количество продукции, которая находится в запасе, совпадает с размером заказа q.

11.2.2. Издержки хранения запасов

Если у внешнего поставщика заказывается партия продукции, процессы подачи и поставки заказа повлекут за собой определенного рода затраты. Необходимо создать соответствующие условия по хранению и управлению запасами. Поэтому в данной области также возникают определенные издержки. В отдельных случаях может возникнуть необходимость и в иных затратах, таких, например, как издержки вследствие нехватки запасов или хранения резервного запаса. Эти понятия будут введены нами по ходу изложения материала. Здесь мы рассмотрим только понятия стоимости подачи заказа, издержек хранения, или складирования, запасов, а также стоимости покупки продукции.

Если говорить о выпуске продукции в форме производственных партий, а не о покупке продукции извне, возникают аналогичные издержки. В данном случае стоимость подачи заказа эквивалентна стоимости организации технологического процесса по выпуску партии продукции, а стоимость покупки — издержкам производства продукции. Поэтому схема анализа останется неизменной. Каждый вид стоимости или издержек включает в себя постоянную и переменную компоненты. С точки зрения анализа основной модели управления запасами нас интересует только переменная ее часть. На этом этапе возникает необходимость введения в модель новой предпосылки: переменные издержки подачи заказа, или организации процесса выпуска партии продукции, известны, являются постоянными и не зависят от размера заказа.

11.2.3. Уравнение общей стоимости

Необходимо построить модель, которая описывает издержки, связанные с наличием запасов, за весь период их хранения. Длительность этого периода значения не имеет: это может быть один день, месяц, год и т.д. В данном случае мы выберем период, равный одному году. Введем следующую систему обозначений:

- D – ежегодный спрос на запас продукции;
- C_0 – переменная стоимость подачи одного заказа, ф. ст./ 1 заказ;
- C_h – переменная стоимость хранения единицы продукции в запасе, ф. ст. на единицу продукции в год;
- C – цена покупки единицы продукции в запасе, ф. ст. в год;
- q – объем заказа, единиц продукции/заказ.

Общая стоимость запасов в год = Общая стоимость подачи заказа в год +
+Общая стоимость хранения запасов в год.

Рассмотрим каждую из составляющих данного уравнения в отдельности.

ЕЖЕГОДНАЯ СТОИМОСТЬ ПОДАЧИ ЗАКАЗА

Если потребность в продукции составляет D единиц в год, а каждый заказ подается на партию в q единиц, тогда ежегодное количество заказов составит (D/q).

Ежегодная стоимость подачи заказов = Стоимость подачи одного заказа ×
× Число заказов, подаваемых ежегодно = $C_0 \times (D/q)$ (ф. ст. в год).

ЕЖЕГОДНАЯ СТОИМОСТЬ ХРАНЕНИЯ ЗАПАСОВ

При расчете этой стоимости обычно исходят из среднего количества продукции, которая составляет запас в течение одного цикла. В простейшей ситуации, которую мы рассматриваем, уровень запасов изменяется линейно и принадлежит промежутку от q до нуля, следовательно, средний уровень запасов равен $(q/2)$. В более сложных ситуациях для расчета среднего уровня запасов используются более сложные математические методы.

Стоимость хранения единицы продукции C_h определяется либо как фиксированная величина на весь год, либо как процент от общей стоимости единицы продукции за год. В различных компаниях применяются самые разнообразные методы расчета издержек в этой сфере, однако в целом C_h характеризует величину процентов с денежных ссуд, замороженных в форме запасов, стоимость повреждения или сохранности запасов, а также определенную часть общей стоимости системы хранения запасов.

Ежегодная стоимость хранения запасов =
= Стоимость хранения единицы продукции в год × Средний размер запаса =
= $C_h \times (q/2)$ (ф. ст. в год).

Из этого следует, что общая стоимость запаса единицы продукции в год определяется следующим образом:

$$TC = C_0 (D/q) + C_h (q/2) \text{ (ф. ст. в год).}$$

Данное уравнение называется уравнением общей стоимости основной модели управления запасами. Теперь мы должны определить значение q, при котором значение общей стоимости наименьшее.

11.2.4. Оптимальный размер заказа q_0

Для определения оптимального значения q используем операцию дифференцирования следующим образом:

$$TC = C_0 (D/q) + C_h (q/2),$$

TC принимает минимальное значение, когда

$$\frac{dTC}{dq} = 0 \text{ и } \frac{d^2 TC}{dq^2} > 0;$$

$$\frac{dTC}{dq} = -C_0 \frac{D}{q^2} + C_h \frac{1}{2}$$

и

$$\frac{d^2 TC}{dq^2} = -2 C_0 \frac{D}{q^3} + 0 > 0 \text{ , если } q > 0.$$

Положим.

$$\frac{DTC}{dq} = 0, \text{ тогда } -C_0 \frac{D}{q^2} + C_h \frac{1}{2} = 0,$$

следовательно

$$C_0 \frac{D}{q^2} = C_h \frac{1}{2};$$

$$q^2 = \frac{2\,C_0 D}{C_h};$$

$$q_0 = \pm\sqrt{\frac{2\,C_0 D}{C_h}}.$$

Таким образом, ТС принимает минимальное значение, если $q_0 = +\sqrt{2C_0\,D/C_h}$. Полученный объем заказа называют **экономичным размером заказа (EOQ)**. Если в течение года с равными интервалами заказывать данное количество продукции, то стоимость хранения будет минимальной. В настоящее время стало уже традиционным непосредственное применение формулы модели EOQ, а не получение ее каждый раз из уравнения общей стоимости.

Полезно воспользоваться графическим представлением уравнения общей стоимости и его компонент. Издержки хранения пропорциональны размеру заказа, следовательно, их график представляет собой прямую, проходящую через начало координат. Стоимость подачи заказа пропорциональна величине $1/q$. Ниже приводится графическое изображение указанных видов издержек и их суммы.

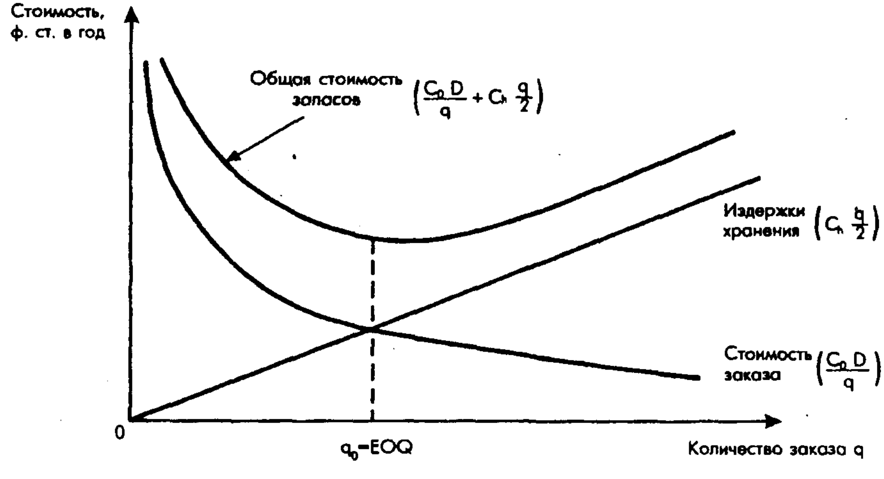


Рис. 11.3. Графическое изображение стоимости подачи заказа, издержек хранения и общей стоимости запасов

Нетрудно заметить, что если размер заказа невелик, то стоимость подачи заказа является доминирующей — в этом случае заказы подаются часто, но на небольшое количество продукции. Если размер заказа является достаточно большим, основной компонентой становится стоимость хранения — делается небольшое число заказов, размер которых достаточно велик. Экстремальная точка на графике уравнения общей стоимости соответствует ситуации, когда оба вида издержек равны друг другу. Этот факт может оказаться полезным при проверке расчетов EOQ. Кроме того, можно отметить, что в критической точке кривая общей стоимости заметно выравнивается. Это означает, что в данной области общая стоимость не обладает высокой чувствительностью по отношению к изменениям в размере заказа. После того как получено значение EOQ, остается еще, как правило, несколько значений, поэтому можно выбрать необходимый размер заказа, не приводящий к значительному увеличению общей его стоимости.

11.2.5. Уровень и интервал повторного заказа

Итак, мы знаем, каким должен быть размер заказа, но нам по-прежнему неизвестно, когда следует осуществлять его подачу.

Если время поставки заказа от поставщика составляет L недель, то в течение поставки будет использоваться L × (D/52) единиц продукции, составляющей запас, в предположении, что в году 52 недели. Поскольку величина спроса постоянна, количество продукции, которое используется в течение поставки заказа, является одновременно и уровнем повторного заказа. Таким образом, новый заказ следует подавать, когда уровень запасов снижается до величины L × (D/52). В этом случае новый заказ будет получен в тот момент, когда уровень запасов станет равным нулю.

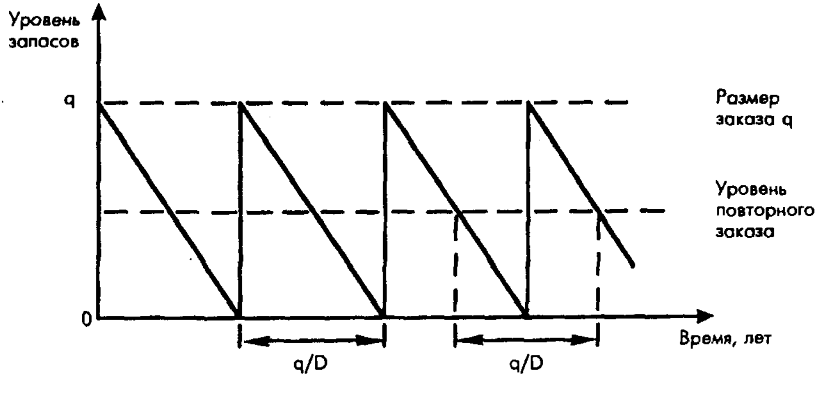


Рис. 11.4. Уровень и интервал повторного заказа

В течение года потребуется D/q заказов с равными интервалами, следовательно, новый цикл заказа всегда начинается в точке

$$\frac{1\text{год}}{(D/q)\ \text{заказов}} = q/D \text{ лет.}$$

Так как все циклы заказов одинаковы, интервал повторного заказа также будет равен (q/D) лет.

Пример 11.1. Объем продажи некоторого магазина составляет 500 упаковок пакетного супа в год. Величина спроса равномерно распределяется в течение года. Цена покупки одного пакета равна 2 ф. ст. За один заказ владелец магазина должен заплатить 10 ф. ст. Время доставки заказа от поставщика составляет 12 рабочих дней (при 6-дневной рабочей неделе). По оценкам специалистов, издержки хранения составляют 20% среднегодовой стоимости запасов. Сколько пакетов должен заказывать владелец магазина каждый раз, если его цель состоит в минимизации общей стоимости запасов? Предположим, что магазин работает 300 дней в году, определим, с какой частотой следует осуществлять подачу заказов и уровень повторного заказа.

Решение.

Экономичный размер заказа равен:

$$q_0 = \sqrt{2C_0 D/C_h},$$

где D = 500 пакетов в год;

C_0 = 10 ф. ст. за один заказ;

C_h = 20% в год от стоимости запаса размером в одну упаковку, или $0{,}2 \times 2$ ф. ст. в год за одну упаковку.

Следовательно,

$$q_0 = \sqrt{\frac{2 \times 10 \times 500}{0{,}2 \times 2}} = 158{,}11.$$

Количество заказываемых пакетов должно быть целым числом, поэтому в качестве EOQ выберем значение, равное 158 пакетам. В дальнейшем мы можем попытаться определить размер заказа более точно. Минимальное значение общей стоимости заказа в год определяется по следующей формуле:

$$TC = C_0 D/q_0 + C_h q_0/2.$$

Следовательно,

$$TC = 10 \times 500/158 + 0{,}2 \times 2 \times 158/2 = 31{,}6 + 31{,}6 = 63{,}2 \text{ ф. ст. в год.}$$

Общая стоимость купленных владельцем магазина 500 упаковок пакетного супа в год составляет:

Стоимость запасов + *Стоимость покупки* =
= 63,2 ф. ст.+ 2 ф. ст. × 500 = 1063,2 ф. ст. в год.

Таким образом, стоимость запасов составляет 6% общей стоимости покупки в год.

Если бы владелец магазина подавал заказы на партии в 150 упаковок, то величина общей стоимости запасов за год составила бы:

$$TC_{150} = 10 \times 500/150 + 0{,}2 \times 2 \times 150/2 = 33{,}33 + 30{,}0 = 63{,}33 \text{ ф. ст. в год.}$$

По сравнению со стоимостью, соответствующей найденному значению EOQ, данное увеличение стоимости является небольшим и составляет 13 пенсов в год.

Подачу нового заказа владелец магазина должен осуществлять каждый раз по истечении периода, равного 158/500 лет. Поскольку в году 300 рабочих дней, интервал повторного заказа будет равен

$$\frac{158 \times 300}{500} = 94{,}8 \approx 95 \text{ рабочих дней.}$$

Объем продажи пакетных супов за 12 дней поставки заказа составит:

(Спрос/Число дней) × Время поставки = (500/300) × 12 = 20 упаковок.

Следовательно, уровень повторного заказа равен 20 упаковкам. Таким образом, подача нового заказа производится в тот момент, когда уровень запасов равен 20 пакетам.

11.2.6. Модель экономичного размера партии

Компании, специализирующиеся на выпуске различных видов товаров, могут организовать технологический процесс не на непрерывной основе, а на основе производства партий продукции. Например, на хлебопекарном предприятии может быть принято решение о производстве партии больших батонов из непросеянной муки, затем — партии маленьких булочек, за которой должна следовать партия ячменных лепешек. Если в компании используется производство продукции партиями, то приходится решать вопрос о размере партии продукции, производимой в течение одного производственного цикла, и о том, с какой частотой следует производить партию определенной продукции. Возникающие трудности аналогичны проблемам, связанным с определением экономичного размера заказа. Вместо заказа определенного количества продукции у внешнего поставщика рассматривается объем производства определенной продукции. Таким образом, стоимости заказа, которая фигурировала в изложенной выше модели, соответствует стоимость организации процесса производства партии продукции.

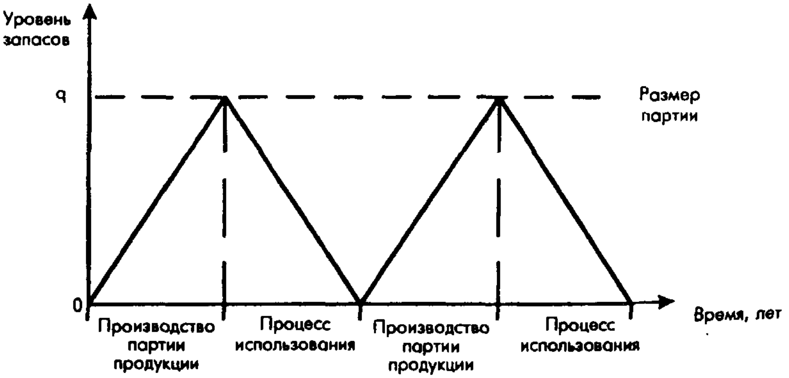


Рис. 11.5. Модель экономичного размера партии

Общая ежегодная стоимость производства =
=Ежегодная стоимость организации технологического процесса +
+ Годовая сумма издержек хранения.

Если через C_s обозначить стоимость организации каждого производственно цикла, то тогда

$$TC = C_s \times (D/q) + C_h \times (q/2) \text{ (ф. ст. в год)},$$

где q — размер партии продукции. Очевидно, что по аналогии с предыдущей задачей TC принимает свое минимальное значение, если

$$q_o = \sqrt{2C_s\, D/C_h}.$$

Полученное оптимальное количество продукции в партии называют **экономичным размером партии (EBQ)**.

❒ **Пример 11.2.** Компания, производящая изделия из керамики, выпускает несколько видов кофейников. Производственный процесс организован по принципу выпуска партий кофейников общим объемом 500 штук в неделю. Спрос на наиболее популярную модель, которую мы обозначим через X, составляет 2500 изделий в год и равномерно распределяется в течение года. Вне зависимости от того, в какой момент времени возникает необходимость в производстве партии кофейников модели X, стоимость производственного процесса составляет 200 ф. ст. По оценкам специалистов компании стоимость хранения кофейников составляет 1,50 ф. ст. за единицу.

Какова должна быть партия кофейников, чтобы затраты на производство и хранение были минимальными? Как часто следует возобновлять производственный цикл и какова его длительность? Предполагается, что в году 50 рабочих недель.

Решение

D_s = 2500 кофейников в год;

C_q = 200 ф. ст. на один производственный цикл;

C_h = 1,50 ф. ст. за один кофейник в год.

Экономичный размер партии можно определить следующим образом:

$$q_o = \sqrt{2C_sD/C_h} = \sqrt{2 \times 200 \times 2500/1{,}50} = 816{,}5.$$

Поскольку кривая общей стоимости не обладает высокой чувствительностью по отношению к небольшим изменениям значений q, вполне вероятно, что выбранное в качестве EBQ значение, равное 820, не приведет к значительному увеличению общей стоимости. Это утверждение можно легко проверить.

Для q = 816,5 единиц имеем:

TC = 200 × 2500/816,5 + 1,5 × 816,5/2 = 612,37 + 612,37 = 1224,74 ф. ст. в год.

Для q = 820 единиц имеем:

TC = 200 × 2500/820 + 1,5 × 820/2 = 609,76 + 615 = 1224,76 ф. ст. в год.

Для q = 800 единиц имеем:

TC = 200 × 2500/800 + 1,5 × 800/2 = 625 + 600 = 1225 ф. ст. в год.

Наиболее удобный размер партии, равный 800 кофейникам, по сравнению с оптимальным размером приводит к увеличению общей стоимости производства и хранения кофейников на 26 пенсов.

Примем в качестве EBQ значение, равное 800 кофейникам. Число производственных циклов в год составит: 2500/800 = 3,125 (т.е. 25 циклов за каждые 8 лет), следовательно, интервал между двумя любыми производственными циклами равен: 800 × 50/2500 = 16 недель. Если объем производства в неделю равен 500 кофейникам, то процесс производства одной партии займет 800/500 = 1,6 недели.

11.3. СКИДКА НА КОЛИЧЕСТВО

При подаче заказа внешнему поставщику цена, назначаемая на тот или иной товар, может зависеть от объема покупки. На заказы большого объема обычно предоставляются скидки. Необходимо выяснить, как повлияет предоставление скидки на общую стоимость. Заказы на более крупные партии продукции повлекут за собой увеличение стоимости запасов (стоимость заказа плюс издержки хранения), однако данное увеличение может быть до некоторой степени компенсировано снижением закупочной цены.

Если принять во внимание стоимость закупки продукции, то уравнение общей стоимости примет вид:

$$\text{Общая стоимость закупки и запасов } \frac{C_0 D}{q} + \frac{C_h q}{2} + CD \text{ (ф. ст. в год)}.$$

где C — закупочная цена единицы продукции. Если цена закупки постоянна и не зависит от q, ее включение в уравнение общей стоимости приводит к перемещению графика этого уравнения параллельно оси q, не изменяя при этом его формы (см. рис. 11.6).

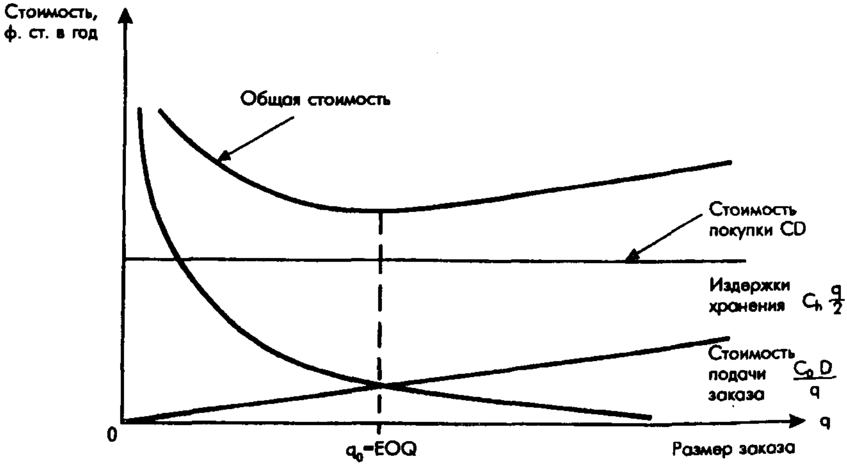


Рис. 11.6. Ежегодная стоимость покупки и запасов продукции

Как правило, стоимость покупки значительно превосходит по величине общую стоимость запасов.

Если обычно товар реализуется по цене C за единицу, но для заказов, размер которых превышает некоторую величину q_1, предоставляется скидка, в соответствии с которой цена за единицу продукции снижается до величины C_1, то изменение общей стоимости будет происходить по схеме, изображенной на рис. 11.7.

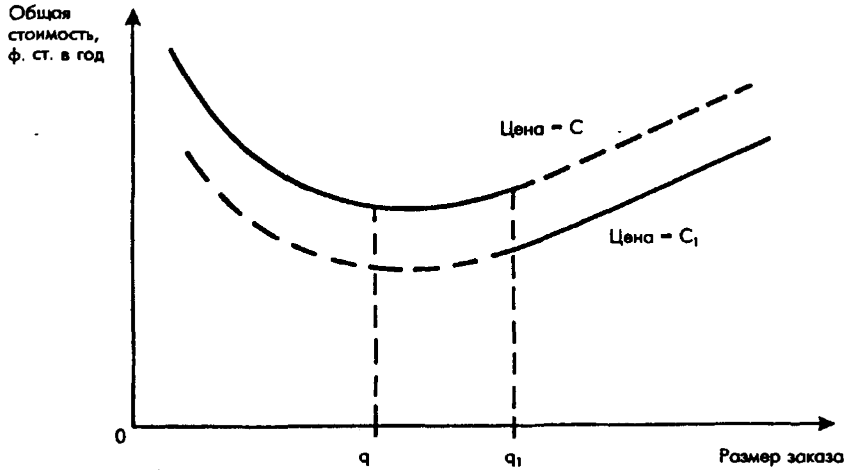


Рис. 11.7. Влияние скидки на ежегодную стоимость заказа и запасов продукции

Если для заказов, размер которых превышает величину q_2, существует дополнительная скидка, позволяющая снизить цену за единицу продукции до величины C_2, общая картина будет примерно такой, как показано на рис. 11.8.

Очевидно, предоставление скидок выгодно для определенного интервала размера заказа. Уровень заказа, начиная с которого устанавливается скидка, называется **уровнем, нарушающим цену**.

На рис. 11.8 изображены три кривые, каждая из которых соответствует определенной цене закупки единицы продукции — C, C_1 и C_2 (ф. ст.), однако использоваться могут лишь некоторые части данных кривых. Если значение q в экстремальной точке кривой не включается в интервал предоставления скидки, то данная экстремальная точка уже не соответствует оптимальному размеру заказа. Чтобы определить оптимальное значение q в данном случае, на первом этапе, не принимая во внимание ограничения на величину q, для каждого уровня цен найдем размер заказа, которому соответствует минимальное значение стоимости. Если полученное значение q попадает в интервал предоставления скидки, то оно является оптимальным размером заказа. Если же значение q в экстремальной точке меньше нижней границы интервала предоставления скидки, то производится пересчет общей стоимости для наименьшего возможного значения q, которое принадлежит интервалу предоставления скидки на цену закупки.

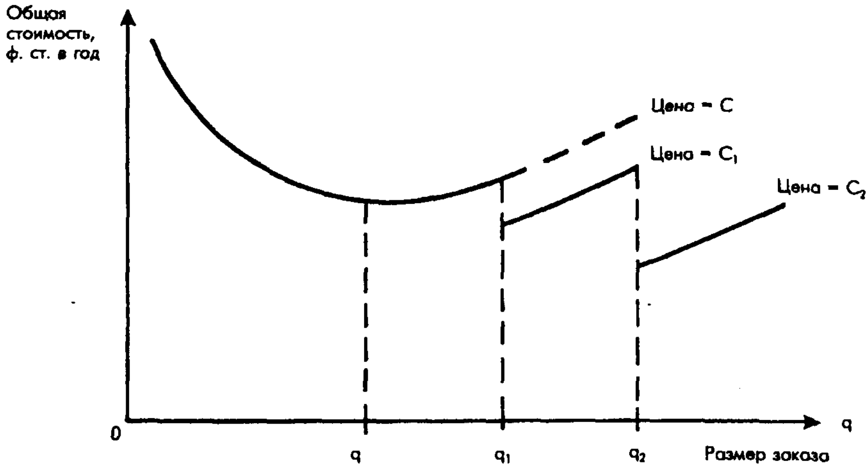


Рис. 11.8. Влияние на ежегодную стоимость заказа и запасов продукции двух скидок на количество

❐ **Пример 11.3.** Вернемся к примеру **11.1**, в котором рассматривалась покупка владельцем магазина пакетных супов. Закупка производилась партиями по 158 упаковок по 2 ф. ст. за единицу. В настоящее время поставщик предоставляет следующие скидки на закупочные цены:

Таблица 11.1. **Скидки на количество, предоставляемое поставщиком**

Размер заказа	*Скидка, %*	*Цена за упаковку, ф.ст.*
0–199	0	2,00
200–499	2	1,96
500 и более	4	1,92

Следует ли владельцу магазина воспользоваться одной из скидок?

Решение.

Мы знаем, что если владелец магазина захочет получить одну из скидок, то размер запасов увеличится, поскольку в этом случае он должен будет заказать не менее 200 упаковок супа, тогда как на настоящий момент уровень запасов составляет 158 упаковок. Будет ли скомпенсировано увеличение издержек хранения снижением закупочных цен и стоимости подачи заказа?

Поскольку размер заказа возрастает, изменение общей стоимости определяется сопоставлением соответствующих частей трех кривых, обозначающих цены за единицу продукции, равные 2 ф. ст., 1,96 и 1,92 ф. ст. (см. рис. 11.9).

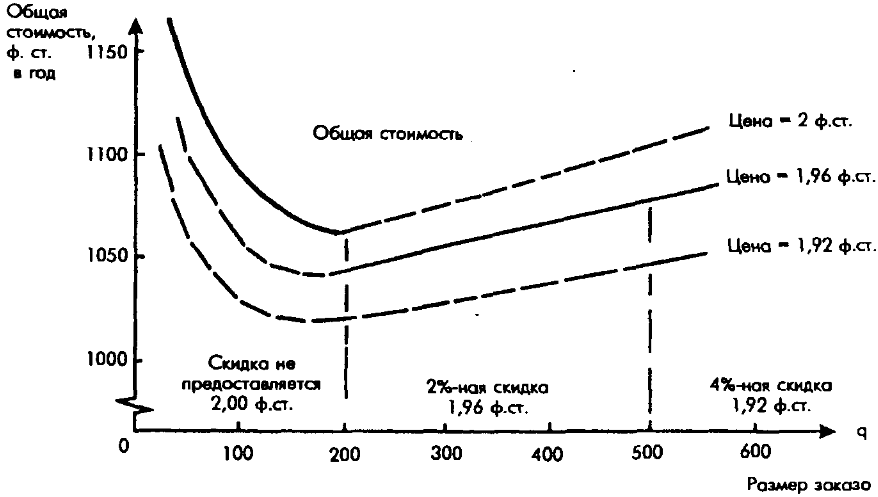


Рис. 11.9. Влияние скидок на ежегодную стоимость запасов пакетного супа

Из примера 11.1 нам известно, что если закупочная цена равна 2 ф. ст. за упаковку, то размер заказа в экстремальной точке кривой общей стоимости равен 158 упаковкам, а соответствующее минимальное значение общей стоимости покупки 500 пакетов в год составляет 1063,2 ф. ст.

Рассмотрим случай, когда закупочная цена равна 1,96 ф. ст. за упаковку. Тогда издержки хранения составят:

C_h = 20% от 1,96 ф. ст. = 0,392 ф. ст. за упаковку ежегодно.

Оптимальный размер заказа будет равен:

$$EOQ = \sqrt{2C_0\,D/C_h} = \sqrt{2 \times 10 \times 500/0{,}392} = 159{,}72.$$

Данное значение меньше, чем нижняя граница интервала предоставления первой скидки, 200–499, следовательно, абсцисса экстремальной точки кривой, соответствующей закупочной цене в 1,96 ф. ст., не является допустимым размером заказа. Минимальная возможная стоимость будет получена, если q = 200 упаковкам.

Минимальная общая стоимость за год (при закупочной цене в 1,96 ф. ст. за упаковку) =

$$= \frac{10 \times 500}{200} + 0{,}392 \times \frac{200}{2} + 1{,}96 \times 500 = 25 + 39{,}20 + 980 = 1044{,}20 \text{ ф. ст. в год}.$$

Издержки хранения при цене, равной 1,92 ф. ст., будут равны:

C_h = 20% от 1,92 ф. ст. = 0,384 ф. ст. за упаковку ежегодно.

Оптимальный размер заказа составляет:

$$EOQ = \sqrt{2C_0\,D/C_h} = \sqrt{2 \times 10 \times 500/0{,}384} = 161{,}37.$$

Данное значение меньше, чем нижняя граница интервала предоставления второй скидки 500 и более, следовательно, абсцисса экстремальной точки кривой, соответствующей закупочной цене в 1,92 ф. ст., не является допустимым размером заказа. Минимальная возможная стоимость будет получена, если q = 500 упаковкам.

Минимальная общая стоимость за год (при закупочной цене в 1,92 ф. ст. за упаковку) =

$$\frac{10 \times 500}{500} + 0{,}384 \times \frac{500}{2} + 1{,}92 \times 500 = 10 + 96 + 960 = 1066 \text{ ф. ст. в год.}$$

Сравнивая три полученных решения, имеем:

Таблица 11.2. **Сравнение минимальных значений общей стоимости, соответствующих трем уровням закупочных цен**

Стоимость одной упаковки, ф. ст.	*Размер заказа*	*Минимальная общая стоимость, ф. ст.*
2,00	158	1063,20
1,96	200	1044,20
1,92	500	1066,00

Таким образом, владельцу магазина выгодна только первая из предоставляемых скидок, для получения которой он должен подать заказ на 200 упаковок супа. Эта мера приведет к снижению общей стоимости на 19 ф. ст. в год.

11.4. ДРУГИЕ МОДЕЛИ УПРАВЛЕНИЯ ЗАПАСАМИ

Основную модель, созданную нами для моделирования процесса закупки продукции у внешнего поставщика, можно модифицировать и применять в иных ситуациях. Мы уже показали возможности использования этой модели в исследовании проблемы производства партий продукции. В данном случае принималась предпосылка о том, что вся продукция, производимая в партии, образует запас. Максимальный уровень запаса совпадал с размером партии продукции. Однако для многих технологических процессов производства партий продукции возникает ситуация, когда продукция, образующая запас, используется в течение всего времени данного технологического процесса, поэтому максимальный уровень запаса оказывается меньше, чем размер партии производимой продукции.

В исследуемых ранее ситуациях предполагалось, что отсутствие запаса недопустимо. Между тем во многих случаях гораздо дешевле допустить отсутствие запаса, чем поддерживать его уровень, необходимый для того, чтобы избежать отсутствия продукции в запасе. В двух следующих разделах рассматриваются пути адаптирования основной модели к различным изменениям исходных условий, включая моменты, изложенные выше.

11.4.1. Модель производства партии продукции

Предположим, что на некотором станке производится партия деталей, часть которых сразу же используется на другом станке, имеющем более низкую производительность. Оставшаяся часть деталей находится в запасе до тех пор, пока эти детали не понадобятся для другого станка. В данном случае не происходит единовременного пополнения всего запаса, и его уровень не изменяется скачкообразно от 0 до q, напротив, запас равномерно возрастает в течение периода работы первого станка, а затем, по мере использования запасов для работы второго станка, начинает убывать. Производительность первого станка равна P, а темп использования запасов равен D, причем $P \geq D$. Как показано на рис. 11.10, уровень запасов изменяется во времени.

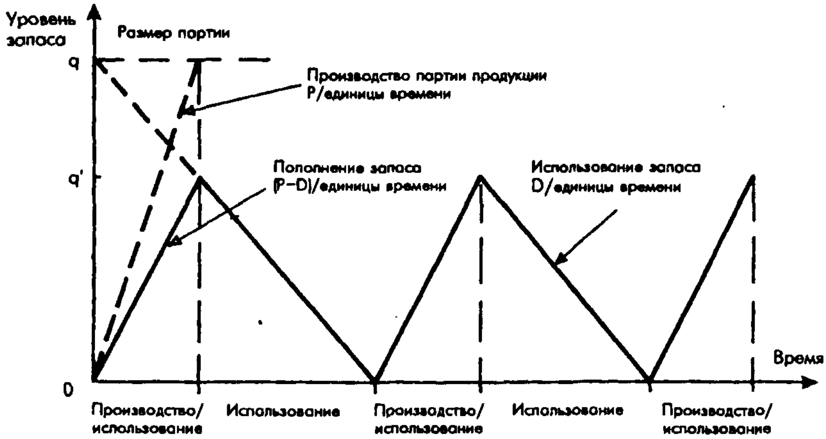


Рис. 11.10. Изменение уровня запасов в модели производства партии продукции

Каково оптимальное значение размера партии продукции q для первого станка? С какой частотой следует выпускать партии продукции? Общая переменная стоимость партии продукции за год TC включает в себя стоимость производственного цикла и издержки хранения. Следовательно,

$$TC = C_s \times \text{Число партий продукции в год} + C_h \times \text{Средний уровень запаса},$$

$$\text{Число партий продукции в год} = \text{Ежегодный спрос} / \text{Размер партии} = D/q.$$

Для того чтобы найти средний уровень запаса, рассмотрим более подробно один цикл запаса.

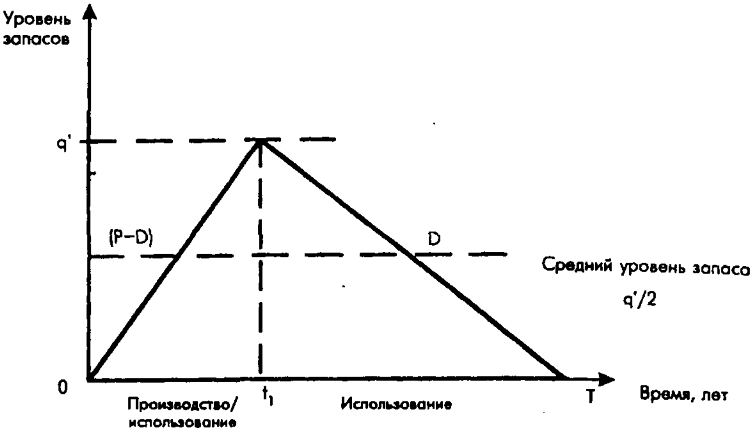


Рис. 11.11. Средний уровень запаса в модели производства партии продукции

Размер партии деталей равен q, однако, поскольку детали используются по мере их изготовления, максимальный уровень запасов q' меньше, чем q. Если выпуск деталей осуществляется с ежегодной производительностью P, а потребление — с ежегодным темпом D ($P \geq D$), то темп пополнения запасов равен ($P - D$). Как и в модели EOQ, средний уровень запаса составляет половину его максимального уровня.

Если производственный цикл длится t_1 лет, то общий объем продукции, производимой в течение цикла, определяется по формуле:

$$q = Pt_1,$$

следовательно,

$$t_1 = q/P \text{ лет.}$$

Максимальный уровень запасов равен $(P - D) \times t_1$ деталей. Подставив в данное соотношение найденное выражение для t_1, получим, что максимальный уровень запасов составляет $(P - D) \times (q/P)$ деталей. Таким образом, средний уровень запасов равен

$$\frac{(P - D)q}{2P} \text{ деталей.}$$

Теперь мы можем вывести уравнение общей переменной стоимости:

$$TC = C_s\frac{D}{q} + C_h\frac{(P - D)\,q}{2P} \text{ (ф. ст. в год).}$$

Минимальное значение TC достигается, когда

$$\frac{dTC}{dq} = 0 \text{ и } \frac{d^2TC}{dq^2} > 0,$$

$$\frac{dTC}{dq} = \frac{-C_sD}{q^2} + \frac{C_h(P-D)}{2P} \text{ и } \frac{d^2TC}{dq^2} = \frac{2C_sD}{q^3} > 0, \text{ если } q > 0 .$$

Если

$$\frac{dTC}{dq} = 0, \text{ то } \frac{C_sD}{q^2} = \frac{C_h(P-D)}{2P},$$

следовательно,

$$q_2 = \frac{2C_sDP}{C_h(P-D)}.$$

Теперь можно найти экономичный размер партии, минимизирующий общую переменную стоимость производства:

$$q = \sqrt{\frac{2C_sD}{C_h}\frac{P}{(P-D)}} = \sqrt{\frac{P}{(P-D)}} \times EBQ \text{ (см. раздел 11.2.6)}$$

□ **Пример 11.4.** На некотором станке производятся детали в количестве 2000 единиц в месяц. Эти детали используются для производства продукции на другом станке производительностью 500 единиц в месяц; оставшиеся детали образуют запас. По оценкам специалистов компании, издержки хранения составляют 20% средней стоимости запасов в год. Стоимость производства одной детали равна 2,50 ф. ст.

1. Каким должен быть размер партии деталей, производимой на первом станке, и с какой частотой следует организовывать циклы для производства этих деталей?
2. Если бы можно было снизить издержки производства до 500 ф. ст., каков был бы ответ на вопрос 1?
3. Как изменился бы ответ на вопрос 1, если бы произошло дальнейшее снижение стоимости производства до 250 ф. ст.?

Решение

1. C_s = 1000 ф. ст. за производство одной партии деталей;
   D = 500 деталей в месяц, или 6000 деталей в год;
   P = 2000 деталей в месяц, или 24000 деталей в год;
   C_h = 0,2 × 2,50 ф. ст. = 0,50 ф. ст. за единицу детали в год.

Экономичный размер партии деталей определяется по формуле:

$$q = \sqrt{\frac{2C_sD}{C_h}\frac{P}{(P-D)}},$$

следовательно,

$$q = \sqrt{\frac{2 \times 1000 \times 6000}{0,5} \times \frac{24000}{(24000 - 6000)}} = 5656,85.$$

Оптимальный размер партии составляет 5657 деталей. В практических целях полученное значение для удобства можно и далее округлять. Количество партий деталей, которое необходимо произвести в течение года, составит:

$$D / q = 6000 / 5657 = 1,06.$$

Следовательно, частота производства партий деталей равна q / D лет:

$$D / q = 5657 / 6000 = 0,94 \text{ лет, или } 11,24 \text{ месяца.}$$

Общая переменная стоимость производства определяется следующим образом:

$$TC = C_s D / q + C_h (P - D) \times q / (2P) =$$

$$= 1000 \times 6000/5657 + 0,5 \times 18000 \times 5657 \times (2 \times 24000) =$$

$$= 1060,63 + 1060,69 = 2121,32 \text{ ф. ст. в год.}$$

Если бы в течение одного года производилась только одна партия деталей объемом 6000 единиц, общая переменная стоимость составила бы:

$$TC = 1000 \times 6000/6000 + 0,5 \times 18000 \times 6000/(2 \times 24000) =$$
$$=1000 + 1125 = 2125 \text{ ф. ст. в год.}$$

Очевидно, что компания выберет тот вариант производства, в котором предусматривается выпуск в течение каждого года только одной партии деталей объемом в 6000 единиц.

2. Если стоимость организации производственного цикла снижается до 500 ф. ст., экономичный размер партии определяется следующим образом:

$$q = \sqrt{\frac{2C_s D}{C_h} \times \frac{P}{(P - D)}},$$

следовательно,

$$q = \sqrt{\frac{2 \times 500 \times 6000}{0,5} \times \frac{24000}{24000 - 6000}} = 4000,$$

Оптимальный размер партии равен 4000 деталей. Количество партий деталей, необходимое в течение года, составит $D / q = 6000 / 4000 = 1,5$. Следовательно, частота выпуска партий деталей равна q / D лет:

$$q / D = 4000/6000 = 2/3 \text{ года, или } 8 \text{ месяцев.}$$

Общая переменная стоимость производства определяется по формуле:

$$TC = C_s D / q + C_h (P-D) q / (2P) =$$
$$= 500 \times 6000 / 4000 + 0,5 \times 18000 \times 4000 / (2 \times 24000) =$$
$$= 750 + 750 = 1500 \text{ ф. ст. в год.}$$

Если можно снизить стоимость производства наполовину, то экономия общей переменной стоимости составляет 625 ф. ст. в год. В этом случае производство деталей будет осуществляться партиями по 4000 штук каждые 8 месяцев.

3. Если стоимость организации производственного цикла снижается до 250 ф. ст., то

$$q = \sqrt{\frac{2C_sD}{C_h} \times \frac{P}{(P-D)}},$$

следовательно,

$$q = \sqrt{\frac{2 \times 250 \times 6000}{0{,}5} \times \frac{24000}{24000 - 6000}} = 2828{,}43.$$

Оптимальный размер партии равен 2828 деталей. Количество партий деталей, необходимое в течение года, составит $D/q = 6000/2828 = 2{,}12$. Следовательно, частота выпуска партий деталей равна q/D лет.

$$q/D = 2828/6000 = 0{,}47 \text{ года, или } 5{,}64 \text{ месяцев.}$$

Общая переменная стоимость производства определяется по формуле:

$$\begin{aligned} TC &= C_sD/q + C_h(P-D)q/(2P) = \\ &= 250 \times 6000 / 2828 + 0{,}5 \times 18000 \times 2828 / (2 \times 24000) = \\ &= 530{,}41 + 530{,}25 = 1060{,}66 \text{ ф. ст. в год.} \end{aligned}$$

Если бы и в данном случае можно было сократить стоимость производственного цикла наполовину, то можно было бы получить дополнительную экономию общей переменной стоимости в 439 ф. ст. в год.

11.4.2. Модель планирования дефицита

Обратимся вновь к закупке продукции у внешнего поставщика. В некоторых случаях издержки хранения продукции являются гораздо более высокими, чем любые издержки, связанные с отсутствием запаса в течение небольшого промежутка времени. Можно построить модель управления запасами, в которой предусматриваются регулярные периоды, в течение которых запас отсутствует.

Возможны два случая. В первом из них спрос на продукцию, возникающий в период отсутствия запаса, остается неудовлетворенным. Администрация супермаркета, к примеру, может принять решение о снижении уровня запасов крупногабаритной продукции, которая хранится на складах, такой, например, как пакетные супы или хлебные завтраки. Это решение приведет к тому, что в каждом цикле в течение нескольких дней запасов данной продукции не будет. Из-за снижения объемов продаж и в некотором смысле потери доверия клиентов появятся определенные издержки. Администрация супермаркета вынуждена будет сопоставить эти издержки и величину экономии, полученной вследствие отсутствия запасов продукции. Можно привести и другой пример. Пусть в магазине по продаже электротоваров принимается решение о сокращении запасов определенного вида стиральных машин, так как в этих запасах замораживается большое количество капитала. Однако в данном случае, если запасов не будет, а покупателю понадобится именно эта стиральная машина, то владелец магазина, вероятнее всего, выразит готовность принять заказ покупателя и обеспечить его необходимым товаром сразу же после получения следующей партии стиральных машин. Владелец магазина понесет некоторые затраты, связанные с поддержанием системы заказов, но и в данном случае их следует сопоставить с величиной экономии стоимости хранения запасов.

Основное различие между двумя описанными случаями состоит в том, что в первом из них после получения новых поставок заказы покупателей не выполняются, следовательно, максимальный уровень запасов совпадает с размером получаемого заказа. Во втором случае часть продукции из новой поставки идет на удовлетворение заказов клиентов, поэтому максимальный уровень запасов представляет собой разницу между размером заказа и максимальным спросом, возникающим при отсутствии запасов (см. рис. 11.12 и 11.13).

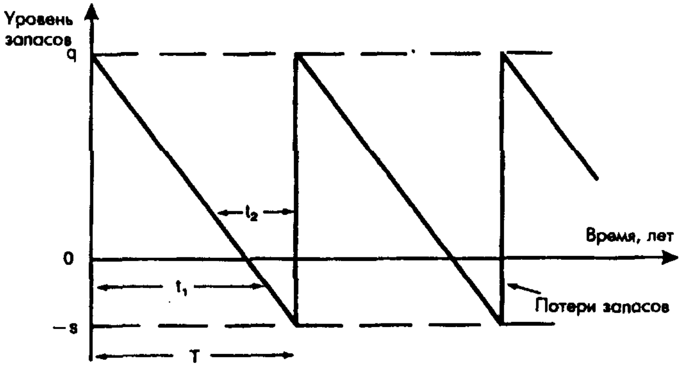


Рис. 11.12. Планирование дефицита — после получения новых поставок заказы покупателей не выполняются

Рассмотрим сначала второй случай, предусматривающий выполнение заказов покупателей. Максимальный уровень запаса представляет собой размер заказа q за вычетом максимального значения спроса в течение периода отсутствия заказа S. Следовательно, максимальный уровень запаса равен (q − S).

Общая переменная стоимость запасов за год ТС включает в себя три компоненты:

1. Годовая стоимость подачи заказов = Стоимость подачи одного заказа × Число заказов в год;
2. Годовые издержки хранения = Стоимость хранения единицы продукции в год × Средний уровень запасов;
3. Годовые издержки отсутствия запасов = Стоимость отсутствия запаса единицы продукции × Средний размер дефицита.

Введем использовавшуюся ранее систему обозначений:

C_0 – стоимость подачи одного заказа (ф. ст. за один заказ);

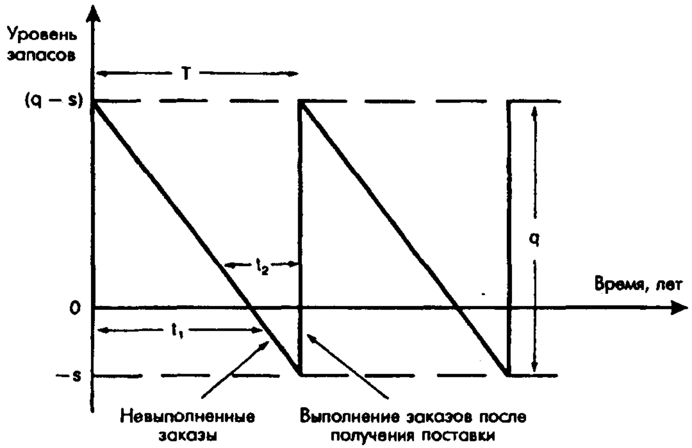


Рис. 11.13. Планирование дефицита — после получения новых поставок заказы покупателей выполняются

C_h – издержки хранения единицы продукции в течение одного года (ф. ст. в год);
D – годовой спрос;
C_b – стоимость отсутствия запаса единицы продукции в течение одного года (ф. ст. в год);
q – размер заказа;

$$TC = C_0 \times D/q + C_h \times \text{Средний размер запаса} + $$
$$+ C_b \times \text{Средний размер дефицита (ф. ст. в год).}$$

Для расчета среднего размера запасов рассмотрим один цикл запаса продолжительностью в T лет. Пусть имеющийся запас потребляется в течение t_1 лет, а в течение t_2 лет запас отсутствует: $t_1+t_2 = T$.

В период существования запаса t_1 средний уровень запаса равен $(q - S)/2$. Следовательно, на складах хранится $(q - S)/2$ единиц продукции в среднем в течение периода t_1. В итоге получаем $(q - S)t_1/2$ единиц продукции. Для оставшейся части цикла, т.е. для времени t_2 на складах хранится 0 единиц продукции; в итоге получаем $0 \times t_2$ единиц продукции. Требуется найти среднее число единиц продукции, которое хранится в запасе в течение всего цикла T. Следовательно, среднее число единиц продукции, которое хранится в запасе в течение цикла запаса, составит:

$$\frac{(q - S)\,t_1/2 + 0\,t_2}{T} = \frac{(q-S)\,t_1}{2T}.$$

Теперь мы можем выразить темп использования запасов D (единиц продукции в год) следующим образом:

$$D = (q - S)/t_1 \text{ или } D = q/T.$$

Следовательно,

$$t_1 = (q - S)/D \text{ и } T = q/D.$$

Подставив найденные соотношения для t_1 и T в формулу среднего уровня запасов в течение одного цикла, получим:

$$\frac{(q - S) \times (q - S)/D}{2q/D} = \frac{(q - S)^2}{2q}.$$

Для того чтобы рассчитать средний уровень дефицита, можно использовать описанный выше алгоритм. В течение периода t_2 средний размер дефицита составит S/2 единиц продукции, а в период t_1 его значение будет равно нулю. Поэтому среднее число недостающих единиц продукции в течение всего цикла определяется как

$$\frac{0 \times t_1 + S \times t_2/2}{T} = \frac{S\,t_2}{2T},$$

$D = S/t_2$, следовательно, $t_2 = S/D$. Таким образом, средний размер дефицита равен:

$$\frac{S \times (S/D)}{2\,q/D} = \frac{S^2}{2q}.$$

Теперь можно определить вид уравнения общей стоимости:

$$TC = C_0\frac{D}{q} + C_h\frac{(q - S)^2}{2q} + C_b\frac{S^2}{2q} \text{ ф. ст. в год.}$$

Данное уравнение отличается от полученных нами ранее тем, что оно содержит две независимые переменные: q и S. Минимальное значение TC можно найти, используя математическую процедуру дифференцирования по частям. Мы ограничимся лишь приведением ее результатов.

Оптимальный размер заказа равен:

$$q = \sqrt{\frac{2\,C_0D}{C_h}\;\frac{C_h + C_b}{C_b}} = EOQ\sqrt{\frac{C_h + C_b}{C_b}},$$

а максимальный размер дефицита составит

$$S = \sqrt{\frac{2\,C_0\,D}{C_b}\;\frac{C_h}{C_h + C_b}}.$$

Если рассматривать первый случай, в котором заказы клиентов не выполняются, то процедура анализа будет аналогична приведенному выше алгоритму, за исключением того, что максимальный размер запасов окажется равным q. Поэтому можно просто произвести замену (q − S) на q, а q — на (q+S), подставив указанные значения в формулы расчета среднего уровня запасов и среднего размера дефицита. В этом случае уравнение общей переменной стоимости примет вид:

$$TC = \frac{C_0 D}{(q + S)} + \frac{C_h q^2}{2(q + S)} + \frac{C_b q^2}{2(q + S)} \text{ ф. ст. в год.}$$

Как и в предыдущем случае, применив операцию дифференцирования по частям, можно показать, что оптимальный размер заказа определяется по следующей формуле:

$$q = \sqrt{\frac{2 C_0 D}{C_h} \frac{C_b}{C_h + C_b}} = EOQ \sqrt{\frac{C_h}{C_h + C_b}},$$

а максимальный размер дефицита составит:

$$S = \sqrt{\frac{2 C_0 D}{C_b} \frac{C_h}{C_h + C_b}}.$$

Пример 11.5. Компания "Greens Ltd" – крупный универмаг по продаже электронной и аудиоаппаратуры. Одним из наиболее популярных товаров является стереоплейер со встроенным радиоприемником. Спрос на эту продукцию, равный 2000 единиц, равномерно распределяется в течение года. Закупка плейеров у непосредственного производителя обходится универмагу в 50 ф. ст. за единицу. Стоимость подачи заказа составляет 50 ф. ст., а издержки хранения – 15% среднегодовой стоимости запасов.

Администратор компании рассматривает вопрос о сокращении запасов данной продукции, что позволило бы улучшить движение потоков наличности. По его оценке система заказов, предусматривающая отсутствие запасов, включая расходы, связанные со снижением объемов продаж и утратой доверия клиентов, составляет 5 ф. ст. в год за один плейер.

Требуется:

1. Определить минимальное значение общей переменной стоимости запасов плейеров при условии, что отсутствие запасов является недопустимым. Каков оптимальный размер заказа?
2. Найти величину экономии, которая достигается при введении системы планирования отсутствия запасов. Принимается предпосылка о покрытии размера дефицита из новых поставок. Каков оптимальный размер заказа?

Решение.

1. C_0 = 50 ф. ст. за заказ;
   D = 2000 плейеров в год;
   C = 50 ф. ст. за один плейер;
   C_h = 0,15 × 50 ф. ст. = 7,50 ф. ст. за единицу продукции в год.

Экономичный размер заказа равен:

$$q = \sqrt{2 C_0 D/C_h} = \sqrt{2 \times 50 \times 2000/7,5} = 163,3.$$

Таким образом, в течение каждого цикла заказа компания "Greens Ltd" должна подавать заказ на 163 стереоплейера.

Годовая общая переменная стоимость запасов определяется в соответствии с формулой:

$$TC = C_0 D/q + C_h q/2 \text{ (ф. ст. в год)};$$

$$TC = 50 \times 2000/163 + 7,5 \times 163/2 = 613,5 + 611,25 = 1224,75 \text{ ф. ст. в год.}$$

Плановый дефицит составит: $C_b = 5$ ф. ст. за плейер в год.

Оптимальный размер заказа равен:

$$q = \sqrt{\frac{2\,C_0 D}{C_h}\ \frac{C_h + C_b}{C_b}},$$

$$q = \sqrt{\frac{2 \times 50 \times 2000}{7,5}\ \frac{(7,5 + 5)}{5}} = 258,2.$$

В данной ситуации "Greens" должна подавать заказы на партии плейеров размером в 258 единиц. Максимальный размер дефицита равен:

$$S = \sqrt{\frac{2\,C_0\,D}{C_b}\ \frac{C_h}{C_h + C_b}},$$

$$S = \sqrt{\frac{2 \times 50 \times 2000 \times 7,5}{5\,(7,5 + 5)}} = 154,9.$$

После округления получим, что максимальный дефицит составляет 155 плейеров. Общая переменная стоимость за год определяется следующим образом:

$$TC = \frac{C_0\,D}{q} + \frac{C_h\,(q - S^2)}{2q} + \frac{C_b\,S^2}{2q},$$

$$TC = \frac{50 \times 2000}{258} + \frac{7,5\,(258 - 155)^2}{2 \times 258} + \frac{5 \times 155^2}{2 \times 258} =$$

$$= 387,6 + 154,2 + 232,8 = 774,6 \text{ ф. ст. в год.}$$

По сравнению с основной моделью величина экономии составляет 1224,75 − 774,6 = 450,15 ф. ст. в год. Таким образом, если компания "Greens" будет использовать модель планирования дефицита, она сможет достичь экономии общей переменной стоимости запасов, равной 450 ф. ст. в год.

11.5. НЕОПРЕДЕЛЕННОСТЬ И ОСНОВНАЯ МОДЕЛЬ УПРАВЛЕНИЯ ЗАПАСАМИ

Все рассмотренные нами модели основывались на предположении, что спрос и время поставки заказа являются постоянными. Однако на практике многие системы управления запасами содержат элемент неопределенности как по отношению ко времени поставки, так и относительно спроса. Нетрудно также показать, что спрос изменяется во времени, т.е. среднее значение спроса колеблется в течение года. Проблемы, связанные с неопределенностью времени поставки заказа и изменением значения спроса во времени, являются особенно сложными. В таких ситуациях вряд ли можно применять математические модели, которые использовались нами ранее. Необходимо привлечение других методов, например, имитационного моделирования (см. гл. 14). Однако, если ограничить возрастание сложности модели, вызванное неопределенностью значений времени поставки заказа или спроса, то можно построить математическую модель, достаточно верно отражающую изложенную ситуацию. Кроме того, следует все же сделать некоторые предположения, касающиеся поведения системы. Если значение спроса не определено, предполагается, что он изменяется в соответствии с характеристиками. Эти характеристики можно получить на основе эмпирических данных, содержащих фактические значения спроса, либо можно предположить, что спрос определяется стандартными статистическими моделями, например, распределением Пуассона или нормальным распределением. Если значения спроса и времени поставки изменяются, может возникнуть ситуация, когда запас будет отсутствовать. Если же уровень повторного заказа определяется только исходя из удовлетворения среднего спроса в течение среднего времени поставки заказа, отсутствие запаса может появиться во многих циклах запаса, функционирующих в течение года.

Пусть вероятность отсутствия запасов для любого цикла запаса равна 0,2. Если продукция, интересующая клиента, заказывается только один раз в год, то возможность нехватки запасов для каждого года небольшая. Математическое ожидание числа нехваток запаса в течение года рассчитывается следующим образом:

Е (число нехваток запаса в год) = Число циклов запаса в году ×
× Вероятность отсутствия запаса в каждом цикле = 1 × 0,2 = 0,2.

Однако если в течение года подача заказа производится 50 раз, то Е (число нехваток запаса за год) = 50 × 0,2 = 10.

Одинаковое значение вероятности нехватки запасов для одного цикла может соответствовать всем циклам одной системы, но механически переносить его на другие системы нельзя. Необходимо определить, когда такая вероятность нехватки запасов приемлема, а когда — нет. Чтобы это выяснить, мы должны решить, какого **уровня обслуживания** мы намерены достичь. Если вероятность нехватки запасов для одного цикла равна 0,2, т.е. 20%, то уровень обслуживания равен 80%. Если же это не так, то следует снизить значение вероятности нехватки запасов. Это можно сделать, изменив **уровень повторного заказа**. Уровень повторного заказа можно увеличить, если добавить к среднему спросу в течение среднего времени поставки размер **буферного**, или **резервного запаса**.

Уровень повторного заказа =
= Средний спрос в течение поставки заказа + Резервный запас.

Чем выше размер резервного запаса, тем ниже вероятность нехватки запасов, но выше издержки их хранения. Снижение стоимости нехватки запасов должно быть компенсировано увеличением стоимости их хранения.

Выбор соответствующего размера резервного запаса зависит от конкретной цели, которую необходимо достичь. Она может состоять в достижении минимального уровня обслуживания независимо от величины связанных с этим дополнительных затрат. С другой стороны, нехватка запаса может привести к нарушению выпуска товаров первой необходимости; она может повлечь за собой дополнительные издержки производства, закупку продукции у другого поставщика по более высоким ценам, увеличение стоимости новых заказов, меньшее удовлетворение потребителя и, как следствие, более низкий спрос. Стоимость нехватки запасов можно определить. Затем, в соответствии с критерием минимизации общей переменной стоимости запасов, можно выбрать нужное количество резервного запаса. Как правило, выделяются два типа моделей, учитывающих неопределенность:

1. **Уровневая модель повторного заказа** — заказывается фиксированное количество продукции с переменными временными интервалами, т.е. в те моменты времени, когда уровень запаса уменьшается до заранее заданного значения.
2. **Циклическая модель повторного заказа** — в фиксированные временные интервалы заказывается различное количество продукции.

Следует отметить, что основная модель (см. 11.2) сочетает в себе обе эти особенности — фиксированное количество продукции заказывается в ней с фиксированными интервалами времени.

Выбор той или иной системы определяется только изменением значений времени поставки заказа и спроса. В следующих разделах мы рассмотрим применение каждого типа моделей.

11.5.1. Уровневая система повторного заказа

МОДЕЛЬ I: ДОСТИЖЕНИЕ МИНИМАЛЬНОГО УРОВНЯ ОБСЛУЖИВАНИЯ

Необходимо принять решение по двум следующим вопросам:

1. Каково значение фиксированного размера заказа q?
2. При каком уровне запасов следует сделать новый заказ? Эта величина называется уровнем повторного заказа R.

Суть алгоритма состоит в том, чтобы с помощью модели EOQ зафиксировать размер повторного заказа, а затем на этой основе выбрать соответствующее значение уровня повторного заказа. Данный алгоритм не всегда приводит к получению наилучшего решения, однако он позволяет найти достаточно хорошее решение. Для того чтобы зафиксировать уровень повторного заказа необходимо знать, как меняется величина спроса в течение исполнения заказа и ожидаемое значение уровня обслуживания. Общее решение покажем на следующем примере.

❐ **Пример 11.6.** Промышленная компания "James" в одном из технологических процессов использует деталь X. Эти детали закупаются у внешнего поставщика. Спрос компании на детали X периодически меняется, однако приблизительно его можно описать с помощью нормального распределения со средним значением, равным 80 деталям в день. Стандартное отклонение спроса составляет 10 деталей в день. Стоимость каждой детали равна 0,50 ф. ст. Как было оценено, за каждый заказ поставщик взимает плату в 25 ф. ст. Время поставки заказа поставщиком фиксировано и составляет 8 дней. По оценкам специалистов компании "James" издержки хранения составляют 20% среднегодовой стоимости запасов. Компания работает 5 дней в неделю в течение 50 недель в году.

Какое количество деталей должна заказывать компания "James" каждый раз и каким должен быть уровень повторного заказа, если нехватка запасов в среднем более чем в 20 циклах нежелательна для компании? Каков размер резервного запаса, соответствующего данному уровню повторного заказа?

Решение.

Для определения нужного размера предположим, что спрос является постоянным и зафиксирован на уровне среднего значения.

C_0 = 25 ф. ст. за один заказ;

D = 80 × 5 × 50 = 20000 деталей в год (в среднем);

C_h = 20% от 0,50 ф. ст. = 0,10 ф. ст. за одну деталь в год.

Если предполагается, что спрос постоянен, то экономичный размер заказа определяется по следующей формуле:

$$q_0 = \sqrt{2C_0\,D/C_h} = \sqrt{2 \times 25 \times 20000/0{,}1} = 3162{,}3.$$

В качестве размера заказа примем значение, равное 3162 деталей. Максимально допустимый уровень нехватки запасов, как было задано априорно, составляет 1 из 20 циклов, т.е. в среднем только в 5% циклов допускается нехватка запасов. Следовательно, уровень обслуживания равен 95%. Поскольку спрос за день аппроксимируется нормальным распределением, спрос в течение поставки также распределен по нормальному закону при условии, что предполагается независимость спроса в любой день от его величины в другие дни.

Среднее значение спроса в течение 8 дней времени поставки составляет: 80 × 8 = 640 деталей. Дисперсия спроса в течение поставки заказа равна: 8 × дисперсия спроса за день = 8×10^2, следовательно, стандартное отклонение спроса в течение поставки составит: $\sqrt{8 \times 10^2}$ = 28,28 деталей. Распределение спроса в течение поставки показано на рис. 11.14.

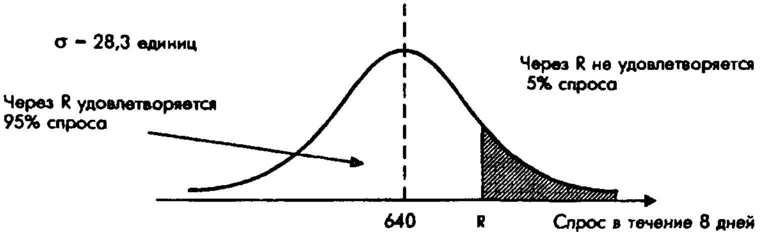


Рис. 11.14. Распределение спроса в течение поставки

Уровень повторного заказа R выбирается с тем условием, чтобы вероятность, что величина спроса в течение поставки окажется меньше уровня повторного заказа, была не менее 0,95, т.е. P (величина спроса в течение поставки < R) > 0,95.

R представляет собой z стандартных отклонений от среднего, где

$$z = \frac{R-640}{28{,}3}.$$

По таблицам стандартного нормального распределения находим, что если P (z > (R − 640)/28,3) = 0,05, то z = 1,645. Следовательно,

$$1,645 = \frac{R - 640}{28,3}$$

и

$$R = 686,55.$$

Зафиксируем уровень повторного заказа на уровне 687 деталей. Следовательно, резервный запас составит 47 деталей. Этот запас необходим для обеспечения колеблемости спроса и требуемого уровня обслуживания. Предполагается, что 47 деталей находятся в запасе в течение всего периода, таким образом, в данном случае среднегодовой уровень запаса равен $(q/2 + 47)$ деталям.

Если не учитывать величину стоимости нехватки запасов, то общая годовая переменная стоимость определяется как

$$\begin{aligned} TC &= C_0\,D/q + C_h\,(q/2 + \text{резервный запас}) = \\ &= 25 \times 20000/3162 + 0,10 \times (3162/2 + 47) = \\ &= 158,13 + 162,8 = 320,93 \text{ ф. ст. в год.} \end{aligned}$$

Стоимость резервного запаса равна: $(0,1 \times 47) = 4,70$ ф. ст. в год.

МОДЕЛЬ II: ДОСТИЖЕНИЕ МИНИМАЛЬНОЙ СТОИМОСТИ

Необходимо принять решение по тем же вопросам, которые были сформулированы для модели I, с использованием такого же алгоритма. Исследуем проблему, поставленную в примере **11.6**, с точки зрения минимизации общей переменной стоимости за год.

❑ **Пример 11.7.** Вернемся к примеру **11.6**. Если возникает нехватка запасов, то процесс производства в компании "James" останавливается, следовательно, при приближении кризиса компания посылает местному поставщику багажный фургон для закупки дополнительной партии деталей. По оценкам фирмы, дополнительная стоимость этой операции составляет 1 ф. ст. за одну деталь.

Какое количество деталей должна заказывать компания "James" единовременно и каким должен быть уровень повторного заказа, если ее целью является минимизация общей переменной стоимости за год? Каков размер резервного запаса, соответствующий данному уровню повторного заказа?

Решение

Общая переменная стоимость за год = Годовая стоимость подачи заказа +
+ Годовые издержки хранения стандартного запаса +
+ Годовые издержки хранения резервного запаса +
+ Годовая стоимость нехватки запасов.

Фиксированный размер запаса является таким же, как и в примере **11.6**, т.е. составляет 3162 деталей в одном заказе, следовательно,

$$\begin{aligned} TC &= C_0\,D/q + C_h\,q/2 + C_h \times (\text{резервный запас}) + C_b \times \\ &\times (\text{математическое ожидание единиц продукции, составляющих} \\ &\quad \text{нехватку запасов, в год}) = 25 \times 20000/3162 + 0,1 \times 3162/2 + 0,1 \times \\ &\times (\text{резервный запас}) + 1 \times (\text{математическое ожидание количества} \\ &\quad \text{единиц продукции, составляющей нехватку запасов, в год}). \end{aligned}$$

Мы должны выбрать значение резервного запаса, минимизирующее суммарное значение последних двух компонент общей стоимости. По мере увеличения резервного запаса издержки хранения также возрастают, а математическое ожидание количества единиц продукции, составляющей нехватку запаса, снижается, следовательно, снижается и стоимость отсутствия запасов, и наоборот. Необходимо определить размер резервного запаса, обеспечивающий наилучшее соотношение этих двух величин. Метод, который будет нами использоваться, основан на теории "проб и ошибок".

В данном примере спрос в течение поставки заказа аппроксимируется непрерывным распределением, следовательно, мы должны найти особую точку данного распределения, в которой следует рассчитать и сопоставить стоимость нехватки запасов и издержек хранения резервного запаса. Проверку указанных значений будем производить с интервалом в 10 деталей. Данный интервал выбирается для удобства расчетов, кроме того, поскольку детали являются относительно недорогими, шаг в 10 деталей является достаточно надежным и обоснованным.

Если спрос в течение поставки не превосходит своего среднего значения, нехватки запасов не появится. Проблемы возникают только в том случае, если значение спроса в течение поставки выше среднего.

Таблица 11.3. **Расчет вероятности различных значений спроса в течение поставки с использованием нормального распределения с $\mu = 640$ и $\sigma = 28{,}3$ единиц за 8 дней**

Приближенное значение спроса в течение поставки	*Вероятность появления этого значения**	*Резервный запас, требующийся для удовлетворения этого спроса*
640	0,135	0
650	0,134	10
660	0,109	20
670	0,082	30
680	0,052	40
690	0,030	50
700	0,016	60
710	0,007	70
720	0,003	80

*При оценке вероятности значение спроса в 640 деталей представляет промежуток от 635 до 645 деталей. Вероятность $P(635 \le \text{спрос} \le 645) = 0{,}135$ находится с помощью таблиц стандартного нормального распределения. Остальные значения рассчитываются аналогично.

Для каждого из выбранных значений резервного запаса вычисляется математическое ожидание количества нехваток запаса в течение цикла. Затем данное значение умножается на число циклов запаса в год, что дает нам математическое ожидание количества нехваток запаса в течение года. Учитывая издержки хранения дополнительного запаса (0,10 ф. ст. за единицу) и ожидаемую стоимость нехватки запасов (ф. ст. за единицу), мы можем получить ожидаемую величину общей годовой стоимости, соответствующую данному уровню резервного запаса. Как правило, эти две стоимости равномерно уменьшаются, пока не достигнут минимального значения, а затем снова начинают возрастать. Как только значения стоимости начинают возрастать, необходимость в дальнейших расчетах отпадает. Число циклов запаса в год составит: 20000/3162 = 6,3 (**Примечание**: в приведенных ниже расчетах предполагается, что вероятность того, что спрос за время поставки превысит 720 деталей, равна нулю).

Таблица 11.4. **Издержки, соответствующие различным уровням резервного запаса**

Резервный запас	*Удовлетворенный спрос*	*Математическое ожидание числа нехваток запасов*		*Стоимость, ф. ст. в год*		
		в течение цикла	*в течение года*	*Нехватки запасов*	*Резервного запаса*	*Общая*
80	720	0	0	0	80×0,10=8	8,00
70	710	10×0,003= =0,03	0,03×6,3= =0,19	0,19×1	70Ф0,10=7	7,19
60	700	20×0,003+ +10×0,007= =0,13	0,13×6,3= =0,82	0,82×1	60Ф0,10=6	6,82
50	690	30×0,003+ +20×0,007+ +10×0,016= =0,39	0,39×6,3= =2,46	2,46×1	50Ф0,10=5	7,46

Поскольку общая ожидаемая стоимость за год возрастает, можно предположить, что свое минимальное значение она принимает, когда резервный запас равен 60 деталям. Общая переменная стоимость за год равна:

TC = 25 × 20000/3162 + 0,1 × 3162/2 + 0,1 × резервный запас +
+ 1 × математическое ожидание количества нехваток запасов в год =
= 158,1 + 158,1 + 0,1 × 60 + 1 × 0,82 = 323,02 ф. ст. в год.

Это значение получено с достаточной степенью приближенности, однако вероятнее всего оно является наилучшим значением, которое можно получить довольно просто. В данном случае переменная стоимость запасов достаточно мала по сравнению со стоимостью закупки продукции (0,50 ф. ст. × 20000 = 10000 ф. ст. в год).

❐ **Пример 11.8.** Компания P&R – крупный оптовый магазин по продаже электротоваров за наличный и безналичный расчет. Магазин производит закупку телевизоров наиболее популярной марки у непосредственного производителя по цене 250 ф. ст. за единицу. Средний объем продаж за 300 дней года составляет 475 телевизоров. Подача каждого заказа обходится компании в 50 ф. ст. Было оценено, что издержки хранения составляют 15% среднегодовой стоимости запасов.

Время поставки заказа – три дня. По данным о последних 50 циклах заказа было получено следующее распределение частот для спроса:

Таблица 11.5

Спрос на телевизоры в течение поставки, шт.	0	1	2	3	4	5	6	7	8
Число циклов запаса	1	2	6	8	10	8	8	5	2

Каждый раз, когда запас товаров исчерпывается, администрация оптового магазина подает срочный заказ. Дополнительная стоимость этого заказа, включая издержки выполнения заказов покупателей, оценивается приблизительно в 20 ф. ст. за телевизор.

Какое количество телевизоров компания P&R должна заказывать единовременно и каким должен быть уровень повторного заказа в условиях, когда цель администрации компании состоит в минимизации стоимости? Насколько велик размер резервного запаса, соответствующий данному уровню повторного заказа?

Решение

Общая переменная стоимость за год = Годовая стоимость подачи заказа +
+ Годовые издержки хранения стандартного запаса +
+ Годовые издержки хранения резервного запаса +
+ Годовые издержки отсутствия запасов.

C_0 = 50 ф. ст. за заказ;
D = 475 телевизоров в среднем за год;
$C_h = 0{,}15 \times 250$ ф. ст. = 37,50 ф. ст. за телевизор в год;
C = 250 ф. ст. за телевизор;
C_b = 20 ф. ст. за телевизор.

Используя среднее значение спроса и зафиксировав размер заказа на уровне EOQ, получим:

$$q = \sqrt{2 \times 50 \times 475/37{,}5} \approx 35{,}6.$$

В качестве фиксированного размера заказа выберем значение, равное 36 телевизорам.

$$TC = C_0\, D/q + C_h\, q/2 + C_h \times (\text{резервный запас}) + C_b \times$$
$$\times\ (\text{математическое ожидание числа изделий, составляющих нехватку запасов, за год}) = 50 \times 475/36 + 37{,}5 \times 36/2 + 37{,}5 \times$$
$$\times\ \text{резервный запас} + 20 \times \text{математическое ожидание размера нехватки запасов за год} = 1334{,}72 + 37{,}5\ (\text{резервный запас}) + 20 \times$$
$$\times\ (\text{математическое ожидание размера нехватки запасов за год})\ (\text{ф. ст. в год}).$$

Необходимо определить такой размер резервного запаса, при котором достигается минимум двух последних видов издержек. Если спрос в течение поставки не превышает среднего значения, нехватка запасов отсутствует. Проблема возникает только в том случае, когда спрос в течение поставки превышает его средний уровень.

Средний спрос за день составляет: 475/300 = 1,58 телевизоров. Таким образом, средний спрос в течение поставки определяется как 1,58 × 3 = 4,75. Допустив некоторую погрешность в округлении, решим, что спрос равен 4 телевизорам. Распределение вероятностей значений спроса за время поставки можно найти из соответствующего распределения частот:

Таблица 11.6. **Распределение вероятностей значений спроса за период поставки**

Спрос на телевизоры за время поставки, шт.	0	1	2	3	4	5	6	7	8
Число циклов запаса	1	2	6	8	10	8	8	5	2
Вероятность	0,02	0,04	0,12	0,16	0,20	0,16	0,16	0,10	0,04

Таблица 11.7. **Размер резервного запаса, необходимый для соответствующих значений спроса за время поставки**

Спрос в течение поставки	Вероятность появления такого спроса	Резервный запас, необходимый для удовлетворения данного спроса
4	0,20	0
5	0,16	1
6	0,16	2
7	0,10	3
8	0,04	4

Для каждого из указанных значений резервного запаса вычислим математическое ожидание размера нехватки запасов в течение одного цикла. Затем, умножив полученное значение на количество циклов в течение года, получим математическое ожидание размера нехватки запасов в течение года. Для того чтобы получить ожидаемое значение общей стоимости, соответствующее данному размеру резервного запаса, необходимо принять во внимание издержки хранения дополнительного запаса (37,5 ф. ст. за единицу продукции) и расходы, связанные с нехваткой запасов (20 ф. ст. за единицу продукции). Количество циклов запаса составит за год: 475/36 = 13,2.

Таблица 11.8. **Значения стоимости, соответствующие различному размеру резервного запаса**

Резервный запас	Удовлетворенный спрос	Математическое ожидание нехватки запасов		Стоимость, ф. ст. в год		
		в течение цикла	в течение года	Нехватки запасов	Резервного запаса	Общая
4	8	0	0	0	4×37,5=15	150,0
3	7	1×0,4= =0,4	0,04×13,2= =0,528	0,528×20= =10,56	3×37,5= =112,5	123,1
2	6	2×0,04+ +1×0,01= =0,18	0,18×13,2= =2,376	2,376 × 20= = 47,52	2 × 37,5 = =75	122,5
1	5	3×0,04+2× ×0,1+1×0,16= =0,48	0,48×13,2= =6,336	6,336 × 20= = 126,72	1 × 37,5 = =37,5	164,2

Ожидаемое значение общей стоимости возрастает, следовательно, можно предположить, что ее минимум достигается в случае, когда резервный запас состоит из 2 телевизоров. Если предположить, что среднее значение спроса в течение времени поставки равно 4, уровень повторного заказа составит: 4+2 = 6 телевизоров. В этом случае общая переменная стоимость за год будет равна:

TC = 1334,72 + 37,5 × (резервный запас) + 20 ×
+ (математическое ожидание размера нехватки запаса за год) =
= 1334,72 +75,0 + 47,52 = 1457,24 ф. ст. в год.

Чтобы минимизировать годовой показатель общей переменной стоимости запасов, администрация компании должна периодически заказывать партии телевизоров объемом 36 штук, когда уровень запасов снижается до 6 единиц.

11.5.2. Циклическая система повторного заказа

Основная циклическая модель повторного заказа предназначена для принятия решений по следующим двум вопросам:

1. Каковы границы фиксированного интервала, в котором следует осуществлять подачу заказа?
2. Какое количество продукции необходимо заказывать?

Как и в предыдущем случае, рассмотрим данную задачу в два этапа, что позволит нам получить достаточно хорошее, однако не обязательно оптимальное, решение. Зафиксируем продолжительность цикла T, не принимая во внимание колеблемость значений спроса и времени поставки заказа. Значение T следует округлить до соответствующей величины. Система управления запасами должна быть построена таким образом, чтобы ею можно было легко управлять, поэтому совершенно нежелательно, чтобы лицу, осуществляющему управление запасами, приходилось проводить проверку запасов с неудобными для него интервалами времени. Критерием выбора размера заказа должна служить цель создания системы управления запасами. Как и в предыдущем случае, исследуем данную проблему с точки зрения минимального уровня обслуживания и минимальной стоимости.

МОДЕЛЬ I: ДОСТИЖЕНИЕ МИНИМАЛЬНОГО УРОВНЯ ОБСЛУЖИВАНИЯ

Для определения фиксированного интервала повторного заказа, не учитывая каких-либо изменений значений спроса или времени поставки, найдем интервал повторного заказа, в котором достигается минимальное значение общей переменной стоимости подачи заказа и хранения запасов:

$$\text{Общая переменная стоимость за год} = \text{Годовая стоимость подачи заказа} + \text{Годовые издержки хранения}$$

Если интервал повторного заказа равен T лет, число подаваемых за год заказов составит 1/T. Размер каждого заказа равен q, где D = q/T, следовательно, q = DT. Если не учитывать резервного запаса, то средний уровень запаса составит q/2 = DT/2. Таким образом, общая переменная стоимость за год определяется по следующей формуле:

$$TC = C_0 \times (1/T) + C_h \times (DT/2) \text{ (ф. ст. в год)}.$$

Минимум TC достигается, если

$$\frac{dTC}{dT} = 0 \text{ и } \frac{d^2TC}{dT^2} > 0$$

$$\frac{dTC}{dT} = \frac{-C_0}{T^2} + \frac{C_h D}{2} \text{ и } \frac{d^2TC}{dT^2} = \frac{2C_0}{T^3} > 0, \text{ если } T > 0 .$$

Если

$$\frac{dTC}{dT} = 0, \quad \frac{-C_0}{T^2} + \frac{C_h D}{2} = 0,$$

следовательно,

$$T = \sqrt{\frac{2C_0}{C_h D}} \text{ ; сравните с } EOQ = \sqrt{\frac{2C_0 D}{C_h}}.$$

После того как значение T найдено, производится его корректировка в соответствии с наиболее удобным интервалом проверки наличия запасов. Если, например, расчеты показали бы, что T = 4,2 дня, найденное значение было бы скорректировано на интервал проверки запасов, равный одной неделе.

Теперь мы должны найти уровень запасов, который будет определять размер подаваемого заказа. Например, можно принять решение, что размер заказа на момент его подачи должен быть выбран таким образом, чтобы уровень запасов увеличился до 100 единиц продукции при условии, что поставка заказа осуществляется незамедлительно. Следовательно, если уровень запасов равен 35, размер заказа будет равен 65, если же уровень запасов равен 43, размер заказа составит 57 единиц продукции.

❐ **Пример 11.9.** Предположим, что для некоторого вида продукции уровень обслуживания совпадает с размером одной нехватки продукции при условии, что цикл повторного заказа составляет четыре рабочих недели. Предположим также, что год состоит из 50 рабочих недель. Зафиксируем время поставки заказа на уровень двух недель. Спрос на данную продукцию в неделю аппроксимируется нормальным распределением, среднее значение которого равно 300 единицам продукции в неделю, а стандартное отклонение — 50 единиц продукции в неделю.

Решение

Число циклов запаса в течение года составит: 50/4 = 12,5. Вероятность нехватки запаса в течение цикла определяется как 1/12,5 = 0,08. Следовательно, уровень обслуживания, которого необходимо достичь, равен 0,92.

Переменный спрос, который нужно учесть в процессе решения, — это спрос, предъявляемый с момента принятия решения о подаче заказа до момента получения новой партии повторного заказа, т.е. спрос, возникающий в течение всего цикла повторного заказа, а также спрос в течение поставки, как было в уровневой модели повторного заказа. Предположим, что распределение спроса в течение 6 недель (продолжительность цикла — 4 недели плюс время поставки заказа – 2 недели) является нормальным и имеет среднее значение: $6 \times 300 = 1800$ единиц продукции, и стандартное отклонение: $\sqrt{6 \times 50^2} = 122{,}5$ единиц продукции. Соответствующий график распределения спроса см. на рис. 11.15.

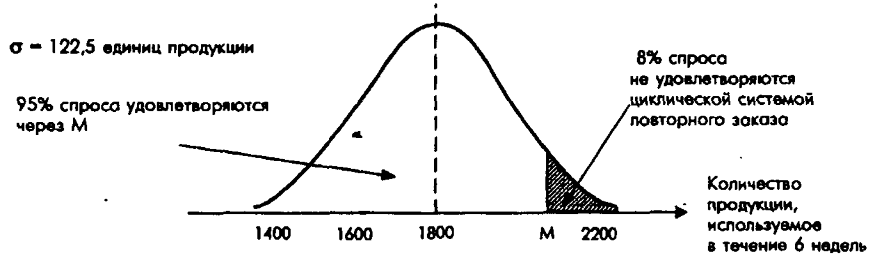


Рис. 11.15. Изменение спроса в течение повторного заказа и времени поставки

Размер заказа выбирается таким образом, чтобы уровень запасов возрос до величины M; M, в свою очередь, выбирается так, чтобы вероятность удовлетворения спроса в продолжении цикла запаса составляла 92%. M представляет собой z стандартных отклонений от среднего, где

$$z = \frac{M - 1800}{122{,}5}.$$

Следовательно, из таблицы для стандартного нормального распределения находим, что при $P(z > (M - 1800/122{,}5) = 0{,}08$; $z = 1{,}405$.

Таким образом,

$$1{,}405 = \frac{M - 1800}{122{,}5}.$$

Следовательно,

$$M = 1800 + (1{,}405 \times 122{,}5) = 1972{,}1.$$

Итак, во время каждой проверки наличия запасов, проводимой один раз в 4 недели, будет сделан новый заказ, размер которого позволит увеличить уровень запасов до 1972 единиц продукции, при условии незамедлительного получения заказа. Такая политика позволит обеспечить уровень обслуживания, равный 92%, или в среднем одну нехватку запасов в год.

МОДЕЛЬ II: ДОСТИЖЕНИЕ МИНИМАЛЬНОЙ СТОИМОСТИ

Алгоритм, который применялся в модели I, можно использовать также и для определения наиболее приемлемой продолжительности цикла повторного заказа. Уровень запасов M, при котором достигается минимум общей переменной стоимости за год, можно определить по аналогии с методом, рассмотренным нами в разделе **11.5.1**, в котором определялся размер необходимого резервного запаса. Используя данные примера **11.8**, определим фиксированный интервал повторного заказа:

$C_0 = 50$ ф. ст. за заказ;

D = 475 телевизоров в среднем за год;
C_h = 0,15 × 250 ф. ст. = 37,5 ф. ст. за телевизор в год;
C = 250 ф. ст. за телевизор;
C_b = 20 ф. ст. за телевизор;
L = 3 дня.

Продолжительность рабочего года = 300 дней.

Оптимальный интервал повторного заказа определяется следующим образом:

$$T = \sqrt{\frac{2C_0}{C_h D}} = \sqrt{\frac{2 \times 50}{475 \times 37{,}5}} = 0{,}07 \text{ года.}$$

Оптимальный интервал повторного заказа составляет: 0,07 × 300 = 21 рабочий день. Предположим, что рабочий год продолжительностью в 300 дней состоит из 6-дневных рабочих недель, тогда наиболее приемлемым для подачи повторных заказов будет интервал, равный 4 неделям. Размер заказа, определяемый каждый раз в момент его подачи, должен быть таким, чтобы уровень запасов возрос до величины M при условии незамедлительного получения заказа, где M минимизирует издержки хранения резервного запаса и стоимость нехватки запасов за год. Размер резервного запаса определяется как:

B = (M − среднее значение спроса в течение поставки и цикла повторного заказа).

Цикл повторного заказа составляет 4 × 6 рабочих дней, а время поставки — 3 рабочих дня. Поэтому необходимо учитывать спрос, возникающий в течение 27 рабочих дней. Нам известно, что годовой спрос равен 475 телевизорам за 300 рабочих дней, поэтому среднее значение спроса за 27 дней составит: (475/300) × 27 = 42,75 телевизоров. Резервный запас равен M − 42,75, а издержки его хранения за год — (M − 42,75) × 37,5 ф. ст. в год. Ожидаемые издержки, связанные с отсутствием запаса в течение года, зависят от колеблемости спроса в течение исследуемых 27 дней. К сожалению, мы не можем произвести расчеты ввиду недостатка информации, имеющейся в нашем распоряжении. На практике нам бы пришлось сгенерировать соответствующее распределение и проверить его надежность, собрав дополнительные данные. После того как этот этап закончится, мы могли бы приступить к расчетам, в которых используются те же методы, что и в разделе **11.5.1**, за исключением того момента, что проведение этих расчетов потребовало бы гораздо больше времени.

11.6. ДРУГИЕ АСПЕКТЫ ТЕОРИИ УПРАВЛЕНИЯ ЗАПАСАМИ

Большинство систем управления запасами, используемых на практике, включает в себя сотни и даже тысячи наименований продукции. Типичными примерами являются крупный универмаг или завод-изготовитель. Совершенно естественно, что в подобной ситуации различные виды продукции будут или должны использоваться по-разному. Поэтому целесообразно ограничить исследование теми товарами, которые обладают высокой годовой стоимостью, не принимая во внимание продукцию с низкой годовой стоимостью. Одним из способов практической реализации этого положения является составление списка всех видов продукции, представляющей собой запасы, в порядке убывания годовой стоимости ее продажи. Весьма вероятно, что в данном списке появится **эффект Парето**, т.е. около 20% товаров составят 80% общей стоимости. Именно этим 20% видов продукции следует уделить первоочередное внимание, поскольку, как ожидается, эти товары позволят получить наибольшую отдачу от исследований в области моделирования систем управления запасами. Внутри рассматриваемой группы товаров, имеющих высокую годовую стоимость, можно выделить различные виды продукции. Одни товары попадают в данную группу, поскольку они используются в достаточно больших количествах, другие — ввиду довольно высокой стоимости единицы продукции.

Например, при обслуживании авиалиний используется огромное количество топлива, один литр которого стоит всего несколько пенсов, однако годовой показатель стоимости топлива достаточно высок из-за большого объема его потребления. Кроме этого, имеются запасные двигатели, годовая стоимость которых также высока, поскольку каждый из двигателей является достаточно дорогим. И тот, и другой товар вероятнее всего окажутся в числе наименований продукции из списка, составляющих 20% товаров с наиболее высокими показателями годовой стоимости, однако, способы моделирования задач управления запасами для каждого из указанных товаров будут различны.

Проблемы, связанные с наличием нескольких видов продукции, могут еще более осложняться при ограничении на складские мощности или финансовые ресурсы, отпущенные на создание запасов продукции. Практически любой магазин — это крупный склад. Наличие прилавков или свободной площади является лимитирующим фактором, строго определенным с точки зрения планировки каждого конкретного магазина. В этом случае администрация должна решить, какое пространство следует выделить для каждого вида продукции. Исследования по совершенствованию систем управления запасами и изучения спроса потребителя в крупных универмагах ведутся уже в течение многих лет. Например, полосные коды, автоматизированные контрольно-кассовые пункты, микрокалькуляторы — все это результаты усилий администрации магазинов по обеспечению наличия на своих местах необходимых товаров в определенные моменты времени. Мы рассматривали в основном модели, в которых минимизируется общая годовая переменная стоимость запасов, а между тем различные торговые предприятия чаще всего организуют работу своих магазинов таким образом, чтобы получать максимальную прибыль.

Большинство систем управления запасами включает в себя сразу несколько магазинов, например, центральный универмаг, осуществляющий поставки продукции в более мелкие, подчиненные ему магазины. В данной ситуации администрации приходится принимать решение о том, какие товары следует хранить и продавать только в центральном универмаге, какие — только в мелких магазинах, а какие и в центральном, и в подчиненных ему магазинах. Кроме того, администрация центрального универмага должна решить, в каком объеме и с какой частотой следует заказывать каждый вид товаров. Необходимо сопоставить издержки хранения запасов на различных уровнях с административными и транспортными расходами, связанными с частой доставкой товаров от центрального универмага в подчиненные ему магазины. Математическую модель, описывающую подобного рода проблемы, можно построить только при условии принятия достаточно большего числа упрощающих предположений. Если система управления запасами является столь сложной, гораздо более полезными при ее моделировании могут оказаться не математические модели, рассмотренные в данной главе, а имитационные методы.

РЕЗЮМЕ

Простейшая модель управления запасами основывается на уравнении стоимости запасов для определенного периода времени:

$$TC = \text{Стоимость подачи заказа} + \text{Издержки хранения (ф. ст. в единицу времени)} = \frac{C_0 D}{q} + \frac{C_h D}{2}.$$

В данной модели предполагается полный контроль за всеми аспектами системы управления запасами. Экономичный размер заказа, которому соответствует минимальное значение TC, рассчитывается как:

$$EOQ = \sqrt{\frac{2C_0 D}{C_h}}.$$

Существует множество способов модификации основной модели для применения ее в различных ситуациях. Включив в уравнение общую стоимость покупки, можно оценить влияние скидок на количество. Используя стоимость C_b, характеризующую затраты при нехватке единицы продукции на определенный момент времени, в модель можно ввести показатель плановой нехватки запасов. Нехватку можно исследовать либо с точки зрения потерянных заказов, либо с точки зрения удовлетворения заказов по мере получения новых поставок продукции.

Данная модель применима не только к описанию процесса получения заказов у внешнего поставщика, но и к описанию процесса серийного производства продукции и создания запасов в рамках отдельного предприятия. В данном случае в качестве EOQ выступает экономичный размер партии EBQ, а стоимость подачи заказа C_0 заменяется на стоимость организации производственного цикла C_s. В некоторых случаях, после того как произведена партия продукции, часть ее остается на складах, образуя тем самым запасы, а оставшаяся часть непосредственно используется в производственном процессе. Основная модель управления запасами может быть адаптирована к данной ситуации. Введение в модель неопределенности (приближение ее к реальной действительности) усложняет процесс ее решения и исследования. При изучении неопределенности спроса и времени поставки заказа могут быть разработаны различные виды моделей, соответствующие различным целям создания систем управления запасами.

В уровневой системе повторного заказа определяется уровень запасов R, на котором производится заказ строго определенного количества продукции EOQ. В циклической системе повторного заказа определяется фиксированный интервал времени T, в течение которого при заказе различного количества продукции уровень запасов поднимается до величины M. В каждой из этих систем можно использовать один из двух критериев: либо достижение заранее заданного уровня обслуживания, либо минимизация общей стоимости подачи заказов, хранения и отсутствия (нехватки) запасов.

УПРАЖНЕНИЯ

Упражнение 11.1

Компания с ограниченной ответственностью "Dekkers" занимается розничной продажей электротоваров. Одним из видов продукции являются калькуляторы. Спрос на них составляет 25 калькуляторов в неделю, причем его величина равномерно распределяется в течение недели. Компания производит закупку калькуляторов по 9 ф. ст. за единицу. Стоимость подачи одного заказа составляет 15 ф. ст., а издержки хранения — 50 пенсов за единицу среднего размера запаса в течение года плюс 15% среднегодовой стоимости запасов. Предполагается, что в году 50 недель.

Требуется:

1. Найти оптимальный размер заказа.
2. В настоящее время администрация "Dekkers" заказывает калькуляторы партиями в 300 штук. Какой будет величина экономии, если заказы будут подаваться в соответствии с размером, найденным в п.1?
3. Если бы стоимость подачи одного заказа снизилась до 5 ф. ст., каким образом администрация компании изменила бы решение, принятое в п.1?

Упражнение 11.2

Некоторой фирме необходимо иметь в своем штате 1000 инженеров, темп увольнения которых с работы является постоянным и составляет 150 человек в год. Перед тем как приступить к работе, вновь принятые инженеры объединяются в группы и проходят обучение на специальных курсах, организуемых компанией. Проведение каждого цикла обучения обходится компании в 25000 ф. ст. Если нет возможности предоставить инженерам работу немедленно, то компания теряет 500 ф. ст. на человека в месяц.

Требуется:

1. Определить, сколько инженеров следует принимать на каждый курс обучения?
2. С какой частотой следует организовывать подобные курсы? Каково годовое значение общей переменной стоимости обучения инженеров?
3. Как повлияет ограничение количества инженеров, обучающихся в течение одного цикла, до 25 человек на решение, полученное в п. 2?

Упражнение 11.3

Компания "Systems" — крупная консалтинговая фирма по компьютерным системам в бизнесе. Фирме необходимо иметь диски под системные программы. Покупка дисков осуществляется у внешнего поставщика и, как было оценено, в ближайшем будущем использование дисков составит 20000 штук в год. Стоимость подачи одного заказа на партию дисков равна 32 ф. ст. По оценкам специалистов фирмы годовые издержки хранения одного диска составляют 1% его стоимости. Стоимость каждого диска равна 0,80 ф. ст. Предполагается, что коэффициент использования дисков является постоянным; отсутствие запасов недопустимо.

Требуется:

1. Определить оптимальный размер одного заказа и количество заказов, которое следует подавать в течение года.

2. Найти соответствующее значение годовой стоимости запасов.
3. Предположим, что оценка спроса оказалась заниженной, и фактическое значение спроса составило 24200 дисков в год. Как при этом условии повлияет сохранение размера заказа, найденного в п.1 и по-прежнему удовлетворяющего спрос, на решение задачи по сравнению с использованием нового оптимального значения уровня заказа?
4. Воспользовавшись результатами п.3, сформулируйте выводы о чувствительности данной модели к изменениям спроса.

Упражнение 11.4

Объем продаж демонстрационного зала автомобилей составляет 200 автомашин в год. Стоимость подачи каждого заказа равна 500 ф. ст., а издержки хранения – 30% среднегодовой стоимости запасов. Если размер заказа меньше, чем 50 автомобилей, то цена покупки одного автомобиля составляет 6000 ф. ст. Для заказов, размер которых колеблется от 50 до 99 автомашин, предоставляется скидка на закупочную цену в 1,5%, а заказам, размер которых составляет 100 и более автомобилей, соответствует скидка, равная 3%.

Требуется:
1. Определить размер заказ.
2. Как повлияет на ответ, полученный в п.1, тот факт, что поставщик увеличит размер скидки с 3 до 5%?

Упражнение 11.5

Небольшой магазин, специализирующийся на продаже слесарных станков, продает в среднем за неделю 3 станка определенного вида. Можно предположить, что значение спроса за неделю подчиняется распределению Пуассона. Время поставки заказа от поставщика является фиксированным и составляет 2 недели. Закупка каждого станка обходится магазину в 40 ф. ст. Стоимость подачи одного заказа – 50 ф. ст. Издержки хранения составляют 30% среднегодовой стоимости запасов, а расходы, связанные с нехваткой запасов, – 100 ф. ст. за каждый станок. Предполагается, что год состоит из 50 недель.

Определите как должна действовать администрация магазина, если цель ее состоит в минимизации общей переменной стоимости запасов станков данного вида за весь год.

Упражнение 11.6

Некоторая компания производит определенный товар, годовой спрос на который равен 4800 единицам. Было оценено, что издержки хранения составляют 20 ф. ст. за единицу товара в год, а подача одного заказа независимо от размера обходится компании в 30 ф. ст. Величина спроса в течение фиксированного времени поставки не постоянна, однако ее можно достаточно точно аппроксимировать нормальным распределением, среднее значение которого равно 100 единицам товара, а стандартное отклонение – 10 единицам. Было оценено, что стоимость нехватки запасов в течение времени поставки составляет 10 ф. ст. на единицу товара.

Требуется:

1. Определить экономичный размер заказа.
2. Найти уровень повторного заказа и размер резервного запаса при условии, что администрация компании намерена обеспечить 3%-ную вероятность нехватки запасов в течение каждого цикла заказа.
3. Если менеджер компании установит уровень повторного заказа, равный 115 единицам товара, каким будет соответствующее значение вероятности нехватки запаса в течение каждого цикла заказа? При условии, что используется данный уровень повторного заказ, сколько раз можно ожидать возникновение нехватки запаса в течение года?
4. Определить уровень повторного заказа, при котором ожидаемые издержки являются минимальными.

Упражнение 11.7

Компания с ограниченной ответственностью Oxydon Office Supplies — современная фирма, имеющая в своем штате оптовиков, торгующих бумажными и канцелярскими принадлежностями, которая работает в течение 50 недель в году и специализируется на розничной продаже общего оборудования для офисов. В число ее покупателей входят финансовые институты, юридические организации и страховые компании. Однако непрерывно возрастающие текущие затраты истощили финансовые резервы компании, что побудило главного бухгалтера разработать рекомендации по сокращению общего объема запасов. Если ранее запасов компании по различным наименованиям продукции хватало более чем на 12 месяцев, что позволяло гарантировать наличие того или иного товара в любой момент, то в настоящее время для обеспечения ликвидности возникла потребность в сокращении уровня запасов.

Продукцией, составляющей основной объем продаж компании, является высококачественная бумага для пишущих машин, тенденция изменения спроса на которую достаточно неустойчива, однако можно предположить, что спрос имеет нормальное распределение со средним значением 800 упаковок в неделю и стандартным отклонением 250 упаковок в неделю. Поставщиком этой бумаги является компания "Tiara Paper", закупочная цена одной упаковки составляет 2,50 ф. ст. Было установлено, что за последнее время срок доставки бумаги от поставщика был относительно постоянным и равнялся 3 неделям.

По оценкам специалистов компании, годовые издержки хранения составляют 15% стоимости запасов соответствующего вида продукции и рассчитываются на основе общей стоимости запасов и величины общего капитала компании. При оценке стоимости поставки бумаги компанией "Tiara" одновременно учитывается стоимость подачи и получения заказа, кроме того, решение задач, связанных с учетом и управлением запасами, требует в общей сложности около 12 чел.-ч, причем средняя норма заработной платы при 40-часовой рабочей неделе составляет 160 ф. ст. в неделю.

Требуется:

1. Выработать основные принципы политики управления запасами и объяснить, почему правильная политика управления запасами имеет для компании "Oxydon" большое значение.
2. Определить для данного вида продукции экономичный размер заказа и среднюю величину интервала пополнения запасов.
3. Определить рекомендуемый уровень повторного заказа при условии, что вероятность появления нехватки запасов в любом цикле запаса составляет не более 1%.
4. Используя найденные значения экономичного размера заказа и уровня повторного заказа, определить общую величину годовых издержек хранения (издержки хранения как таковые и стоимость поставки).

(АССА, декабрь 1987 г.).

Упражнение 11.8

Корпорация "Plutonic Pharmaceutical" для изготовления большинства видов своей продукции использует специфический химикат, который хранится в специальных рефрижераторных установках, за аренду которых компания платит 40 ф. ст. ежемесячно. Величина спроса на данный химикат является относительно постоянной и составляет около 1000 л в месяц. В настоящее время компания арендует одну рефрижераторную установку, вместимость которой равна 1000 л, поэтому подача повторных заказов производится ежемесячно в тот момент, когда уровень запасов опускается до нуля. Процедура пополнения запасов предусматривает очистку и стерилизацию рефрижераторной установки, что обходится компании в 50 ф. ст.

Вследствие расширения компанией ассортимента выпускаемой продукции ожидается, что спрос на данный химикат увеличился до 2500 л, поэтому главному бухгалтеру поручили разработать рекомендации по проведению соответствующей политики закупки и хранения запасов химиката. Теоретически возможно увеличить запасы, однако это повлечет за собой повышение стоимости аренды рефрижераторных установок на 40 ф. ст. в месяц за каждую дополнительную единицу. Однако в данном случае можно получить некоторую экономию на стоимости очистки и стерилизации, поскольку ее увеличение составит только 25 ф. ст. на каждую единицу, привлекаемую дополнительно.

Стоимость 1 л химиката равна 5 ф. ст., а темпы роста капитала компании составляют 24% в год.

Требуется:

1. Доказать, что существующая на настоящий момент политика, при которой производится ежемесячная подача заказов на 1000 л химиката, является наиболее выгодной в условиях существующего спроса и наличия только одной рефрижератоной установки. Каково значение общей годовой стоимости, соответствующее данному уровню запасов химиката?
2. В условиях предполагаемого увеличения спроса определить, целесообразно ли компании арендовать дополнительную рефрижераторную установку, если ее целью является минимизация общей стоимости запасов.
3. Показать, что аренда второй рефрижераторной установки целесообразна лишь в том случае, если величина спроса возрастает до 7200 л в месяц.

(АССА, июнь 1987 г.).

Упражнение 11.9

1) Объясните значение терминов:
   а) детерминированный;
   б) стохастический применительно к теории планирования и управления запасами.
2) Менеджер крупного магазина спортивных товаров, который открыт в течение 50 недель в году, имеет в своем распоряжении некоторый запас высококачественных шаров для гольфа. Несмотря на то, что менеджер вынужден производить закупку коробок для хранения шаров по 9,60 ф. ст. за единицу, причем вместимость каждой коробки составляет 12 шаров, он намерен осуществлять продажу шаров для гольфа поштучно. В течение последнего года средний объем продаж составлял 12 коробок шаров в неделю, причем вероятнее всего данный объем продаж сохранится и в ближайшем будущем.

Как было оценено, стоимость получения каждого заказа, включая затраты на телефонные переговоры, а также административные и транспортные расходы, составляет 16 ф. ст. в месяц. Годовая стоимость хранения запасов в соответствии с проведенными оценками составляет 20% общей стоимости запасов данного товара и рассчитывается на основе общей стоимости складских помещений и темпов роста капитала компании. Менеджер магазина определяет величину цены единицы товара как сумму стоимости покупки и приблизительной величины издержек хранения (стоимость складских помещений и поставки товаров), приходящихся на единицу продукции, а затем делает торговую накидку, составляющую 50% полученной стоимости.

Требуется:

а) Определить оптимальное число коробок под шары для гольфа, которое должен заказывать менеджер в одной партии продукции, и оптимальное количество заказов в течение года.

Показать, что цена продажи одного шара, соответствующая данной оптимальной политике, составляет 1,24 ф. ст.

б) Поставщик предоставляет 4%-ную скидку на цену каждой коробки под шары для гольфа, если менеджер подает заказ на партию коробок размером не менее 500 штук (можно предположить, что цена не оказывает влияния на спрос).

Показать, является ли данная скидка экономически выгодной для потребителя, через цену продажи, устанавливаемую магазином.

в) Какую скидку должен предоставлять поставщик на заказ размером в 500 коробок, чтобы она была выгодна магазину как потребителю?

(АССА, июнь 1990 г.).

Упражнение 11.10

Компания "Krispy Crisps Ltd" ежегодно осуществляет закупку крупной партии деревянных поддонов, которые используются при хранении и транспортировке продукции для предотвращения возможных потерь или повреждений изделий во время перевозки. Среднегодовой спрос в течение последних двух лет составил

3000 поддонов, причем можно предположить, что в течение данного года спрос не изменится. Потребность в поддонах, обеспечивающих сохранность продукции, является относительно постоянной, стоимость подачи и оформления заказа равна 15 ф. ст. Политика управления запасами, которая традиционно применялась в "Krispy Crisps", предусматривает, что издержки хранения единицы продукции составляет 18% ее закупочной цены. Номинальная цена, которую устанавливает компания-производитель, равна 8 ф. ст. за поддон.

Требуется:

1. Определить оптимальный размер заказа и интервал времени между двумя последовательными подачами заказа.
2. Описать систему предпосылок, которая была принята в п.1, и оценить их обоснованность в рамках данного вопроса.
3. Производитель предоставляет скидку в 3,125%, если "Krispy Crisps Ltd" подает заказ не менее чем на 2000 поддонов единовременно. Покажите, что скидка данного размера не является экономически выгодной для компании. Какой процент скидки необходимо предоставлять компании при условии, если она подаст заказ на 2000 или более поддонов единовременно?
4. Определить воздействие на политику управления запасами, проводимую компанией и описанную в п.1, того факта, что время поставки поддонов становится переменным.

(АССА, декабрь 1988 г.).

Упражнение 11.11

Администрация компании "Leandler Products Ltd проявляет все возрастающий интерес к капиталу, замороженному в запасах, особенно в форме произведенных деталей. Все детали выпускаются в машинном цехе партиями и при проверке отчетности по запасам было обнаружено, что уровень запасов одной из деталей XT/24 является необоснованно высоким. Данная деталь выпускается на специальном станке, управление которым осуществляет оператор, причем размер его основного заработка — 96 ф. ст. за 40-часовую рабочую неделю. Кроме стоимости работы оператора, которую можно считать постоянной, при работе станка возникают определенные эксплуатационные издержки, составляющие 1,40 ф. ст. в час.

Ожидаемый спрос на деталь XT/24 составит 50 единиц в неделю, причем значительные его изменения маловероятны. Стоимость сырья, необходимого для выпуска одной детали, равна 7,40 ф. ст., а время производства единицы детали составляет 30 мин, причем подготовка станка к выпуску деталей данного вида занимает 6 ч. Учитывая труд специалиста, обслуживающего данный станок, было оценено, что стоимость последнего равна 10 ф. ст. в час.

По окончании производственного цикла все детали поступают на специальный склад, за который несет ответственность товаровед, заработок которого является фиксированным и составляет 80 ф. ст. в неделю. Получение страховки на данный склад обходится компании в 230 ф. ст. в год, общая сумма других видов расходов (все они фиксированные) равна 2600 ф. ст. в год. Продолжительность одной рабочей недели равна 5 дням, компания работает в течение 50 недель, а темпы роста ее капитала составляют 18% годовых.

Требуется:

1. Определить, указывая причины, какие из приведенных выше видов затрат необходимо учитывать при нахождении оптимального размера производственной партии и уровня запасов детали ТХ/24.
2. Формула для определения "экономичного размера производственной партии" деталей любого вида, которые выпускаются с коэффициентом R, а потребляются с коэффициентом D в неделю, имеет следующий вид:

$$Q\sqrt{\frac{R}{R-D}},$$

где Q — обычный "экономичный размер заказа" в условиях незамедлительного пополнения запасов. В предположении, что процесс производства всех деталей начинается, когда уровень запасов достигает нуля, определить оптимальный размер производственной партии деталей ХТ/24 и максимальный уровень запасов деталей данного вида.
3. В предположении, что все детали потребляются из запаса в соответствии с принципом "одна за другой, в порядке поступления", определить максимальное и минимальное значения срока, в течение которого каждая единица детали составляет запас.

(АССА, июнь 1988 г.).

Часть 4

МОДЕЛИРОВАНИЕ В БИЗНЕСЕ

В части IV мы продолжим рассматривать применение методов моделирования в принятии решений в бизнесе. В данном разделе излагаются метод линейного программирования, алгоритм решения транспортной задачи и алгоритмы оценки

Глава 12. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

12.1. ВВЕДЕНИЕ

Существует множество форм деятельности предприятий, которые связаны с распределением ресурсов. Эти ресурсы включают труд, сырье, оборудование и денежные средства. Иногда процесс распределения ресурсов называют **программированием**. Поскольку обычно размеры ресурсов ограничены, возникают определенные проблемы. Если компания выпускает продукцию нескольких видов с использованием одного и того же оборудования и трудовых ресурсов, то ее администрация должна решить, какое количество продукции каждого вида производить. Принятое решение будет направлено на удовлетворение определенной цели администрации. Администрация может задаться целью наладить производство таким образом, чтобы максимизировать общий выпуск продукции за месяц, максимизировать время использования оборудования за неделю или минимизировать еженедельные затраты труда. Переменные решения — это количество продукции каждого вида, которое необходимо произвести за данный период времени.

Аналогично, если компания обладает определенным капиталом для инвестирования ряда проектов, распределение денежных сумм по каждому проекту будет подчинено некоторой цели. Она может заключаться в минимизации риска или максимизации темпов роста капитала. Переменные решения в данном случае — это денежные суммы, помещаемые в каждый проект.

В общем случае цель состоит в определении наиболее эффективного метода такого распределения ресурсов по соответствующим переменным, которое оптимизирует некоторый результат функционирования системы. Очень часто полезным инструментом в процессе распределения ресурсов являются методы моделирования. Математическим программированием называется использование математических моделей и методов для решения проблем программирования. Существует ряд различных методов, основанных на идеях математического программирования, однако мы рассмотрим только один из них, который нашел наиболее широкое применение, — **линейное программирование**.

Линейное программирование является подходящим **методом** для моделирования распределения ресурсов, если цель и ограничения на ресурсы можно выразить количественно в форме линейных взаимосвязей между переменными. Этот метод включает в себя ряд шагов:

1. Необходимо осуществить математическую формализацию задачи линейного программирования. Это означает, что нужно идентифицировать управляемые

переменные и цель задачи. Затем с помощью этих переменных цель и ограничения на ресурсы описываются в форме линейных соотношений.

2. После завершения формулировки задачи линейного программирования рассматриваются все допустимые сочетания переменных. Из них выбирается то, которое оптимизирует целевую функцию задачи. Если исследуемая задача содержит только две переменные, ее можно решить графически. Однако в случае исследования задачи со многими переменными необходимо прибегнуть к одному из алгебраических методов решения задач линейного программирования, для использования которых существуют пакеты прикладных программ.
3. Когда оптимальное решение получено, производится его оценка. Она включает в себя анализ задачи на чувствительность.

Решение задачи линейного программирования, как и любой иной математический инструмент, применяемый в теории принятия решений, является лишь одним из факторов, влияющих на конечное решение, принимаемое администрацией. Рассмотрение линейного программирования мы начнем с проблемы формулировки задачи.

12.2. ФОРМУЛИРОВКА ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Основная процедура является общей для формулирования всех задач линейного программирования:

Шаг 1. Определение переменных задачи, значения которых нужно получить в пределах существующих ограничений.

Шаг 2. Определение цели и ограничений на ресурсы.

Шаг 3. Описание цели через переменные задачи.

Шаг 4. Описание ограничений через переменные задачи.

Хотя на применение данной процедуры не влияет число переменных в задаче линейного программирования, рассмотрим сначала задачу с двумя переменными.

❐ **Пример 12.1.** Небольшая семейная фирма производит два широко популярных безалкогольных напитка — "Pink Fizz" и "Mint Pop". Фирма может продать всю продукцию, которая будет произведена, однако объем производства ограничен количеством основного ингредиента и производственной мощностью имеющегося оборудования. Для производства 1 л "Pink Fizz" требуется 0,02 ч работы оборудования, а для производства 1 л "Mint Pop" — 0,04 ч. Расход специального ингредиента составляет 0,01 кг и 0,04 кг на 1 л "Pink Fizz" и "Mint Pop" соответственно. Ежедневно в распоряжении фирмы имеется 24 ч времени работы оборудования и 16 кг специального ингредиента. Доход фирмы составляет 0,10 ф. ст. за 1 л "Pink Fizz" и 0,30 ф. ст. за 1 л "Mint Pop". Сколько продукции каждого вида следует производить ежедневно, если цель фирмы состоит в максимизации ежедневного дохода?

Решение

Шаг 1. Определение переменных. В рамках заданных ограничений фирма должна принять решение о том, какое количество каждого вида напитков следует выпускать. Пусть p — число литров "Pink Fizz", производимое за день. Пусть m — число литров "Mint Pop", производимое за день.

Шаг 2. Определение цели и ограничений. Цель состоит в максимизации ежедневного дохода. Пусть P — ежедневный доход, ф. ст. Он максимизируется в рамках ограничений на количество часов работы оборудования и наличие специального ингредиента.

Шаг 3. Выразим цель через переменные:

$$P = 0{,}10\ p + 0{,}30\ m \text{ (ф. ст. в день).}$$

Это целевая функция задачи — количественное соотношение, которое подлежит оптимизации.

Шаг 4. Выразим ограничения через переменные. Существуют следующие ограничения на производственный процесс:

а) Время работы оборудования. Для производства p литров "Pink Fizz" и m литров "Mint Pop" требуется: (0,02 p + 0,04 m) часов работы оборудования ежедневно. Максимальное время работы оборудования в день составляет 24 ч, следовательно, объем производства должен быть таким, чтобы число затраченных часов работы оборудования было меньше либо равно 24 ч ежедневно. Таким образом,

$$0{,}02\ p + 0{,}04\ m \le 24 \text{ ч/день.}$$

б) Специальный ингредиент. Производство p литров "Pink Fizz" и m литров "Mint Pop" требует (0,01 p + 0,04 m) кг ингредиента ежедневно. Максимальный расход ингредиента составляет 16 кг в день, следовательно, объем производства должен быть таким, чтобы требуемое количество специального ингредиента составляло не более 16 кг в день. Таким образом,

$$0{,}01\ p + 0{,}04\ m \le 16 \text{ кг/день.}$$

Других ограничений нет, однако разумно предположить, что фирма не может производить напитки в отрицательных количествах, поэтому:

в) Условие неотрицательности:

$$p \ge 0,\ m \ge 0.$$

Окончательная формулировка задачи линейного программирования имеет следующий вид. Максимизировать:

$$P = 0{,}10\ p + 0{,}30\ m \text{ (ф. ст. в день)}$$

при ограничениях:

время работы оборудования:	$0{,}02\ p + 0{,}04\ m \le 24$ ч/день;
специальный ингредиент:	$0{,}01\ p + 0{,}04\ m \le 16$ кг/день;
	$p,\ m \ge 0.$

❒ **Пример 12.2.** Завод-производитель высокоточных элементов для автомобилей выпускает два различных типа деталей: X и Y. Завод располагает фондом рабочего времени в 4000 чел.-ч. в неделю. Для производства одной детали типа X требуется 1 чел.-ч, а для производства одной детали типа Y — 2 чел.-ч. Производственные мощности завода позволяют выпускать максимум 2250 деталей типа X и 1750 деталей типа Y в неделю. Каждая деталь типа X требует 2 кг металлических стержней и 5 кг листового металла, а для производства одной детали типа Y необходимо 5 кг металлических стержней и 2 кг листового металла. Уровень запасов каждого вида металла составляет 10000 кг в неделю. Кроме того, еженедельно завод поставляет 600 деталей типа X своему постоянному заказчику. Существует также профсоюзное соглашение, в соответствии с которым общее число производимых в течение одной недели деталей должно составлять не менее 1500 штук.

Сколько деталей каждого типа следует производить, чтобы максимизировать общий доход за неделю, если доход от производства одной детали типа X составляет 30 ф. ст., а от производства одной детали типа Y — 40 ф. ст.?

Решение

Сначала необходимо сформулировать задачу линейного программирования.

Шаг 1. Идентификация переменных. Необходимо произвести x деталей типа X и y деталей типа Y в неделю.

Шаг 2. Какова цель задачи? Каковы ограничения на процесс производства? Цель состоит в максимизации общего дохода за неделю. Производственный процесс ограничивается уровнем:

а) фонда рабочего времени — максимально возможный фонд рабочего времени составляет 4000 чел. -ч. в неделю.

б) производственной мощности — для каждого типа деталей существует отдельное ограничение по производственной мощности. Оборудование позволяет выпускать не более 2250 деталей типа X и 1750 типа Y в неделю.

в) металлических стержней — максимальный их уровень составляет 10000 кг в неделю.

г) листового металла — максимальный уровень этого ресурса равен 10000 кг в неделю.

Кроме того, существуют ограничения на минимальный объем производства деталей каждого вида:

а) постоянные заказы — число произведенных деталей X должно быть достаточным для удовлетворения размера постоянных заказов.

б) Профсоюзное соглашение — общее число деталей (x + y) не должно быть ниже объема, предусмотренного соглашением.

Шаг 3. Целевая функция. Пусть P — общий доход за неделю, ф. ст., где

$$P = 30x + 40y \text{ (ф. ст. в неделю).}$$

Шаг 4. Ограничения на производственный процесс. Для каждого ограничения на ресурсы, необходимые для производства x деталей типа X и y деталей типа Y в неделю, ниже приведены количества и соответствующие им максимальные уровни наличных ресурсов.

Требуемый фонд рабочего времени:	$1x + 2y$	≤ 4000 чел.-ч.
Требуемая производственная мощность:	x	≤ 2250 деталей
	y	≤ 1750 деталей

Требуемое количество металлических стержней:	$2x + 5y \le 10000$ кг
Требуемое количество листового металла:	$5x + 2y \le 10000$ кг
Постоянные заказы:	$x \ge 600$ деталей
Профсоюзное соглашение:	$x + y \ge 1500$ деталей
Условие неотрицательности:	$x, y \ge 0.$

Окончательная формулировка задачи линейного программирования имеет вид:
Производится x деталей типа *X* и y деталей типа *Y* в неделю.
Максимизировать:

$$P = 30x + 40y \text{ (ф. ст.)}$$

при ограничениях:

Фонд рабочего времени:	$1x + 2y \le 4000$ чел.-ч.
Производственная мощность:	$x \le 2250$ деталей
	$y \le 1750$ деталей
Металлические стержни:	$2x + 5y \le 10000$ кг
Листовой металл:	$5x + 2y \le 10000$ кг
Постоянные заказы:	$x \ge 600$ деталей
Профсоюзное соглашение:	$x + y \ge 1500$ деталей
Условие неотрицательности:	$x, y \ge 0.$

Теперь рассмотрим задачу, число переменных в которой больше двух. Общая схема формулировки и в этом случае остается неизменной.

❑ **Пример 12.3.** Завод по производству электронного оборудования выпускает персональные компьютеры и системы подготовки текстов. В настоящее время освоены четыре модели:

а) "Юпитер" — объем памяти 512 Кбайт, одинарный дисковод;
б) "Венера" — объем памяти 512 Кбайт, двойной дисковод;
в) "Марс" — объем памяти 640 Кбайт, двойной дисковод;
г) "Сатурн" — объем памяти 640 Кбайт, жесткий диск.

В производственный процесс вовлечены три цеха завода — цех узловой сборки, сборочный и испытательный. Распределение времени, требуемого для обработки каждой модели в каждом цехе, а также максимальные производственные мощности цехов приведены в табл. 12.1. Отдел исследований рынка производит периодическую оценку потребительского спроса на каждую модель. Максимальные прогнозные значения спроса и доходы от реализации единицы продукции каждой модели также содержатся в табл. 12.1.

Построить задачу линейного программирования для изложенной проблемы производства изделий в ассортименте, если цель состоит в максимизации общего ежемесячного дохода.

Решение

Шаг 1. Выбор переменных. Производится:

j	единиц "Юпитера" в месяц,
v	единиц "Венеры" в месяц,
m	единиц "Марса" в месяц,
s	единиц "Сатурна" в месяц.

Таблица 12.1. **Время, требуемое на обработку каждой модели в каждом цехе**

Цех	*Время на единицу продукции, ч*				*Максимальная производственная мощность, ч/мес.*
	"Юпитер"	*"Венера"*	*"Марс"*	*"Сатурн"*	
Узловой сборки	5	8	20	25	800
Сборочный	2	3	8	14	420
Испытательный	0,1	0,2	2	4	150
Максимальное прогнозное значение спроса, за месяц	100	45	25	20	
Доход, ф.ст.	15	30	120	130	

Шаг 2. Какова цель задачи? Каковы ограничения на производственный процесс? Цель состоит в максимизации общего дохода за месяц. Объем производства ограничен размером фонда рабочего времени по каждому цеху и возможностью продажи компьютеров каждой модели.

Шаг 3. Целевая функция задачи. Пусть P (ф. ст.) — общий доход в месяц, тогда:

$$P = 15\,j + 30\,v + 120\,m + 130\,s \text{ (ф. ст. в месяц).}$$

Шаг 4. Ограничения на производственный процесс. Для каждого цеха время, требуемое для производства j, v, m и s единиц продукции соответствующих моделей увязывается с максимальной производственной мощностью данного цеха.

Цех узловой сборки	$5j + 8v + 20m + 25s$	≤ 800 (ч/мес.)
Сборочный цех:	$2j + 3v + 8m + 14s$	≤ 420 (ч/мес.)
Испытательный цех:	$0{,}1j + 0{,}2v + 2m + 4s$	≤ 150 (ч/мес.)
Спрос на "Юпитер":	j	≤ 100 (ед./мес.)
Спрос на "Венеру":	v	≤ 45 (ед./мес)
Спрос на "Марс":	m	≤ 25 (ед./мес.)
Спрос на "Сатурн":	s	≤ 20 (ед./мес.)
Условие неотрицательности:	j, v, m, s	≥ 0

Окончательная формулировка задачи линейного программирования такова: каждый месяц производится j, v, m и s единиц компьютеров типа "Юпитер", "Венера", "Марс" и "Сатурн" соответственно. Максимизировать:

$$P = 15\,j + 30\,v + 120\,m + 130\,s \text{ (ф. ст. в месяц)}$$

в условиях ограничений, указанных выше.

❑ **Пример 12.4.** Менеджер по ценным бумагам намерен разместить 100000 ф. ст. капитала таким образом, чтобы получать максимальные годовые проценты с дохода. Его выбор ограничен четырьмя возможными объектами инвестиций: A, B, C и D. Объект A позволяет получать 6% годовых, объект B — 8% годовых, объект

С – 10%, а объект D – 9% годовых. Для всех четырех объектов степень риска и условия размещения капитала различны. Чтобы не подвергать риску имеющийся капитал, менеджер принял решение, что не менее половины инвестиций необходимо вложить в объекты А и В. Чтобы обеспечить ликвидность, не менее 25% общей суммы капитала нужно поместить в объект D. Учитывая возможные изменения в политике правительства, предусматривается, что в объект С следует вкладывать не более 20% инвестиций, тогда как особенности налоговой политики требуют, чтобы в объект А было вложено не менее 30% капитала. Сформулируем для изложенной проблемы распределения инвестиций модель линейного программирования.

Решение

Вкладывается a ф. ст. в объект А, b ф. ст. – в объект В, с ф. ст. – в объект С и d ф. ст. – в объект D. Целью является максимизация общей суммы годовых процентов с дохода. На распределение инвестиций наложены ограничения, связанные с отсутствием риска, ликвидностью, политикой правительства и системой налогообложения. Обозначим через R общую сумму годового процентного дохода, тогда:

$$R = 0{,}06\,a + 0{,}08\,b + 0{,}10\,c + 0{,}09\,d \text{ (ф. ст. в год).}$$

Максимизация целевой функции осуществляется в условиях ограничений на

Общую сумму инвестиций:	$a + b + c + d \le 100000$ ф. ст.
Отсутствие риска:	$a + b \ge 0{,}05\,(a + b + c + d)$
Ликвидность:	$d \ge 0{,}25\,(a + b + c + d)$
Правительственную политику:	$c \le 0{,}2\,(a + b + c + d)$
Систему налогообложения:	$a \ge 0{,}3\,(a + b + c + d)$
Неотрицательность:	$a,\ b,\ c,\ d \ge 0.$

Чтобы решить задачу линейного программирования, ограничения обычно преобразовывают таким образом, чтобы переменные находились только в левой части любого неравенства. Результаты этого преобразования представлены ниже. Окончательная *форма задачи* линейного программирования имеет следующий вид:

Вкладывается a ф. ст. в объект А,
b ф. ст. в объект В,
с ф. ст. в объект С,
d ф. ст. в объект D.

Максимизируется общая сумма годового процентного дохода, т.е.:

$$R=0{,}06\,a + 0{,}08\,b + 0{,}10\,c + 0{,}09\,d \text{ (ф. ст. в год)}$$

в условиях следующих ограничений (ф. ст.):

Общая сумма инвестиций:	$a + b + c + d \le 100000$
Отсутствие риска:	$0{,}5\,a + 0{,}5\,b - 0{,}5\,c - 0{,}5\,d \ge 0$
Ликвидность:	$-0{,}25\,a - 0{,}25\,b - 0{,}25\,c + 0{,}75\,d \ge 0$
Правительственная политика:	$-0{,}2\,a - 0{,}2\,b + 0{,}8\,c - 0{,}2\,d \le 0$
Система налогообложения:	$0{,}7\,a - 0{,}3\,b - 0{,}3\,c - 0{,}3\,d \ge 0$
Условие неотрицательности:	$a, b, c, d \ge 0$

Во всех четырех приведенных выше примерах целевую функцию требовалось максимизировать. На стадии постановки задачи процедура не меняется, если целью является минимизация некоторого показателя. Примеры таких задач приводятся в конце данной главы.

12.3. РЕШЕНИЕ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

В данном разделе будет рассмотрен процесс нахождения значений переменных, которые удовлетворяют системе ограничений и оптимизируют целевую функцию задачи. Однако гораздо удобнее исследовать не систему неравенств, а систему уравнений. Процесс преобразования неравенств в уравнения достаточно прост. Для этого в левую часть неравенства вводится дополнительная переменная. Эта переменная призвана отразить величину разности между правой и левой частями неравенства. Чтобы продемонстрировать этот алгоритм, обратимся к примеру **12.2**, в котором рассматривается производство деталей типов X и Y к автомобилям. Для получения системы уравнений в каждое ограничение введем дополнительную переменную. Обозначим данную переменную через s, таким образом, в первое ограничение вводится переменная s_1, во второе — s_2 и т.д. Кроме того, примем предпосылку о неотрицательности значений этих переменных, т.е. $s_i \geq 0$. Это значит, что дополнительные переменные прибавляются к левым частям всех ограничений знака "≤" и вычитаются из левых частей ограничений знака "≥". Задача линейного программирования в данном случае принимает следующий вид: производится x деталей типа X и y деталей типа Y в неделю. Цель состоит в максимизации общего дохода в неделю. Максимизировать:

$$P = 30x + 40y \text{ (ф. ст. в неделю)}$$

при ограничениях:

Фонд рабочего времени:	$1x + 2$	y	$+$	s_1	$=$	4000 (чел.-ч/нед.)
Производственная мощность:		x	$+$	s_2	$=$	2250 (деталей/нед.)
		y	$+$	s_3	$=$	1750 (деталей/нед.)
Металлические стержни:	$2x + 5$	y	$+$	s_4	$=$	10000 (кг/нед.)
Листовой металл:	$5x + 2$	y	$+$	s_5	$=$	10000 (кг/нед.)
Постоянные заказы:	x		$-$	s_6	$=$	600 (деталей/нед.)
Профсоюзное соглашение:	$x +$	y	$-$	s_7	$=$	1500 (деталей/нед.)
Условие неотрицательности:			$x,$	y	$\geq$	0

Такие вспомогательные переменные для ограничений со знаком "≤" называются **остаточными переменными**. Они представляют собой количество недоиспользуемого ресурса, т.е. разность между используемым количеством ресурса и его максимальным объемом. Рассмотрим, например, ограничение на фонд рабочего времени, указанное выше. Предположим, что в течение недели выпускается 1000 деталей каждого типа, тогда используемое число человеко-часов составит: $1\times1000 + 2\times1000 = 3000$. Поскольку максимальный фонд рабочего времени равен 4000 чел.-ч., резерв времени, или остаток, составит: $4000 - 3000 = 1000$ чел.-ч. Следовательно, для данной комбинации x – y и s_1 принимают значение, равное 1000.

Вспомогательные переменные, используемые в ограничениях типа "≤", называются **избыточными переменными**, так как они показывают количество ресурса, используемое сверх минимального его объема. Рассмотрим, к примеру, ограничение на постоянные заказы в случае, когда выпускается 1000 деталей типа X. Минимальное число деталей типа X составляет в соответствии с данным ограничением 600 штук, следовательно, уровень производства, равный 1000 деталей, порождает излишек в 400 штук сверх минимального количества. Таким образом, s_6 принимает значение, равное 400.

Итак, мы получили систему уравнений. Однако мы не можем решить ее с применением традиционных алгебраических методов и получить единственное множество значений переменных (единственное решение), поскольку число переменных превосходит число уравнений системы. Единственное множество решений можно получить только в случае, если число переменных и число уравнений системы совпадают. В лучшем случае мы можем определить множество допустимых решений системы уравнений. Данное множество содержит все сочетания переменных, которые удовлетворяют системе ограничений. Затем из этого множества можно будет выбрать одно или несколько решений, оптимизирующих целевую функцию задачи.

Как следует поступать при определении множества допустимых решений? Если задача содержит только две переменные, это можно сделать графически. Однако в случае решения задачи с множеством переменных необходимо прибегнуть к алгебраическому методу решения.

12.3.1. Графическое решение задачи линейного программирования

Линейное уравнение описывает множество точек, лежащих на одной прямой. Линейное неравенство описывает некоторую область на плоскости. Например, неравенство $x \le 7$ означает, что x принимает значения, которые либо меньше 7, либо равны 7. Графически эту ситуацию можно проиллюстрировать следующим образом. Проведем прямую $x = 7$. Обратимся к графику, изображенному на рис. 12.1 слева. Данная прямая разделяет плоскость на три множества точек: точки, для

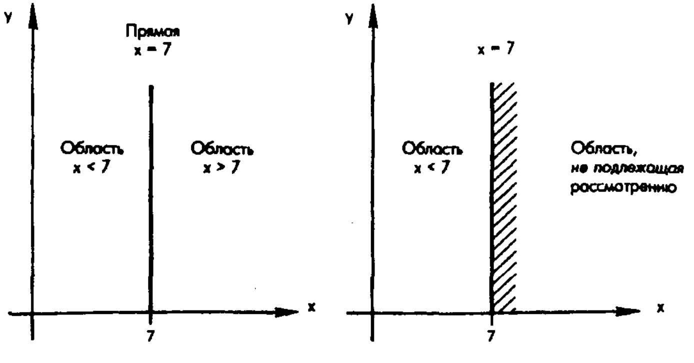


Рис. 12.1. Графическое изображение неравенства $x \le 7$

которых $x = 7$, т.е. точки, лежащие на самой прямой; точки, для которых $x < 7$, область слева от прямой; и точки, для которых $x > 7$, т.е. точки, принадлежащие области, лежащей справа от прямой. Последнее множество нас не интересует. Область, не подлежащую рассмотрению, обычно принято заштриховывать. Обратите внимание на график, изображенный на рис. 12.1 справа.

Предположим, что $x + y \le 10$. Какую часть плоскости описывает данное неравенство? Схема поиска ответа на этот вопрос аналогична схеме, используемой в предыдущем примере. Во-первых, проведем прямую $x + y = 10$. Обратимся к графику, изображенному на рис. 12.2 слева. Как и в предыдущем примере, проведенная прямая разделяет плоскость на три множества точек: точки, для которых $x + y = 10$, принадлежащие прямой; точки, для которых $x + y < 10$, принадлежащие области, лежащей ниже прямой; и точки, для которых $x + y > 10$, принадлежащие области, лежащей выше указанной прямой.

Полезным приемом при определении недопустимой области на графике является следующая процедура. Необходимо выбрать любую точку на графике, не принадлежащую прямой, и подставить ее координаты в неравенство. Если неравенство будет выполняться, то данная точка является допустимым решением. Если неравенство не выполняется, то точка является недопустимой и принадлежит, следовательно, недопустимой области. Удобной для использования при подстановке в неравенство точкой является начало координат. Подставим $x = y = 0$ в неравенство $x + y \le 10$. Получим $0 + 0 \le 10$. Данное утверждение является верным, следовательно, начало координат — допустимое решение, и недопустимой областью является часть плоскости, лежащая по другую сторону прямой. Это отражено на графике, изображенном на рис. 12.2 справа.

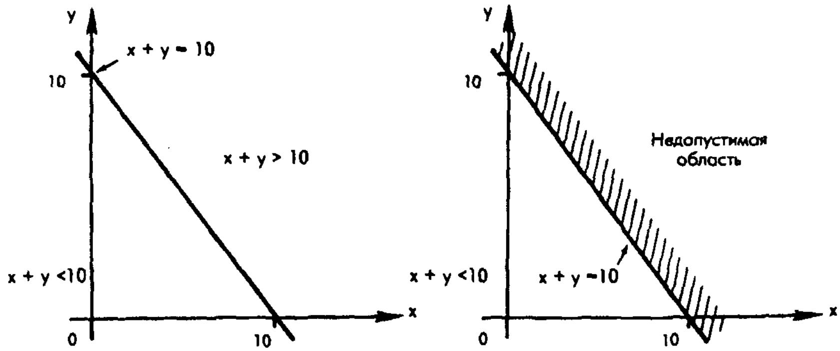


Рис. 12.2. Графическое изображение неравенства $x + y \le 10$

Аналогично можно изобразить графически каждое ограничение задачи линейного программирования и определить недопустимую область. Если все ограничения задачи изобразить на одном графике, то область, которая останется незаштрихованной, будет содержать множество точек, удовлетворяющих системе ограничений.

Данная область называется **допустимым множеством**. Порядок расположения переменных на осях координат в задаче линейного программирования значения не имеет. На графике следует всегда отмечать начало координат. Смещённое начало координат использовать нельзя.

Применим теперь данную процедуру к задаче линейного программирования, сформулированной в примере **12.1**, в котором рассматривается производство двух видов безалкогольных напитков. Ограничения задачи можно изобразить графически.

Время работы оборудования: $0{,}02\ p + 0{,}04\ m \le 24$ ч/день.

Проведем прямую $0{,}02\ p + 0{,}04\ m = 24$. Простейшим способом нанесения прямой на график является нахождение точек пересечения данной прямой с осями координат p и m. Подставив $p = 0$ в уравнение и рассчитав значение m, получим, что при $p = 0$ $m = 600$. Подставив $m = 0$ в уравнение и рассчитав значение p, получим, что при $m = 0$ $p = 1200$. Нанесем эти две точки на график и соединим их прямой. Этим приемом можно пользоваться всегда, за исключением случая, когда прямая проходит через начало координат. В последнем случае применяется иная процедура подстановки в уравнение любого другого значения p и нахождения соответствующего значения m.

Для определения области, которую следует заштриховать, подставим $p = 0$ и $m = 0$ в неравенство:

$$0{,}02 \times 0 + 0{,}04 \times 0 \le 24.$$

Данное утверждение является верным, таким образом, начало координат принадлежит допустимой области.

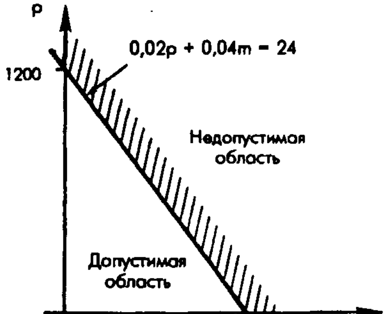


Рис. 12.3. Графическое изображение неравенства
$0{,}02\ p + 0{,}04\ m \le 24$

Специальный ингредиент: $0{,}01\ p + 0{,}04\ m \le 16$.

Проведем прямую: $0{,}01\ p + 0{,}04\ m = 16$.

Как и в предыдущем ограничении, начало координат принадлежит допустимой области, поэтому следует заштриховать область, лежащую выше прямой.

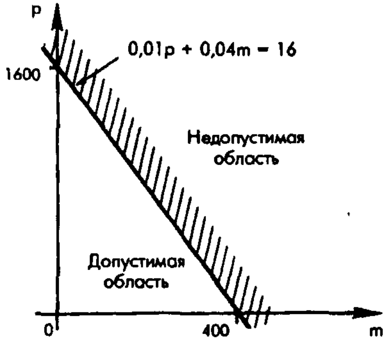


Рис. 12.4. Графическое изображение неравенства 0,01 p + 0,04 m ≤ 16

Рис. 12.5. Графическое изображение условий неотрицательности переменных

Условие неотрицательности: $p \geq 0$ и $m \geq 0$.

Заштриховываются области, содержащие отрицательные значения каждой переменной (см. рис. 12.5).

Нанеся все ограничения задачи на один график, получим:

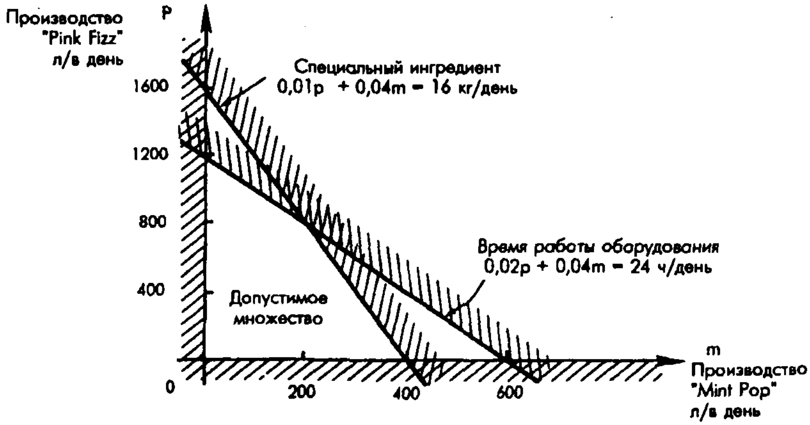


Рис. 12.6. Графическое изображение ограничений примера 12.1

Область, оставшаяся незаштрихованной для всех ограничений, – это **допустимое множество**, которое содержит все возможные сочетания объемов производства, удовлетворяющие данным ограничениям. Координаты любой точки, принадлежащей допустимому множеству, являются возможным сочетанием объемов производства двух видов прохладительных напитков, выпускаемых фирмой.

Рассмотрим алгоритм выбора объема производства, максимизирующего ежедневный общий доход фирмы. Целевая функция задачи имеет следующий вид:

$$P = 0{,}10\ p + 0{,}30\ m \text{ (ф. ст. в день)}$$

Если задать P = 100 ф. ст. в день, целевую функцию можно проиллюстрировать графически. Если затем придать P другое значение, то новая прямая будет параллельна прямой, соответствующей значению P = 100 ф. ст. в день (рис. 12.7.)

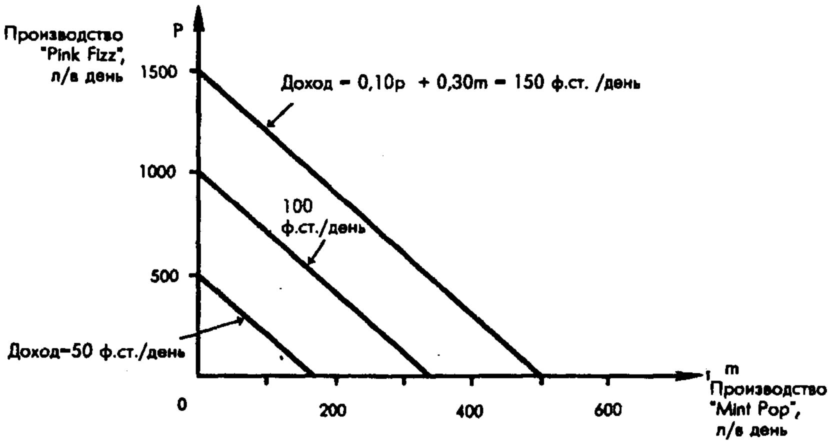


Рис. 12.7. Целевая функция

Нанеся на график какую-либо линию уровня и двигаясь в пределах допустимого множества параллельно данной линии, можно создать целое семейство линий уровня возможного дохода. Чем дальше линия уровня от начала координат, тем больше величина дохода, которой она соответствует.

Если построить на графике линию уровня задачи линейного программирования так, как показано на рис. 12.8, можно двигаться параллельно этой линии вдоль допустимого множества в направлении увеличения дохода до тех пор, пока не будет достигнуто последнее допустимое решение (или решения), т.е. до тех пор, пока все точки линии уровня не окажутся за пределами допустимого множества.

Нетрудно заметить, что последним допустимым решением является точка A. Координаты этой точки соответствуют оптимальному сочетанию объемов производства двух напитков. Приближенные значения координат точки A можно найти непосредственно из графика, а точные их значения можно получить, решив систему из двух уравнений, описывающих те ограничения, на пересечении которых находится точка A.

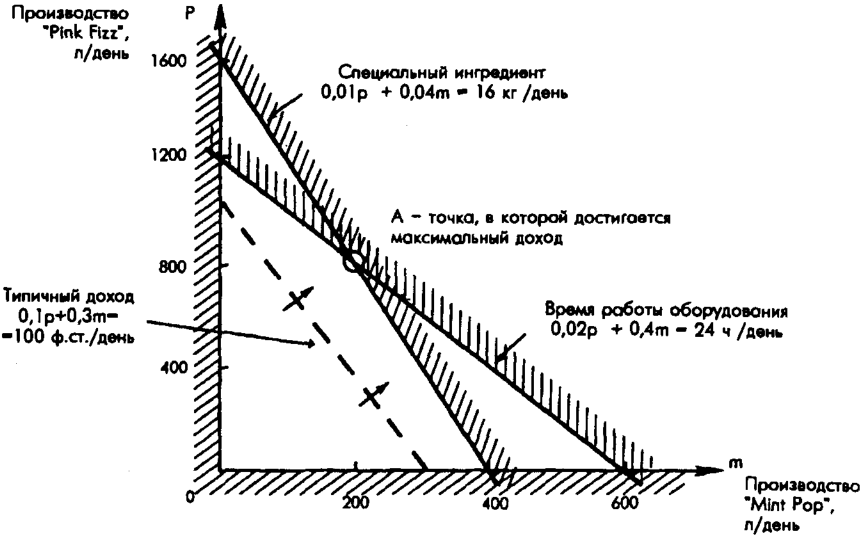


Рис. 12.8. Задача линейного программирования для примера 12.1

Эти два ограничения называются **связывающими**, или **лимитирующими** ограничениями. Они соответствуют тем ресурсам, которые в процессе производства используются полностью и, следовательно, препятствуют дальнейшему увеличению ежедневного дохода. Оптимальное решение задачи — это точка пересечения прямых.

$$0{,}02p + 0{,}04m = 24; \quad (1)$$

$$0{,}01p + 0{,}04m = 16. \quad (2)$$

Вычтем уравнение (2) из уравнения (1):

$$0{,}01\ p = 8.$$

Следовательно, p = 800 л в день. Подставив найденное значение p в уравнение (2), найдем m:

$$0{,}01 \times 800 + 0{,}04\ m = 16.$$

Следовательно, m = 200 л в день.

Таким образом, чтобы получать максимальный ежедневный доход, фирма должна производить 800 л "Pink Fizz" и 200 л "Mint Pop" в день. Это сочетание объемов производства дает максимальное значение дохода: 0,10 × 800 + 0,30 × 200 = = 140 ф. ст. в день.

При таких объемах выпуска продукции максимально используется время работы оборудования и полностью расходуется запас специального ингредиента в день. В обоих ограничениях резервный запас или остаток ресурсов отсутствует.

Этот метод определения оптимальной крайней точки зависит от того, насколько правильно была построена линия уровня дохода. Ниже излагается практический прием, который может помочь при нанесении на график линии уровня, которая будет являться основой для определения оптимальной крайней точки. Выберем любую удобную точку, лежащую приблизительно в середине допустимого множества. Предположим, что в примере, изложенном выше, выбрана точка с координатами p = 200, m = 200. Ежедневный доход от выпуска продукции в таком объеме составит:

$$P = 0{,}10\ p + 0{,}30\ m = 0{,}10 \times 200 + 0{,}30 \times 200 = 80\ \text{ф. ст. в день}.$$

Все остальные сочетания объемов производства, позволяющие получать 80 ф. ст. ежедневного дохода, принадлежат прямой

$$80 = 0{,}10\ p + 0{,}30\ m\ (\text{ф. ст. в день}).$$

Одна точка этой прямой уже известна, это — точка с координатами p = 200, m = 200. Для определения другой точки можно положить m = 0, тогда p = 800. Теперь мы можем построить на графике эту линию уровня ежедневного дохода и, используя вышеизложенный алгоритм, определить оптимальное решение (или решения). Очевидно, что в результате применения данного алгоритма оптимальное решение всегда будет представлено либо крайней точкой множества, либо в случае, если целевая функция параллельна одному из ограничений, множеством точек отрезка, соединяющего две крайние точки.

Мы предполагали, что переменные в задаче линейного программирования непрерывны или, если это условие не выполняется, могут принимать дробные значения. И действительно, часто имеет место ситуация, когда в течение временного промежутка, рассматриваемого в задаче, допустимы дробные значения объемов выпускаемой продукции. Если, например, производятся две модели автомобилей, а цель задачи линейного программирования состоит в максимизации использования оборудования в неделю, может оказаться, что оптимальное решение предполагает наличие к концу недели незавершенного производства. При такой постановке задачи, когда рассматриваемый период времени равен одной неделе, незавершенное производство вполне допустимо.

Если, однако, необходимо осуществить распределение рабочих по определенным видам работ, дробные значения для числа рабочих недопустимы. В этом случае оптимальное решение должно включать только целые значения. Допустимыми решениями являются все точки допустимого множества, в которых переменные принимают целые значения. В качестве оптимального решения выбирается последняя точка допустимого множества, координаты которой являются целыми числами, однако в данном случае она может уже не быть крайней точкой допустимого множества.

В задаче линейного программирования с двумя переменными процедура нахождения оптимального решения в условиях целочисленности переменных не составляет особого труда. Вместо допустимого множества рассматривается множество допустимых точек, лежащих в пределах заданных ограничений. Движение типичной линии уровня целевой функции осуществляется не вдоль всего допустимого множества, а только через данные точки. Однако при решении задачи с множеством переменных используется один из методов целочисленного программирования, рассмотрение которых выходит за рамки этой книги.

❐ **Пример 12.5.** Обратимся к примеру **12.2**, в котором рассматривалось производство двух типов деталей к автомобилям. Необходимо определить объемы производства, при которых достигается максимальное значение общего дохода за неделю.

Решение.

Допустимые области для каждого из ограничений задачи выглядят следующим образом:

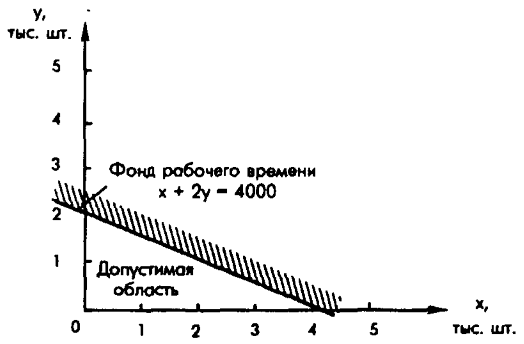


Рис. 12.9. Ограничение на фонд рабочего времени: $x + 2y \le 4000$ чел.-ч. в неделю

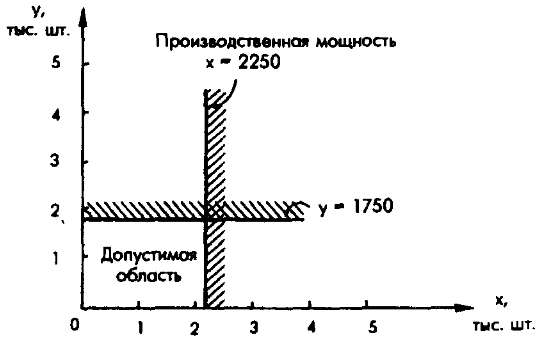


Рис. 12.10. Ограничение на производственные мощности: $x \le 2250$ деталей в неделю и $y \le 1750$ деталей в неделю

Рис. 12.11. Ограничение на металлические стержни: $2x + 5y \leq 10000$ кг в неделю

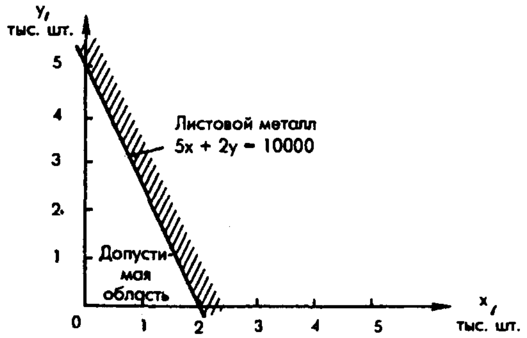


Рис. 12.12. Ограничение на листовой металл: $5x + 2y \leq 10000$ кг в неделю

Рис. 12.13. Ограничение на постоянные заказы и неотрицательность: $x \geq 600$ деталей в неделю и $x \geq 0$, $y \geq 0$

Рис. 12.14. Ограничение на профсоюзное соглашение: $x + y \geq 1500$ деталей в неделю

Допустимое множество, содержащее все возможные ассортиментные наборы продукции для данной задачи, осталось на графике (рис. 12.15) незаштрихованным.

Необходимо определить оптимальный ассортиментный набор, максимизирующий общий доход завода за неделю. Целевая функция задачи имеет вид:

$$P = 30x + 40y \text{ (ф. ст. в неделю).}$$

Чтобы построить линию уровня этой функции для типичного значения дохода, выберем принадлежащую допустимому множеству точку с координатами $x = 1000$, $y = 1000$. Для этого асссортиментного набора общий доход за неделю составит:

$$P = 30 \times 1000 + 40 \times 1000 = 70000 \text{ ф. ст. в неделю.}$$

В качестве контрольной линии уровня будем использовать прямую
$70000 = 30x + 4y$ (ф. ст. в неделю).

Эта прямая проходит также через точку с координатами $x = 0$, $y = 1750$. На рис. 12.15 она изображена пунктирной линией. Движение параллельно этой прямой в направлении увеличения дохода приводит нас в точку А, которая является последним допустимым решением.

Лимитирующими являются ограничения на:

Фонд рабочего времени:	$x + 2y \leq 4000$	ч в неделю;
Листовой металл:	$5x + 2y \leq 10000$	кг в неделю.

Решив соответствующую систему уравнений, получим:

$$\begin{array}{rrcll} & x + 2y & = & 4000; & (2) \\ & \underline{5x + 2y} & = & \underline{10000;} & (1) \\ (2) - (1) & 4x & = & 6000, & \end{array}$$

следовательно, $x = 1500$ и после подстановки значения x в систему получим значение $y = 1250$.

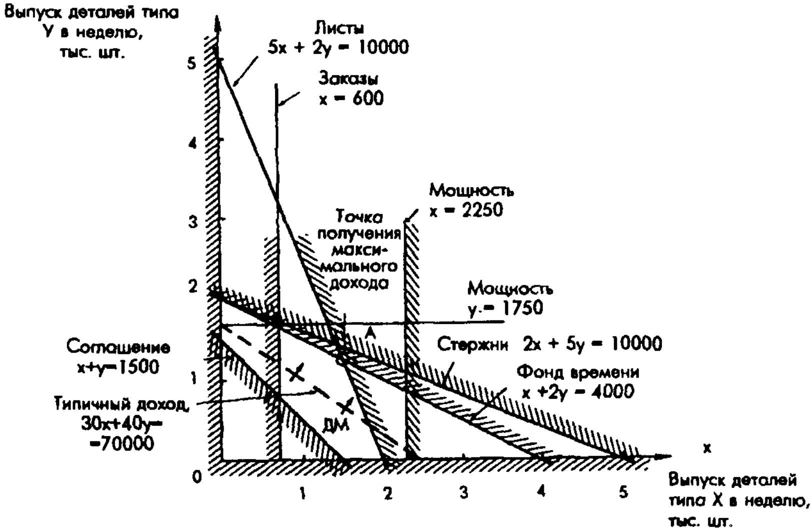


Рис. 12.15. Задача линейного программирования для выпуска деталей типа X и Y к автомобилям в неделю

Оптимальным ассортиментным набором является выпуск 1500 деталей типа X и 1250 деталей типа Y в неделю. Таким образом, максимальный доход за неделю составит:

$$P = 30 \times 1500 + 40 \times 1250 = 95000 \text{ ф. ст. в неделю.}$$

В процессе производства данного ассортиментного набора полностью расходуются два ресурса — фонд рабочего времени и листовой металл. Эти ограничения являются лимитирующими. Однако для обоих типов деталей остается свободной часть производственных мощностей, а также не полностью используются металлические стержни. Объемы производства деталей также превышают минимальный уровень, требуемый для удовлетворения постоянных заказов, и минимальный уровень, оговоренный в профсоюзном соглашении.

Нетрудно установить, что значения остаточных переменных в ограничениях на производственные мощности равны 750 для деталей типа X и 500 для деталей типа Y, а именно:

$$1500 + s_2 = 2250, \text{ следовательно } s_2 = 750 \text{ деталей в неделю.}$$
$$1250 + s_3 = 1750, \text{ следовательно } s_3 = 500 \text{ деталей в неделю.}$$

Остаточная переменная ограничения на металлические стержни:

$$2 \times 1500 + 5 \times 1250 + s_4 = 10000,$$

следовательно,

$$s_4 = 750 \text{ кг в неделю.}$$

Избыточная переменная ограничения на постоянные заказы:

$$1500 - s_6 = 600,$$

следовательно,

$$s_6 = 900 \text{ деталей в неделю}$$

сверх минимального количества, необходимого для удовлетворения постоянных заказов. Избыток по профсоюзному соглашению составил:

$$1500 + 1250 - s_7 = 1500,$$

следовательно,

$$s_7 = 1250 \text{ деталей в неделю}$$

сверх минимального количества деталей, оговоренного в профсоюзном соглашении.

Как уже отмечалось выше, оптимальным решением обычно является крайняя точка допустимого множества. Следовательно, после построения графика определить оптимальную крайнюю точку можно путем подсчета значений целевой функции во всех крайних точках допустимого множества. Множество значений переменной, соответствующей крайней точке допустимого множества, называется **базисным решением**. Переменные, принимающие ненулевые значения в некоторой крайней точке, называются **базисными переменными**.

Иногда в процессе решения задачи линейного программирования возникают некоторые трудности. Задача может оказаться **несовместной**. В этом случае допустимое множество задачи является пустым. Ни одно сочетание переменных не удовлетворяет всем ограничениям задачи одновременно и задача не имеет решений. Если в данной ситуации все же необходимо найти решение задачи, чтобы построить допустимое множество, то необходимо исключить одно или несколько ограничений.

Проблемы иного рода возникают, если задача линейного программирования является **неограниченной**. В этом случае решение задачи может неограниченно улучшаться, не нарушая при этом ни одного ограничения задачи. Обычно это означает, что задача линейного программирования сформулирована некорректно, и некоторые ограничения в ней отсутствуют.

О возможности существования целого множества оптимальных решений мы уже говорили. Это случается, если целевая функция параллельна лимитирующему ограничению задачи. Оптимальное значение целевой функции будет достигаться в любой точке этого ограничения, лежащей между двумя оптимальными крайними точками, и соответственно любая из этих точек является оптимальным решением модели. Эта ситуация имеет ряд преимуществ, поскольку предоставляет администрации некоторую свободу в принятии решений.

12.4. АНАЛИЗ ЧУВСТВИТЕЛЬНОСТИ

В большинстве случаев принятия решений полезно составить предварительный план действий, чтобы проанализировать, какое воздействие окажут изменения, вносимые в исходную задачу, на принятое решение. Линейное программирование не является исключением. Существуют три аспекта решения задач линейного программирования, которые необходимо подвергнуть тщательному изучению:

- воздействие дополнительного количества лимитирующего ресурса;
- воздействие дополнительного количества нелимитирующих ресурсов;
- воздействие изменений в коэффициентах целевой функции. Рассмотрим каждую из этих ситуаций в отдельности. При дальнейшем рассмотрении примем предпосылку о том, что в любой момент времени изменяется только один параметр.

12.4.1. Воздействие изменений в обеспечении лимитирующим ресурсом на решение задачи линейного программирования

Поскольку один или несколько ресурсов используются полностью, значение целевой функции ограничено. Если появляется дополнительное количество лимитирующего ресурса, то оптимальное решение может быть улучшено. Однако необходимо принять во внимание, что изменение оптимального решения приведет к улучшению значения целевой функции только в том случае, если сумма дополнительных издержек по обеспечению дополнительным количеством ресурса не превышает сумму прибыли, полученной в результате его использования.

С увеличением объема лимитирующего ресурса соответствующее ограничение становится менее жестким. Так как жесткость лимитирующего ограничения постепенно снижается, его график будет перемещаться параллельно своему начальному положению, одновременно будет происходить перемещение оптимальной крайней точки в направлении, которое улучшает значение целевой функции. Этот процесс будет продолжаться до тех пор, пока какой-либо другой ресурс не будет полностью использован и рассматриваемое ограничение перестанет быть лимитирующим. Величина, на которую увеличивается значение целевой функции при снижении жесткости лимитирующего ограничения на единицу, т.е. при увеличении количества лимитирующего ресурса на единицу, называется **теневой ценой** ресурса. Теневая цена ресурса — это стоимость единицы данного ресурса в оптимальном решении. Увеличение объема лимитирующего ресурса на единицу целесообразно только в том случае, если существует возможность его получения по стоимости, которая ниже, чем теневая цена данного ресурса.

❐ **Пример 12.6.** Обратимся к примерам **12.2** и **12.5**. Из примера **12.5** мы знаем, что лимитирующими являются ограничения на фонд рабочего времени и на листовой металл. Рассмотрим сначала последнее из указанных ограничений. Обратимся к графику, изображенному на рис. 12.16. Жесткость ограничения на листовой металл снижается по мере перемещения линии ограничения параллельно ее исходному положению в противоположном направлении от начала координат. Допустимое множество расширяется, а оптимальная крайняя точка перемещается вниз по линии ограничения на фонд рабочего времени, что увеличивает x и уменьшает y. Снижение жесткости ограничения на листовой металл является эффективным до тех пор, пока линия ограничения не достигнет точки пересечения ограничений на фонд рабочего времени и производственные мощности для деталей типа X, т.е.

точки В. Если и далее снижать жесткость ограничения на листовой металл, оно перестанет быть лимитирующим, что приведет к появлению остатка в виде неиспользованного листового металла.

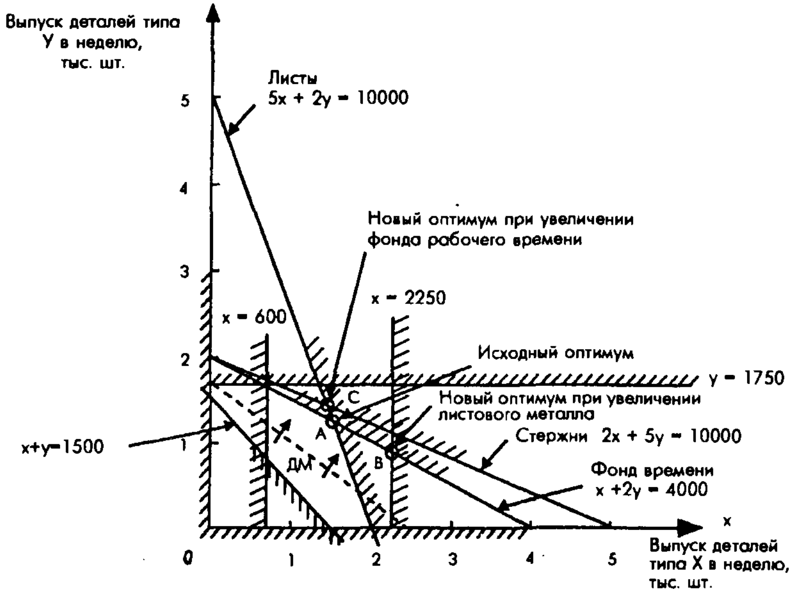


Рис. 12.16. Задача линейного программирования для производства деталей типа X и Y в неделю

Новой оптимальной крайней точкой является теперь точка В. Координаты точки В можно определить, решив систему уравнений для ограничений на фонд рабочего времени и производственные мощности для детали X.

Фонд рабочего времени: $x + 2y = 4000$ чел.-ч в неделю;

Производственные мощности для X: $x = 2250$ деталей в неделю.

Так как $x = 2250$, то подставив его в первое уравнение, найдем значение y:

$$2250 + 2y = 4000,$$

следовательно, $y = 875$ деталей.

Новым оптимальным ассортиментным набором является производство 2250 деталей типа X и 875 деталей типа Y в неделю. Этот ассортиментный набор дает максимальный доход, равный $30 \times 2250 + 40 \times 875 = 102500$ ф. ст. в неделю, таким образом, увеличение дохода составит: $102500 - 95000 = 7500$ ф. ст. в неделю. Количество листового металла, используемое для производства данного ассортиментного набора, равно:

$$5 \times 2250 + 2 \times 875 = 13000 \text{ кг в неделю.}$$

Оно превышает начальное количество на 3000 кг в неделю. В новой оптимальной точке фонд рабочего времени и производственные мощности для деталей X также используются максимально.

Дополнительное количество листового металла — 3000 кг — позволяет получать дополнительный доход, равный 7500 ф. ст. в неделю, следовательно, теневая цена данного ресурса составит: 7500 : 3000 = 2,50 ф. ст. за 1 кг. Каждый дополнительный килограмм листового металла ведет к увеличению еженедельного дохода в 2,50 ф. ст. Из этого следует, что сверхнормативный запас этого ресурса целесообразен только в случае, если стоимость получения любого дополнительного количества ресурса не превышает 2,50 ф. ст. за 1 кг ресурса.

Предположив, что ограничение на листовой металл остается неизменным, применим аналогичную процедуру ко второму лимитирующему ограничению.

Если жесткость ограничения на фонд рабочего времени снизилась на единицу, т.е. появилась возможность использовать 1 чел.-ч рабочего времени дополнительно, то тогда данное ограничение принимает вид:

$$x + 2y \leq 4001.$$

Данное ограничение параллельно первоначальному, но его линия находится дальше от начала координат по сравнению с исходной линией. Из приведенного выше графика легко видеть, что точка пересечения ограничения на листовой металл и нового ограничения на фонд рабочего времени все еще является оптимальной крайней точкой. В данном случае оптимальным решением является точка с координатами $x = 1499{,}75$ и $y = 1250{,}625$, что приводит и к значению целевой функции $P_{max} = 95017{,}50$ ф. ст. Таким образом, значение целевой функции увеличилось на 17,50 ф. ст. Теневая цена фонда рабочего времени составляет 17,50 ф. ст. за 1 чел.-ч. Если можно получить один дополнительный час рабочего времени за дополнительные 17,50 ф. ст. или менее, то это необходимо использовать. Если же стоимость 1 чел.-ч. превышает 17,50 ф. ст., то дополнительное количество рабочего времени использовать нецелесообразно.

Какое количество дополнительного рабочего времени следует купить? Поскольку линия ограничения на фонд рабочего времени движется параллельно своему исходному положению в направлении от начала координат, она стремится к точке пересечения ограничений на листовой металл и металлические стержни к точке C. Если и далее снижать жесткость ограничения на фонд рабочего времени, то оно перестанет быть лимитирующим, и дальнейшее привлечение дополнительного рабочего времени нецелесообразно. Максимальное число дополнительных человеко-часов можно определить, решив систему ограничений, линии которых пересекаются в точке C:

Листовой металл:	$5x + 2y = 10000;$
Металлические стержни:	$2x + 5y = 10000.$

Ее решением являются следующие значения переменных:

$x = 10000/7$ и $y = 10000/7$ деталей в неделю.

Число используемых в точке C человеко-часов равно:

$$x + y = 10000/7 + 2\,(10000/7) = 4285{,}7 \text{ чел.-ч в неделю.}$$

Это значение на 285,7 чел.-ч превосходит первоначальное максимальное значение 4000 чел.-ч. Получение максимального сверхнормативного запаса в 285,7 чел.-ч в неделю целесообразно при условии, что стоимость единицы дополнительного человеко-часа не превосходит 17,50 ф. ст. в неделю. Если сверхнормативное количество часов рабочего времени используется максимально, то новое максимальное значение еженедельного дохода составит:

$$P_{max} = 30 \times 10000/7 + 40 \times 10000/7 - \text{дополнительная стоимость} =$$
$$= (100000 - \text{дополнительная стоимость}) \text{ ф. ст. в неделю.}$$

Важно уяснить, что из дохода вычитаются только те издержки, которые не фигурируют в исходной постановке задачи. Предположим, например, что при производстве автомобильных деталей стоимость одного часа рабочего времени равна 4,00 ф. ст. При расчете дохода, приходящегося на единицу деталей каждого типа, бухгалтер будет использовать именно эту стоимость. Если привлекается дополнительный фонд рабочего времени, к примеру, сверхурочная работа, по 6,00 ф. ст. за 1 чел.-ч, то из них 4,00 ф. ст. уже учтены в показателях единичного дохода. Отдельно в счет нужно внести только дополнительные 2 ф. ст. за час.

12.4.2. Воздействие на оптимальное решение изменений в обеспечении нелимитирующими ресурсами

В примере **12.6** рассматривались два лимитирующих ограничения на труд и листовой металл. Остальные ограничения в первоначальном оптимальном решении не являются лимитирующими. Это ограничения на:

1. Производственные мощности для выпуска деталей типа X.
2. Производственные мощности для выпуска деталей типа Y.
3. Металлические стержни.
4. Постоянные заказы.
5. Профсоюзное соглашение.

Что происходит при изменении каждого из этих ограничений? Первые три ресурса используются в меньших или равных максимальному количествах. Любое увеличение запаса этих ресурсов не будет оказывать влияния на оптимальное решение задачи. Однако на него может влиять уменьшение запасов, соответствующих трем указанным ограничениям. Увеличение жесткости одного из нелимитирующих ограничений приведет к перемещению его линии в сторону начала координат. Сначала единственным изменением будет сокращение размеров допустимого множества. Однако когда линия ограничения переместится ниже исходной оптимальной крайней точки, данное ограничение станет лимитирующим, что приведет к появлению нового оптимального решения.

Предельные значения для этих ограничений ниже максимального уровня. Так, производственные мощности для деталей типа X можно сократить с 2250 до 1500 ч, т.е. на 750 ч, прежде чем это ограничение начнет оказывать воздействие на решение задачи. Производственные мощности для деталей типа Y могут быть сокращены с 1750 до 1250 ч, т.е. на 500 ч. Запас металлических стержней можно уменьшить с 10000 до 9250 кг, т.е. на 750 кг в неделю. Количественные выражения этих сокращений есть не что иное, как значения остаточных переменных, о которых мы упоминали выше. С ограничениями, для которых количество ресурсов больше либо равно минимальному, все наоборот. Любое сокращение минимального количества ресурсов приведет к увеличению размеров допустимого множества, но не окажет воздействия на оптимальное решение.

Любое увеличение правой части этих ограничений сначала приведет к сокращению размеров допустимого множества, а затем повлияет и на оптимальное решение. Если постоянные заказы на детали типа *X* возрастут на 900 и достигнут 1500 деталей в неделю, оптимальное решение начнет изменяться. Если предусмотренное профсоюзным соглашением число деталей возрастет не менее чем на 1250 и превысит 2750 деталей, допустимое множество станет пустым, и задача не будет иметь решения. Количественные выражения увеличения ресурсов — это значения избыточных переменных для этих ресурсов, о которых говорилось выше.

12.4.3. Воздействие на оптимальное решение изменений в коэффициентах целевой функции

Условия, для которых составлялась задача линейного программирования, неизбежно изменяются. Чаще всего эти изменения предполагают повторное выполнение формализации задачи, но должна существовать возможность идентифицировать воздействие незначительных изменений на решение исходной задачи. В этом разделе мы рассмотрим изменения коэффициентов целевой функции. Если цель состоит в максимизации еженедельного дохода, то изменение стоимости сырья приведет к изменению значений коэффициентов целевой функции.

В задаче о портфеле ценных бумаг, когда целью является максимальная ежегодная отдача инвестиций, на коэффициенты целевой функции может воздействовать изменение процентной ставки, происшедшее в одном из объектов вложения инвестиций.

Рассмотрим ситуацию, когда один из коэффициентов целевой функции изменяется во времени. Предположим, что

$$P = ax + 4y \text{ (ф. ст. в неделю)}$$

целевая функция, максимизирующая прибыль в задаче линейного программирования, где 4 — прибыль от выпуска единицы продукции Y (ф. ст.), a — прибыль от выпуска единицы продукции X (ф. ст.). Прибыль от продукции X может меняться. Предположим, что существует графическое изображение данной задачи, в котором значения переменных x и y отложены на соответствующих осях координат. Полезно переписать целевую функцию таким образом, чтобы y являлось зависимой переменной:

$$y = P/4 - (a/4)x.$$

Линия уровня целевой функции пересекает ось ординат в точке $P/4$, а тангенс угла ее наклона равен $-(a/4)$. Точка пересечения линии уровня с осью ординат не зависит от значения параметра a, однако угол ее наклона увеличивается с ростом a, и наоборот. Иными словами, с изменением параметра a линия уровня поворачивается. Незначительные перемещения ее в любом направлении обычно не приводят к изменению оптимальной крайней точки. Однако в результате более значительных изменений угла наклона линии уровня может появиться новая оптимальная крайняя точка. Полезно определить промежуток значений параметра a, для которых некоторая крайняя точка допустимого множества является оптимальной. Аналогичным образом можно поступать, если фиксированным является коэффициент целевой функции при переменной x, а коэффициент при y подвержен изменениям.

❐ **Пример 12.7.** Обратимся к примерам **12.2** и **12.5**, в которых рассматривается производство деталей к автомобилям. Допустимое множество выглядит следующим образом:

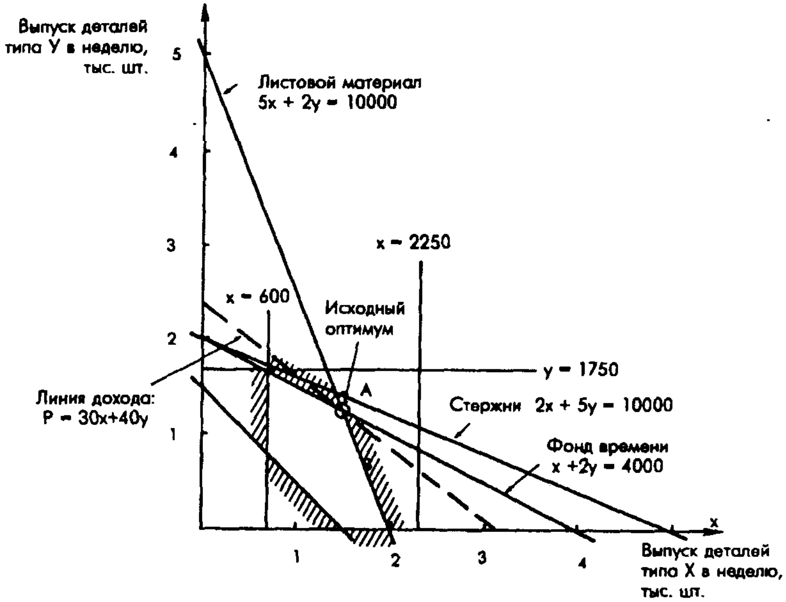


Рис. 12.17. Задача линейного программирования для производства деталей типа X и Y в неделю

Линия уровня еженедельного дохода имеет вид:

$$P = 30\,x + 40\,y \text{ (ф. ст. в неделю).}$$

На рис. 12.17 она проходит через оптимальную крайнюю точку A. Теперь вспомним, что доход от выпуска единицы деталей типа X может меняться. Каков промежуток значений единичного дохода, для которых A остается оптимальной крайней точкой? Единичный доход от выпуска деталей типа Y остается неизменным.

Решение

Перепишем уравнение дохода за неделю в следующем виде:

$$P = ax + 40\,y \text{ (ф. ст. в неделю),}$$

где a — единичный доход от выпуска деталей типа X. Преобразовав это уравнение, получим:

$$y = P/40 - (a/40)x.$$

Тангенс угла наклона линии дохода за неделю равен – (a/40). В исходном положении при a = 30 ф. ст. за единицу тангенс угла наклона равен – (30/40) = = – (3/4).

Если a меньше 30 ф. ст. за единицу, то наклон линии еженедельного дохода становится более пологим. В точке A линия уровня будет отклоняться в сторону лимитирующего ограничения на фонд рабочего времени. Это ведет к уменьшению оптимального значения функции P, дохода за неделю. Обратите внимание на рис 12.18. Если сильно уменьшать значение параметра a, то линия уровня еженедельного дохода совпадет с ограничением на фонд рабочего времени. Это показано на рис. 12.19.

Рис. 12.18. Уменьшение дохода от выпуска деталей типа X

Если и далее уменьшать значение параметра a, оптимум переместится из точки A в точку D (см. рис. 12.20). Следовательно, граничным является положение линии уровня дохода, при котором она совпадает с линией лимитирующего ограничения на фонд рабочего времени. Этому положению соответствует наименьшее значение a, для которого A является оптимальной крайней точкой.

Угол наклона линии ограничения на фонд рабочего времени можно найти, преобразовав данное ограничение к виду:

$$y = 4000/2 - 1/2\ x.$$

Тангенс угла наклона лимитирующего ограничения равен – (1/2). Нижний предел значений находится из условия – (a/40) = – (1/2), таким образом, a = 20 ф. ст. за единицу. Следовательно, единичный доход от выпуска деталей типа X может уменьшаться до 20 ф. ст. до того, как оптимум переместится из точки A в точку D.

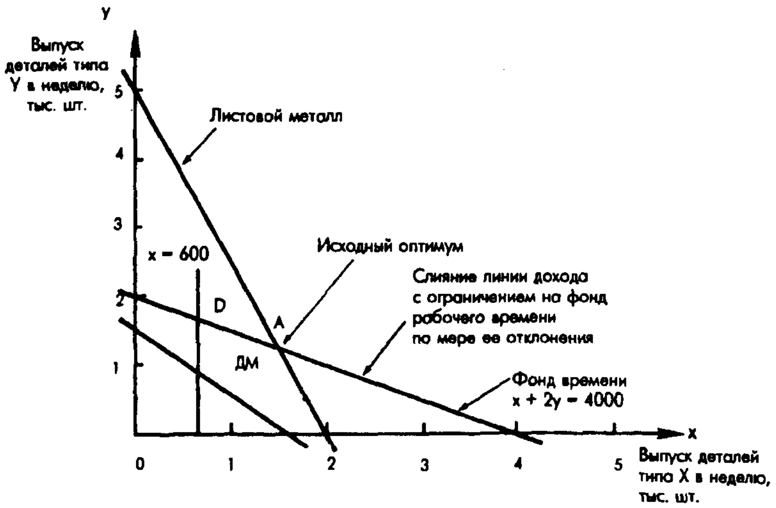


Рис. 12.19. Граничное положение линии дохода по мере его уменьшения

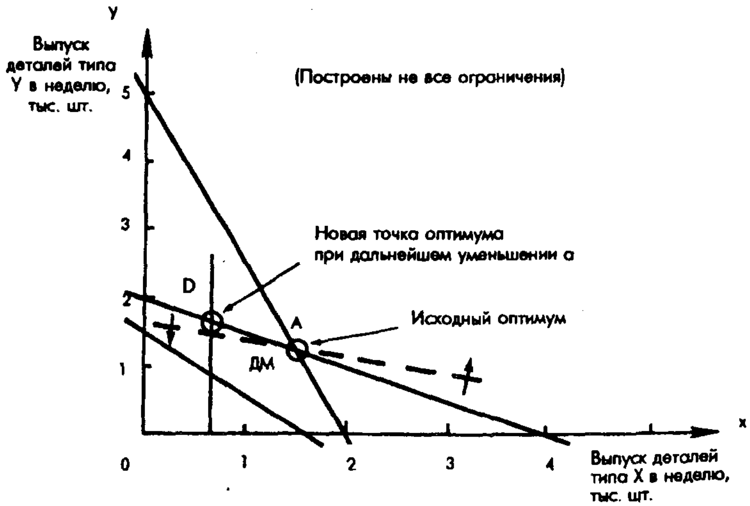


Рис. 12.20. Влияние дальнейшего сокращения дохода от выпуска деталей типа X

Причем оптимальный доход будет сокращаться, но оптимальный ассортиментный набор не изменится до тех пор, пока значение параметра a не опустится ниже 20 ф. ст. Аналогичным образом можно найти верхний предел значений a. С увеличением значения a линия еженедельного дохода становится все менее пологой и в конечном итоге окажется параллельной линии другого лимитирующего ограничения, а именно на листовой металл. Любое дальнейшее увеличение значения a вызовет изменение оптимальной крайней точки и перемещение ее в точку E. Это показано на рис. 12.21 и 12.22.

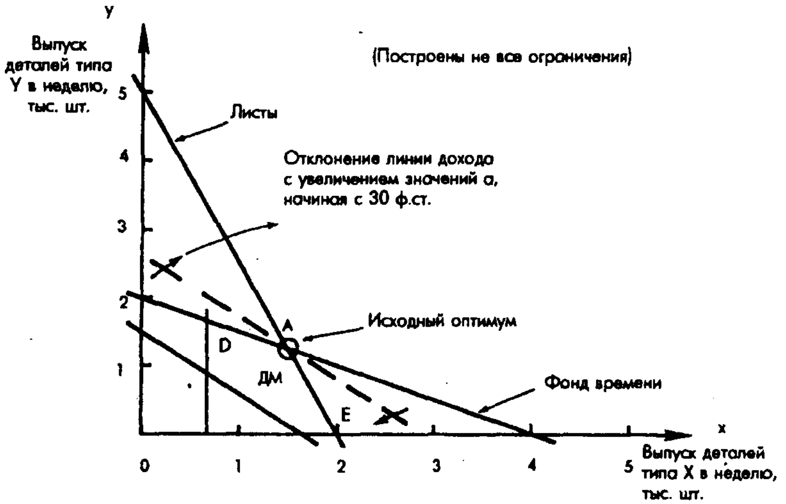


Рис. 12.21. Увеличение дохода от выпуска деталей типа X

Граничное положение линии уровня еженедельного дохода достигается в момент ее совпадения с ограничением на листовой металл. Этому положению соответствует верхний предел значений параметра a, для которых точка A является оптимальной крайней точкой допустимого множества. Угол наклона ограничения на листовой металл можно найти, преобразовав это уравнение к виду:

$$y = 10000/2 - (5/2)x.$$

Тангенс угла наклона лимитирующего ограничения равен $-(5/2)$, а верхний предел параметра a находится из условия $-(a/40) = -(5/2)$, следовательно, $a = 100$ ф. ст. за единицу. Таким образом, до того как оптимальный ассортиментный набор переместится из точки A в точку E, единичный доход от выпуска деталей типа X может возрастать до 100 ф. ст.

Два соответствующих предела значения единичного дохода от выпуска деталей типа Y можно найти аналогичным образом, если в изложенной схеме расчетов заменить x на y. Предположим, что значение коэффициента целевой функции при х является неизменным, тогда:

$$P = 30\,x + b\,y \text{ (ф. ст. в неделю)}$$

и

$$x = P/30 - (b/30)y.$$

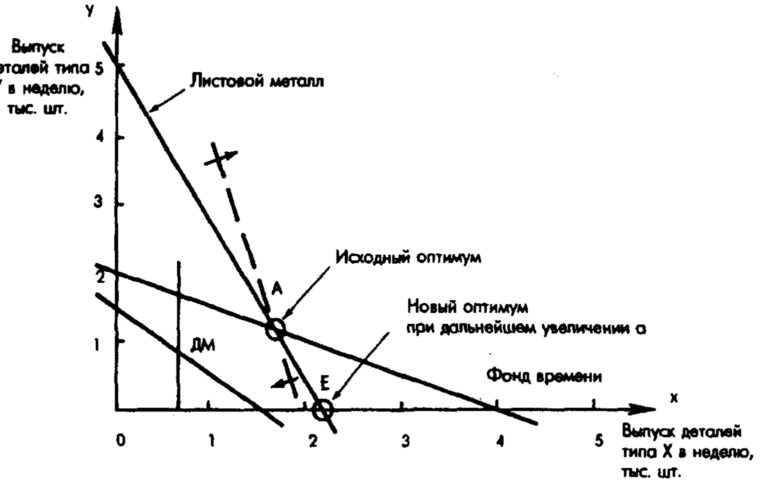


Рис. 12.22. Воздействие увеличения дохода от выпуска деталей типа X после прохождения линией уровня ее лимитирующего положения

По мере увеличения или уменьшения параметра b граничные положения линии уровня еженедельного дохода определяются теми же двумя ограничениями, что и в предыдущем случае. Теперь необходимо записать уравнения этих ограничений так, чтобы х выступал в качестве зависимой переменной:

Фонд рабочего времени: $x = 4000 - 2y$.

Тангенс угла наклона равен -2, следовательно, предельное значение достигается при условии $-(b/30) = -2$, т.е. $b = 60$ ф. ст. за единицу.

Листовой металл: $x = 10000/5 - (2/5)y$.

Тангенс угла наклона равен $-(2/5)$, для предельного значения выполняется условие: $-(b/30) = -(2/5)$, следовательно, $b = 12$ ф. ст. за единицу.

Крайняя точка А соответствует оптимальному ассортиментному набору только до тех пор, пока доход от выпуска деталей типа Y изменяется в пределах от 12 до 60 ф. ст. за единицу. В случае, если показатели единичных доходов от выпуска деталей типа X или Y будут изменяться по сравнению с их исходными значениями, значение оптимального дохода также будет отличным от 95000 ф. ст.

12.5. СИМПЛЕКС-МЕТОД РЕШЕНИЯ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ С МНОЖЕСТВОМ ПЕРЕМЕННЫХ

Если задача линейного программирования содержит более двух переменных, то ее решение требует применения некоторого алгебраического метода. Принцип, лежащий в основе решения задачи с множеством переменных, достаточно прост. Предполагается, что оптимальному решению соответствует одна из крайних точек допустимого множества. Следовательно, необходимо провести оценку значений целевой функции во всех крайних точках допустимого множества и выбрать ту из них, в которой достигается оптимальное значение целевой функции. Нами используются методы матричной алгебры и такой алгоритм перехода от одной крайней точки допустимого множества к другой, при котором переход осуществляется только в случае, когда значение целевой функции улучшается. Если оказывается, что некоторое базисное решение улучшить уже нельзя, то оно является оптимальным планом задачи. Этот алгоритм получил название **симплекс-метода**. Нет необходимости вдаваться в детали алгоритма симплекс-метода, поскольку для решения задач линейного программирования с множеством переменных используется, как правило, один из компьютерных пакетов прикладных программ, которые общедоступны и широко применяются для этих целей. Однако для более полной интерпретации и всесторонней оценки решения задачи линейного программирования, полученного с использованием пакета прикладных программ, с основными принципами этого метода полезно ознакомиться.

В обычном симплекс-методе принимается предпосылка о максимизации целевой функции задачи линейного программирования в условиях системы ограничений со знаком "$\le$". Это означает, что при реализации данного алгоритма в качестве начальной крайней точки может быть выбрано начало координат. Поиск оптимального решения всегда начинается со значения целевой функции, равного нулю.

Симплекс-метод можно применять также и в решении задач минимизации, и в решении задач, система ограничений которых содержит ограничения со знаком "$\ge$" или уравнения. Эта процедура предусматривает введение в задачу искусственных, а также избыточных и остаточных переменных. Мы не будем подробно останавливаться на подобных усложнениях, поскольку обычно задачи линейного программирования решаются с помощью пакетов прикладных программ, в которых указанные переменные вводятся в модель автоматически.

Базовую модель, с которой мы будем работать в дальнейшем, формально можно представить следующим образом:

Максимизировать $Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$.

Здесь c_i – константы. Данная функция максимизируется в условиях системы m линейных ограничений:

$$\begin{array}{l}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n \le b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2n}x_n \le b_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \ldots + a_{3n}x_n \le b_3 \\
\quad\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\qquad \vdots
\end{array}$$

$$a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \ldots + a_{mn}x_n \le b_m$$

$$x_i \ge 0$$

Данная система содержит n переменных и m ограничений. Первая цифра двойных индексов коэффициентов в левой части системы ограничений соответствует номеру ограничения, вторая — номеру переменной. Например, a_{32} принадлежит ограничению 3 и является коэффициентом при переменной x_2. Проиллюстрируем применение симплекс-метода на примере простой задачи с двумя переменными, решение которой было получено нами ранее с помощью графического метода. Этот прием позволит нам сравнить решение, полученное графическим и алгебраическим методами.

❐ **Пример 12.8**. Некоторая фирма производит два вида продуктов X и Y в условиях ограничений на три вида сырья: RM1, RM2 и RM3. Целью фирмы является выбор такого ассортиментного набора, при котором достигается максимум прибыли в неделю. Задача линейного программирования имеет вид:

1. Выпускается x единиц продукта X в неделю и y единиц продукта Y в неделю.
2. Максимизируется еженедельная прибыль P (ф. ст.), где $P = 2x + y$.
3. Максимизация осуществляется в условиях ограничений:

RM1:	$3x \le 27$ кг в неделю;
RM2:	$2y \le 30$ кг в неделю;
RM3:	$x + y \le 20$ кг в неделю;
	$x, y \ge 0$.

4. Найти оптимальный ассортиментный набор и максимальную прибыль за неделю. Определить свободный запас каждого ресурса.

Решение

Графический метод. В каждое ограничение модели вводятся остаточные переменные s_i. Максимизировать:

$$P = 2x + y \text{ (ф. ст. в неделю)}$$

при ограничениях:

RM1:	$3x + s_1 = 27$ кг в неделю;
RM2:	$2y + s_2 = 30$ кг в неделю;
RM3:	$x + y + s_3 = 20$ кг в неделю;
	$x, y, s \ge 0$.

Изобразим систему ограничений графически (см. рис. 12.23)

Точка с координатами $x = 5$, $y = 5$ принадлежит допустимому множеству. Еженедельная прибыль в этой точке составит:

$$P = 2 \times 5 + 5 = 15 \text{ ф. ст. в неделю.}$$

В качестве типичной линии уровня прибыли выберем прямую

$$15 = 2x + y \text{ (ф. ст. в неделю).}$$

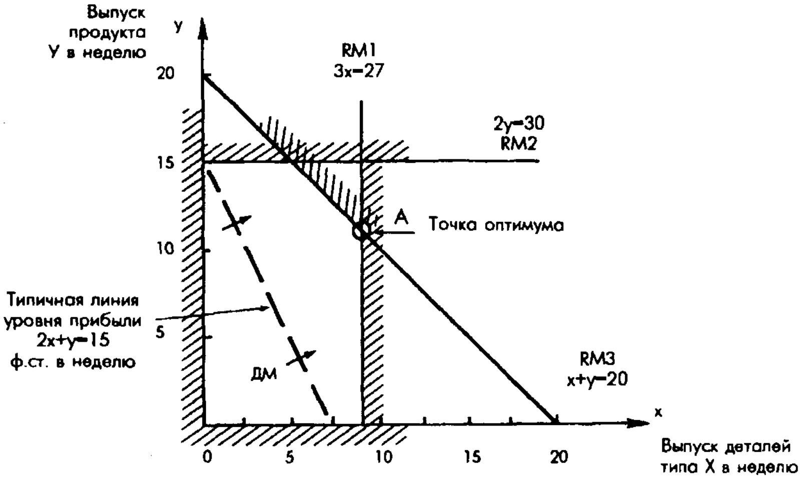


Рис. 12.23. Задача линейного программирования для объемов выпуска продуктов X и У в неделю

Точка с координатами x = 0, y = 15 также принадлежит этой прямой. Линия уровня изображена на приведенном выше графике пунктиром. Если осуществлять перемещение линии уровня параллельно ее начальному положению (отмеченному пунктиром), то легко можно убедиться, что оптимум находится в крайней точке А. Эта точка лежит на пересечении линий ограничений RM1 и RM3. Решение системы этих уравнений дает следующие результаты:

$$3x = 27, \text{ следовательно, } x = 9.$$

Подставив это значение во второе уравнение системы, получим:

$$x + y = 20, \text{ следовательно, } y = 11.$$

Оптимальным ассортиментным набором является производство 9 единиц продукта X и 11 единиц продукта Y в неделю. Таким образом, максимальная прибыль, получаемая за неделю, составит:

$$P_{max} = 2 \times 9 + 11 = 29 \text{ ф. ст. в неделю.}$$

Сырье типа 1 и 3 используется полностью, однако существует свободный запас сырья типа 2, т.е.

$$2 \times 11 + s_2 = 30,$$

следовательно, $s_2 = 8$ кг в неделю.

Решение

Симплекс-метод. Представим коэффициенты, стоящие в левой части системы ограничений, в матричной форме. За обозначения столбцов примем переменные, которым они соответствуют. Значения правой части ограничений запишем в отдельном столбце матрицы справа. За обозначения строк примем обозначения соответствующих переменных, которые являются базисными (имеющими ненулевые значения) переменными в начальной крайней точке (начале координат). Наконец, введем в таблицу дополнительную строку, соответствующую коэффициентам целевой функции γ. Существует несколько немного отличных друг от друга вариантов решения задачи симплекс-методом. В том из них, который излагается ниже, требуется вводить коэффициенты целевой функции с отрицательным знаком. Полученная в результате применения данной процедуры матрица называется **симплекс-таблицей**, а сама эта процедура представляет собой *Шаг 1* в алгоритме симплекс-метода.

Таблица 12.2. **Первая симплекс-таблица**

Базисные переменные	*Переменные*					*Правая часть*
	x	y	s_1	s_2	s_3	*b*
s_1	3	0	1	0	0	27
s_2	0	2	0	1	0	30
s_3	1	1	0	0	1	20
Целевая функция, P	−2	−1	0	0	0	0

Шаг 2. В строке коэффициентов целевой функции найдем наибольшее отрицательное значение (− 2). Столбец, соответствующий этому значению, называется ведущим. Разделим значения правой части на соответствующие значения ведущего столбца. В результате получим ряд отношений.

Таблица 12.3. **Первая симплекс-таблица с учетом отношений**

Базисные переменные	*Переменные*					*Правая часть*	*Отношения b/элемент ведущего столбца*
	x	y	s_1	s_2	s_3	*b*	
s_1	3*	0	1	0	0	27	27/3=9 ← ведущая строка
s_2	0	2	0	1	0	30	30/0 = ∞
s_3	1	1	0	0	1	20	20/1 = 20
Целевая функция P	−2	−1	0	0	0	0	

↑

Ведущий столбец x

Шаг 3. Выберем среди полученных отношений наименьшее положительное отношение. В нашем случае оно равно 9. Соответствующая ему строка s_1 является ведущей. Пересечение ведущего столбца и ведущей строки дает ведущий элемент 3, в приведенной выше табл. 12.3 он отмечен знаком "*".

Шаг 4. Разделим все элементы ведущей строки на ведущий элемент 3. Заменим все элементы ведущей строки на полученные новые значения (табл. 12.4). Обозначение ведущей строки s_1 заменим на обозначение ведущего столбца x. Новые переменные, соответствующие обозначениям строк, — это базисные переменные второго базисного решения.

Таблица 12.4. **Вторая симплекс-таблица**

Базисные переменные	*Переменные* x	y	s_1	s_2	s_3	*Правая часть* b	
x	1	0	1/3	0	0	9	Новая R_1 = Прошлая R_1 + + ведущий элемент (3)
s_2	0	2	0	1	0	30	Новая R_2 = Прошлая R_2 − − 0 × Новая R_1
s_3	0	1	−1/3	0	1	11	Новая R_3 = Прошлая R_3 − − 1 × Новая R_1
Целевая функция P	0	−1	2/3	0	0	18	Новая P = Прошлая P − − (−2) × Новая R_1

Шаг 5. Применив к строкам матрицы арифметические операции (строчные операции в матричной алгебре), приведем все остальные элементы ведущего столбца x к нулю. В качестве базиса в этих арифметических операциях должна использоваться только ведущая строка.

Обозначим через R_i i-ю строку. Соотношение "Новая R_3 = Прошлая R_3 − − Новая R_1" означает, что новые элементы строки 3 были получены вычитанием элементов новой ведущей строки (строка 1) из соответствующих элементов ведущей строки 3 предыдущего шага. Выполненные операции перечислены в крайнем правом столбце табл. 12.4.

Шаг 6. Шаги 2–5 повторяются до тех пор, пока не будет достигнута неотрицательность всех элементов в строке целевой функции.

Таблица 12.5. **Вторая симплекс-таблица с отношениями**

Базисные переменные	*Переменные* x	y	s_1	s_2	s_3	*Правая часть* b	*Отношения b/элемент ведущего столбца*
x	1	0	1/3	0	0	9	9/0 = ∞
s_2	0	2	0	1	0	30	30/2 = 15
s_3	0	1*	−1/3	0	1	11	11/1=11 ← ведущая строка
Целевая функция P	0	−1	2/3	0	0	18	

↑

Ведущий столбец y

Таблица 12.6. **Третья, итоговая, симплекс-таблица**

Базисные переменные	Переменные x	y	s_1	s_2	s_3	Правая часть b	
x	1	0	1/3	0	0	9	Новая R_1 = Прошлая R_1 – – 0 × Новая R_3
s_2	0	0	2/3	1	–2	8	Новая R_2 = Прошлая R_2 – – 2 × Новая R_3
y	0	1	–1/3	0	1	11	Новая R_3 = Прошлая R_3 + + ведущий элемент (1)
Целевая функция P	0	0	1/3	0	1	29	Новая P = Прошлая P – – (–1) × Новая R_3

Теперь все элементы в строке целевой функции либо положительны, либо равны нулю, следовательно, представленное в данной таблице решение является оптимальным.

Чтобы дать интерпретацию итоговой симплекс-таблице, обратим сначала внимание на значения ее крайних элементов.

Таблица 12.7. **Интерпретация итоговой симплекс-таблицы**

Базисные переменные	Переменные x	y	s_1	s_2	s_3	Правая часть b
x s_2 y						9 = значение x 8 = значение ресурса $RM_2=s_2$ 11 = значение y
Целевая функция P	0	0	1/3	0	1	29 = максимальное значение прибыли

↓ *Стоимость введения в решение одной небазисной переменной* ↓ *Теневые цены*

Базисными называются переменные, которые имеют ненулевые значения в крайней точке допустимого множества. Значения базисных переменных находятся в соответствующих строках столбца b. Следовательно,

x = 9 единиц в неделю;

y = 11 единиц в неделю,

а значение остаточной переменной для сырья 2 составило 8 кг в неделю.

Все остальные переменные имеют нулевые значения, т.е. остаточные переменные ограничений 1 и 3, s_1 и s_3 соответственно равны нулю. Это означает, что данные ограничения являются лимитирующими, а сырье типов 1 и 3 расходуется полностью. Оптимальное значение целевой функции указано в строке целевой функции столбца b. Максимальное значение получаемой за неделю прибыли составляет 29 ф. ст. Полученное решение полностью совпадает с графическим решением задачи, найденным ранее. Элементы, стоящие на пересечении строки целевой функции и столбцов остаточных переменных, соответствуют, как показано в табл. 12.7, **теневым ценам** ресурсов. Теневая цена для ограничения 1, т.е. цена ресурса RM1, составляет 1/3 ф. ст. за 1 кг, а теневая цена для ограничения 3 – 1 ф. ст. за 1 кг. Это означает, что если реализуется 1 кг ресурса RM1 сверх нормативного запаса, прибыль за неделю возрастает на 33 пенса (минус любые издержки сверх обычной стоимости RM1). Аналогичным образом, если используется сверхнормативное количество ресурса RM3 в 1 кг, рост прибыли за неделю составит 1 ф. ст. (минус любые дополнительные издержки). Найденные значения теневых цен можно проверить, вычислив их с помощью графического метода. Чтобы проиллюстрировать эту процедуру, обратимся к ограничению 1.

Ограничение 1, соответствующее сырью 1, имеет вид: $3x = 27$ кг в неделю. Снизив жесткость этого ограничения на 1 кг, получим: $3x = 28$. Оптимальная крайняя точка по-прежнему будет находиться на пересечении линий ограничений 1 и 3. Чтобы убедиться в справедливости этого положения, достаточно взглянуть на соответствующий график. Новая точка оптимума имеет следующие координаты:

$$x = 28/3 = 9\frac{1}{3},$$

а соотношение

$$28/3 + y = 20$$

приводит к тому, что

$$y = 32/3 = 10\frac{2}{3}.$$

Новое максимальное значение прибыли за неделю составит:

$$2 \times (28/3) + (32/3) = 88/3 = 29{,}33 \text{ ф. ст. в неделю.}$$

Увеличение прибыли на 33 пенса произошло в результате использования 1 кг ресурса RM1 дополнительно. Следовательно, теневая цена ресурса RM1 равна 33 пенсам за 1 кг.

Оставшиеся крайние значения итоговой таблицы — это элементы, лежащие на пересечении строки "целевая функция" и столбцов переменных задачи. В нашем примере значения, соответствующие столбцам x и y, равны нулю. Эти элементы были бы ненулевыми, если бы соответствующие переменные в оптимальном решении не являлись базисными. Например, если бы в оптимальном решении утверждалось, что следует производить только продукт X, переменная y была бы небазисной, т.е. выполнялось бы соотношение $y = 0$, то в этом случае значение, указанное на пересечении строки "целевая функция" и столбца b, показало бы, на сколько уменьшится максимальное значение целевой функции при выпуске единицы продукта y.

Предположим, что в результате решения данной задачи мы получили итоговую таблицу следующего вида:

Таблица 12.8. **Модификация итоговой таблицы**

Базисные переменные	*Переменные* x	y	s_1	s_2	s_3	*Правая часть* b
x						9
s_2						8
s_3						11
Целевая функция P	0	0,5	1/3	0	0	18 = максимальное значение прибыли

Оптимальный ассортиментный набор для этого решения — это выпуск только продукта X в количестве 9 единиц. Если по тем или иным причинам некоторое количество продукта У все же необходимо произвести, значение целевой функции уменьшится на 0,5 ф. ст., приходящихся на каждую производимую единицу продукта У.

12.6. АНАЛИЗ ЧУВСТВИТЕЛЬНОСТИ И СИМПЛЕКС-МЕТОД

Итоговую таблицу симплекс-алгоритма можно использовать для проведения анализа чувствительности решения задачи линейного программирования. Значения остаточных переменных в столбцах, соответствующих лимитирующим ограничениям, представляют собой изменение значений базисных переменных при использовании дополнительной единицы лимитирующего ресурса.

❐ **Пример 12.9.** В качестве итоговой симплекс-таблицы будем пользоваться таблицей 12.6 примера **12.8**. С помощью данных этой таблицы необходимо определить:

1. Влияние на оптимальное решение задачи сверхнормативного запаса ресурса RM1 в количестве 1 кг;
2. Влияние на оптимальное решение задачи сверхнормативного запаса ресурса RM1 в количестве 2 кг;
3. Влияние на оптимальное решение задачи сверхнормативного запаса ресурса RM3 в количестве 5 кг;
4. Максимальное дополнительное количество ресурса RM3, которое используется полностью и не приводит к созданию излишка ресурса;
5. Влияние на оптимальное решение задачи уменьшения запаса ресурса RM1 на 2 кг.

Решение.

Воспроизведем формулировку задачи линейного программирования и данные итоговой симплекс-таблицы.

Максимизировать еженедельную прибыль P, где $P = 2x + y$ (ф. ст. в неделю)

в условиях ограничений на

$$\begin{array}{lrl} \text{RM1:} & 3x & \le 27 \text{ кг в неделю;} \\ \text{RM2:} & 2y & \le 30 \text{ кг в неделю;} \\ \text{RM3:} & x + y & \le 20 \text{ кг в неделю;} \\ & x,\ y & \ge 0. \end{array}$$

Таблица 12.9. **Итоговая симплекс-таблица**

Базисные переменные	*Переменные* x	y	s_1	s_2	s_3	*Правая часть* b
x	1	0	1/3	0	0	9
s_2	0	0	2/3	1	−2	8
s_3	0	1	−1/3	0	1	11
Целевая функция P	0	0	1/3	0	1	29

1. Если существует сверхнормативный запас ресурса RM1 в количестве 1 кг, жесткость соответствующего лимитирующего ограничения снижается на 1 кг. Элементы столбца s_1 — это изменения базисных переменных, вызванные снижением жесткости данного ограничения. Ниже приводится итоговая таблица, в которой представлены только значения соответствующих элементов и процедура их расчета.

Таблица 12.10. **Модифицированные элементы итоговой симплекс-таблицы (при наличии 1 кг RM1 дополнительно)**

Базисные переменные	*Переменные* x	y	s_1	s_2	s_3	*Правая часть, модифицированные* b
x			1/3			$9+(1/3)=9\frac{1}{3}$
s_2			2/3			$8+(2/3)=8\frac{2}{3}$
y			−1/3			$11-(1/3)=10\frac{2}{3}$
Целевая функция P			1/3			$29+(1/3)=29\frac{1}{3}$

Один дополнительный килограмм ресурса RM1 ведет к увеличению значения x на 1/3 единицы, росту остатка ресурса RM2 на 2/3 кг, снижению значения y на 1/3 единицы и к увеличению значения максимальной прибыли за неделю на 1/3 ф. ст., что соответствует значению теневой цены на ресурс RM1. Новое оптимальное решение состоит в производстве $9\frac{1}{3}$ продукта X и $10\frac{2}{3}$ продукта У в неделю. При этом значение остаточной переменной, т.е. неиспользуемое количество ресурса 2, равно $8\frac{2}{3}$ кг. Остальные переменные принимают нулевые значения. Равенство нулю остаточных переменных для ограничений 1 и 3 означает полное использование ресурсов RM1 и RM3. Следовательно, данные ограничения являются лимитирующими. Максимальное значение прибыли за неделю составляет 29,33 ф. ст. Приведенные значения соответствуют графическому решению задачи (рис. 12.24).

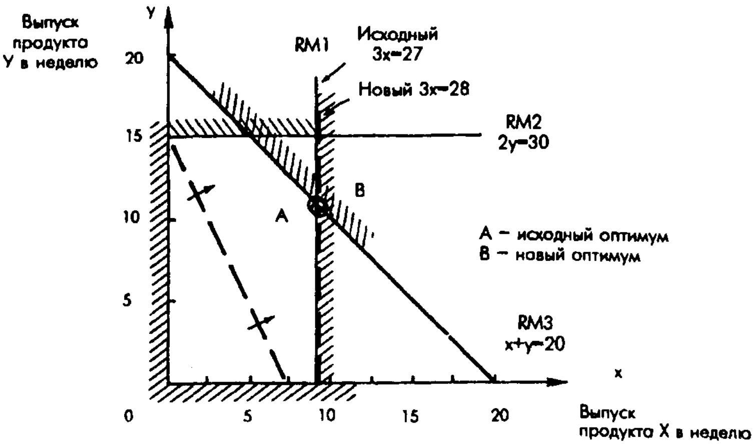


Рис. 12.24. Производство продуктов X и У в неделю при наличии 1 кг ресурса RM1 дополнительно

2. Если имеется сверхнормативный запас RM1 в количестве 2 кг, жесткость соответствующего лимитирующего ограничения также снижается на 2 кг. Элементы столбца s_1 умножаются на 2. Полученные новые значения характеризуют изменения в базисных переменных, происшедшие в связи с использованием 2 кг ресурса RM1 дополнительно. В табл. 12.11 показаны значения соответствующих элементов итоговой таблицы и процедура их расчета.

Таблица 12.11. **Модифицированные элементы итоговой симплекс-таблицы (при наличии 2 кг RM1 дополнительно)**

Базисные переменные	*Переменные*					*Правая часть, модифицированные b*
	x	y	s_1	s_2	s_3	
x			1/3 × 2			9+(2/3)= $9\frac{2}{3}$
s_2			2/3 × 2			8+(4/3)= $9\frac{1}{3}$
y			−1/3 × 2			11−(2/3)=$10\frac{1}{3}$
Целевая функция P			1/3 × 2			29+(2/3)= $29\frac{2}{3}$

Новое оптимальное решение состоит в выпуске $9\frac{2}{3}$ и $10\frac{1}{3}$ единиц продуктов X и У соответственно в неделю. Остаток, соответствующий ограничению 2, равен $9\frac{1}{3}$ кг.

Значения других остаточных переменных s_1 и s_2 являются нулевыми. Это означает, что соответствующие им ограничения являются лимитирующими. Максимальное значение получаемой за неделю прибыли равно 29,67 ф. ст. Указанные компоненты оптимального решения можно проиллюстрировать графически по аналогии с п.1.

3. В случае, если имеется сверхнормативный запас ресурса RM3 в размере 5 кг, жесткость соответствующего ему лимитирующего ограничения также понижается на 5 кг. Элементы столбца s_3 умножаются на 5. В модифицированной итоговой таблице (табл. 12.12) показаны изменения значений базисных переменных, связанные с использованием дополнительных 5 кг ресурса RM3.

   В данном случае возникает новая проблема. Значение остаточной переменной s_2 для ресурса 2 становится отрицательным. Это недопустимо, так как по условиям задачи значения переменных должны быть положительными или равными нулю. Если обратиться к графическому решению задачи, легко можно понять, почему так происходит. Жесткость ограничения RM3 снижается настолько, что оно перестает быть лимитирующим. В симплекс-таблицу вводится точка, не принадлежащая допустимому множеству. Это означает, что дополнительное привлечение всех 5 кг ресурса RM3 невозможно. Данная проблема рассматривается в п. 4.

Таблица 12.12. **Модифицированные элементы итоговой симплекс-таблицы (в условиях наличия 5 кг RM3 дополнительно)**

Базисные переменные	*Переменные*					*Правая часть, модифицированные b*
	x	y	s_1	s_2	s_3	
x					0×5	9 + (0) = 9
s_2					−2×5	8 +(−10) = −2
y					1×5	11 + (5) = 16
Целевая функция P					1×5	29 + (5) = 34

4. Ограничение RM3 представлено в итоговой таблице столбцом s_3. Единственным отрицательным значением в столбце s_3 является итоговое значение переменной s_2, равное −2. При снижении жесткости ограничения RM3 на единицу значение s_2 снижается на 2 единицы, но оно не может стать отрицательным. Лимитирующее положение линии ограничения на RM3 возникает, когда s_2 достигает нуля. Предположим, лимитирующее положение достигается при снижении жесткости ограничения на RM3 на r кг, тогда s_2 примет значение, равное нулю, следовательно,

$$8 + (-2 \times r) = 0.$$

Таким образом, $r = 4$ кг. До того как ограничение RM3 перестанет быть лимитирующим, его жесткость может быть снижена на 4 кг, с 20 до 24 кг. Граничного положения линия данного ограничения достигает при прохождении через точку пересечения ограничений RM1 и RM2, для которой $x = 9$, $y = 15$, а линия ограничения RM3 имеет вид: $x + y = 24$.

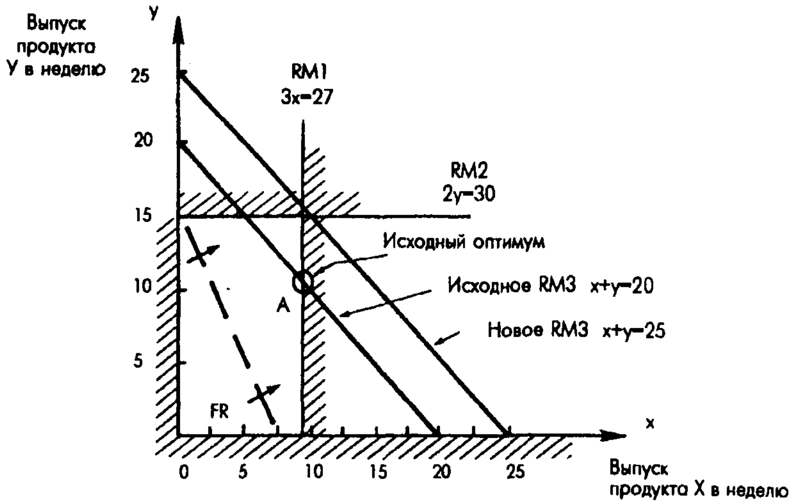


Рис. 12.25. Задача линейного программирования для производства продуктов X и У при наличии 5 кг ресурса RM3 дополнительно

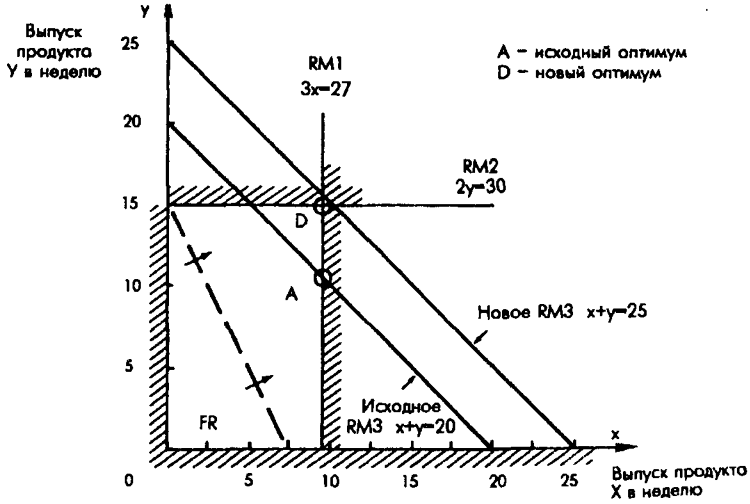


Рис. 12.26. Задача линейного программирования для производства продуктов X и У в неделю в условиях максимального использования ресурса RM3

5. Если количество ресурса RM1, имеющееся в распоряжении производителя, уменьшается на 2 кг, жесткость лимитирующего ограничения возрастает на 2 кг. Значения элементов столбца s_1 умножаются на 2. Полученные значения вычитаются из соответствующих значений базисных переменных. Полученная в результате описанной процедуры итоговая таблица представлена ниже.

Таблица 12.13. **Модифицированные элементы итоговой симплекс-таблицы (в условиях уменьшения запаса RM1 на 2 кг)**

Базисные переменные	*Переменные* x	y	s_1	s_2	s_3	*Правая часть, модифицированные b*
x			1/3 × 2			9–(2/3) = $8\frac{1}{3}$
s_2			2/3 × 2			8–(4/3) = $6\frac{2}{3}$
y			–1/3 × 2			11–(–2/3)= $11\frac{2}{3}$
Целевая функция P			1/3 ×2			29–(2/3)= $28\frac{1}{3}$

Новое оптимальное решение состоит в выпуске $8\frac{1}{3}$ и $11\frac{2}{3}$ единиц продуктов X и Y в неделю соответственно. Остаточная переменная ограничения 2 равна $6\frac{2}{3}$ кг.

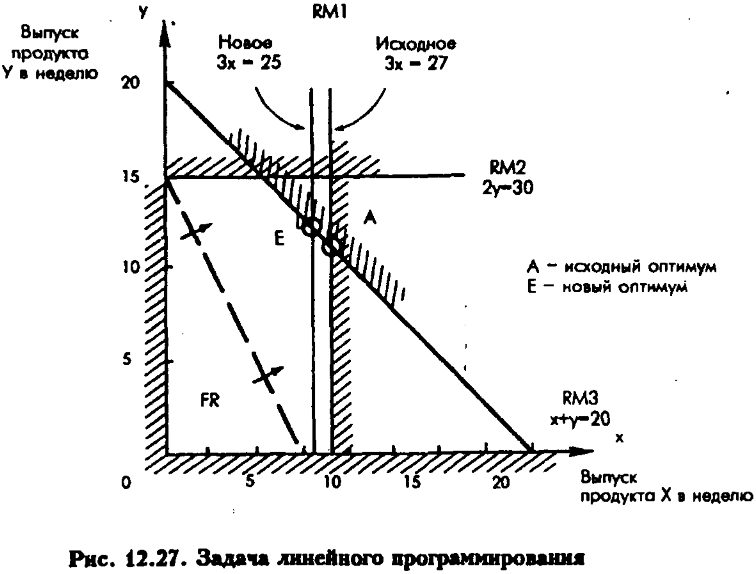


Рис. 12.27. Задача линейного программирования для производства продуктов X и Y в неделю (в условиях снижения запаса ресурса RM1 на 2 кг)

Остаточные переменные, соответствующие ограничениям 1 и 3, принимают нулевые значения. Это значит, что данные ограничения являются лимитирующими. Максимальное значение прибыли за неделю равно 28,33 ф. ст. На рис. 12.27 представлено графическое решение данного варианта задачи.

Проведение подобного анализа вручную довольно утомительно, даже если симплекс-метод используется для решения простейшей задачи линейного программирования с двумя переменными. Обычно всю необходимую информацию можно почерпнуть из стандартных пакетов прикладных программ по линейному программированию. На практике анализ чувствительности многомерных задач осуществляется именно таким путем. Однако основные принципы подобного анализа полностью совпадают с принципами анализа чувствительности задачи линейного программирования с двумя переменными, изложенными выше.

12.7. ДВОЙСТВЕННАЯ МОДЕЛЬ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Двойственная модель линейного программирования используется для изучения поставленной проблемы с точки зрения, отличной от той, которая исследуется в обычной прямой задаче. Прямая и двойственная модели приводят к одному и тому же решению и к получению одинаковой информации о чувствительности модели. Единственная причина, по которой предпочтение отдается той или иной модели, состоит в том, что одну из них решить, как правило, легче, чем другую. Однако по мере все более широкого распространения пакетов прикладных программ альтернативное использование прямой или двойственной задачи становится менее существенным. Переменные двойственной модели являются для исходной, или прямой, модели теневыми ценами ресурсов. Структура двойственной и прямой задачи одинакова. Если прямая модель линейного программирования построена, из нее легко получить соответствующую двойственную модель. В общем виде задачу линейного программирования можно сформулировать следующим образом:

Максимизировать

$$Z = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

в условиях системы из m линейных ограничений:

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n &\le b_1;\\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n &\le b_2;\\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n &\le b_3;\\
\vdots \qquad \vdots \qquad \vdots \qquad &\vdots\\
a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n &\le b_m;\\
x_i &\ge 0.
\end{aligned}$$

Сформулированная выше задача линейного программирования является задачей максимизации, а все ее ограничения имеют знак "≤". К этому виду можно привести любую модель линейного программирования, а затем построить двойственную к ней, как это будет показано ниже. Двойственная модель имеет следующий вид:

Минимизировать

$$G = b_1y_1 + b_2y_2 + \ldots + b_my_m$$

в условиях системы из n линейных ограничений:

$$\begin{aligned}
&a_{11}y_1 + a_{21}y_2 + a_{31}y_3 + \ldots + a_{m1}y_m \ge c_1;\\
&a_{12}y_1 + a_{22}y_2 + a_{32}y_3 + \ldots + a_{m2}y_m \ge c_2;\\
&a_{13}y_1 + a_{23}y_2 + a_{33}y_3 + \ldots + a_{m3}y_m \ge c_3;\\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots\\
&a_{1n}y_1 + a_{2n}y_2 + a_{3n}y_3 + \ldots + a_{mn}y_m \ge c_n;\\
&\qquad\qquad\qquad\qquad\qquad y_i \ge 0
\end{aligned}$$

В этой задаче m двойственных переменных y, каждая из которых соответствует одному из m ограничений прямой задачи, и n ограничений, каждое из которых связано с одной из n переменных x прямой задачи. Коэффициенты целевой функции прямой задачи c и значения правой части ограничений b в двойственной задаче меняются местами. Строки коэффициентов левой части системы ограничений прямой модели становятся столбцами в двойственной, а столбцы — строками. Двойственные переменные y являются теневыми ценами ресурсов в прямой задаче, и наоборот. В данном случае целевая функция двойственной задачи минимизируется, а целевая функция прямой задачи — максимизируется. Если ограничения прямой задачи имеют знак "≤", то ограничения двойственной задачи записываются со знаком "≥".

❒ **Пример 12.10.** Некоторая фирма выпускает два продукта R и Q, каждый из которых требует двух видов сырья RM1 и RM2. Для выпуска 1 кг продукта R необходимо 2 кг сырья RM1 и 3,5 кг сырья RM2. Производство 1 кг продукта Q требует 3 кг RM1 и 1,5 кг RM2. В распоряжении фирмы имеются 10 кг RM1 и 12 кг RM2 в неделю, трудовые ресурсы и производственные мощности – в неограниченном количестве, кроме того, фирма может реализовать всю произведенную продукцию. Прибыль от выпуска единицы продукта R составляет 5 ф. ст., а от выпуска единицы продукта Q – 8 ф. ст.

1. Для изложенной проблемы сформулируем задачу линейного програмирования, в которой максимизируется прибыль.
2. Построим двойственную модель линейного программирования.
3. Объясним взаимосвязи между моделями, построенными в п.1 и 2.
4. Для обеих моделей нужно найти оптимальное решение графическим методом.

Решение

1. Производится x_1 кг продукта R и x_2 кг продукта Q в неделю. Максимизируется полученная за неделю прибыль P (ф. ст.), где

$$P = 5\,x_1 + 8\,x_2 \text{ (ф. ст. в неделю)}$$

в условиях следующей системы ограничений:

$$\begin{aligned} &\text{RM1:} & 2\,x_1 + 3\,x_2 &\le 10 \text{ кг в неделю;} \\ &\text{RM2:} & 3{,}5\,x_1 + 1{,}5\,x_2 &\le 12 \text{ кг в неделю;} \\ & & x_1, x_2 &\ge 0. \end{aligned}$$

2. Используя формулировку прямой модели, построим двойственную модель:
Минимизировать $G = 10\,y_1 + 12\,y_2$ (ф. ст. в неделю)

в условиях следующей системы ограничений:

$$\begin{aligned} &\text{Продукт R:} & 2\,y_1 + 3{,}5\,y_2 &\ge 5 \text{ ф. ст. за единицу;} \\ &\text{Продукт Q:} & 3\,y_1 + 1{,}5\,y_2 &\ge 8 \text{ ф. ст. за единицу;} \\ & & y_1, y_2 &\ge 0. \end{aligned}$$

3. **Прямая модель**. Переменные модели — это количество каждого продукта, которое необходимо производить каждую неделю. Целевая функция задачи — это общая прибыль, получаемая в неделю от производства продуктов R и Q. Каждое ограничение соответствует одному виду сырья. Левая часть каждого ограничения представляет собой общее количество сырья одного вида, требуемое для производства обоих продуктов. Правая часть ограничений содержит общее количество сырья каждого вида, которое фирма может использовать в течение недели.

Двойственная модель. Переменные модели — это теневые цены ресурсов для прямой модели, т.е. величины, на которые увеличилось бы значение целевой функции при росте имеющегося запаса сырья соответствующего вида на единицу. Теневые цены характеризуют стоимость единицы сырья каждого вида. Целевая функция задачи — это общая еженедельная стоимость всех видов сырья, используемых при производстве R и Q. Каждое ограничение связано с одним из продуктов. В левой части каждого ограничения дана общая стоимость всех видов сырья, используемых при выпуске 1 кг соответствующего продукта; в правой – прибыль от выпуска единицы соответствующего продукта. Обратимся вновь к формулировке двойственной модели и попытаемся дать интерпретацию отдельным ее компонентам (см. стр. 447).

Из каждого ограничения следует, что общая стоимость сырья, используемого для производства данного продукта, должна быть больше либо равна прибыли от производства единицы этого продукта. Из решения прямой или двойственной модели можно получить решение обратной модели.

4. Графическое решение прямой задачи приведено на рис. 12.28.

Оптимальным решением задачи является точка A, лежащая на пересечении линии ограничения на сырье 1 и оси Q. Чтобы получать максимальную прибыль, следует производить только продукт Q в количестве $3\frac{1}{3}$ кг. При этом RM1 будет использоваться полностью, а RM2 — нет. Максимальная прибыль составит: $3\frac{1}{3} \times 8 = 26{,}67$ ф. ст. в неделю. Ниже приводится графическое решение двойственной задачи (рис. 12.29).

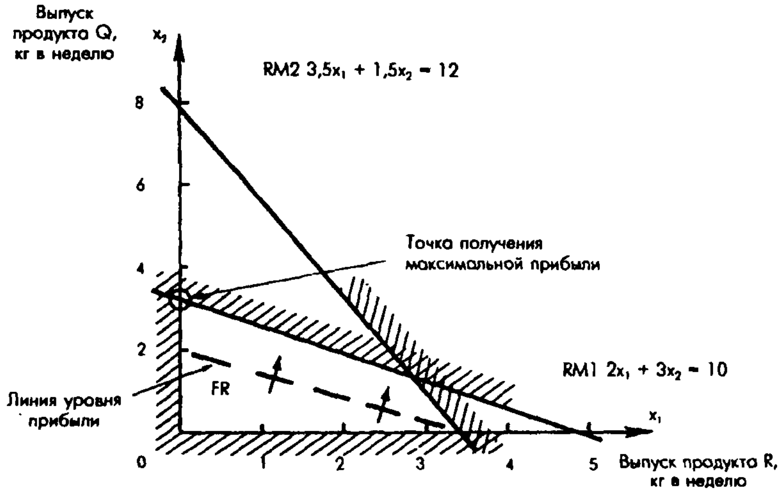


Рис. 12.28. Прямая модель

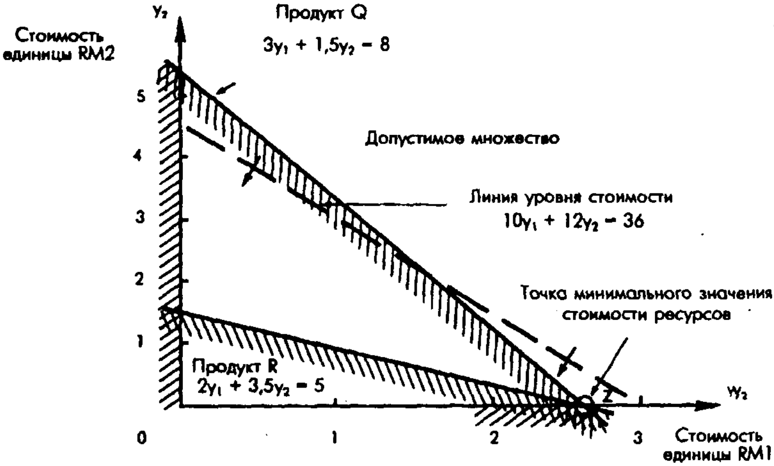


Рис. 12.29. Двойственная модель

Минимизировать G:

$$\begin{array}{ccccccccc} 10 & & y_1 & + & 12 & & y_2 & & \text{(ф. ст.} \\ & & & & & & & & \text{в неделю)} \\ \begin{pmatrix}\text{наличие RM1}\\ \text{в неделю}\end{pmatrix} & \times & \begin{pmatrix}\text{количество}\\ \text{RM1 (кг)}\end{pmatrix} & + & \begin{pmatrix}\text{наличие RM2}\\ \text{в неделю}\end{pmatrix} & \times & \begin{pmatrix}\text{количество}\\ \text{RM2 (кг)}\end{pmatrix} & & \end{array}$$

Продукт R:

$$\begin{array}{ccccccccc} 2 & & y_1 & + & 3{,}5 & & y_2 & & \ge 5 \text{ ф.ст.} \\ & & & & & & & & \text{прибыли} \\ & & & & & & & & \text{на единицу} \\ \begin{pmatrix}\text{RM1 (кг)}\\ \text{для R}\end{pmatrix} & \times & \begin{pmatrix}\text{количество}\\ \text{RM1 (кг)}\end{pmatrix} & + & \begin{pmatrix}\text{RM2 (кг)}\\ \text{для R}\end{pmatrix} & \Phi & \begin{pmatrix}\text{количество}\\ \text{RM2 (кг)}\end{pmatrix} & & \end{array}$$

Продукт Q:

$$\begin{array}{ccccccccc} 3 & & y_1 & + & 1{,}5 & & y_2 & & \ge 8 \text{ ф.ст.} \\ & & & & & & & & \text{прибыли} \\ & & & & & & & & \text{на единицу} \\ \begin{pmatrix}\text{RM1}\\ \text{для Q (кг)}\end{pmatrix} & \times & \begin{pmatrix}\text{количество}\\ \text{RM1 (кг)}\end{pmatrix} & + & \begin{pmatrix}\text{RM2}\\ \text{для Q (кг)}\end{pmatrix} & \times & \begin{pmatrix}\text{количество}\\ \text{RM2 (кг)}\end{pmatrix} & & \end{array}$$

Данная задача является задачей минимизации. Необходимо уменьшить значение целевой функции настолько, насколько это возможно, следовательно, перемещение линии уровня целевой функции осуществляется параллельно ее исходному положению в направлении начала координат. Точка Z является последней крайней точкой допустимого множества, через которую проходит линия уровня, и, таким образом, оптимальным решением двойственной задачи. Z является пересечением линии ограничения для продукта Q и оси y_1 т.е.: $y_2 = 0$

и

$$3y_1 + 1{,}5y_2 = 8,$$

следовательно,

$$y_2 = 0 \text{ и } y_1 = 2\frac{2}{3}.$$

Минимальная стоимость ресурсов в двойственной задаче имеет вид:

$$G = 10 \times 2\frac{2}{3} + 12 \times 0 = 26{,}67 \text{ ф. ст. в неделю.}$$

Это значение совпадает со значением целевой функции прямой задачи.

Обобщая полученные решения, можно сделать вывод, что максимальное значение прибыли, равное 26,27 ф. ст. в неделю, достигается, если продукт Q выпускать в количестве $3\frac{1}{3}$ кг, а продукт R не производить вообще. Стоимость сырья, т.е. теневые цены ресурсов, составила 2,67 ф. ст. за 1 кг RM1 и ноль для RM2. Эту же информацию можно было бы получить через проведение полного анализа только прямой задачи.

РЕЗЮМЕ

Модели линейного программирования используются в решении проблемы распределения ограниченных ресурсов для достижения своих целей в бизнесе. Целью может являться максимизация прибыли за неделю или минимизация ежедневных издержек. Формулировка задачи линейного программирования требует последовательного выполнения следующих шагов:

Шаг 1. Определение переменных решения.
Шаг 2. Определение линейной целевой функции и линейных ограничений.
Шаг 3. Выражение целевой функции через переменные задачи.
Шаг 4. Выражение ограничений через переменные задачи.

При формулировке задач с двумя или со множеством переменных применяется одна и та же процедура. Однако задачу с двумя переменными можно решить графически. Ограничения, которые обычно представлены неравенствами знака "≤" или "≥", изображаются на графике с помощью прямых и областей на плоскости. Каждое ограничение разделяет плоскость графика на допустимую и недопустимую области. Область, точки которой удовлетворяют всем ограничениям задачи, называется допустимым множеством. Допустимое множество содержит все возможные решения задачи.

Оптимальное решение, которое всегда находится в крайней точке допустимого множества, можно найти после нанесения на график линии уровня целевой функции. Целевая функция перемещается параллельно этой линии в направлении, противоположном началу координат, в случае максимизации целевой функции, или в сторону начала координат в случае ее минимизации. Координаты последней крайней точки, через которую проходит линия уровня перед тем, как она всецело окажется вне пределов допустимого множества, являются значениями переменных, которые оптимизируют целевую функцию задачи.

Поскольку практическая реализация модели может осуществляться в условиях неопределенности, большое место в линейном программировании занимает анализ чувствительности модели. Этот метод позволяет учесть вариацию и неопределенность коэффициентов целевой функции и значений правой части ограничений задачи.

Задачи линейного программирования со множеством переменных решаются на компьютерах с помощью симплекс-метода. Итоговая таблица алгоритма симплекс-метода содержит оптимальное значение целевой функции, соответствующие ему значения переменных решения и значения остаточных или избыточных переменных. Кроме того, в ней указываются теневые цены на ресурсы. Итоговую таблицу симплекс-метода можно использовать также в анализе чувствительности, чтобы выявить общее воздействие изменений в запасах лимитирующих ресурсов на целевую функцию и каждое из ограничений.

Для каждой исходной задачи линейного программирования существует ее двойственная формулировка. Решения прямой и двойственной задачи одинаковы. Двойственную модель можно получить непосредственно из исходной прямой модели, поменяв местами ее коэффициенты. Иногда более простая формулировка двойственной задачи дает существенные преимущества в процессе решения по сравнению со сложной постановкой прямой задачи.

УПРАЖНЕНИЯ

Упражнение 12.1

Фабрика "GRM plc" выпускает два вида каш для завтрака – "Crunchy" и "Chewy". Используемые для производства обоих продуктов ингредиенты в основном одинаковы и, как правило, не являются дефицитными. Основным ограничением, накладываемым на объем выпуска, является наличие фонда рабочего времени в каждом из трех цехов фабрики.

Управляющему производством Джою Дисону необходимо разработать план производства на месяц. В приведенной ниже таблице указаны общий фонд рабочего времени и число человеко-часов, требуемое для производства 1 т продукта.

Цех	*Необходимый фонд рабочего времени, чел.-ч/т*		*Общий фонд рабочего времени, чел.-ч. в месяц*
	"Crunchy"	*"Chewy"*	
А. Производство	10	4	1000
В. Добавка приправ	3	2	360
С. Упаковка	2	5	600

Доход от производства 1 т "Crunchy" составляет 150 ф. ст., а от производства "Chewy" – 75 ф. ст. На настоящий момент нет никаких ограничений на возможные объемы продаж. Имеется возможность продать всю произведенную продукцию.

Требуется: сформулировать модель линейного программирования, максимизирующую общий доход фабрики за месяц.

Упражнение 12.2

Оливер А. Петерс скоро выйдет на пенсию, и ему предстоит решить, как поступить с единовременным пособием, которое в соответствии с пенсионной программой будет предоставлено ему фирмой. М-р Петерс и его супруга намерены предпринять длительный визит в Австралию к своей дочери сроком на два года, поэтому любые сделанные в настоящий момент инвестиции будут свободны для использования на данный период. Очевидно, цель м-ра Петерса состоит в максимизации общего дохода от вложений, полученного за двухлетний период.

Мистера Петерса проконсультировали, что наилучшим вариантом вложения инвестиций был бы инвестиционный фонд, и в настоящее время он рассматривает возможность помещения инвестиций в один из таких фондов, состоящий из инвестиций трех типов – А, В и С. Сумма единовременного пособия составит 25000 ф. ст., однако, мистер Петерс считает, что нет необходимости вкладывать в данный инвестиционный фонд все деньги; часть из них он намерен перевести на свой счет жилищно-строительного кооператива, который гарантирует ему 9% годовых.

По мнению бухгалтера фирмы, мистеру Петерсу следует попытаться распределить свои инвестиции таким образом, чтобы обеспечить как получение дохода, так и рост капитала. Поэтому ему посоветовали не менее 40% от общей суммы вложить в вариант А и перевести на свой счет. Для обеспечения значительного роста капитала не менее 25% общей суммы денежных средств, вложенных в инвестиционный фонд, необходимо поместить в проект В, однако, вложения в В не должны превышать 35% общего объема вложений в инвестиционный фонд ввиду высокой вероятности риска, соответствующей проекту В. Кроме того, для сохранности капитала в проекты А и С следует вложить не менее 50% средств, помещаемых в инвестиционный фонд.

В настоящее время проект А позволяет получать 10 % годовых и обеспечивает 1% роста капитала; проект В предполагает рост капитала в 15%; проект С дает 4% годовых и 5%-ный рост капитала.

Требуется: учитывая цель м-ра Петерса, сформулировать модель линейного программирования, показывающую, как следует распределить сумму единовременного пособия между различными проектами инвестиций.

Упражнение 12.3

Китайская компания с ограниченной ответственностью по производству гусеничных механизмов выпускает пять сходных друг с другом товаров — A, B, C, D, и E. В нижеследующей таблице представлены расходы ресурсов, необходимых для выпуска единицы каждого товара, а также недельные запасы каждого ресурса и цены продажи единицы каждого продукта.

Ресурсы	*Товар*					*Недельный запас ресурсов*
	A	*B*	*C*	*D*	*E*	
Сырье, кг	6,00	6,50	6,10	6,10	6,40	35000
Сборка, ч	1,00	0,75	1,25	1,00	1,00	6000
Обжиг, ч	3	4,50	6	6	4,50	30000
Упаковка, ч	0,50	0,50	0,50	0,75	1,00	4000
Цена продажи, ф.ст.	40	42	44	48	52	

Известны также издержки, связанные с использованием каждого вида ресурсов:

сырье — 2,10 ф. ст. за 1 кг;
сборка — 3,00 ф. ст. за 1 ч;
обжиг — 1,30 ф. ст. за 1 ч;
упаковка — 8,00 ф. ст. за 1 ч.

Требуется:

а) сформулировать задачу линейного программирования таким образом, чтобы в качестве переменных как целевой функции, так и ограничений выступали ресурсы;

б) кратко сформулировать предпосылки применения модели. Для максимизации элементов, составляющих прибыль за неделю, следует использовать компьютерный пакет прикладных программ.

(АССА, декабрь 1987).

Упражнение 12.4

Используя модель линейного программирования, построенную в упражнении **12.1**, нужно помочь Джою Дисону найти оптимальный ассортиментный набор на следующий месяц, если политика компании состоит в максимизации общего дохода за месяц. Каково значение максимального дохода?

Упражнение 12.5

Нефтяная компания "РТ" для улучшения эксплуатационных качеств и снижения точки замораживания дизельного топлива, которое она производит, добавляет в него определенные химикаты. В каждом бензобаке объемом 1000 л должно содержаться не менее 40 мг химической добавки X, не менее 14 мг химической добавки Y и не менее 18 мг химической добавки Z. Необходимые химические добавки в форме готовых смесей поставляют "РТ" две химические компании А и В. В нижеследующей таблице приведено содержание химических добавок в каждом продукте, поставляемом указанными компаниями.

Продукт	*Химические добавки, мг/л*		
	X	*Y*	*Z*
А	4	2	3
В	5	1	1

Стоимость продукта А — 1,50 ф. ст. за 1 л, а продукта В — 3,00 ф. ст. за 1 л.

Требуется: найти ассортиментный набор продуктов А и В, минимизирующий общую стоимость добавленных в топливо химикатов.

Упражнение 12.6

Обратимся вновь к упражнениям **12.1** и **12.4**. Исполнительный директор корпорации "GRM" принял решение об увеличении фонда рабочего времени посредством введения сверхурочной работы служащих корпорации. Джой Дисон собрал следующую информацию о стоимости сверхурочной работы и вероятной ее продолжительности по каждому из трех цехов:

Наименование цеха	*Стоимость одного обычного человеко-часа, ф.ст.*	*Стоимость 1 чел.-ч. сверхурочной работы, ф.ст.*	*Максимально возможное число сверхурочных человеко-часов в месяц*
Производство	4,50	6,50	150
Добавка приправ	4,75	6,50	100
Упаковка	3,50	4,50	80

Для того, чтобы свести к минимуму административную работу и издержки, Джой Дисон принял решение о том, что сверхурочную работу следует ввести только в одном из цехов, по крайней мере на первое время.

Требуется: используя приведенную выше информацию, принять решение о том, в каком цехе следует ввести сверхурочную работу, и каково максимальное число сверхурочных человеко-часов для данного цеха.

Упражнение 12.7

Один из заводов легкой промышленности производит порошок для изготовления солодовых напитков трех видов. Один из них продается в качестве напитка здоровья, поскольку имеет низкое содержание сахара; другой напиток поставляется в медицинские учреждения в качестве продукции для больных, поскольку он содержит витаминные добавки; наконец, третий является стандартным товаром.

В приведенной ниже таблице для каждого напитка указаны основные ингредиенты, их стоимость и размер недельного запаса, а также оценки максимального спроса на соответствующие товары за неделю.

	Расход ингредиентов на 1 кг продукта, кг			*Оценка максимального спроса за неделю, кг*	*Цена продажи 1 кг напитка, ф.ст.*
	Сахар	*Солодовый экстракт*	*Сухие сливки*		
Стандартный напиток	0,30	0,30	0,35	2000	1,00
Напиток здоровья	0,15	0,25	0,55	1800	1,20
Напиток для больных	0,15	0,30	0,25	1200	1,50
Стоимость 1 кг ингредиента, пенсов	20	60	50		
Размер недельного запаса ингредиентов, кг	1000	1250	2200		

Запас витаминных добавок неограничен. Издержки производства остальных переменных имеют следующие значения: 10 пенсов за 1 кг стандартного напитка, 9 пенсов за 1 кг напитка здоровья и 12 пенсов за 1 кг напитка для больных.

Требуется:

1. Для изложенной проблемы сформулировать модель линейного программирования, целевая функция которой максимизирует общий доход, получаемый за неделю.
2. Ниже приведена итоговая симплекс-таблица, полученная при решении данной задачи:

	m	*h*	*i*	s_1	s_2	s_3	s_4	s_5	s_6	*b*
s_1	0	0	0	1	−1	0	0	0,1	0,15	110
m	1	0	0	0	3,333	0	0	−0,833	−1	1466,67
s_3	0	0	0	0	−1,167	1	0	−0,258	0,1	396,67
s_4	0	0	0	0	−3,333	0	1	0,833	1	533,33
h	0	1	0	0	0	0	0	1	0	1800
i	0	0	1	0	0	0	0	0	1	1200
P	0	0	0	0	1,617	0	0	0,251	0,56	3144,33

В данной таблице переменные m, h, i связаны со стандартным напитком, напитком здоровья и напитком для больных соответственно. Переменные s_1, s_2, s_3 связаны с ограничениями на сахар, солод и сливки соответственно. Переменные s_4, s_5, s_6 связаны с ограничениями для максимального спроса на стандартный напиток, напиток здоровья и напиток для больных соответственно. Определить:

а) оптимальный ассортиментный набор;

б) максимальное значение дохода за неделю;

в) значения резервного запаса для ограничений задачи.

3. Используя приведенную в п.2 таблицу, дать ответы на следующие вопросы:

а) Последние исследования потребительского рынка показали, что напиток здоровья приобретает все большую популярность. Новое значение максимального спроса составило 2500 кг в неделю. Каково воздействие этого процесса на оптимальный ассортиментный набор?

б) Администрация компании обдумывает решение о покупке некоторого дополнительного количества солодового экстракта. Однако компания будет вынуждена иметь дело с новым поставщиком и покупать сырье по 80 пенсов за 1 кг. Позволит ли такая мера увеличить еженедельный доход? Если это так, то каково максимальное количество сырья, которое следует закупить у нового поставщика?

Упражнение 12.8

По условиям упражнения **12.1** требуется:

1. Построить для изложенной проблемы двойственную модель линейного программирования;
2. Дать интерпретацию двойственных переменных в контексте поставленного выше вопроса.

Упражнение 12.9

"Princetown Paints Ltd" выпускает три основных типа румян — жидкие, перламутровые и матовые — с использованием одинаковых смесеобразующих машин и видов работ. Главному бухгалтеру фирмы было поручено разработать для компании план производства на неделю. Информация о ценах продаж и стоимости 100 л товара приведена в таблице (ф. ст.).

	Румяна		
	Жидкие	*Перламутровые*	*Матовые*
Цена продажи на 100 л	120	126	110
Издержки производства товаров на 100 л:			
Стоимость сырья	11	25	20
Стоимость трудозатрат	30	36	24
Стоимость приготовления смеси	32	20	36
Другие издержки	12	15	10

Стоимость 1 чел.-ч составляет 3 ф. ст. а стоимость 1 ч приготовления смеси — 4 ф. ст. Фонд рабочего времени ограничен 8000 чел.-ч. в неделю, а ограничение на фонд работы смесеобразующих машин равно 5900 ч в неделю.

В соответствии с контрактными соглашениями компания должна производить 25000 л матовых румян в неделю. Максимальный спрос на жидкие румяна равен 35000 л в неделю, а на перламутровые румяна — 29000 л в неделю.

Требуется:

1. Сформулировать задачу линейного программирования, позволяющую определить объемы производства жидких и перламутровых румян в неделю, при которых достигается максимальное значение получаемой за неделю прибыли.
2. Решить эту задачу графически. Определить оптимальные объемы производства в неделю и соответствующее значение прибыли.

3. Рассчитать, на сколько нужно изменить цену продажи жидких румян, чтобы получить новое оптимальное решение задачи.
4. Предположим, что рабочие готовятся к сверхурочной работе за дополнительное вознаграждение в 1 ф. ст. за каждый сверхурочно отработанный час. Будет ли целесообразным введение сверхурочной работы на таких условиях? Если это так, то каковы ваши рекомендации по поводу количества часов сверхурочной работы, которое следует ввести, и какова будет дополнительная прибыль от применения сверхурочной работы?

(АССА, июнь 1988 г.)

Упражнение 12.10

Администрация компании "Nemesis Company", осуществляя рационализаторскую программу корпорации, приняла решение о слиянии двух своих заводов в Аббатсфилде и Берчвуде. Предусматривается закрытие завода в Аббатсфилде и за счет этого — расширение производственных мощностей предприятия в Берчвуде. На настоящий момент распределение рабочих высокой и низкой квалификации, занятых на обоих заводах, является следующим:

Квалификация рабочих	*Аббатсфилд*	*Берчвуд*
Высокая	200	100
Низкая	300	200
Итого	500	300

В то же время после слияния завод в Берчвуде должен насчитывать 240 рабочих высокой и 320 рабочих низкой квалификации.

После проведения всесторонних переговоров с привлечением руководителей профсоюзов были выработаны следующие финансовые соглашения:

1. Все рабочие, которые попали под сокращение штатов, получат выходные пособия следующих размеров:
   Квалифицированные рабочие – 2000 ф. ст.;
   Неквалифицированные рабочие – 1500 ф. ст.
2. Рабочие завода в Аббатсфилде, которые должны будут переехать, получат пособие по переезду в размере 2000 ф. ст.
3. Во избежание каких-либо преимуществ для рабочих Берчвудского завода доля бывших рабочих завода в Аббатсфилде на новом предприятии должна совпадать с долей бывших рабочих Берчвудского завода.

Требуется:

1. Построить модель линейного программирования, в которой определяется, как осуществить выбор работников нового предприятия из числа рабочих двух бывших заводов таким образом, чтобы минимизировать общие издержки, связанные с увольнением и переменой места жительства части рабочих. В процессе формализации следует использовать следующие переменные:

S_1 – число квалифицированных рабочих, переведенных на новую работу с завода в Аббатсфилде;

S_2 – число квалифицированных рабочих, переведенных на новую работу с завода в Берчвуде;

U_1 – число неквалифицированных рабочих, переведенных на новую работу с завода в Аббатсфилде;

U_2 – число неквалифицированных рабочих, переведенных на новую работу с завода в Берчвуде.

2. Используя два из ограничений-уравнений, элиминируйте влияние двух из четырех переменных модели и решите полученную задачу графическим методом. Каковы минимальные издержки увольнения и перемены места жительства части рабочих?

(АССА, июнь 1987).

Упражнение 12.11

а) Выявите преимущества и недостатки графического метода решения задач линейного программирования по сравнению с симплекс-методом.

б) Менеджер международной банковской организации по инвестициям располагает 550000 ф. ст., находящимися на счете банка, которые необходимо инвестировать, и рассматривает четыре общих типа инвестиций, а именно:

Тип 1: государственные ценные бумаги;
Тип 2: ценные бумаги корпораций;
Тип 3: обыкновенные акции отраслей сферы обслужнвания;
Тип 4: обыкновенные акции отраслей производственной сферы.

Целью менеджера по инвестициям является максимизация нормы отдачи вложений, причем размер годовых процентов от инвестиций равен 8, 9, 10 и 12% для типов 1, 2, 3 и 4 соответственно. Денежные средства, не инвестированные ни по одному из указанных выше типов, остаются на банковском счете и приносят 4% годовых.

Менеджер по инвестициям принял решение, что не менее 50000 ф. ст. следует поместить в ценные бумаги корпораций, а в инвестиционные проекты с элементами риска (т.е. ценные бумаги корпораций и все виды обыкновенных акций) следует вложить не более 300000 ф. ст. Кроме того, он считает, что по крайней мере половину общей суммы денежных средств, инвестированных в соответствии с указанными выше типами инвестиций, следует вложить в обыкновенные акции, но в акции отраслей производственной сферы следует поместить не более одной четверти общей суммы инвестиций.

Требуется: сформулировать для данной проблемы задачу линейного программирования, целевая функция и ограничения которой будут содержать четыре переменных таким образом, чтобы ввод информации и анализ задачи можно было осуществить с использованием пакета прикладных программ по линейному программированию.

После ввода исходных данных и анализа целевой функции и ограничений с помощью ППП *линейного программирования*, использующего *симплекс-метод*, была получена следующая выходная информация:

Итоговое решение, достигнутое через 5 шагов:							
Переменная	*Значение*	*Остаточная переменная*	*Значение*	*Избыточная переменная*	*Значение*	*Ограничение*	*Теневая цена*
X_1	200000	S_1	50000	S_2	0	C_1	0,1
X_2	50000	S_3	0			C_2	0
X_3	125000	S_4	0			C_3	0,11
X_4	125000	S_5	0			C_4	0,045
						C_5	0,005

где X_i – сумма, вложенная в i-й тип инвестиций (i = 1, 2, 3, 4), ф. ст.

Основываясь на полученной выходной информации, кратко пояснить, каково значение терминов "остаточные и избыточные переменные".

Используя выходную информацию, определить оптимальный план инвестиций, сумму денежных средств, оставленных на банковском счете, и ежегодный доход от реализации данного плана, выраженный в процентах.

Теневая цена ограничения, связанного с тем, что в акции или ценные бумаги с элементами риска следует вкладывать не более 300000 ф. ст., принимает значение, равное 0,11. Интерпретируйте данное значение.

(АССА, декабрь 1989 г.)

Упражнение 12.12

По данным упражнения **12.3** требуется:

1. В результате применения пакета прикладных программ была получена следующая итоговая таблица решения данной задачи симплекс-методом:

Базис	*A*	*B*	*C*	*D*	*E*	*X*	*S*	*T*	*U*	*Значение*
A	1	1,18	1,04	0,46	0	0,36	0	0	–2,29	3,357
S	0	–0,34	0,23	0,02	0	–0,18	1	0	0,14	321
T	0	1,37	2,97	2,28	0	–0,27	0	1	–2,79	9,482
E	0	–0,09	–0,02	0,52	1	–0,18	0	0	2,14	2,321
	0	1,26	1,06	0,51	0	2,02	0	0	8,81	105,791

Здесь A, B, C, D и E — объемы производства пяти продуктов в неделю; X — количество неиспользуемого сырья, которое остается от максимального запаса; S, T, U — соответствующие количества неиспользуемого фонда рабочего времени на стадиях производства, обжига и упаковки, которое остается от максимального резерва фонда рабочего времени в неделю.

а) Используя информацию, представленную в таблице, определить для компании по производству гусеничных механизмов оптимальный план производства продукции на неделю.

б) Описать последствия реализации этого плана с точки зрения недоиспользуемых ресурсов и вклада отдельных компонент в общую прибыль.

2. На примере данной задачи объяснить значение термина "двойственная оценка или теневая цена ресурса".
3. Есть предложение, что компании следует производить дополнительный продукт, цена продажи которого составит 50 ф. ст. за единицу. Производство

единицы этого продукта требует 6 кг сырья, а также затрат рабочего времени в 1 ч для производства, 5 ч для обжига и 1 ч для упаковки. Целесообразна ли реализация этого предложения на практике?

(АССА, декабрь 1987 г.)

Упражнение 12.13

Компания "Bermuda Paint" — частная промышленная фирма, специализирующаяся на производстве технических лаков. Представленная ниже таблица содержит информацию о ценах продажи и соответствующих издержках производства единицы полировочного и матового лаков.

Лак	*Цена продажи 1 галлона, ф. ст.*	*Издержки производства 1 галлона, ф. ст.*
Матовый	13,0	9,0
Полировочный	16,0	10,0

Для производства 1 галлона матового лака необходимо затратить 6 мин трудозатрат, а для производства одного галлона полировочного лака — 12 мин. Резерв фонда рабочего времени составляет 400 чел.-ч. в день. Размер ежедневного запаса необходимой химической смеси равен 100 унциям, тогда как ее расход на один галлон матового и полировочного лаков составляет 0,05 и 0,02 унции соответственно. Технологические возможности завода позволяют выпускать не более 3000 галлонов лака в день.

В соответствии с соглашением с основным оптовым покупателем компания должна поставлять ему 5000 галлонов матового лака и 2500 галлонов полировочного лака за каждую рабочую неделю (состоящую из 5 дней). Кроме того, существует профсоюзное соглашение, в котором оговаривается минимальный объем производства в день, равный 2000 галлонов. Администрации данной компании необходимо определить ежедневные объемы производства каждого вида лаков, которые позволяют получать максимальный общий доход.

Требуется:

а) Построить линейную модель для производственной проблемы, с которой столкнулась компания.

б) Используя графический метод, определить ежедневный оптимальный план производства и соответствующую ему величину дохода.

в) Профсоюз компании требует увеличения оплаты 1 ч. сверхурочных работ на 20 ф. ст.

Обосновать, сочтет ли администрация компании целесообразным такое предложение?

Если указанный размер оплаты сверхурочных работ является выгодным, какое количество часов сверхурочных работ в день целесообразно использовать?

г) Для исходной задачи (не учитывающей сверхурочные работы) определить промежуток изменений показателя единичного дохода за 1 галлон полировочного лака, в котором исходное оптимальное решение остается прежним.

Глава 13. ТРАНСПОРТНАЯ ЗАДАЧА И ЗАДАЧА О НАЗНАЧЕНИЯХ

13.1. ВВЕДЕНИЕ

Методы линейного программирования, рассмотренные нами в гл. 12, являются хорошим инструментом для решения ряда проблем распределения ресурсов. Применение пакетов прикладных программ позволяет значительно упростить решение задачи. Поэтому лицо, принимающее решение, получает возможность уделить большее внимание интерпретации и оценке решения задачи. Однако применение прикладных пакетов предполагает предварительную формализацию модели линейного программирования. В процессе решения большинства проблем эта задача является основной. При построении модели необходимо идентифицировать ее переменные и сформулировать систему ограничений.

При решении некоторых видов проблем распределения ресурсов использование специально созданных для этих целей алгоритмов упрощает процесс построения исходной модели. Данная глава будет посвящена рассмотрению двух примеров таких алгоритмов, созданных для решения транспортной задачи и задачи о назначениях.

В обоих случаях проблема распределения ресурсов связана с продуктами, которые в соответствии с определенной целью перевозятся из пунктов производства в пункты потребления. Целью часто является минимизация общей стоимости транспортировки. Пусть, например, некоторой компании принадлежат три завода и пять пунктов распределения продукции, находящиеся в одном регионе. Администрация компании должна организовать перевозку конечной продукции с заводов в пункты распределения с минимальной стоимостью. В этой ситуации наиболее подходящими могли бы стать методы решения транспортной задачи.

Частным случаем транспортной задачи является задача о назначениях. Предполагается, что из каждого пункта производства в каждый пункт потребления перевозится только один товар. Например, в машинном цехе имеется шесть токарных станков различного срока службы и различной конструкции. Каждое утро начальник цеха должен распределить по этим станкам шесть видов работ. Продолжительность выполнения каждой работы на различных станках неодинакова. Начальник цеха намерен распределить по каждому станку работу таким образом, чтобы свести к минимуму общее время выполнения работ. В процессе решения этой и подобных проблем можно использовать алгоритм решения задачи о назначениях.

В настоящей главе мы рассмотрим применение указанных алгоритмов для решения задач небольшой размерности. Однако следует принять во внимание, что на практике размерность таких задач гораздо больше, поэтому решаются они с использованием пакетов прикладных программ. Более того, очень часто решение транспортной задачи осуществляется в несколько этапов, например, при перевозках типа "завод–склад–розничная продажа". В таких случаях приходится модифицировать основной алгоритм и использовать более сложные методы решения.

13.2. ТРАНСПОРТНАЯ ЗАДАЧА И АЛГОРИТМ ЕЕ РЕШЕНИЯ

Данная проблема связана с распределением товаров между поставщиками (находящимися в пунктах производства) и потребителями (находящимися в пунктах назначения) таким образом, чтобы общая стоимость этого распределения была минимальной. Эта задача может быть решена либо с помощью методов линейного программирования, либо специального алгоритма решения транспортной задачи. Применение методов линейного программирования проиллюстрировано в примере **13.1**.

13.2.1. Транспортная задача

❑ **Пример 13.1.** Компания с ограниченной ответственностью "Ace Foods Ltd" осуществляет производство прохладительных напитков на двух заводах — А и В. Поставкой бутылок на каждый из заводов занимаются две фирмы — Р и Q. На ноябрь заводу А требуется 5000 бутылок, а заводу В — 3500 бутылок. Фирма Р может поставить максимум 7500 бутылок, а фирма Q — 4000 бутылок. Табл. 13.1 содержит информацию о стоимости перевозки одной бутылки от каждого поставщика каждому заводу.

Таблица 13.1. **Стоимость перевозки бутылок, показатели спроса и предложения**

Поставщик	*Стоимост перевозки одной бутылки на завод, пенсов*		*Максимальный объем поставки*
	A	*B*	
P	4	4	7500
Q	3	2	4000
Спрос на бутылки	5000	3500	

Как следует организовать доставку бутылок на заводы, чтобы общая стоимость перевозки была минимальной?

Решение.

При решении транспортной задачи всегда полезно проверить, не существует ли очевидного решения. Теоретически было бы желательно использовать для перевозок только наиболее дешевые маршруты. Для обоих заводов Q был бы наиболее предпочтительным поставщиком, так как стоимость перевозки для него ниже, чем для Р. Однако максимальный объем перевозок для Q составляет только 4000 бутылок, тогда как общий спрос равен 8500. Вероятно, наиболее дешевым вариантом было бы использование маршрута из Q в В стоимостью 2 пенса за единицу, удовлетворяющее весь спрос завода В (3500). Остаток запаса (500) следует направить из Q в A по стоимости 3 пенса за единицу. Остальной спрос завода A следует удовлетворить через поставщика P, причем стоимость перевозки составит 4 пенса за единицу. Общая стоимость транспортировки при таком распределении будет иметь вид:

$$0{,}02 \times 3500 + 0{,}03 \times 500 + 0{,}04 \times 4500 = 265 \text{ ф. ст. в месяц.}$$

Однако мы не можем доказать, что данное распределение ресурсов является наиболее экономичным. Основные аспекты исследования транспортной модели состоят в следующем: доказательство того, что сформулированная задача имеет решение; обоснование положения о том, что это решение является оптимальным; изучение влияния на полученное решение любых изменений условий задачи.

Построив соответствующую модель линейного программирования, решим сформулированную выше проблему графическим методом.

Пусть фирма P поставляет x бутылок для завода A и y бутылок для завода B. Тогда для полного удовлетворения спроса фирма должна поставлять оставшиеся (5000 − x) бутылок на завод A и (3500 − y) бутылок на завод B. Цель состоит в минимизации общей стоимости транспортировки C (в пенсах), где

$$C = 4x + 4y + 3(5000 - x) + 2(3500 - y),$$

следовательно,

$$C = x + 2y + 22000,$$

а целевая функция задачи имеет вид:

$$Z = C - 22000 = x + 2y.$$

Z принимает свое минимальное значение тогда, когда C принимает минимальное значение. Значения x и y, которые минимизируют Z, минимизируют также и C. Минимизация целевой функции осуществляется в условиях следующей системы ограничений:

Спрос завода	A:	x	≤ 5000 бутылок
Спрос завода	B:	y	≤ 3500 бутылок
Поставки из	P:	$x + y$	≤ 7500 бутылок
Поставки из	Q:	$(5000 - x) + (3500 - y)$	≥ 4000 бутылок
	т.е.:	$x + y$	≤ 4500 бутылок
		x, y	≥ 0

Графическое изображение системы ограничений представлено на рис. 13.1.

Точка с координатами x = 4000, y = 2000 принадлежит допустимому множеству. Значение функции в этой точке

$$Z = 4000 + 2 \times 2000 = 8000 \text{ пенсов.}$$

Типичная линия уровня целевой функции имеет вид: 8000 = x + 2y. На рис. 13.1 она изображена пунктиром. Перемещение линии уровня в сторону уменьшения значений целевой функции приводит нас в крайнюю точку A, которая является

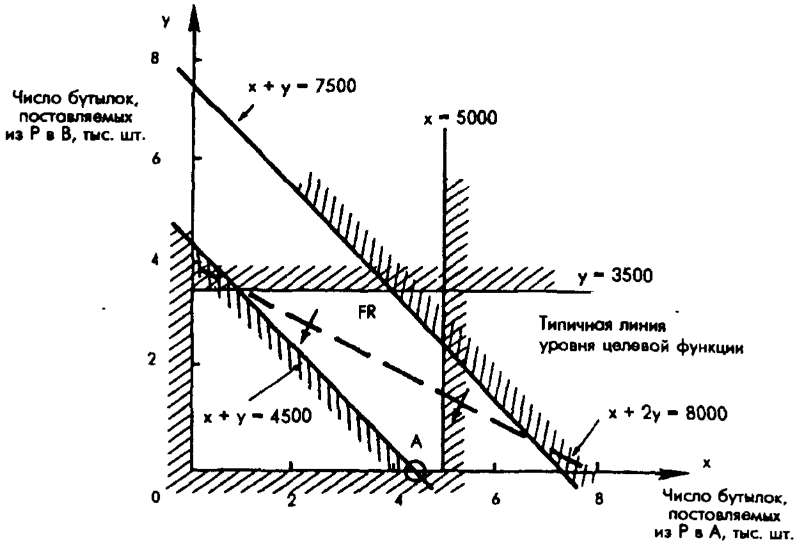


Рис. 13.1. Задача линейного программирования поставки бутылок

оптимальной. В этой точке x = 4500, а y = 0. Следовательно, оптимальное решение состоит в поставке из P в A 4500 бутылок, в отсутствии поставок из P в B, в поставке из Q в A 500 бутылок, а из Q в B — 3500 бутылок. Минимальная стоимость транспортировки для этого решения равна:

$$C_{min} = 4500 + 2 \times 0 + 22000 = 26500 \text{ пенсов} = 265 \text{ ф. ст.}$$

Резервный запас остается только на фирме P и составляет 3000 единиц. Начиная решать задачу, мы предполагали, что именно это решение минимизирует стоимость перевозки. Теперь мы доказали, что это действительно так.

13.2.2. Алгоритм решения транспортной задачи

Задачу, рассмотренную в 13.2.1, можно решить, используя **алгоритм решения транспортной задачи**. Применение этого алгоритма требует соблюдения ряда предпосылок:

1. Должна быть известна стоимость перевозки единицы продукта из каждого пункта производства в каждый пункт назначения.
2. Запас продуктов в каждом пункте производства должен быть известен.
3. Потребности в продуктах в каждом пункте потребления должны быть известны.
4. Общее предложение должно быть равно общему спросу.

Приведенная в примере **13.1** задача удовлетворяет предпосылкам 1–3, однако предпосылка 4 для этой задачи не выполняется. Тем не менее, можно ввести

фиктивный завод, потребность которого определяется разностью между общим предложением и общим спросом. Потребность фиктивного завода по данным примера **13.1** составила бы (11500 – 8500) = 3000 бутылок. Любые продукты, которые подлежат распределению в фиктивный пункт назначения, на деле не вывозятся из пункта производства. В случае, если общее предложение меньше общего спроса, поступают аналогичным образом, т.е. в модель вводится фиктивный поставщик, максимальный объем поставок которого равен величине неудовлетворенного спроса. Количество товаров, вывозимых из фиктивного пункта производства, характеризует величину недостающих поставок.

Алгоритм решения транспортной задачи состоит из четырех этапов:

Этап 1. Представление данных в форме стандартной таблицы и поиск любого допустимого распределения ресурсов. Допустимым называется такое распределение ресурсов, которое позволяет удовлетворить весь спрос в пунктах назначения и вывезти весь запас продуктов из пунктов производства.

Этап 2. Проверка полученного распределения ресурсов на оптимальность.

Этап 3. Если полученное распределение ресурсов не является оптимальным, то ресурсы перераспределяются, снижая стоимость транспортировки.

Этап 4. Повторная проверка оптимальности полученного распределения ресурсов.

Данный итеративный процесс повторяется до тех пор, пока не будет получено оптимальное решение.

13.2.3. Поиск начального распределения ресурсов

Начальное распределение ресурсов может быть получено с помощью любого метода, позволяющего найти допустимое решение задачи. Однако при систематическом решении таких задач можно разработать методы, позволяющие получать более выгодные начальные решения. Мы остановимся на двух методах нахождения начального распределения ресурсов — **методе минимальной стоимости** и **методе Вогеля**. Алгоритмы этих методов рассматриваются в примере **13.2**.

❒ **Пример 13.2.** Три торговых склада — P, Q, R — могут поставлять некоторое изделие в количестве 9, 4 и 8 единиц соответственно. Величины спроса трех магазинов розничной торговли, находящихся в пунктах A, B и C, на это изделие равны 3, 5 и 6 единицам соответственно. Какова минимальная стоимость транспортировки изделий от поставщиков потребителям? Единичные издержки транспортировки приведены в табл. 13.2.

Таблица 13.2. **Издержки транспортировки, объемы потребностей и предложения**

Поставщик	*Транспортные издержки для магазинов, ф. ст. за единицу*			*Общий объем предложения*
	A	*B*	*C*	
P	10	20	5	9
Q	2	10	8	4
R	1	20	7	8
Общий объем спроса	3	5	6	

Решение

В нашем распоряжении имеется информация об издержках, предложении изделий и потребностях в них, но общее предложение превышает общий спрос. Общее количество изделий, которое могут поставить все склады, равно 21, однако розничным магазинам необходимо только 14 изделий. Следовательно, необходимо ввести фиктивный розничный магазин, потребность которого будет равна 7 изделиям, определяющим избыток предложения. Фактически эти 7 изделий не будут вывезены с торговых складов, поэтому предполагается, что издержки транспортировки для них будут равны нулю. Ниже приводится первая транспортная таблица:

Таблица 13.3. **Сбалансированная транспортная таблица**

Поставщик	*Транспортные издержки для магазинов, ф. ст. за единицу*				*Общий объем предложения*
	A	*B*	*C*	*Фиктивный*	
P Q R	10 2 1	20 10 20	5 8 7	0 0 0	9 4 8
Общий объем спроса	3	5	6	7	21

Для нахождения начального допустимого распределения ресурсов будем использовать метод минимальной стоимости, а затем метод Вогеля. Тем не менее следует иметь в виду, что на практике требуется применение только одного из методов.

МЕТОД 1. МЕТОД МИНИМАЛЬНОЙ СТОИМОСТИ

1. В клетку с минимальной единичной стоимостью записывают наибольшее возможное количество продукта.
2. Производится корректировка оставшихся объемов предложения и потребностей.
3. Выбирается следующая клетка с наименьшей стоимостью, в которую помещается наибольшее возможное количество продукта, и т.д. до тех пор, пока спрос и предложение не станут равными нулю.
4. Если наименьшее значение стоимости соответствует более чем одной клетке таблицы, выбор осуществляется случайным образом.

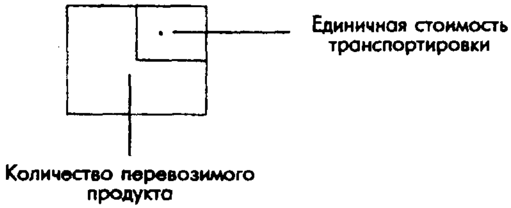
В табл. 13.4 стоимость транспортировки находится в верхнем правом углу каждой клетки, внутри прямоугольника. Индексы, соответствующие количествам продукта, характеризуют последовательность распределения ресурсов и облегчают читателю понимание процедуры распределения. Прочерки в клетках — отсутствие предложения или спроса, соответствующих этим клеткам.

1. Наименьшая стоимость транспортировки равна нулю. Следовательно, можно выбрать любую из клеток (P, фиктивный), (Q, фиктивный) или (R, фиктивный). Пусть выбрана клетка(P, фиктивный), в соответствии с алгоритмом в ней помещается максимальное количество продукта, равное 7 единицам. Предложение в P и спрос фиктивного магазина уменьшаются на 7. Затем в клетках, которые уже нельзя использовать в дальнейшем распределении перевозок, ставится прочерк; в нашем случае это клетки (Q, фиктивный) и (R, фиктивный).

Таблица 13.4. **Начальное распределение ресурсов, полученное методом минимальной стоимости**

Торговый склад	*Розничный магазин*				*Общее предложение*
	A	*B*	*C*	*Фиктивный*	
P	10 –	20 –	5 2_3	0 7_1	~~9~~ ~~2~~ 0
Q	2 –	10 4_5	8 –	0 –	~~4~~ 0
R	1 3_2	20 1_6	7 4_4	0 –	~~8~~ ~~5~~ ~~1~~ 0
Общая потребность	~~3~~ 0	~~5~~ ~~1~~ 0	6 ~~4~~ 0	7 0	21

Ключ:



2. Клеток с нулевой стоимостью больше нет, поэтому выбирается клетка (R,A), которой соответствует наименьшая стоимость, равная 1. В данной клетке размещается наибольшее возможное количество продукта, равное 3. Затем производится корректировка итоговых значений спроса и предложения, соответствующих данным строке и столбцу, а в клетках (P,A) и (Q,A), которые нельзя использовать в дальнейшем, ставится прочерк.
3. Наименьшая стоимость перевозки равна 5 и соответствует клетке (P,C). В данной клетке размещаются две единицы изделия, оставшиеся на складе P. Производится корректировка итоговых значений соответствующих строки и столбца, а в остальных клетках строки P ставится прочерк.
4. Наконец, оставшееся количество продукта распределяется последовательно в клетки (R,C), (Q,B) и (R,B).

Если распределение является допустимым, то объемы предложения на складах и объемы потребностей во всех магазинах должны быть равны нулю. Полученное выше распределение перевозок является допустимым.

Стоимость = (3 × 1) + (4 × 10) + (1 × 20) + (2 × 5) + (4 × 7) + (7 × 0) = 101 ф. ст.

Мы еще не можем сказать, является ли данное распределение перевозок наиболее дешевым, однако оно позволяет получить некоторую реальную стоимость.

МЕТОД 2. МЕТОД ВОГЕЛЯ

В данном методе используется штрафная стоимость. Штрафная стоимость для каждой строки и столбца — разность между наиболее дешевым маршрутом и следующим за ним с точки зрения критерия минимизации стоимости перевозок.

Суть метода состоит в минимизации этих штрафов.

1. Чтобы вычислить значения штрафной стоимости для каждой строки и столбца, необходимо найти клетки с наименьшей стоимостью и ближайшим к ним значением стоимости. Для каждой строки и столбца наименьшее значение стоимости вычитается из ближайшего к нему значения, найденного по критерию минимизации стоимости. Такая процедура позволяет получить значения штрафов за отсутствие перевозок в клетках с наименьшей стоимостью.
2. Выбирается строка или столбец с наибольшим значением штрафной стоимости, и в клетку с наименьшим значением стоимости перевозки для данной строки и столбца помещается наибольшее возможное количество продукта. Такая процедура позволяет избежать назначения высоких штрафов.
3. Как и в предыдущем методе, производится корректировка итоговых значений по строкам и столбцам таблицы.
4. В строках или столбцах, в которых предложение или спрос приняли нулевые значения, ставится прочерк во всех клетках, в которых отсутствуют перевозки, так как эти клетки нельзя использовать в процессе дальнейшего распределения перевозок.
5. Производятся возврат к шагу 1 и перерасчет штрафных стоимостей без учета клеток, в которых указаны перевозки, или клеток, в которых стоит прочерк.

Указанные шаги повторяются до тех пор, пока весь спрос не будет удовлетворен. Индексы, соответствующие количествам перевозок, отражают порядок выбора штрафных стоимостей и распределения перевозок.

После третьего распределения продукта оставшееся его количество распределяется по клеткам транспортной таблицы однозначно. Оставшийся продукт помещается в клетки (Р,В), (Р,С) и (Р, фиктивный).

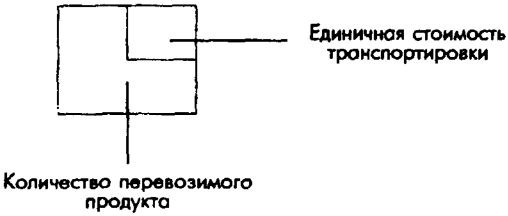
Стоимость = (1 × 20 + 6 × 5 + 2 × 0 + 4 × 10 + 3 × 1 + 5 × 0)= 93 ф. ст.

Как и в предыдущем случае, мы еще не знаем, является ли данное решение оптимальным, однако, можно с уверенностью утверждать, что план перевозок, полученный методом Вогеля, более дешевый по сравнению с планом, стоимость транспортировки для которого составила 101 ф. ст., полученная методом минимальной стоимости.

Таблица 13.5. **Начальное распределение перевозок, полученное методом Вогеля**

Торговый склад	*Розничный магазин* A	B	C	*фиктивный*	*Общее предложение*	*Штрафная стоимость* 1 2 3
	10	20	5	0		5 5 5
P	–	1	6	2	~~9~~ ~~8~~ ~~2~~ 0	
	2	10	8	0		2 – –
Q	–	4_1	–	–	~~4~~ 0	
	1	20	7	0		1 1 7_3
R	3_2	–	–	5_3	~~8~~ ~~5~~ 0	
Общая потребность	~~3~~ 0	~~5~~ ~~1~~ 0	6 0	~~4~~ ~~2~~ 0	21	
1-й штраф 2-й штраф 3-й штраф	1 9_2 –	10_1 0 0	2 2 2	0 0 0		

Ключ:



13.2.4. Проверка на оптимальность

Чтобы осуществить проверку оптимальности, необходимо определить, является ли начальное распределение перевозок базисным, т.е. находится ли полученное решение в крайней точке допустимого множества. Представленное в таблице 13.4. распределение перевозок является допустимым решением, т.е. лежит внутри или на границе допустимого множества. Если распределение перевозок является базисным, каждому ограничению должна соответствовать одна базисная переменная. Задача для m торговых складов и n розничных магазинов (включая фиктивный) содержит (m + n − 1) независимых ограничений. Следовательно, базисное решение должно размещаться в (m + n − 1) клетках транспортной таблицы. Все (m + n − 1) переменные должны занимать независимые позиции. Однако на данной стадии нет необходимости проявлять беспокойство по поводу независимости переменных, поскольку в процессе проверки решения на оптимальность любые нарушения будут выявлены.

Если распределение перевозок включает $(m + n - 1)$ независимую переменную, то к нему непосредственно можно применять методы проверки оптимальности. Если же число переменных меньше указанного количества, то критерий проверки оптимальности необходимо модифицировать так, как это будет показано в 13.2.6. Однако если число переменных превышает $(m + n - 1)$, процедура распределения перевозок проведена некорректно. В этом случае должны существовать варианты такого перераспределения перевозок, которые при меньшей стоимости содержат требуемое число переменных.

Обратимся в данным примера **13.2** и проверим каждое из полученных распределений перевозок на базисность. В нашей таблице 3 строки и 4 столбца, следовательно, базисное решение должно содержать $(3 + 4 - 1) = 6$ заполненных клеток. Можно легко убедиться, что это верно для обоих методов распределения перевозок. Кроме того, переменные решения, полученные с помощью обоих методов, находятся в различных точках допустимого множества. Следовательно, процедуру проверки можно применять, не прибегая к каким-либо модификациям.

Проверка исходного распределения перевозок производится для того, чтобы определить, является ли данный вариант наиболее дешевым для транспортировки, и, если это не так, какие изменения следует внести в данное распределение. Ниже будут изложены два метода проверки решения на оптимальность. В **методе ступенек** рассчитываются значения стоимости неиспользованных клеток, или теневые издержки. Сама процедура довольно длительная и кропотливая, однако, понимание ее сущности не представляет затруднений. Метод МОДИ (**модифицированных распределений**) — это математический алгоритм, позволяющий получить те же значения теневых издержек, причем гораздо быстрее, однако, этот метод более сложен для понимания. В обоих методах в случае, если распределение перевозок является неоптимальным, для перехода к следующему базисному распределению используется ступенчатая процедура. Как только получено базисное решение, алгоритм позволяет осуществить переход от одной крайней точки допустимого множества к другой до тех пор, пока не будет достигнуто оптимальное решение.

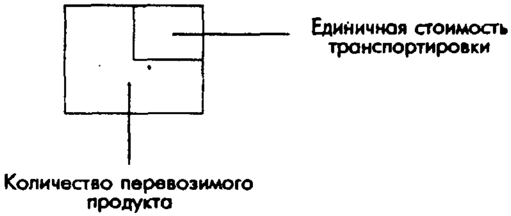
❐ **Пример 13.3.** Для иллюстрации применения данного алгоритма используем распределение перевозок, полученное методом минимальной стоимости. Данное распределение приводится в табл. 13.6.

Ступеньками называются точки, в которые производится распределение перевозок — (P,C), (P, фиктивный), (Q,B), (R,A), (R,B) и (R,C). Выбирается одна из пустых клеток и предполагается, что в нее перемещается одна единица продукта. Такая процедура нарушает баланс итоговых значений столбца или строки, на пересечении которых лежит данная клетка. Затем для восстановления баланса производится корректировка количества перевозимого продукта в некоторых заполненных клетках. Эти заполненные клетки, или ступеньки, используются при вычислении стоимости перевозки единицы продукта.

Таблица 13.6. **Начальное распределение, полученное методом минимальной стоимости**

Торговый склад	Розничный магазин				Общее предложение
	A	B	C	фиктивный	
P	10 –	20 –	5 2	0 7	9
Q	2 –	10 4	8 –	0 –	4
R	1 3	20 1	7 4	0 –	8
Общая потребность	3	5	6	7	21

Ключ:



Если значение стоимости положительное, то привлечение пустой клетки увеличит общую стоимость транспортировки, а это невыгодно. Если же значение стоимости отрицательное, использование пустой клетки, напротив, снижает общую стоимость транспортировки. Последнее означает, что полученное распределение перевозок является неоптимальным, и при использовании данной незаполненной клетки можно получить лучшее решение задачи.

Какая из пустых клеток будет выбрана в начале процедуры, значения не имеет. Выберем клетку (P,A). Добавим в нее одну единицу изделия. Теперь полученное распределение является несбалансированным. Розничный магазин A получает 4 единицы изделия, в то время как его потребность — 3. Торговый склад P является поставщиком 10 изделий, тогда как максимальный объем его предложения равен 9. Необходимо произвести корректировку столбца A и строки P. Для восстановления баланса в столбце A необходимо вычесть одно изделие из ступеньки (R,A). Эта мера корректирует столбец A, но нарушает баланс строки R, уменьшая соответствующее предложение с 8 до 7 единиц.

Можно осуществить перебалансировку строки P вычитанием одного изделия либо из клетки (P,C), либо из клетки (P, фиктивный). Если мы выберем клетку (P, фиктивный), то в фиктивном столбце нет больше заполненных клеток, которые можно было бы использовать в дальнейшей корректировке этого столбца, следовательно, данный выбор неприемлем. Корректировку можно осуществлять только с помощью тех клеток, которые уже заполнены на настоящий момент. Поэтому мы должны выбрать клетку (P,C). Из (P,C) вычитаем одно изделие. Это корректирует баланс по строке P, но нарушает его по столбцу C. На данном этапе проблема несбалансированности связана со строкой R и столбцом C. Их можно скорректировать одновременно, добавив одно изделие в (R,C). Схематично процесс заполнения пустой клетки (P,A) и восстановления баланса распределения перевозок показан в табл. 13.7.

Денежный эффект от перемещения одного изделия в клетку (P,A) рассчитывается следующим образом:

+ 1 × стоимость (P,A) − 1 × стоимость (R,A) + 1 × стоимость (R,C) − 1 × стоимость (P,C) = + (1 × 10) − (1 × 1) + (1 × 7) − (1 × 5) = +11 ф. ст. за 1 изделие.

Таблица 13.7. **Проверка пустой клетки (P,A)**

Изменение натурального объема, изделий

	A	C
P	Клетка, подвергнутая проверке + 1	Заполненная клетка − 1
R	Заполненная клетка − 1	Заполненная клетка + 1

Использование клетки (P,A) увеличило бы стоимость транспортировки на 11 ф. ст. за каждое изделие, перевозимое из P в A. Значение теневой цены является положительным, следовательно, использование данной клетки нежелательно.

Мы возвращаемся к исходному распределению перевозок и проводим последовательную проверку остальных пустых клеток. Выберем клетку (R, фиктивный), а для иллюстрации натуральных и стоимостных изменений, связанных с перемещением одной единицы изделия в клетку (R, фиктивный), используем ступеньки (P, фиктивный), (P,C) и (R,C).

Таблица 13.8. **Проверка пустой клетки (R, фиктивный)**

Изменения натурального объема – изделий

	C	*Фиктивный*
P	Заполненная клетка + 1	Заполненная клетка − 1
R	Заполненная клетка − 1	Клетка, подвергнутая проверке + 1

Таблица 13.9. **Проверка пустой клетки (R, фиктивный)**

Стоимостные изменения, ф. ст.

	C	*Фиктивный*
P	Заполненная клетка + 5	Заполненная клетка − 0
R	Заполненная клетка − 7	Клетка, подвергнутая проверке + 0

Стоимостные изменения от дополнения одного изделия в клетку (R, фиктивный) составили:

$$+0-0+5-7=-2 \text{ ф. ст. за 1 изделие.}$$

Размещение перевозок в клетке (R, фиктивный) дает возможность снизить издержки транспортировки, следовательно, начальное распределение перевозок оптимальным не является. Используя клетку (R, фиктивный) и указанный ступенчатый маршрут, можно найти более дешевое решение, позволяющее сэкономить 2 ф. ст. за каждую единицу изделия, помещаемую в данную клетку. Однако проверку пустых клеток необходимо завершить, поскольку могут существовать клетки, использование которых позволяет получить еще большую экономию.

Теперь построим ступенчатый путь для пустой клетки (Q, фиктивный). Необходимо учитывать, что для последующего осуществления балансировки движение можно осуществлять только через заполненные клетки. В этом случае цикл из четырех шагов построить уже невозможно. Нам приходится выбирать более сложный маршрут. В клетку (Q, фиктивный) поместим одно изделие. Строка Q и фиктивный столбец содержат только по одной заполненной клетке. Предположим, что мы приняли решение двигаться из (Q, фиктивный) в (Q,B). Для того, чтобы сбалансировать строку Q, из этой клетки вычтем одно изделие. Восстановить баланс для столбца B можно только с помощью клетки (R,B), следовательно, в нее необходимо добавить одно изделие. Балансировку строки R можно осуществить через клетки (R,A) и (R,C), но поскольку (R,A) — единственная заполненная клетка в столбце A, ее использовать нельзя. Если бы маршрут проходил через данную клетку, мы не могли бы сбалансировать столбец A. Объем перевозок в (R,C) уменьшается на одно изделие. Оставшаяся часть маршрута очевидна. Восстановление баланса в столбце C производится увеличением перевозок в (P,C) на одну единицу, а баланс строки P достигается вычитанием одного изделия из (P, фиктивный). Последний шаг позволяет также сбалансировать фиктивный столбец и замкнуть цикл. Следует помнить, что построение замкнутого цикла внутри транспортной таблицы, который начинается и заканчивается в выбранной пустой точке, возможно только в случае, если исходное распределение перевозок является базисным. Натуральные и стоимостные изменения, соответствующие построенному циклу, показаны в табл. 13.10 и 13.11.

Чистый стоимостный эффект от размещения в пустой клетке (Q, фиктивный) составит:

$$+0-0+5-7+20-10=+8 \text{ ф. ст. за изделие.}$$

В случае заполнения данной пустой клетки общая стоимость транспортировки увеличится на 8 ф. ст. за 1 изделие. Поэтому мы не будем вводить рассмотренные изменения. Теневые цены для оставшихся пустых клеток рассчитываются аналогичным образом. В табл. 13.12 показано все множество значений теневых цен (они обведены в кружочки).

Таблица 13.10. **Проверка пустой клетки (Q, фиктивный)**

Натуральные изменения, изделий

	B	C	*Фиктивный*
P	Пустая	Заполненная +1	Заполненная −1
Q	Заполненная −1	Пустая	Проверяемая +1
R	Заполненная +1	Заполненная −1	Пустая

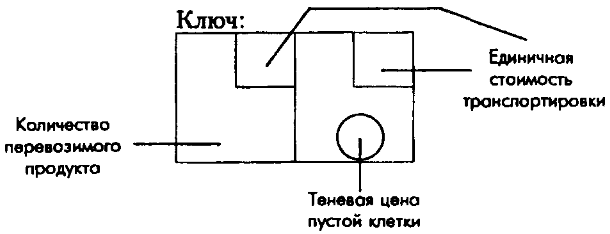
Таблица 13.11. **Проверка пустой клетки (Q, фиктивный)**

Стоимостные изменения, ф. ст.

	B	C	*Фиктивный*
P	Пустая	Заполненная +5	Заполненная −0
Q	Заполненная −10	Пустая	Проверяемая +0
R	Заполненная +20	Заполненная −7	Пустая

Таблица 13.12. **Проверка начального распределения перевозок на оптимальность — метод ступенек**

Торговый склад	*Розничный магазин*				*Общее предложение*
	A	*B*	*C*	*фиктивный*	
P	10 +11	20 +2	5 2	0 7	9
Q	2 +11	10 4	8 +11	0 +8	4
R	1 3	20 1	7 4	0 −2	8
Общая потребность	3	5	6	7	21



Это решение является неоптимальным, так как клетке (R, фиктивный) соответствует отрицательная теневая цена, равная −2 ф. ст. Стоимость транспортировки в 101 ф. ст. можно уменьшить, если ввести эту клетку и соответствующий ступенчатый цикл в распределение перевозок, что позволит достичь экономии стоимости в 2 ф. ст. на 1 изделие.

Мы продолжим решать этот пример и найдем оптимальное распределение перевозок в **13.2.5**, но сначала рассмотрим метод МОДИ вычисления теневых цен. Алгоритм метода ступенек является довольно трудоемким, и в процессе его реализации легко допустить ошибки. Использование оптимальности метода МОДИ в данном случае является гораздо более разумным. Хотя его алгоритм не позволяет выявить натуральные изменения, однако с его помощью можно получить те же значения теневых цен, затратив при этом гораздо меньше усилий.

Для начала рассмотрим только заполненные клетки. Для этих клеток каждое значение единичной стоимости c_{ij} разделяется на две компоненты — u_i для строк и v_j для столбцов. Например, единичная стоимость для клетки (R,B), лежащей на пересечении строки 3 и столбца 2, равна c_{32} = 20 ф. ст. В ней можно выделить компоненту u_3, соответствующую строке, и компоненту v_2, соответствующую столбцу, т.е.

$$c_{32} = 20 = u_3 + v_2.$$

Теневые цены для каждой пустой (небазисной) клетки можно найти из соотношения

$$s_{ij} = c_{ij} - (u_i + v_j).$$

Эта теневая цена отражает дополнительную стоимость транспортировки единицы изделия из пункта i в пункт j. Если все теневые цены положительны или равны нулю, т.е. $s_{ij} \geq 0$, то полученное решение является оптимальным. В этом случае перемещение единицы изделия в пустую клетку, которой соответствует положительная теневая цена, только увеличит общую стоимость транспортировки. Если же соответствующая теневая цена имеет нулевое значение, то общая стоимость транспортировки не изменится.

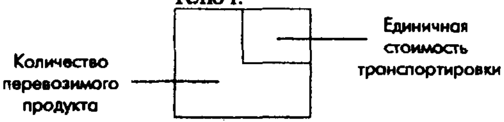
☐ **Пример 13.4.** Обратимся вновь к начальному распределению перевозок, полученному методом минимальной стоимости. Проведем проверку данного распределения на оптимальность с помощью метода МОДИ. Ниже воспроизведено начальное распределение перевозок (см. табл. 13.13).

Расчет компонент для строк u_i и компонент для столбцов v_j производится с помощью заполненных клеток. Заполненные клетки (P,C), (P, фиктивный), (Q,B), (R,A), (R,B) и (R,C) приводят к системе из шести уравнений. Эти шесть уравнений содержат семь переменных, поэтому система имеет не одно решение. Поскольку множество значений переменных является совместимым, фактические значения, присваиваемые компонентам, не играют никакой роли.

Таблица 13.13. **Начальное распределение перевозок, полученное методом минимальной стоимости**

Торговый склад	*Розничный магазин*				*Общее предложение*
	A	*B*	*C*	*фиктивный*	
P	10 –	20 –	5 2	0 7	9
Q	2 –	10 4	8 –	0 –	4
R	1 3	20 1	7 4	0 –	8
Общая потребность	3	5	6	7	21

Ключ:



$c_{13} = 5 = u_1 + v_3$ для заполненной клетки (P,C);
$c_{14} = 0 = u_1 + v_4$ для заполненной клетки (P, фиктивный);
$c_{33} = 7 = u_3 + v_3$ для заполненной клетки (R,C);
$c_{31} = 1 = u_3 + v_1$ для заполненной клетки (R,A);
$c_{32} = 20 = u_3 + v_2$ для заполненной клетки (R,B);
$c_{22} = 10 = u_2 + v_2$ для заполненной клетки (Q,B).

Какой-либо из компонент присваивается некоторое значение, по которому из соответствующих уравнений рассчитываются значения остальных компонент. Положим $u_1 = 0$. Из этого следует, что $v_3 = 5$, $v_4 = 0$, $u_3 = 2$, $v_1 = -1$, $v_2 = 18$ и $u_2 = -8$. Теперь, пользуясь соотношением

$$s_{ij} = c_{ij} - (u_i + v_j),$$

мы можем найти значения теневых цен, соответствующих незаполненным клеткам.

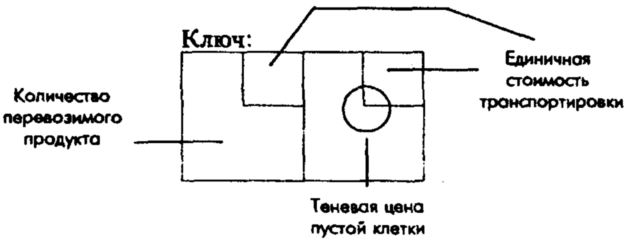
Подставив найденные значения компонент u_i и v_j, получим следующие теневые цены:

$s_{11} = 10 - (0 - (-1)) = +11$ для пустой клетки (P,A);
$s_{12} = 20 - (0 + 18) = +2$ для пустой клетки (P,B);
$s_{21} = 2 - (-8 - 1) = +11$ для пустой клетки (Q,A);
$s_{23} = 8 - (-8 + 5) = +11$ для пустой клетки (Q,C);
$s_{24} = 0 - (-8 + 0) = +8$ для пустой клетки (Q, фиктивный);
$s_{34} = 0 - (2 + 0) = -2$ для пустой клетки (R, фиктивный).

Эти значения заносятся в транспортную таблицу так, как это показано в табл. 13.14.

Таблица 13.14. Применение метода МОДИ для проверки на оптимальность начального распределения перевозок

Торговый склад	*Розничный магазин*				*Общее предложение*
	A	*B*	*C*	*фиктивный*	
P	10 (+11)	20 (+2)	5 2	0 7	$u_1=0$ 9
Q	2 (+11)	10 4	8 (+11)	0 (+8)	$u_2=-8$ 4
R	1 3	20 1	7 4	0 (−2)	$u_3=2$ 8
Общая потребность	3	5	6	7	21
	$v_1=-1$	$v_2=18$	$v_3=5$	$v_4=0$	



Теневые цены совпадают с теми значениями, которые были найдены методом ступенек и представлены в табл. 13.12. Маршрут (R, фиктивный) имеет отрицательную теневую цену − 2 ф. ст., следовательно, полученное решение является неоптимальным. Необходимо осуществить перераспределение перевозимых изделий с использованием указанной клетки и соответствующего ей ступенчатого цикла, что позволит снизить стоимость транспортировки.

13.2.5. Поиск оптимального решения

Итеративная процедура нахождения оптимального распределения перевозок может быть представлена следующим образом:

1. Если транспортная таблица содержит более одной пустой клетки с отрицательным значением теневой цены, то выбирается та из них, которой соответствует наибольшее значение по абсолютной величине.
2. Построение для этой клетки ступенчатого цикла аналогично описанному выше.
3. Выявление клеток, количество перевозок в которых необходимо сократить, и определение величины этих сокращений таким образом, чтобы ни одно из значений перевозок не оказалось отрицательным. Максимальное количество изделий, соответствующее выбранной клетке, определяется минимумом из этих значений. Перераспределение производится только для клеток, входящих в построенный цикл.
4. Нет никаких гарантий, что в полученном распределении нельзя предпринять никаких улучшений. Поэтому новое решение необходимо проверить на оптимальность с использованием метода МОДИ. Утверждать, что найденная стоимость транспортировки является минимальной, можно только в том случае, если все теневые цены положительны или равны нулю.

Продолжение примера 13.4. Единственной клеткой с отрицательным значением теневой цены, равным − 2 ф. ст., является клетка (R, фиктивный). В эту клетку желательно разместить максимально возможное количество изделий.

Ниже приведен ступенчатый цикл для клетки (R, фиктивный), которая имеет значение теневой цены, равное − 2 ф. ст., а также исходное распределение перевозок и единичные издержки.

Таблица 13.15. **Ступенчатый цикл для (R, фиктивный)**

	C			*Фиктивный*		
P	+	5	2	−	0	7
R	−	7	4	+	0	-2

Знак "+" означает увеличение количества перевозимых изделий в данной клетке; знак "−" — уменьшение соответствующего количества изделий.

Клетки со знаком "−" — это клетки (R, фиктивный) и (R,C), объем перевозок в которых равен 7 и 4 изделиям соответственно. Минимальным значением для клеток, отмеченных знаком "−", является 4, что означает, что внутри цикла можно осуществлять перемещение четырех изделий, добавляя их в клетки со знаком "+" и вычитая из клеток со знаком "−". Общая экономия стоимости транспортировки составит в данном случае (2 × 4) = 8 ф. ст. Изменения, внесенные в транспортную таблицу, отражены в табл. 13.16.

Таблица 13.16. **Перераспределение перевозок**

Торговый склад	*Розничный магазин*				*Общее предложение*
	A	*B*	*C*	*фиктивный*	
P	10 –	20 –	5 2 + 4	0 7 – 4	9
Q	2 –	10 4	8 –	0 –	4
R	1 3	20 1	7 4 – 4	0 0 + 4	8
Общая потребность	3	5	6	7	21

Данное решение по-прежнему является базисным, так как число заполненных клеток равно 6. Проверим данное решение на оптимальность с использованием метода МОДИ. Обратившись к заполненным клеткам (P,C), (P, фиктивный), (Q,B), (R,A), (R,B) и (R, фиктивный), получим:

$c_{13} = 5 = u_1 + v_3$ Положим $u_1 = 0$, тогда $v_3 = 5$;
$c_{14} = 0 = u_1 + v_4$ $v_4 = 0$;
$c_{34} = 0 = u_3 + v_4$ $u_3 = 0$;
$c_{31} = 1 = u_3 + v_1$ $v_1 = 1$;
$c_{32} = 20 = u_3 + v_2$ $v_2 = 20$;
$c_{22} = 10 = u_2 + v_2$ $u_2 = -10$.

Таким образом, теневые цены соответствующие пустым клеткам, будут равны:

$$
\begin{aligned}
s_{ij} &= c_{ij} - (u_i + v_j);\\
s_{11} &= 10 - (0 + 1) = +9;\\
s_{12} &= 20 - (0 + 20) = 0;\\
s_{21} &= 2 - (-10 + 1) = +11;\\
s_{23} &= 8 - (-10 + 5) = +13;\\
s_{24} &= 0 - (-10 + 0) = +10;\\
s_{33} &= 7 - (0 + 5) = +2.
\end{aligned}
$$

Поскольку ни одно из значений теневых цен не отрицательно, полученное решение является оптимальным.

Минимальная стоимость равна:

$$101 + (4 \times (-2)) = 93 \text{ ф. ст.}$$

Таблица 13.17. **Проверка распределения перевозок на оптимальность с использованием метода МОДИ**

Торговый склад	*Розничный магазин*				*Общее предложение*	
	A	*B*	*C*	*Фиктивный*		
P	10 +9	20 0	5 6	0 3	9	$u_1 = 0$
Q	2 +11	10 4	8 +13	0 +10	4	$u_2 = -10$
R	1 3	20 1	7 +2	0 4	8	$u_3 = 0$
Общая потребность	3	5	6	7	21	
	$v_1 = 1$	$v_2 = 20$	$v_3 = 5$	$v_4 = 0$		

Решение.

- Шесть изделий перевозятся со склада P в розничный магазин C, три изделия остаются на складе P.
- Четыре изделия перевозятся со склада Q в магазин B.
- Со склада R перевозятся три изделия в магазин A, одно – в магазин B, а четыре изделия остаются на складе.

В случае если и повторное распределение перевозок не является оптимальным, процедуру перераспределения повторяют необходимое число раз.

Следует отметить, что минимальная стоимость была достигнута еще в исходном распределении перевозок, полученном методом Вогеля. Такая ситуация в задачах небольшой размерности бывает довольно часто. Обычно метод Вогеля позволяет получить наилучшее начальное решение, однако нет никаких гарантий, что применение этого метода сразу обеспечивает получение оптимального решения. Следует также отметить, что распределение перевозок, полученное методом Вогеля, несколько отличается от распределения, найденного выше (см. пример **13.2**). Данная задача имеет альтернативное оптимальное решение:

- Со склада P одно изделие вывозятся в магазин B, шесть – в магазин C, а два – остаются на складе;
- Со склада Q четыре изделия вывозится в магазин B;
- Со склада R три изделия вывозятся в магазин A, а пять остаются на складе.

О существовании альтернативного оптимального решения говорит и нулевое значение теневой цены, соответствующей клетке (P,B). Нулевые значения теневых цен всегда связаны с существованием альтернативных оптимальных распределений перевозок, которым соответствует одно значение общей стоимости транспортировки.

13.2.6. Анализ чувствительности

Итоговое распределение перевозок, а также значения теневых цен, соответствующие пустым клеткам, можно использовать при проведении анализа модели на чувствительность. Теневая цена показывает, на сколько увеличится общая стоимость, если в пустую клетку поместить одну единицу продукта. Если нам придется осуществить перевозку одного изделия с торгового склада Q в розничный магазин C, увеличение стоимости составит 13 ф. ст., что гораздо выше, чем стоимость самого маршрута (Q,C), равная 8 ф. ст. Дополнительное увеличение стоимости появляется в связи с перебалансировкой распределения перевозок, при которой применяется нижеследующий ступенчатый цикл.

Таблица 13.18.
Ступенчатый цикл для (Q,C)

Натуральные изменения, изделий

	B	C	*Фиктивный*
P	Пустая	Заполненная −1	Заполненная +1
Q	Заполненная −1	Проверяемая +1	Пустая
R	Заполненная +1	Пустая	Заполненная −1

Таблица 13.19.
Ступенчатый цикл для (Q,C)

Стоимостные изменения, ф. ст.

	B	C	*Фиктивный*
P	Пустая	Заполненная −5	Заполненная +0
Q	Заполненная −10	Проверяемая +8	Пустая +8
R	Заполненная +20	Пустая	Заполненная −0

Чистые изменения стоимости составят:

$$+ 8 - 5 + 0 - 0 + 20 - 10 = 13 \text{ ф. ст. за изделие.}$$

Максимальное число изделий, которое можно перемещать внутри цикла, — это минимальное из значений, стоящих в клетках со знаком "−", т.е.

$$(P,C) = 6,\ (R,\text{ фиктивный}) = 4 \text{ и } (Q,B) = 4.$$

Следовательно, максимальное количество изделий, подлежащих перемещению, равно 4.

О нулевом значении теневой цены в клетке (P,B) мы уже упоминали в предыдущем разделе. Ступенчатый цикл для данной пустой клетки имеет следующий вид:

Таблица 13.20. **Ступенчатый цикл для (P,B)**

Натуральные изменения изделий

	B	Фиктивный
P	Проверяемая +1	Заполненная −1
R	Заполненная −1	Заполненная +1

Таблица 13.21. **Ступенчатый цикл для (P,B)**

Стоимостные изменения, ф. ст.

	B	Фиктивный
P	Проверяемая +20	Заполненная −0
R	Заполненная −20	Заполненная −0

Можно поместить некоторое число изделий в клетку (P,B), причем чистый стоимостный эффект будет равен нулю. Это означает, что существует альтернативное распределение перевозок, которое также позволяет получить минимальную стоимость в 93 ф. ст. Максимальное количество изделий, которое можно добавить в клетку (P,B), — это минимум из значений, указанных в клетках со знаком "−": (R,B) = 1 и (P, фиктивный) = 3. Следовательно, только одно изделие можно, перемещая по циклу, поместить в клетку (P,B).

Теневые цены можно использовать также в качестве индикаторов изменений стоимости транспортировки, соответствующей пустой клетке, которые оказывают воздействие на оптимальное распределение перевозок. Например, теневая цена пустой клетки (R,C) равна 2 ф. ст., а фактическая стоимость транспортировки — 7 ф. ст. за 1 изделие. Следовательно, для того, чтобы использование данной клетки в распределении перевозок привело к снижению общей стоимости транспортировки, фактическую единичную стоимость, соответствующую этой клетке, необходимо снизить как минимум до (7 − 2) = 5 ф. ст.

Действие стоимостных изменений в заполненных клетках выявить гораздо сложнее. При снижении издержек увеличение числа изделий в данной клетке выгодно. Если же издержки, стоящие в заполненных клетках, возрастает, то при достижении ими определенного значения использование этой клетки является нежелательным, и необходимо осуществить переход к иному маршруту.

Рассмотрим заполненную клетку (P,C). Соответствующая ей фактическая стоимость перевозок составляет 5 ф. ст. за изделие. Уменьшение этой стоимости не повлияет на объем перевозок, поскольку количество изделий, указанное в данной клетке, удовлетворяет всю потребность магазина C.

Если стоимость перевозки становится больше 5 ф. ст. то следует обратить внимание на ступенчатые циклы, в которых задействована клетка (P,C). Эти циклы дают значения теневых цен: 13 ф. ст. для (Q,C) и 2 ф. ст. для (R,C). В обоих циклах клетка (P,C) помечена знаком "−", и любое увеличение стоимости на 5 ф. ст. повлечет за собой снижение теневых цен указанных пустых клеток.

Изменение натурального объема перевозок будет иметь место в случае, если единичная стоимость транспортировки для клетки (P,C) возрастет более чем на 2 ф. ст. и превысит 7 ф. ст. При этом теневая цена клетки (R,C) станет отрицательной. В данной ситуации использование пустой клетки (R,C) окажется выгодным, что приведет к изменению объема перевозок для (P,C).

Таким образом, для полученного оптимального распределения перевозок верхним пределом стоимости, соответствующей (P,C), является значение 7 ф. ст., а нижним пределом — 0. Внутри указанного промежутка происходит изменение лишь общей стоимости транспортировки, тогда как в натуральном выражении распределение перевозок не меняется.

13.2.7. Модификации транспортной задачи

НЕДОПУСТИМЫЕ ПЕРЕВОЗКИ

Если перевозка из некоторого пункта производства в некоторый пункт назначения по той или иной причине невозможна, то в алгоритме решения задачи данное ограничение можно учесть, присвоив соответствующей клетке достаточно большое значение стоимости. Точное значение в данном случае неважно, однако, оно должно быть больше, чем остальные значения стоимости, указанные в таблице. Таким образом, алгоритм автоматически позволит избежать перевозок через данную клетку.

☐ **Пример 13.5.** В данном примере показано применение алгоритма решения транспортной задачи в решении проблем, связанных с недопустимостью прямых перевозок товаров из пунктов производства в пункты назначения. В примере будет рассмотрено движение продукта во времени. Пусть в нашем распоряжении имеется производственный график сроком на четыре месяца, который необходимо выполнить. Ниже приведены значения спроса на продукцию и производственных мощностей.

Таблица 13.22. **Значения спроса и производственных мощностей**

Месяц	*Производственные мощности, изделий*	*Спрос, изделий*
1	300	300
2	350	275
3	325	400
4	375	300

К началу первого месяца имеется начальный запас изделий объемом 50 шт. Изделия можно производить как для удовлетворения текущего спроса, так и создания запаса для удовлетворения спроса в последующие месяцы. Если спрос на изделия в течение месяца не удовлетворяется полностью, то прибыль от продажи теряется. Издержки производства составляют 100 ф. ст. за единицу изделия. Стоимость хранения запасов равна 2 ф. ст. за единицу изделия. Каков оптимальный план производства?

Решение.

Данную ситуацию можно формализовать, используя транспортную таблицу, в которой строками являются начальный запас и объемы производства изделий в месяц, а столбцы отражают ежемесячный спрос на продукцию. Маршруты (клетки), в которых подразумевается удовлетворение спроса за текущий месяц в следующих месяцах, считаются недопустимыми. В табл. 13.23 этим клеткам соответствуют бесконечные значения стоимости.

Таблица 13.23. **Данные производственного плана для месяцев 1–4**

		Стоимость единицы изделия, ф. ст. Месяцы				*Общее предложение*
		M1	*M2*	*M3*	*M4*	
Запас	M1	2	4	6	8	50
Производство	M1	100	102	104	106	300
	M2	∞	100	102	104	350
	M3	∞	∞	100	102	325
	M4	∞	∞	∞	100	375
Общая потребность		300	275	400	300	

Решение этой транспортной задачи производится с помощью обычного алгоритма, позволяющего минимизировать стоимость выполнения производственного графика (см. пример **13.8**).

ВЫРОЖДЕННОСТЬ

Решение называется вырожденным, если число перевозок в транспортной таблице меньше, чем $(m + n - 1)$. Данную проблему можно разрешить, проставив в независимые клетки очень маленькие, по сути равные нулю объемы перевозок. Число перевозок увеличивается таким образом до $(m + n - 1)$. Выявить клетки, которые следует использовать для этой цели, поможет алгоритм метода МОДИ проверки решения на оптимальность.

❐ **Пример 13.6.** Три торговых склада (X, Y и Z) могут осуществлять поставки 6, 3 и 4 единиц продукта в три магазина (L, M и N), спрос которых равен 4, 5 и 1 единицам соответственно. Значения единичной стоимости транспортировки указаны в приведенной ниже таблице.

Таблица 13.24. **Исходная информация**

Торговый склад	*Магазин, ф. ст./ед.*			*Общее предложение*
	L	*M*	*N*	
X	6	4	9	6
Y	5	3	2	3
Z	2	3	6	4
Общая потребность	4	5	1	

Как следует распределить перевозки, чтобы общая стоимость транспортировки была минимальной?

Решение.

Общее предложение составляет 13 единиц, что превышает общую потребность в 10 единиц, поэтому в задачу вводится фиктивный магазин, потребность которого в продукции балансирует излишек предложения торговых складов. Чтобы найти начальное распределение перевозок, применим метод Вогеля:

Таблица 13.25. **Начальное распределение перевозок, полученное методом Вогеля**

Торговый склад	*Розничный магазин*				*Общее предложение*	*Штрафная стоимость 123*
	L	*M*	*N*	*Фиктивный*		
X	$\boxed{6}$ –	$\boxed{4}$ 3	$\boxed{9}$ –	$\boxed{0}$ 3_1	~~630~~	4_1 2 2
Y	$\boxed{5}$ –	$\boxed{3}$ 2	$\boxed{2}$ 1_2	$\boxed{0}$ –	~~320~~	2 1 2
Z	$\boxed{2}$ 4_3	$\boxed{3}$ –	$\boxed{6}$ –	$\boxed{0}$ –	~~40~~	2 1 1
Общая потребность	~~4~~ 0	~~5~~ ~~3~~ 0	1 0	~~3~~ 0	13	
1-й штраф 2-й штраф 3-й штраф	3 3 3_3	0 0 0	4 4_2 –	0 – –		

Значение стоимости транспортировки составит:

$$4 \times 3 + 0 \times 3 + 3 \times 2 + 2 \times 1 + 2 \times 4 = 28 \text{ ф. ст.}$$

Для того, чтобы решение являлось базисным, оно должно включать $(3 + 4 - 1) = 6$ переменных, тогда как в нашей задаче число перевозок равно лишь 5. Найденное решение является вырожденным. Поступая в соответствии с алгоритмом метода МОДИ, мы должны ввести нулевую перевозку, чтобы использовать в качестве заполненной одну из пустых клеток. Этот прием позволяет получить требуемое число перевозок, равное 6. Затем можно будет рассчитать значения всех компонент u и v, а следовательно, и теневые цены.

Реализацию алгоритма метода МОДИ мы начнем, используя 5 заполненных клеток, соответствующих начальному распределению перевозок. Дополнительная нулевая перевозка будет введена только тогда, когда без нее продолжение алгоритма будет невозможно. Обратимся к данным табл. 13.26.

Таблица 13.26. **Проверка вырожденного решения на оптимальность — метод МОДИ**

Торговый склад	*Розничный магазин*				*Общее предложение*	
	L	*M*	*N*	*Фиктивный*		
X	6 +7	4 3	9 +6	0 3	6	$u_1 = 0$
Y	5 +7	3 2	2 1	0 +1	3	$u_2 = -1$
Z	2 4	3 −4	6 $\boxed{0}$	0 −3	4	$u_3 = 3$
Общая потребность	4	5	1	3	13	
	$v_1 = -1$	$v_2 = 4$	$v_3 = 3$	$v_4 = 0$		

Заполненные клетки используются для расчета соответствующих компонент по строкам и по столбцам из соотношения: $c_{ij} = u_i + v_j$ при условии, что $u_i = 0$. Значения v_2, v_4, u_2 *и* v_3 *можно найти*, не испытывая никаких затруднений, однако, значения u_3 и v_1 рассчитать нельзя. Для этого необходимо иметь дополнительную заполненную клетку.

Нулевую перевозку можно поместить в пустую клетку столбца v_1 или строки u_3. Какая из этих клеток будет выбрана, значения не имеет. Пусть выбрана клетка (Z, N). Теперь можно завершить алгоритм и найти значения теневых цен для пустых клеток из соотношения $s_{ij} = c_{ij} - (u_i + v_j)$. Соответствующие величины приведены в табл 13.26. Как видно из таблицы, в двух клетках теневые цены принимают отрицательные значения. Следовательно, полученное распределение перевозок является неоптимальным, и необходимо осуществить их перераспределение, используя при этом клетки (Z,M) или (Z, фиктивный). Начнем с клетки (Z,M), поскольку ей соответствует большее по абсолютной величине значение теневой цены. Ступенчатый цикл для клетки (Z,M) и движение объемов перевозок по входящим в него клеткам можно представить в виде табл. 13.27.

Чтобы определить число единиц, которые следует перемещать вдоль построенного цикла, обратимся к клеткам (Y,M) и (Z,N), помеченным знаком "−", количество перевозок в которых равно 2 и 0 единицам.

Таблица 13.27. **Ступенчатый цикл для клетки (Z,M)**

	M	N
Y	Заполненная – 2	Заполненная + 1
Z	Проверяемая +	Нулевая перевозка $\boxed{0}$ –

Это означает, что по циклу следует осуществлять перемещение нулевой перевозки таким образом, чтобы клетка (Z, N) снова стала пустой, а клетку (Z, M) предполагается использовать при распределении перевозок, поскольку в нее помещается нулевая перевозка. Остальные перевозки остаются без изменений. При дальнейшей проверке данного распределения на оптимальность выясняется, что значения всех теневых цен положительны. Данное распределение перевозок оптимальное. Это предполагает, что начальное решение, включающее 5 переменных, также оптимально. Обратимся к данным табл. 13.28.

Таблица 13.28. **Проверка оптимального решения — метод МОДИ**

С торгового склада	*В магазин*				*Общее предложение*	
	L	*M*	*N*	*фиктивный*		
X	6 +3	4 3	9 +6	0 3	6	$u_1 = 0$
Y	5 +3	3 2	2 1	0 +1	3	$u_2 = -1$
Z	2 4	3 $\boxed{0}$	6 +4	0 +1	4	$u_3 = -1$
Общая потребность	4	5	1	3	13	
	$v_1 = 3$	$v_2 = 4$	$v_3 = 3$	$v_4 = 0$		

Такие результаты далеко не всегда имеют место в случае вырожденности решения. В некоторых ситуациях при перераспределении перевозок определенное количество единиц продукции помещается в клетку с нулевой перевозкой, и тем самым данная клетка вводится в новое распределение перевозок. Это приводит к исчезновению вырожденности решения. Затем, для получения улучшенного распределения перевозок, применяются обычные алгоритмы.

МАКСИМИЗАЦИЯ

Алгоритм решения транспортной задачи предполагает, что ее целевая функция стремится к минимуму, Однако если некоторая проблема требует максимизации целевой функции перед тем, как применять для решения этой задачи стандартный алгоритм, его необходимо немного модифицировать. Например, мы намерены по условиям примера **13.5** осуществить перевозку товаров таким образом, чтобы максимизировать общий доход. В этом случае нам необходима информация о единичных доходах от транспортировки товаров между всеми пунктами производства и назначения. Модификация заключается в умножении всех значений единичного дохода на (– 1), а затем поступают обычным образом.

13.3. ЗАДАЧА О НАЗНАЧЕНИЯХ

Частным случаем транспортной задачи является задача о назначениях, в которой число пунктов производства равно числу пунктов назначения, т.е. транспортная таблица имеет форму квадрата. Кроме того, в каждом пункте назначения объем потребности равен 1, и величина предложения каждого пункта производства равна 1. Любая задача о назначениях может быть решена с использованием методов линейного программирования или алгоритма решения транспортной задачи. Однако ввиду особой структуры данной задачи был разработан специальный алгоритм, получивший название **Венгерского метода**.

13.3.1. Алгоритм решения задачи о назначениях

Этот алгоритм состоит из трех этапов.

Этап 1:

1. Формализация проблемы в виде транспортной таблицы по аналогии с решением транспортной задачи.
2. В каждой строке таблицы найти наименьший элемент и вычесть его из всех элементов данной строки.
3. Повторить ту же самую процедуру для столбцов.

Теперь в каждой строке и в каждом столбце таблицы есть по крайней мере один нулевой элемент. Представленная в полученной с помощью описанного выше приема "приведенной" транспортной таблице задача о назначениях эквивалентна исходной задаче, и оптимальное решение для обеих задач будет одним и тем же. Сущность Венгерского метода заключается в продолжении процесса приведения матрицы до тех пор, пока все подлежащие распределению единицы не попадут в клетки с нулевой стоимостью. Это означает, что итоговое значение приведенной целевой функции будет равно нулю. Так как существует ограничение на неотрицательность переменных, нулевое значение целевой функции является оптимальным.

Этап 2.

Если некоторое решение является допустимым, то каждой строке и каждому столбцу соответствует только один элемент. Если процесс распределения элементов осуществляется только в клетки с нулевой стоимостью, он приведет к получению минимального значения целевой функции.

1. Найти строку, содержащую только одно нулевое значение стоимости, и в клетку, соответствующую данному значению, поместить один элемент. Если такие строки отсутствуют, допустимо начать с любого нулевого значения стоимости.
2. Зачеркнуть оставшиеся нулевые значения данного столбца.
3. Пункты 1 и 2 повторять до тех пор, пока продолжение описанной процедуры окажется невозможным.

Если на данном этапе окажется, что есть несколько нулей, которым не соответствуют назначения и которые являются незачеркнутыми, то необходимо:

4. Найти столбец, содержащий только одно нулевое значение, и в соответствующую клетку поместить один элемент.
5. Зачеркнуть оставшиеся нули в данной строке.
6. Повторять пункты 4 и 5 до тех пор, пока дальнейшая их реализация окажется невозможной.

Если окажется, что таблица содержит неучтенные нули, повторить операции 1–6. Если решение является допустимым, т.е. все элементы распределены в клетки, которым соответствует нулевая стоимость, то полученное решение одновременно является оптимальным. Если решение является недопустимым, осуществляется переход к этапу 3.

Этап 3.

1. Провести минимальное число прямых через строки и столбцы матрицы (но не по диагоналям) таким образом, чтобы они проходили через все нули, содержащиеся в таблице.
2. Найти наименьший среди элементов, через которые не проходит ни одна из проведенных прямых.
3. Вычесть его из всех элементов, через которые не проходят прямые.
4. Прибавить найденный элемент ко всем элементам таблицы, которые лежат на пересечении проведенных ранее прямых.
5. Все элементы матрицы, через которые проходит только одна прямая, оставить без изменения.

В результате применения данной процедуры в таблице появляется по крайней мере один новый ноль. Необходимо возвратиться к этапу 2 и повторять алгоритм до тех пор, пока не будет получено оптимальное решение.

❐ **Пример 13.7.** Некоторая компания имеет четыре сбытовые базы и четыре заказа, которые необходимо доставить различным потребителям. Складские помещения каждой базы вполне достаточны для того, чтобы вместить один из этих заказов. В табл. 13.29 содержится информация о расстоянии между каждой базой и каждым потребителем. Как следует распределить заказы по сбытовым базам, чтобы общая дальность транспортировки была минимальной?

Таблица 13.29. **Расстояние от сбытовых баз до потребителей**

Сбытовая база	*Расстояние, миль* *Потребители*			
	I	*II*	*III*	*IV*
A	68	72	75	83
B	56	60	58	63
C	38	40	35	45
D	47	42	40	45

Решение

Понимание существа проблемы можно в значительной степени облегчить, если перед тем, как применять Венгерский метод, попытаться решить поставленную задачу, используя один из широко известных методов. Примените метод Вогеля и проследите, насколько он приближает нас к оптимальному решению, которое мы рассмотрим в конце данного раздела. Значения общего спроса и общего предложения для всех строк и столбцов равны единице.

Этап 1 Венгерского метода: В каждой строке находится наименьший элемент.

Таблица 13.30. **Выявление наименьших элементов по строкам**

	Потребители				*Наименьший элемент строки*
	I	*II*	*III*	*IV*	
A	68	72	75	83	68
B	56	60	58	63	56
C	38	40	35	45	35
D	47	42	40	45	40

Наименьший элемент вычитается из всех элементов соответствующей строки

Таблица 13.31. **Вычитание наименьшего элемента по строкам и выявление наименьшего элемента по столбцам**

0	4	7	15	
0	4	2	7	
3	5	0	10	
7	2	0	5	
0	2	0	5	*Наименьший элемент столбца*

Найденный наименьший элемент вычитается из всех элементов соответствующего столбца.

Таблица 13.32. **Вычитание наименьшего элемента по столбцам**

0	2	7	10
0	2	2	2
3	3	0	5
7	0	0	0

В соответствии с процедурой, описанной в этапе 2, осуществляются назначения. Наличие назначения обозначается через $\boxed{0}$.

Таблица 13.33. **Назначения в клетки с нулевыми значениями**

$\boxed{0}$	2	7	10
θ	2	2	2
3	3	$\boxed{0}$	5
7	$\boxed{0}$	θ	θ

На данном этапе мы можем осуществить только три нулевых назначения, тогда как требуемое их количество равно четырем. Полученное распределение является недопустимым. Переходим к этапу 3. Проводим наименьшее число прямых, проходящих через все нули таблицы.

Таблица 13.34. **Проведение прямых через нулевые элементы**

0	2	7	10
0	2	2	2
3	3	0	5
7	0	0	0

Наименьшим элементом, через который не проходит ни одна из прямых, является число 2. Скорректируем таблицу так, как это описано выше в соответствии с этапом 3, т.е. вычтем 2 из каждого элемента, через который не проходит ни одна прямая, и добавим 2 ко всем элементам, лежащим на пересечении двух прямых, оставив без изменения все прочие элементы, через которые проходит только одна прямая. Теперь перераспределим соответствующие назначения сбытовых баз и потребителей.

Таблица 13.35. **Скорректированная таблица с назначениями для нулевых клеток**

	I	*II*	*III*	*IV*
A	$\boxed{0}$	θ	7	8
B	θ	$\boxed{0}$	2	θ
C	3	1	$\boxed{0}$	3
D	9	θ	2	$\boxed{0}$

Теперь требование о размещении четырех назначений в клетки с нулевой стоимостью выполняется, следовательно, полученное решение является оптимальным. Перевозки осуществляются со сбытовой базы A к потребителю I, с базы B — к потребителю II, с базы C — к потребителю III и с базы D — к потребителю IV. Хотя данное решение и является оптимальным, однако оно не единственное. Тем не менее в любом оптимальном решении должен присутствовать маршрут (C,III), поскольку это единственный элемент с нулевой стоимостью в строке C. Два других оптимальных распределения назначений представлены ниже.

Таблица 13.36. **Первое альтернативное оптимальное решение**

	I	*II*	*III*	*IV*
A	$\boxed{0}$	θ	7	8
B	θ	θ	2	$\boxed{0}$
C	3	1	$\boxed{0}$	3
D	9	$\boxed{\theta}$	2	θ

Таблица 13.37. **Второе альтернативное оптимальное решение**

	I	*II*	*III*	*IV*
A	θ	$\boxed{0}$	7	8
B	$\boxed{0}$	θ	2	θ
C	3	1	$\boxed{0}$	3
D	9	θ	2	$\boxed{0}$

Минимальную дальность перевозок для каждого из трех решений можно вычислить из исходной таблицы:

Решение 1: 68 + 60 + 35 + 45 = 208 миль;

Решение 2: 68 + 63 + 35 + 42 = 208 миль;

Решение 3: 72 + 56 + 35 + 45 = 208 миль.

Общая дальность перевозок для всех трех решений одинакова.

Примечание: в задачах большей размерности, чем задача из примера **13.7**, убедиться в том, что проведенное в соответствии с пунктом 1 этапа 3 число прямых является минимальным, гораздо труднее. В этой связи может оказаться полезным так называемое "правило правой руки":

1. Выбирается любая строка или столбец, содержащие только один нулевой элемент.
2. Если выбрана строка, прямая проводится через столбец, в котором находится данный нулевой элемент.
3. Если выбран столбец, прямая проводится через строку, содержащую данный нулевой элемент.
4. Пункы **1–3** повторяются до тех пор, пока не будут учтены все входящие в таблицу нули.

13.3.2. Особые случаи задачи о назначениях

МАКСИМИЗАЦИЯ ЦЕЛЕВОЙ ФУНКЦИИ

Алгоритм решения задачи о назначениях предполагает минимизацию ее целевой функции. Если имеется задача о назначениях, целевую функцию которой нужно максимизировать, то поступают таким же образом, как и в алгоритме решения транспортной задачи: после окончания формирования первой таблицы все ее элементы умножаются на (– 1).

☐ **Пример 13.8.** В распоряжении некоторой компании имеется 6 торговых точек и 6 продавцов. Из прошлого опыта известно, что эффективность работы продавцов в различных торговых точках неодинакова. Коммерческий директор компании произвел оценку деятельности каждого продавца в каждой торговой точке. Результаты этой оценки представлены в табл. 13.38.

Таблица 13.38. **Объемы продаж в различных торговых точках для различных продавцов**

Продавец	*Объемы продаж, ф. ст./тыс. шт. Торговые точки*					
	I	*II*	*III*	*IV*	*V*	*VI*
A	68	72	75	83	75	69
B	56	60	58	63	61	59
C	35	38	40	45	25	27
D	40	42	47	45	53	36
E	62	70	68	67	69	70
F	65	63	69	70	72	68

Как коммерческий директор должен осуществить назначение продавцов по торговым точкам, чтобы достичь максимального объема продаж?

Решение.

Все элементы исходной таблицы умножаются на (–1);

Таблица 13.39. **Модификация исходных данных и выявление минимальных элементов**

Продавец	*Торговые точки*						
	I	*II*	*III*	*IV*	*V*	*VI*	*Минимальный элемент*
A	–68	–72	–75	–83	–75	–69	–83
B	–56	–60	–58	–63	–61	–59	–63
C	–35	–38	–40	–45	–25	–27	–45
D	–40	–42	–47	–45	–53	–36	–53
E	–62	–70	–68	–67	–69	–70	–70
F	–65	–63	–69	–70	–72	–68	–72

Минимальный (наибольший по абсолютной величине) элемент вычитается из всех элементов соответствующей строки.

Таблица 13.40. **Вычитание минимального элемента по строкам и выявление минимальных элементов по столбцам**

15	11	8	0	8	14
7	3	5	0	2	4
10	7	5	0	20	18
13	11	6	8	0	17
8	0	2	3	1	0
7	9	3	2	0	4
7	0	2	0	0	0

Минимальный элемент

Минимальный элемент вычитается из всех элементов соответствующего столбца.

Таблица 13.41. **Вычитание минимального элемента по столбцам**

8	11	6	0	8	14
0	3	3	0	2	4
3	7	3	0	20	18
6	11	4	8	0	17
1	0	0	3	1	0
0	9	1	2	0	4

Дальнейший поиск оптимального решения осуществляется в соответствии с обычным алгоритмом (см. пример **13.9**).

НЕДОПУСТИМЫЕ НАЗНАЧЕНИЯ

Данную проблему можно решить так же, как и транспортную задачу. Если по той или иной причине некоторое назначение является недопустимым, то в соответствующей клетке проставляется значение стоимости, которое заведомо больше любого другого значения. После этого в ходе реализации алгоритма мы сможем избежать данного назначения автоматически.

НЕСООТВЕТСТВИЕ ЧИСЛА ПУНКТОВ ПРОИЗВОДСТВА И НАЗНАЧЕНИЯ

Если исходная таблица не является квадратной, в нее следует включить дополнительные фиктивные строки и столбцы, необходимые для приведения ее к квадратной форме. Значения стоимости, соответствующие фиктивным клеткам, как правило, равны нулю.

Назначения, размещаемые в клетках фиктивных строк, фактически не существуют. Назначения, соответствующие фиктивным столбцам, на деле представляют собой те единицы, которые не подлежат распределению.

РЕЗЮМЕ

Транспортная модель — это частный случай модели линейного программирования. Стандартная задача включает в себя некоторое множество пунктов производства, например, несколько торговых складов, которые осуществляют поставки в некоторое множество пунктов назначения, например, в несколько магазинов. Цель состоит в минимизации общей стоимости транспортировки в рамках ограничений на спрос и предложение. Решение этой задачи может быть найдено с помощью традиционных методов линейного программирования. Относительно простая структура задачи позволяет, однако, разработать специальные алгоритмы, применение которых оказывается более трудоемким, чем применение обычных методов решения задач линейного программирования со множеством переменных.

Первый шаг алгоритма состоит в построении транспортной таблицы, в которой содержится информация об издержках транспортировки. Строкам этой таблицы соответствуют пункты производства, а столбцам — пункты назначения.

Второй шаг алгоритма — это поиск начального распределения перевозок. Нами было описано два метода реализации данной процедуры. В методе минимальной стоимости перевозки распределяются в первую очередь по наиболее дешевым маршрутам. Метод Вогеля предполагает расчет значений штрафной стоимости и такое распределение перевозок, которое позволяет избежать получения высоких штрафов. Однако ни один из методов не гарантирует, что полученное начальное распределение перевозок окажется оптимальным.

Третий шаг состоит в проверке начального распределения перевозок на оптимальность. Мы изложили два метода проверки решения на оптимальность. Оба они основаны на вычислении значений теневых цен для незаполненных клеток. Если эти значения положительны или равны нулю для всех пустых клеток, то полученное распределение перевозок является оптимальным.

В методе ступенек в пустую клетку помещается одна единица продукции. Затем определяются натуральные и стоимостные изменения, происшедшие под воздействием такого размещения. Метод МОДИ в большей степени основан на математической теории. Используя значения стоимости перевозки в каждой заполненной клетке, мы получаем стоимость, соответствующую строке или столбцу:

$$c_{ij} = u_i + v_j.$$

Используя значения компонент u и v, полученных для строк и столбцов соответственно, рассчитывают значения теневых цен, соответствующие всем пустым клеткам. Их расчет производится по формуле:

$$s_{ij} = c_{ij} - (u_i + v_j).$$

Реализация четвертого шага необходима только в случае, если полученное распределение перевозок является неоптимальным. Для осуществления перераспределения применяется ступенчатый цикл, соответствующий клетке с отрицательным значением теневой цены. Полученное решение вновь подвергается проверке на оптимальность.

Транспортная задача может иметь некоторые специфические особенности. Если предложение и спрос несбалансированны, то необходимо ввести в задачу фиктивные пункты производства или назначения. Оптимальное решение должно находиться в крайней точке допустимого множества, иными словами, должно быть базисным. Базисным называется решение, число переменных в котором равно числу строк в таблице плюс число столбцов минус единица. Если число переменных оказывается меньше указанной величины, то решение является вырожденным, и в этом случае следует использовать пустые клетки, размещая в них псевдоперевозки, объем которых равен нулю.

Недопустимые маршруты могут быть блокированы введением в соответствующие клетки таблицы достаточно больших значений стоимости транспортировки. Целевую функцию задачи можно не только минимизировать, но и максимизировать.

Еще более специфической задачей, для которой разработаны особые методы решения, является задача о назначениях. Число пунктов производства в этой задаче совпадает с числом пунктов назначения, причем каждой строке и каждому столбцу должно соответствовать только одно назначение. Для решения этой модифицированной транспортной задачи был разработан Венгерский метод.

УПРАЖНЕНИЯ

Упражнение 13.1

Два торговых склада поставляют продукцию в четыре магазина. Издержки транспортировки продукции с торговых складов в магазины, наличие продукции на складах и потребности магазинов приведены в следующей таблице:

Торговый склад	*Транспортные издержки, ф. ст. за единицу Магазин*				*Предложение продукции, ед.*
	G	*H*	*I*	*J*	
1	4	3	5	6	100
2	8	2	4	7	200
Потребность в продукции, ед	50	100	75	75	

Требуется найти распределение перевозок, позволяющее свести к минимуму общие транспортные издержки.

Упражнение 13.2

Три завода поставляют некоторую разновидность стали на пять торговых складов. Спрос каждого торгового склада в декабре, наличие стали на заводах, а также значения стоимости транспортировки 1 т стали приведены в нижеследующей таблице.

Завод	*Транспортные издержки, ф. ст. за единицу Торговый склад*					*Предложение, т*
	1	*2*	*3*	*4*	*5*	
A	20	27	33	25	34	200
B	22	36	34	28	26	250
C	26	29	27	26	28	300
Потребность, т	100	150	200	100	200	

Требуется определить минимальную стоимость транспортировки на декабрь.

Упражнение 13.3

Три пекарни осуществляют ежедневные поставки хлеба для четырех магазинов. Ниже представлена информация о спросе на продукцию, ее наличии и транспортных издержках.

Пекарня	*Транспортные издержки, пенсов/т Магазин*				*Общее предложение*
	I	*II*	*III*	*IV*	
X	1,5	2,5	1,0	2,0	700
Y	2,0	3,0	2,0	1,5	650
Z	1,0	1,5	2,5	3,0	800
Общая потребность	400	500	350	1000	

Требуется найти распределение поставок из каждой пекарни, минимизирующее общие транспортные издержки.

Упражнение 13.4

Предприятие розничной торговли имеет четыре крупных универмага, расположенных в различных городах — P, Q, R, S. Поставки продукции в эти универмаги осуществляются с двух торговых складов A и B, площади которых вмещают по 40 единиц продукции ежедневно.

В будущем планируется расширить площади универмагов, поэтому их потребности в продукции с торговых складов составят 27, 25, 30 и 35 единиц в день соответственно. Чтобы удовлетворить текущий и будущий спрос, планируется построить третий склад, площади которого позволят хранить в нем 60 единиц продукции ежедневно. Рассматриваются два варианта его размещения. Ниже приведены транспортные издержки, соответствующие перевозке продукции с двух существующих складов, и два варианта размещения нового склада.

Торговый склад	*Транспортные издержки, ф. ст. / ед. Универмаг*			
	P	*Q*	*R*	*S*
A	70	85	55	120
B	110	90	75	110
Вариант 1	115	115	70	90
Вариант 2	135	95	80	75

Требуется оценить две транспортные модели и принять решение о том, какой вариант размещения нового склада лучше. Предполагается, что остальные издержки сохраняют существующие значения.

Упражнение 13.5

Компании "Zeit plc" принадлежат три фермы, где выращиваются овощи, предназначенные для последующей обработки на двух холодильных заводах компании. Одним из выращиваемых овощей являются бобы, которые холодильные заводы продают по 200 ф. ст. за 1 т. Прогнозные значения спроса на следующий сезон равны 2750 т для завода "Craft" и 3250 т для завода "Liver". Ниже приведены издержки производства для каждой фермы и каждого холодильного завода, а также максимальные значения урожая для каждой фермы.

		Издержки производства, ф. ст. за 1 т	*Максимальный урожай, т*
Ферма	"Ascent Hill"	90	2000
	"Midrow Top"	95	3000
	"Alum Up"	87	1500
Завод	"Craft"	20	2750
	"Liver"	23	3250

Стоимость транспортировок следующая:

Ферма	*Холодильный завод, ф. ст./ т*	
	"Craft"	*"Liver"*
"Ascent Hill"	10	15
"Midrow Top"	12	12
"Alum Up"	18	9

Требуется:

1. Для ферм и холодильных заводов найти производственный план на следующий сезон, позволяющий получить максимальный доход.
2. Администрация компании "Zeit" планирует превратить ферму "Midrow Top" в центр производства бобов высокого качества, вследствие чего она обратила внимание на то, что издержки производства на данной ферме являются самыми высокими и составляют 95 ф. ст. за 1 т. На сколько вы порекомендовали бы снизить эти издержки, прежде чем изменение оптимального распределения перевозок будет целесообразным?

Упражнение 13.6

Администрация деревоперерабатывающего предприятия "Vibra" приняла на работу пять человек. Каждый из них имеет различные способности и навыки и затрачивает различное время на выполнение определенной работы. В настоящее время необходимо выполнить пять видов работ. Время выполнения работы каждым работником приведено в таблице:

Работник	*Время выполнения, ч*				
	Работы 1	*Работы 2*	*Работы 3*	*Работы 4*	*Работы 5*
M1	25	16	15	14	13
M2	25	17	18	23	15
M3	30	15	20	19	14
M4	27	20	22	25	12
M5	29	19	17	32	10

Требуется назначить на каждый вид работы одного из работников. Как это нужно сделать, чтобы общее время, необходимое для завершения всех видов работ, было минимальным?

Упражнение 13.7

Предприятие "Vibra" (см. упражнение **13.6**) может принять на работу еще одного рабочего по совместительству, который выполняет каждую работу в течение следующего времени:

Рабочий по совместительству	*Время выполнения, ч*				
	Работы 1	*Работы 2*	*Работы 3*	*Работы 4*	*Работы 5*
M6	28	16	19	16	15

Требуется определить, каким образом данная мера повлияет на назначение рабочих и минимизацию общего времени выполнения работ.

Упражнение 13.8

Завершить решение задачи о составлении плана производства по данным примера **13.5**, приведенного в **13.2.7**.

Упражнение 13.9

Завершить решение задачи о назначениях по данным примера **13.8**, приведенного в **13.3.2**. Провести назначение шести продавцов по шести торговым точкам, позволяющее максимизировать общий объем продаж.

Упражнение 13.10

В Kingdom of the Republik of Jdion имеется пять угольных шахт, показатели объемов выпуска продукции и издержек производства которых приведены в нижеследующей таблице:

Шахта	*Выпуск продукции, т/день*	*Издержки производства, ф. ст. за 1 т*
1	120	25
2	150	29
3	80	34
4	160	26
5	140	28

До того как уголь будет готов к продаже, его необходимо "очистить" и отсортировать на одном из трех углеперерабатывающих заводов. Ниже приведены значения производственных возможностей и эксплуатационных расходов по каждому заводу:

Завод	*Выпуск продукции, т/день*	*Эксплуатационные расходы, ф. ст. за 1 т*
A	300	2
B	200	3
C	200	3

Перевозка угля производится по железной дороге, ее стоимость равна 0,5 ф. ст. за 1 т-км. Расстояние от каждой шахты до каждого углеперерабатывающего завода следующее (км):

Углеперерабатывающий завод	*Шахта*				
	1	*2*	*3*	*4*	*5*
A	22	44	26	52	24
B	18	16	24	42	48
C	44	32	16	16	22

1. Построив транспортную модель, определите, как следует распределить перевозки добытого угля с шахт на каждый из трех перерабатывающих заводов.
2. Ввиду установки нового оборудования на шахте 3 ее издержки производства, как ожидается, снизятся до 30 ф. ст. за 1 т. Окажет ли это изменение воздействие, и если да, то какое на распределение перевозок угля на перерабатывающие заводы?
3. Планируется увеличение объема добычи на шахте 5 до 180 т в день, причем его можно достичь, не увеличивая издержки производства 1 т угля. Как это повлияет на распределение перевозок угля к перерабатывающим заводам?
(АССА, июнь 1986 г.).

Упражнение 13.11

1. Кратко опишите и сравните два метода поиска начального допустимого решения транспортной задачи.

2. Компания "Braintree Electronics Company" выпускает ленты к видеокассетам, предназначенные для продажи населению. Ниже приведены значения спроса (100 м) и производственных возможностей (выпуск продукции, 100 м) за IV квартал.

Месяц	*Спрос*	*Выпуск продукции в урочное время*	*Выпуск продукции в сверхурочное время*
Октябрь	300	400	150
Ноябрь	450	400	150
Декабрь	800	400	150

Отметим, что производственные возможности позволяют производить ленты к видеокассетам как в течение урочного, так и сверхурочного времени работ, причем если показатели производственных возможностей постоянны, то значение спроса возрастает перед Рождеством. Компания не обладает каким-либо запасом продукции на данный момент и не намерена создавать его после декабря.

Издержки производства 100 м ленты к видеокассетам равны 150 ф. ст. в урочное время и 180 ф. ст. в сверхурочное время работы. Было установлено, что стоимость хранения запасов составляет 20 ф. ст. за 100 м ленты в месяц. При ответе на вопросы примите предпосылку о том, что все заказы удовлетворяются точно в срок, а спрос и предложение возникают в середине каждого месяца.

Требуется:

а) Формализовать изложенную ситуацию на производстве в виде транспортной модели, включающей шесть "пунктов производства" и три "пункта назначения", в которой показаны значения единичной стоимости для каждой пары: пункт производства — пункт назначения.

б) Используя алгоритм решения транспортной задачи, найти оптимальный план производства на указанный период. Определить общую стоимость, соответствующую найденному решению.

(АССА, июнь 1988 г.).

Упражнение 13.12

а) Объясните значение терминов:

выpожденность;

неравенство спроса и предложения;

не единственное оптимальное решение применительно к транспортной задаче. Объясните, как можно модифицировать алгоритм ее решения, чтобы преодолеть указанные трудности.

б) Компания "Royal Wedgetoun Pottery" получила заказы на три вида выпускаемой ею продукции (бокалы, чашки и вазы), которые необходимо удовлетворить в течение следующей недели. Размеры заказов следующие:

Продукт	*Размер заказа, единиц*
Бокалы	4000
Чашки	2400
Вазы	1000

В распоряжении компании имеются три станка, на каждом из которых можно производить любой из указанных видов продукции с одинаковой производительностью. Однако единичные затраты по каждому виду продукции варьируют в зависимости от используемого станка. В нижеследующей таблице приведены единичные издержки (ф. ст.) по каждому станку:

Станок	*Бокалы*	*Чашки*	*Вазы*
A	1,20	1,30	1,10
B	1,40	1,30	1,50
C	1,10	1,00	1,30

Кроме того, известно, что производственные мощности станков B и C на следующую неделю составят 3000 единиц, а станка A — 2000 единиц.

Требуется, используя транспортную модель, найти план производства для видов продукции и станков, минимизирующий общую стоимость производства. Определить значение минимальной стоимости.

Если найденное оптимальное решение не единственное, нужно привести другие варианты решений, которым соответствует минимальная стоимость производства. Если бы менеджер по производству захотел, чтобы в производственном плане было как можно меньше изменений в производстве изделий на различных станках, то какое оптимальное решение вы бы порекомендовали?

(АССА, июнь 1989 г.).

Упражнение 13.13

а) Кратко поясните, как можно модифицировать алгоритм решения транспортной задачи, если цель состоит не в минимизации затрат, а в максимизации прибыли.

б) Компания "Orange Computer" производит только один вид продукции — матричные печатающие устройства, которые в настоящее время являются дефицитом. Четыре основных покупателя — это крупные специализированные компьютерные универмаги, расположенные в Аббатстауне, Бесвиче, Карлике и Денстоуне, уже подали заявки, общий размер которых превышает общие производственные мощности трех заводов компании в Рексфорде, Сидоне и Тристроне. Компания должна принять решение о том, как распределить производственные мощности, чтобы получить максимальную прибыль.

После того, как каждый принтер тщательно упакован в мягкую упаковку, предохраняющую его от каких-либо повреждений, его помещают в отдельную коробку. В нижеследующей таблице приведены значения стоимости транспортировки одной единицы от каждого завода-производителя в каждый специализированный универмаг (ф. ст.):

	"Аббатстаун"	*"Бесвич"*	*"Карлик"*	*"Денстоун"*
Рексфорд	22	24	22	30
Сидон	24	20	18	28
Тристрон	26	20	26	24

Поскольку все четыре специализированных универмага расположены в различных частях страны и, следовательно, стоимость транспортировки продукции между заводами-производителями и универмагами различна, а также ввиду некоторых различий и в издержках производства каждого из четырех заводов, существующая структура цен предусматривает возможность установления различных цен для каждого из четырех универмагов. В настоящее время установлены следующие цены за единицу продукции: 230 ф. ст. в Аббатстауне, 235 ф. ст. в Бесвиче, 225 ф. ст. в Карлике и 240 ф. ст. в Денстоуне. Издержки производства на единицу продукции составляют 150 ф. ст. на заводах в Рексфорде и Тристроне и 155 ф. ст. на заводе в Сидоне.

Требуется сформировать матрицу, состоящую из входящих в прибыль единичных доходов, соответствующих каждой паре перевозок с заводов-производителей в универмаги.

Значения спроса в Аббатстауне, Бесвиче, Карлике и Денстоуне равны 850, 640, 380 и 230 единицам соответственно. Производственные мощности позволяют производить на заводе в Рексфорде 625, в Сидоне — 825, а в Тристроне — 450 принтеров. Используя алгоритм решения транспортной задачи, определить оптимальное распределение перевозок.

Определить соответствующую оптимальному решению прибыль.

(АССА, июнь 1990 г.).

Упражнение 13.14

а) Задача о назначениях является частным случаем транспортной задачи.

Опишите специфические особенности этой задачи и объясните, почему при решении задачи о назначениях нежелательно использовать алгоритм решения транспортной задачи.

б) Членов Ассоциации ученых Мидленда недавно уведомили, что их ассоциация получит государственные гранты на проведение исследований в соответствии с четырьмя основными исследовательскими проектами. Исполнительный директор ассоциации должен по каждому проекту назначить научного руководителя. В настоящее время эти обязанности можно возложить на одного из пяти исследователей — Адамс, Браун, Карр, Дэй и Иванс. Время, требуемое для завершения каждого из исследовательских проектов, зависит от опыта и способностей исследователя, которому будет поручено руководство выполнением проекта. Исполнительному директору были представлены оценки времени выполнения проекта каждым из ученых (в днях).

Ученый-исследователь	*Проект*			
	1	*2*	*3*	*4*
Адамс	80	120	60	104
Браун	72	144	48	110
Карр	96	148	72	120
Дэй	60	108	52	92
Иванс	64	140	60	96

Поскольку все четыре проекта обладают равным приоритетом в выполнении, исполнительный директор заинтересован в таком назначении научных руководителей, которое бы позволило свести к минимуму общее время (в днях), требуемое для завершения всех четырех проектов.

Требуется определить оптимальный вариант назначения научных руководителей проектов и, следовательно, общее число дней, необходимое для завершения четырех проектов.

Найти какие-либо другие варианты назначения, которые привели бы к тому же результату. Учитывая, что ученые Браун, Карр и Дэй отдают предпочтение проектам 2 и 3, а ученые Адамс и Иванс — проектам 1 и 4, какой из имеющихся оптимальных вариантов назначения, принятый исполнительным директором, был бы наиболее разумным?

Какие особенности матрицы продолжительности выполнения проектов, сформированной для данной задачи, можно было бы использовать, чтобы упростить поставленную задачу?

(АССА, декабрь 1989 г.).

Глава 14. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

14.1. ВВЕДЕНИЕ

В главе 11, чтобы формализовать ситуации, связанные с хранением и управлением запасами, мы использовали математические модели. Рассмотренные проблемы были достаточно просты и удовлетворяли всем основным предпосылкам, введенным в модели. Однако иногда возникают более сложные задачи, решение которых с помощью изложенных математических моделей не отвечает существу поставленной проблемы. Поэтому мы вынуждены обращаться к иной группе методов, которые можно использовать в ситуациях, выходящих за рамки системы предпосылок, на которых основаны простые модели.

Имитационное моделирование используется в случаях, когда применение математических аналитических моделей неадекватно или является слишком сложным. Хотя методы имитационного моделирования не слишком элегантны, они являются очень гибкими и мощными в применении. Они шаг за шагом воспроизводят процесс функционирования системы. Эта система может включать ряд стохастических переменных. В системе управления запасами, например, неопределенности могут быть подвержены как ежегодный спрос, так и срок реализации заказа.

Используя выборочные данные, можно моделировать поведение системы. Если имитационное моделирование применяется в течение достаточно длительного периода, появляется возможность создавать модели с периодическим циклом или рассчитывать математические ожидания для определенных параметров. Имитационное моделирование может помочь при составлении прогнозов относительно возможного поведения системы в будущем.

Более подробно мы остановимся на одном методе, который известен под названием **Метод Монте-Карло**. В данном методе всем переменным присваиваются дискретные значения, даже если на самом деле эти переменные являются непрерывными. Переменная времени, например, может подразделяться на интервалы в минутах, часах или днях в зависимости от моделируемой системы. Затем рассчитываются вероятности каждого значения, а в отборе значений переменных из распределения вероятности используются случайные числа. С помощью описанной процедуры генерируются ряды значений переменных, которые являются основой для построения имитационной модели.

В имитационном моделировании, как и в большинстве методов исследования операций, рассмотренных нами ранее, при построении моделей и их последующем анализе, как правило, широко используются компьютеры. В этой области применение компьютеров становится особенно важным, поскольку значимую и обоснованную информацию из имитационной модели можно получить только после проведения расчетов для различных случайных чисел. Если мы заинтересованы в нахождении стационарного состояния модели, необходимо сделать расчет за длительный период моделируемой переменной времени и таким образом получить средние значения соответствующих статистических характеристик. Если же моделируемый период слишком мал, то на средние значения переменных могут оказывать воздействие начальные (стартовые) колебания.

Конечно, весь спектр применения имитационных методов и моделей в рамках данной книги продемонстрировать невозможно, однако, принципы, лежащие в их основе, одинаковы.

14.2. ПРИНЦИПЫ ПОСТРОЕНИЯ ДИСКРЕТНЫХ ИМИТАЦИОННЫХ МОДЕЛЕЙ

В основу исследований рынка положена предпосылка о том, что клиент пытается выработать оценку общественного мнения по интересующему его вопросу. Клиент желает знать, какова продолжительность определенной работы и ее стоимость. Первый этап заключается в организации выборочного обследования и разработке анкеты. Вторым этапом является сбор исходных данных. Предположим, что подготовлена соответствующая анкета и выработан план проведения выборочного обследования. Пусть принято решение, что сбор информации будет проводиться интервьюерами путем опроса прохожих на улицах крупного города. Длительность проведения обследования и соответствующие затраты зависят от того, сколько времени понадобится интервьюеру для сбора исходных данных. Каким образом ответственная за проведение обследования организация может оценить, сколько времени потребуется для его проведения?

Необходимо провести анализ ситуации. Интервьюеру придется останавливать прохожих, спрашивать об их желании или нежелании дать интервью и в случае, если они согласны, задать им соответствующие вопросы. Переменными в данной ситуации являются следующие величины:

1. Интервьюеру придется ожидать прохожего, которого можно остановить. Следовательно, нам необходимо знать величину интервала между последовательными моментами появления прохожих (IAT).
2. Желание прохожего дать интервью.
3. Продолжительность самого интервью.

Если нам удастся сгенерировать информацию, отражающую процесс остановки прохожего, и его возможное интервьюирование, то мы сможем построить имитационную модель для данной проблемы и оценить время, требующееся для того, чтобы набрать необходимое число интервью. Причем данные должны отражать стандартные характеристики переменных, которые были идентифицированы выше. Каждая из этих переменных является стохастической, т.е. подверженной неопределенности. Наиболее простой способ состоит в сборе определенных данных через проведение испытаний.

Если в качестве испытания выбрать поток из 100 прохожих, то можно зафиксировать временные интервалы между их последовательным появлением, желание или нежелание быть проинтервьюированным и, если они дадут согласие, продолжительность интервью. Степень точности этих данных зависит от специфики проблемы. В данном случае совершенно неважно, чтобы время было зафиксировано с высокой степенью точности. Кстати, именно на этой стадии принимается решение о том, какие дискретные значения времени следует использовать. Например, между последовательным появлением двух прохожих проходит приблизительно 1 мин., а каждое интервью занимает примерно 2 мин.

После того, как собраны данные для потока из 100 прохожих, для каждой переменной можно построить распределение частот и рассчитать соответствующее значение вероятности. Предположим, что по результатам испытания были зафиксированы следующие данные:

Таблица 14.1. **Модель появления прохожих — интервалы между моментами появления**

Время между появлениями прохожих около интервьюера, мин.	*0*	*1*	*2*	*3*	*4*	*5*
Число появлений, f	25	35	18	10	8	4
Вероятность (f ÷ 100)	0,25	0,35	0,18	0,10	0,08	0,04

Из общего числа опрошенных 75 человек выразили желание дать интервью. Следовательно, вероятность того, что некоторый прохожий будет согласен на интервью, можно оценить как 0,75.

Таблица 14.2. **Продолжительность интервью**

Продолжительность интервью, мин	*2*	*4*	*6*
Количество интервью, f	40	45	15
Вероятность (f ÷ 100)	0,40	0,45	0,15

Как эти данные можно использовать для того, чтобы сгенерировать процесс появления прохожих? Один из методов генерирования — это использование таблицы случайных чисел (см. Приложение 2). Таблица случайных чисел заключает в себе цифры от 0 до 9, выбранные случайным образом. Группировки в таблицах применяются исключительно для удобства чтения. При пользовании таблицей в качестве точки отсчета может быть выбрана любая позиция. В зависимости от требований цифры можно выбирать по одной, по две или по три, двигаясь по таблицам вправо или вниз. Случайные числа используются для того, чтобы множеству значений переменной поставить в соответствие множество случайных чисел (например, 0–9, 00–99). Случайные числа ставятся в соответствие значениям переменной пропорционально значениям вероятностей.

Таким образом, из указанных таблиц выбирается случайное число, и переменной присваивается соответствующее значение. Так как в данной задаче значения вероятностей указаны с точностью до двух десятичных знаков, мы будем пользоваться случайными числами, содержащими две цифры. Распределение интервала случайных чисел 00–99 показано на табл. 14.3.

Таблица 14.3. **Распределение случайных чисел для интервалов между моментами появления прохожих**

Интервал между появлениями, мин.	*Вероятности*	*Кумулятивные вероятности*	*Случайные числа*
0	0,25	0,25	00–24
1	0,35	0,60	25–59
2	0,18	0,78	60–77
3	0,10	0,88	78–87
4	0,08	0,96	88–95
5	0,04	1,00	96–99

Если выбирается случайное число 03, то оно принадлежит промежутку (00–24) и характеризует интервал между появлениями прохожих (IAT) в ноль минут. Случайное число 47 принадлежит промежутку (25–59) и соответствует IAT в одну минуту. Используя последовательные случайные числа и двигаясь вдоль по строке или вниз по столбцу таблицы, а также с помощью приведенных выше данных мы можем поставить в соответствие каждому человеку интервал его появления около интервьюера. Полученные значения IAT накапливаются, начиная с нулевого значения, и в результате позволяют найти время появления каждого прохожего.

Таблица 14.4. **Распределение интервалов случайных чисел для желающих дать интервью**

Согласие прохожего дать интервью	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
Да	0,75	0,75	00–74
Нет	0,25	1,00	75–99

Чтобы установить, согласится ли моделируемый прохожий дать интервью, выбираем случайное число из другого столбца или строки таблицы. Пусть выбрано число 35. Оно находится в промежутке (00–74). Данный прохожий согласен дать интервью. Если следующее число равно 64, то, поскольку оно принадлежит тому же промежутку, следующий прохожий также даст согласие на интервью.

Таблица 14.5. **Распределение интервалов случайных чисел для продолжительности интервью**

Продолжительность интервью, мин	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
2	0,40	0,40	00–39
4	0,45	0,85	40–84
6	0,15	1,00	85–99

Продолжительность интервью устанавливается аналогично, но с использованием отличного от двух предыдущих множества случайных чисел.

Теперь все готово к тому, чтобы начать процесс моделирования. Мы будем продолжать его до тех пор, пока не будет получено 10 интервью. Для каждой переменной выбираются случайные числа, а затем генерируются значения переменных, необходимых для продолжения процесса моделирования (время появления прохожего), а также переменных, необходимых для описания поведения системы (согласие дать интервью и его продолжительность).

Ниже приведены данные из таблиц случайных чисел, которые помогут вам проследить за ходом процесса моделирования.

03 47 43 73 86 97 74 24 67 62 16 76 62 27 66 12 56 85 99 26 55 59 56
35 64 16 22 77 94 39 84 42 17 53 31 63 01 63 78 59 33 21 12 34 29 57

Для интервалов появления прохожего выберем случайные числа с начала списка и будем продвигаться вдоль строки. Данный ряд начинается с чисел: 03, 47, 43. Для согласия дать интервью выберем случайные числа второй строки, которая начинается с чисел: 35, 64, 16. Для продолжительности интервью также выберем числа второй строки, но начнем с конца и будем двигаться справа налево: 57, 29, 34. Предположим, что моделируемый счетчик времени начинается с нулевого момента. Тогда первый прохожий появится в момент времени, равный (0 + первый интервал появления прохожего). Предположим также, что каждое следующее интервью может начаться сразу же после окончания предыдущего.

Десятое интервью завершится через 36 мин. после начала процедуры. Использование иного множества случайных чисел приведет к другому результату. Если потребуется определить время, необходимое для десяти интервью, мы должны будем сделать по имитационной модели расчеты для большего числа интервью — например, для 100 или 200. И только после этого можно будет рассчитать среднее время, требуемое для завершения 10 интервью.

Одна из проблем, возникающих при построении имитационных моделей, состоит в том, что необходимо точно знать, какого рода информацию следует собирать, чтобы процесс моделирования можно было продолжить. В данной ситуации небольшой размерности существует возможность идентифицировать каждый шаг и при необходимости вернуться на предыдущий этап, если возникла потребность в дополнительной информации. В ситуациях большего масштаба, моделирование которых осуществляется с помощью компьютеров, очень важно, чтобы решение о том, какие данные необходимы, и о способах их сбора и представления было принято еще на начальном этапе.

Таблица 14.6. **Моделирование процесса проведения 10 интервью одним интервьюером**

Номер прохожего	*Модель появления*			*Согласие дать интервью*		*Модель интервью*			
	Случайное число	*IAT, мин*	*Время появления, мин*	*Случайное число*	*Да/нет*	*Случайное число*	*Продолжительность, мин*	*Время*	
								Начало	*Окончание*
1	03	0	0	35	Да	57	4	0	4
2	47	1	1	Интервьюер занят					
3	43	1	2	Интервьюер занят					
4	73	2	4	64	Да	29	2	4	6
5	86	3	7	16	Да	34	2	7	9
6	97	5	12	22	Да	12	2	12	14
7	74	2	14	77	Нет		Отказ		
8	24	0	14	94	Нет		Отказ		
9	67	2	16	39	Да	21	2	16	18
10	62	2	18	84	Нет		Отказ		
11	16	0	18	42	Да	33	2	18	20
12	76	2	20	17	Да	59	4	20	24
13	62	2	22	Интервьюер занят					
14	27	1	23	Интервьюер занят					
15	66	2	25	53	Да	78	4	25	29
16	12	0	25	Интервьюер занят					
17	56	1	26	Интервьюер занят					
18	85	3	29	31	Да	63	4	29	33
19	99	5	34	63	Да	01	2	34	36
						10 интервью набрано			

Сбор данных преследует две основные цели. Во-первых, их можно использовать при проверке положения о том, что модель функционирует именно так, как и предполагалось при ее составлении. Эта процедура является составной частью обоснования модели. Например, по данным исходного распределения, математическое ожидание продолжительности интервью составит:

$$\text{E (продолжительность интервью)} = 2 \times 0{,}4 + 4 \times 0{,}45 + 6 \times 0{,}15 = 3{,}5 \text{ мин.}$$

По данным нашей небольшой имитационной модели, на проведение 10 интервью интервьюер затрачивает 28 мин., таким образом, среднее значение продолжительности одного интервью составляет 2,8 мин., что несколько меньше, чем предполагалось изначально. Для выборки такого небольшого размера эта колеблемость неудивительна. Однако если бы мы получили эти же результаты для первых 100 интервью, они означали бы, что модель является некорректной и требует тщательной проверки.

Во-вторых, данные можно использовать для получения некоторой информации непосредственно из модели. Например, сколько времени потребуется, чтобы получить 10 интервью? — 36 мин. Какую часть времени интервьюер бездействует? — 8 мин из 36. Сколько человек прошло мимо интервьюера, пока он получал 10 интервью? — 19: 6 человек прошли, пока интервьюер был занят; 3 человека отказались дать интервью; 10 человек были проинтервьюированы.

Данное исследование можно расширить, если, например, ввести в модель второго интервьюера. Затраты на оплату его работы могут быть компенсированы сокращением времени, необходимого для получения 10 интервью.

Введение в модель второго интервьюера требует принятия дополнительных правил, определяющих функционирование модели. Что произойдет, если оба интервьюера будут свободны? Кто из них подойдет к ближайшему следующему прохожему? Примем предпосылку о том, что первого прохожего всегда берет на себя интервьюер 1. Подобные правила необходимо вводить на начальном этапе формулировки любой модели. В данном случае неважно, какой из интервьюеров будет выбран, но при построении всех имитационных моделей нужно быть последовательным и уметь четко сформулировать правила функционирования моделируемой системы.

Итак, мы установили, что в данном случае получение 10 интервью займет 27 мин. Между тем на практике средние значения вычисляются на основе гораздо более длительного процесса моделирования. Для того, чтобы принять решение о том, пользоваться услугами второго интервьюера или нет, потребуются и другие данные, например, о промежутках времени, когда оба интервьюера свободны.

Таблица 14.7. **Моделирование процесса проведения 10 интервью двумя интервьюерами**

Номер прохожего	*Модель появления*			*Согласие дать интервью*		*Модель интервью*					
								Интервьюер 1		*Интервьюер 2*	
	Случайное число	*IAT, мин*	*Время появления, мин*	*Случайное число*	*Да/нет*	*Случайное число*	*Продолжительность, мин*	*Начало*	*Окончание*	*Начало*	*Окончание*
1	03	0	0	35	Да	57	4	0	4		
2	47	1	1	64	Да	29	2			1	3
3	43	1	2	Оба интервьюера заняты							
4	73	2	4	16	Да	34	2	4	6		
5	86	3	7	22	Да	12	2	7	9		
6	97	5	12	77	Нет	Отказ					
7	74	2	14	94	Нет	Отказ					
8	24	0	14	39	Да	21	2	14	16		
9	67	2	16	84	Нет	Отказ					
10	62	2	18	42	Да	33	2	18	20		
11	16	0	18	17	Да	59	4			18	22
12	76	2	20	53	Да	78	4	20	24		
13	62	2	22	31	Да	63	4			22	26
14	27	1	23	Оба интервьюера заняты							
15	66	2	25	63	Да	01	2	25	27		
								10 интервью набрано			

После того, как имитационная модель построена, необходимо оценить ее надежность. Мы должны быть уверены в том, что модель воспроизводит формализуемую ситуацию с достаточной степенью точности. Простейший способ оценки надежности состоит в использовании ретроспективных данных и сравнении результатов расчетов, полученных для этих данных по модели, с действительным поведением системы во времени. В приведенном выше примере ретроспективные данные можно не использовать, а оценку надежности модели следует основывать на тщательной проверке и оценке используемых распределений вероятности. Этот момент является очень важным для имитационного моделирования, хотя иногда им пренебрегают.

14.3. ПРИМЕНЕНИЕ ИМИТАЦИОННЫХ МОДЕЛЕЙ В СИСТЕМАХ МАССОВОГО ОБСЛУЖИВАНИЯ

Особенности формулировки любой имитационной модели зависят от специфики рассматриваемой проблемы. Для иллюстрации общего алгоритма рассмотрим два типа систем массового обслуживания.

☐ **Пример 14.1.** Доктор Аббот и доктор Буф имеют в совместной собственности кабинет, в котором начиная с 9.00 ведут утренний прием больных. Приемная открывается в 8.30, а закрывается в 10.00 утра. Секретарь сохраняет записи об обращениях пациентов за последние десять недель, кроме того, сами врачи ведут учет пациентов, принятых ими в часы консультаций. Входной поток имеет следующую структуру:

Таблица 14.8. **Модель входного потока пациентов**

Промежуток между моментами появления пациентов, мин	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Вероятность	0,05	0,05	0,10	0,20	0,40	0,10	0,05	0,05

Одна половина пациентов регистрируется у доктора Аббота, другая — у доктора Буфа, причем они образуют две отдельные очереди, которые движутся по принципу "обслуживания в порядке прибытия" (FIFO). Однако если свободен другой доктор, то 90% пациентов высказывают желание обратиться к нему, когда подошла их очередь, а их доктор занят. Распределение времени консультаций обоих докторов имеет следующий вид:

Таблица 14.9. **Распределение времени консультаций — модель обслуживания**

Продолжительность консультаций, мин.	*6*	*8*	*10*	*12*	*14*
Вероятность	0,10	0,20	0,50	0,10	0,10

Для каждого пациента отводится одинаковое время на консультацию независимо от того, какой из докторов его обслуживает. Однако в зависимости от конкретной ситуации можно ввести в модель и два типа распределений времени консультаций отдельно для каждого из врачей.

Используя имитационную модель, оценить входной поток пациентов в часы утреннего приема и ответить на следующие вопросы:

1. Какое число пациентов ожидает в приемной в 9.00 утра?
2. Чему равно среднее время ожидания пациентом приема в очереди?
3. В котором часу каждого из докторов покидает последний пациент?

Решение.

Данная задача включает в себя следующие стохастические переменные:

а) интервалы между последовательными появлениями пациентов, на основе которых рассчитывается время прибытия каждого пациента;

б) доктор, к которому попадает пациент;

в) согласие пациента пойти на прием к другому доктору, если последний свободен;

г) продолжительность консультации, которая, как предполагается, зависит от самого пациента, а не от доктора, к которому он попадает.

Каждому значению переменных поставим в соответствие случайное число.

Таблица 14.10. **Интервалы появления пациентов, мин.**

Количество мин.	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
1	0,05	0,05	00–04
2	0,05	0,10	05–09
3	0,10	0,20	10–19
4	0,20	0,40	20–39
5	0,40	0,80	40–79
6	0,10	0,90	80–89
7	0,05	0,95	90–94
8	0,05	1,00	95–99

Таблица 14.11. **Продолжительность консультации, мин.**

Количество мин.	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
6	0,10	0,10	00–09
8	0,20	0,30	10–29
10	0,50	0,80	30–79
12	0,10	0,90	80–89
14	0,10	1,00	90–99

Таблица 14.12. **Доктор, принимающий пациента**

Доктор	*Вероятность*	*Случайные числа*
A	0,5	0–4
B	0,5	5–9

Таблица 14.13. **Согласие пойти к другому доктору**

	Вероятность	*Случайные числа*
Да	0,9	0–8
Нет	0,1	9

Теперь предварительная работа закончена, и можно начинать непосредственно процесс моделирования. Моделируемый счетчик времени устанавливается на 8.30 утра. Первый пациент приходит в 8.30 + первый интервал появления пациента (IAT).

1. К 9.00 в приемной находятся пять человек.

2. Среднее время ожидания пациентами в очереди составляет:
   У доктора Аббота — 38,9 мин. для 11 пациентов;
   У доктора Буфа — 24,9 мин. для 11 пациентов;
   Итого — 31,9 мин., причем минимальное время ожидания составляет 4 мин., максимальное — 54 мин.
3. Последний пациент уйдет от доктора Аббота в 10 ч 56 мин., а от доктора Буфа — в 10 ч 46 мин.

Прежде чем использовать полученную информацию, докторам следует смоделировать несколько утренних приемов больных и рассчитать средние значения всех статистических характеристик. Начать им следует с того, чтобы задать вопросы о том, как улучшить обслуживание клиентов. Почему приемная открывается именно за 30 мин. до начала приема? Следует ли ввести систему предварительной записи? Нужно ли организовывать отдельную очередь для пациентов, которые хотели бы попасть на прием к любому врачу? Следует провести моделирование каждого из указанных альтернативных вариантов и определить его теоретический эффект прежде, чем ввести его в практику обслуживания.

Таблица 14.14. **Имитационная модель утреннего приема пациентов двумя врачами**

Приход пациента			*Доктор, обслуживающий пациента*		*Пойдете к другому врачу?*		*Консультация*						*Время ожидания в очереди, мин.*
IAT							*Время*		*Доктор Аббот*		*Доктор Буф*		
Случайное число	*Минут*	*Время*	*Случайное число*	*Врач*	*Случайное число*	*Да/Нет*	*Случайное число*	*Минут*	*Начало*	*Окончание*	*Начало*	*Окончание*	
63	5	8:35	5	В	6	Да	69	10			9:00	9:10	25
27	4	8:39	4	А	2	Да	39	10	9:00	9:10			21
15	3	8:42	2	А	0	Да	39	10	9:10	9:20			28
99	8	8:50	2	А	4	Да	27	8			9:10	9:18	20
86	6	8:56	3	А	8	Да	85	12	9:20	9:32			24
71	5	9:01	1	А	3	Да	49	10	9:32	9:42			31
74	5	9:06	3	А	3	Да	90	14	9:42	9:56			36
45	5	9:11	3	А	7	Да	25	8	9:56	10:04			45
11	3	9:14	5	В	1	Да	84	12			9:18	9:30	4
02	1	9:15	7	В	1	Да	47	10			9:30	9:40	15
15	3	9:18	5	В	2	Да	42	10			9:40	9:50	22
14	3	9:21	5	В	5	Да	04	6			9:50	9:56	29
18	3	9:24	3	А	2	Да	83	12	10:04	10:16			40
07	2	9:26	1	А	9	Нет	03	6	10:16	10:22			50
14	3	9:29	7	В	0	Да	78	10			9:56	10:06	27
58	5	9:34	2	А	2	Да	87	12	10:22	10:34			48
68	5	9:39	7	В	1	Да	61	10			10:06	10:16	27
39	4	9:43	3	А	5	Да	82	12	10:34	10:46			51
31	4	9:47	8	В	2	Да	69	10			10:16	10:26	29
08	2	9:49	2	А	6	Да	33	10			10:36	10:46	47
13	3	9:52	4	А	0	Да	40	10	10:46	10:56			54
55	5	9:57	7	В	2	Да	64	10			10:26	10:36	29
Приемная закрыта													

❐ **Пример 14.2.** Компания с ограниченной ответственностью "АМС Tyres" производит продажу и ремонт покрышек к автомобилям в ремонтной мастерской, расположенной в центре города. Приход клиентов носит случайный характер, система предварительной записи отсутствует. Клиентам, которые звонят в мастерскую заранее, отвечают, что они могут прийти в любое удобное для них время. В результате наблюдений за временными интервалами между последовательными моментами прихода клиентов были получены следующие данные:

Таблица 14.15. **Модель интервалов приезда автомобилей в ремонтную мастерскую**

Временные интервалы между прибытием автомобилей, мин	*0*	*5*	*10*	*15*	*20*	*25*	*30*	*35*
Вероятность	0,04	0,08	0,15	0,30	0,20	0,13	0,08	0,02

Время, необходимое для осмотра и замены покрышек, было оценено с точностью до минуты. Оно изменяется в пределах промежутка от 21 до 40 мин, причем появление любого значения равновероятно.

На настоящий момент внутри мастерской компании "АМС" имеются одна оборудованная всем необходимым монтажная площадка, а также место для парковки еще одного автомобиля. Кроме того, вне мастерской есть еще место для парковки только одного автомобиля. Стоянка на близлежащей дороге запрещена, поэтому любой водитель, который подъехал в тот момент, когда заняты как монтажная площадка, так и оба отведенных для парковки места, вынужден будет уехать и по сути является для компании потерянным клиентом. Потеря каждого клиента обходится компании в среднем в 50 ф. ст. Если сделать небольшую реконструкцию, то внутри ремонтной мастерской можно оборудовать вторую монтажную площадку, но при этом место для парковки внутри мастерской придется демонтировать. На самом деле это не представляет особой проблемы, так как в любом случае длина очереди и порядок продвижения клиентов останутся неизменными. Стоимость эксплуатации второй монтажной площадки составляет 35 ф. ст. в час. Построим имитационную модель для ситуации с 25 клиентами. Следует ли "АМС" вводить в эксплуатацию вторую монтажную площадку?

Решение

Переменными в задаче являются:

1. Интервалы времени между последовательными моментами приезда клиентов. Эти данные собрать достаточно трудно, поскольку нам необходима модель прибытия потенциальных клиентов, многие из которых вынуждены уехать в связи с тем, что отведенные для парковки автомобилей места уже заняты.
2. Время обслуживания автомобилей. Оно представляет собой 20 интервалов длиной в одну минуту каждый, от 21 до 40 включительно. Если все они равновероятны, то вероятность появления каждого значения составит: $1/20 = 0{,}05$.

Таблица 14.16. **Распределение интервалов случайных чисел для интервалов приезда клиентов**

IAT, мин	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
0	0,04	0,04	00–3
5	0,08	0,12	04–11
10	0,15	0,27	12–26
15	0,30	0,57	27–56
20	0,20	0,77	57–76
25	0,13	0,90	77–89
30	0,08	0,98	90–97
35	0,02	1,00	98–99

Таблица 14.17. **Распределение интервалов случайных чисел для продолжительности обслуживания**

Время, мин	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
21	0,05	0,05	00–04
22	0,05	0,10	05–09
23	0,05	0,15	10–14
24	0,05	0,20	15–19
25	0,05	0,25	20–24
26	0,05	0,30	25–29
27	0,05	0,35	30–34
28	0,05	0,40	35–39
29	0,05	0,45	40–44
30	0,05	0,50	45–49
31	0,05	0,55	50–54
32	0,05	0,60	55–59
33	0,05	0,65	60–64
34	0,05	0,70	65–69
35	0,05	0,75	70–74
36	0,05	0,80	75–79
37	0,05	0,85	80–84
38	0,05	0,90	85–89
39	0,05	0,95	90–94
40	0,05	1,00	95–99

Процесс моделирования мы начнем с одной монтажной площадкой и максимальной очередью, состоящей из трех автомобилей, установив моделируемый счетчик времени на нулевой отметке. Первый клиент прибудет в момент времени: 0 + первый IAT.

В процессе моделирования потока из 25 клиентов 10 из них уехали из-за отсутствия места для стоянки. Этот факт приводит к потере прибыли, составившей 10 × 50 ф. ст. = 500 ф. ст. Среднее время ожидания окончания обслуживания до того момента, когда получивший услугу клиент покинет мастерскую, составляет 50 мин.

Таблица 14.18. **Имитационная модель сложившейся ситуации по обслуживанию клиентов компанией "AMC Tyres"**

Автомобиль	Модель прибытия клиентов			Размещение автомобилей			Обслуживание, парковка				Ожидание обслуживания, мин
	Случайное число	IAT, мин	Время начиная с 0 мин	Обслуживание	Площадка 1	Площадка 2	Случайное число	Время, мин	Начало	Окончание	
A	44	15	15	A	–	–	95	40	15	55	0
B	22	10	25	A	B	–	23	25	55	80	30
C	78	25	50	A	B	C	10	23	80	103	30
D	84	25	75	B	C	D	76	36	103	139	28
E	26	10	85	C	D	E	30	27	139	166	54
F	04	5	90	C	D	E	F Не может встать в очередь				
G	33	15	105	D	E	G	76	36	166	202	61
H	46	15	120	D	E	G	H Не может встать в очередь				
I	09	5	125	D	E	G	I Не может встать в очередь				
J	52	15	140	E	G	J	28	26	202	228	62
K	68	20	160	E	G	J	K Не может встать в очередь				
L	07	5	165	E	G	J	L Не может встать в очередь				
M	97	30	195	G	J	M	75	36	228	264	33
N	06	5	200	G	J	M	N Не может встать в очередь				
O	57	20	220	J	M	O	39	28	264	292	44
P	16	10	230	M	O	P	30	27	292	319	62
Q	90	30	260	M	O	P	Q Не может встать в очередь				
R	82	25	285	O	P	R	90	39	319	358	34
S	66	20	305	P	R	S	14	23	358	381	53
T	59	20	325	R	S	T	80	37	381	418	56
U	83	25	350	R	S	T	U Не может встать в очередь				
V	62	20	370	S	T	V	52	31	418	449	48
W	64	20	390	T	V	W	91	39	449	488	59
X	11	5	395	T	V	W	X Не может встать в очередь				
Y	12	10	405	T	V	W	Y Не может встать в очередь				

Теперь построим имитационную модель для двух монтажных площадок и одного места парковки. Воспользуемся указанными в табл. 14.18 моментами времени прибытия клиентов. Используемое распределение времени обслуживания останется неизменным, однако, соответствующие случайные числа будут иными. В целях упрощения таблицы их значения опущены.

Легко заметить, что новая модель значительно отличается от построенной ранее. В ней не получают обслуживание только 2 водителя. Значительно сокращается и время обслуживания, а стоянка для парковки часто оказывается пустой. В целом модель отражает большее удовлетворение клиентов.

Что касается затрат, то следует сопоставить значения стоимости, связанной с потерей клиентов, и стоимости эксплуатации второй монтажной площадки. Для потока из 25 клиентов время ее работы составило 461 мин., или 7,68 ч. Общая стоимость эксплуатации второй площадки, соответствующая потоку из 25 клиентов, составила:

$$7{,}68 \times 35 + 2 \times 50 = 368{,}80 \text{ ф. ст.}$$

В данной ситуации по сравнению с задачей для одной монтажной площадки достигается экономия в 131,20 ф. ст. Однако нельзя забывать о том, что процесс моделирования следует продолжить и рассчитать экономию, приходящую на 1 ч работы второй площадки, более точно, прежде чем принять какое-либо решение.

Кроме того, что имитационная модель позволяет получить информацию о возможной сумме экономии, она воспроизводит процесс обслуживания клиентов. Воспользовавшись моделью, можно вычислить среднее время ожидания клиентами окончания обслуживания и время простоя монтажных площадок.

Таблица 14.19. **Имитационная модель предполагаемой системы обслуживания клиентов компанией "AMC Tyres"**

Автомобиль	*Прибытие клиентов, начиная с 0, мин.*	*Размещение автомобилей*			*Обслуживание*					*Ожидание обслуживания, мин.*
		Площадка 1	*Площадка 2*	*Площадка*	*Время, мин.*	*1 Парковка*		*2 Парковка*		
						Начало	*Окончание*	*Начало*	*Окончание*	
A	15	A	–	–	40	15				0
B	25	A	B	–	38			25	63	0
C	50	A	B	C	23	55	78			5
D	75	C	D	–	27			75	102	0
E	85	E	D	–	40	85	125			0
F	90	E	D	F	23			102	125	12
G	105	E	F	G	36	125	161			20
H	120	E	F	G		H	Не может встать в очередь			
I	125	G	I	–	37			125	162	0
J	140	G	I	J	26	161	187			21
K	160	G	I	J		K	Не может встать в очередь			
L	165	J	L	–	29			165	194	0
M	195	M	–	–	36	195	231			0
N	200	M	N	–	24			200	224	0
O	220	M	N	O	37			224	261	4
P	230	M	O	P	27	231	258			1
Q	260	Q	O	–	37	260	297			0
R	285	Q	R	–	39			285	324	0
S	305	S	R	–	23	305	328			0
T	325	S	T	–	36			325	361	0
U	350	U	T	–	32	350	382			0
V	370	U	V	–	35			370	405	0
W	390	W	V	–	39	390	429			0
X	395	W	V	X	21			405	426	10
Y	405	W	X	Y	35			426	461	21

14.4. ПРИМЕНЕНИЕ ИМИТАЦИОННЫХ МОДЕЛЕЙ В УПРАВЛЕНИИ ЗАПАСАМИ

Алгоритмы, аналогичные описанным выше, могут быть использованы в имитационном моделировании сложных проблем, возникающих в управлении запасами. Они позволяют учитывать неопределенность как спроса, так и срока поставки заказа. В таких случаях должна быть собрана информация, на основе которой можно построить распределения вероятностей для соответствующих переменных.

❐ **Пример 14.3.** Корпорация "ELA" занимается производством легковых автомобилей. Аккумуляторы для модели "Lunar" компания закупает на стороне, у внешнего поставщика. На основе прошлого опыта специалисты "ELA" оценили, что спрос на аккумуляторы за неделю можно аппроксимировать нормальным распределением со средним значением 500 и стандартным отклонением 10 для промежутка от 470 до 530.

Начальный запас аккумуляторов составляет 2000 шт., причем администрация компании приняла решение о подачах заказов на партии аккумуляторов размером в 2500 шт. каждый раз, когда их запас опускается ниже уровня в 1500 шт. Кроме того, прошлый опыт показывает, что интервалы времени между подачей заказа и осуществлением поставок изменяются следующим образом:

Таблица 14.20. **Распределение времени поставки заказа, корпорации "ELA"**

Время поставки заказа, недель	1	2	3	4
Вероятность	0,20	0,50	0,25	0,05

Единичная стоимость хранения запасов равна 0,50 ф. ст. в неделю и рассчитывается для общего размера запаса, оставшегося на конец недели. Стоимость заказа — 50 ф. ст., а отсутствие аккумуляторов на складе оценивается в 20 ф. ст. в неделю.

Используя имитационную модель для периода в 20 недель, оценим среднюю стоимость проведения изложенной выше политики в неделю. Принимается предпосылка о том, что все расчеты производятся в конце недели, а подачи заказов и поставки по ним — в начале недели.

Решение.

Переменными являются спрос и время поставки заказа. Так как спрос аппроксимируется непрерывным нормальным распределением, будем моделировать переменную спроса с шагом в 5 аккумуляторов. Например, вероятность спроса, равного 510 аккумуляторам, будет оцениваться с помощью соотношения $P(507{,}5 < \text{спрос} < 512{,}5)$.

Среднее значение спроса — 10,050/20 = 502,5 аккумуляторов в неделю.

Средний размер запаса на конец недели — 22,865/20 = 1143,25 аккумуляторов в неделю.

Средний размер дефицита — 1020/20 = 51,0 аккумуляторов в неделю.

Число заказов, поданных в течение 20 недель, равно, 4, следовательно, среднее число заказов в неделю — 4/20 = 0,2.

Ожидаемая стоимость в неделю = 1143,25 × 0,50 + 51 × 20 + 0,2 × 50 = 1602 ф. ст.

Как и в предыдущих примерах, процесс моделирования следует продолжить, чтобы убедиться, что достигнутые условия действительно характеризуют стационарное состояние модели.

Имитационные модели можно также применять при исследовании поведения системы управления запасами в условиях альтернативных вариантов политики подачи заказов. Это позволит администрации выбрать тот вариант, который наилучшим образом отвечает поставленным целям.

Таблица **14.21. Распределение интервалов случайных чисел для времени поставки заказа**

Время поставки, недель	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
1	0,20	0,20	00–19
2	0,50	0,70	20–69
3	0,25	0,95	70–94
4	0,05	1,00	95–99

Таблица **14.22. Распределение интервалов случайных чисел для спроса за неделю**

Спрос за неделю	*Вероятность*	*Кумулятивная вероятность*	*Случайные числа*
470	0,003	0,003	00–002
475	0,009	0,012	003–011
480	0,028	0,040	012–039
485	0,066	0,106	040–105
490	0,121	0,227	106–226
495	0,175	0,402	227–401
500	0,197	0,599	402–598
505	0,175	0,774	599–773
510	0,121	0,895	774–894
515	0,066	0,961	895–960
520	0,028	0,989	961–988
525	0,009	0,998	989–997
530	* 0,003	1,000	998–999*

* Небольшие отклонения ввиду ошибок округления

Теперь можно осуществить моделирование.

Таблица 14.23. **Моделирование управления запасами**

Неделя	*Запас на начало недели*	*Спрос*		*Запас на конец недели*	*Повторный заказ, Да/нет*	*Время поставки*		*Дефицит*
		Случайное число	*Объем*			*Случайное число*	*Недели*	
1	2000	034	480	1520				
2	1520	743	505	1015				
3	1015	738	505	510	Да	95	4	
4	510	636	505	5				
5	5	964	520	0				515
6	0	736	505	0				505
7	2500	614	505	1995				
8	1995	698	505	1490				
9	1490	637	505	985	Да	73	3	
10	985	162	490	495				
11	495	332	495	0				
12	2500	616	505	1995				
13	1995	804	510	1485				
14	1485	560	500	985	Да	10	1	
15	3485	111	490	2995				
16	2995	410	500	2495				
17	2595	959	515	1980				
18	1980	774	510	1470				
19	1470	246	495	975	Да	76	3	
20	975	762	505	470				
		Итого	10050	22865				1020

РЕЗЮМЕ

Имитационное моделирование является одним из методов, который применяется специалистами в случаях, когда использование математических моделей вызывает определенные трудности или когда лежащие в их основе предпосылки неадекватны реальным условиям. Метод имитационного моделирования можно применять в сложных ситуациях, не принимая никаких предпосылок об исходных данных.

Мы рассмотрели метод Монте-Карло, в котором всем переменным модели ставится в соответствие определенное множество дискретных значений. Данный метод позволяет на основе собранной исходной информации сгенерировать для каждой переменной соответствующее распределение вероятностей. Из этих распределений с помощью случайных чисел получают значения переменных модели, которые используют затем в процессе моделирования. Построение каждой модели начинают с определения входящих в нее переменных и формулирования правил их функционирования. Результаты расчетов по имитационным моделям небольшой размерности обычно представляют в виде таблиц, легко поддающихся количественному анализу.

Существует возможность модификации имитационной модели, по которой вновь производятся расчеты, а затем проводится сравнительный анализ новых результатов с полученными ранее. Методы имитационного моделирования, хотя и не приводят к получению оптимальных решений, как, например, методы линейного программирования, однако, позволяют выработать направления политики, приводящей к лучшим результатам. Но прежде, чем внедрять какой-либо из результатов, полученных по имитационной модели, в практику, необходимо произвести оценку ее надежности и, осуществив расчеты на более длительный период, получить репрезентативные характеристики. Обычно расчеты по имитационным моделям проводятся с помощью пакетов прикладных программ.

УПРАЖНЕНИЯ

Упражнение 14.1

Городская администрация Баттербай Сити контролирует услуги микроавтобусов, которые развозят туристов и покупателей с автобусов и железнодорожного вокзалов в различные районы города. О потоке пассажиров, прибывающих на автобусную остановку, находящуюся около железнодорожного вокзала, были собраны следующие данные:

Время между моментами прибытия пассажиров, мин	0	1	2	3	4	5	6
Вероятность	0,04	0,16	0,24	0,28	0,16	0,10	0,02

По расписанию микроавтобусы должны прибывать каждые 10 мин, однако изменчивость транспортных условий приводит к следующему распределению их прибытия:

Интервал между последовательными прибытиями автобусов, мин	8	10	12	14	16
Вероятность	0,10	0,38	0,28	0,15	0,09

Число мест в автобусе определяется следующим распределением:

Число свободных мест	0	1	2	3	4	5	6
Вероятность	0,06	0,18	0,27	0,34	0,11	0,03	0,01

Требуется:

1. Построить имитационную модель потока из 30 пассажиров, прибывающих на автобусную остановку, в предположении, что моделируемый счетчик времени установлен на нулевой отметке.
2. Оценить среднее время ожидания автобуса пассажиром и среднюю длину очереди.

Упражнение 14.2

Решите пример **14.3** (раздел 14.4) в предположении, что подача повторного заказа производится, если уровень запаса аккумуляторов опускается ниже 1000 шт. Сравните полученные вами результаты с результатами примера **14.3.**

Упражнение 14.3

1. Опишите преимущества и недостатки применения имитационных моделей в исследовании систем массового обслуживания по сравнению с методами теории массового обслуживания.
2. По данным проведенных наблюдений было установлено, что интервалы между поступлениями жалоб от покупателей в крупном универмаге подчиняются следующему распределению:

Время между поступлениями жалоб, мин.	0–4	4–8	8–12	12–16
Вероятность	0,25	0,45	0,20	0,10

Жалобы покупателей рассматриваются одним служащим универмага, однако все покупатели, которые сочли свои претензии "серьезными" или которым пришлось ждать 5 мин и более, прежде чем быть принятыми служащим, который рассматривает жалобы, требуют переговоров с главным администратором, который принимает их жалобы самостоятельно. Продолжительность рассмотрения жалоб клиентов соответствующим служащим аппроксимируется нормальным распределением со средним значением 7 мин и стандартным отклонением, равным 2 мин. Было оценено, что 20% покупателей, имеющих жалобы, считают свои претензии "серьезными".

Требуется:

а) Пользуясь приведенными выше данными, таблицей случайных чисел и нижеследующей таблицей, полученной из таблицы кумулятивного стандартного нормального распределения, описать этапы построения имитационной модели потоков прибытия и обслуживания клиентов в условиях данной системы.

Случайное число	*Число отклонений от среднего значения*	*Случайное число*	*Число отклонений от среднего*
00–01	−2,5	61–77	+0,5
02–04	−2,0	78–88	+1,0
05–10	−1,5	89–94	+1,5
11–21	−1,0	95–97	+2,0
22–38	−0,5	98–99	+2,5
39–60	0		

б) Используя приведенные ниже случайные числа, построить имитационную модель обслуживания 10 имеющих жалобы покупателей, если претензии некоторых из них являются серьезными.

Интервалы между прибытием покупателей, мин	09	06	51	62	83	61	59	20	82	68
Серьезные претензии, кол-во	5	0	7	3	8	2	9	8	1	6
Продолжительность обслуживания	39	60	50	31	02	02	83	90	71	16

в) Используя построенную имитационную модель, оценить долю покупателей, которые действительно встретятся с главным администратором универмага. Затем оценить общее время, которое главный администратор затратит на рассмотрение жалоб в предположении, что у него 8-часовой рабочий день, а время, которое он затрачивает на рассмотрение одной жалобы, подчиняется соответствующему распределению для служащего универмага, рассматривающего жалобы клиентов.

г) Кратко поясните, каким образом можно использовать имитационную модель для принятия решения о целесообразности или нецелесообразности принятия на работу еще одного служащего, рассматривающего жалобы покупателей.

(АССА, июнь 1988 г.).

Упражнение 14.4

Корпорация "Pomulus Products" производит операции со всеми клиентами в 30-дневный срок. Как показывает опыт, 80% всех счетов закрывается в течение одного месяца, а 70% оставшихся счетов — в течение следующего месяца после того, как клиенту посылается стандартное письмо-напоминание о просроченных счетах. Половина счетов остается неоплаченной в течение двух месяцев. После того, как было отправлено письмо, содержащее "последнее предупреждение", выплаты производятся в течение третьего месяца.

Со всеми счетами, которые оказались неоплаченными в течение трех месяцев, поступают одним из двух возможных способов. Если сумма на счету превышает 1000 ф. ст., то для того, чтобы вернуть деньги, компания прибегает к законодательным процедурам. Принимая во внимание связанные с судебными процедурами издержки, доля исходной суммы, которую в конечном счете удается вернуть, варьирует следующим образом:

Доля исходной денежной суммы, которую удается вернуть, %	0–40	40–60	60–80	80–100
Вероятность	0,1	0,3	0,4	0,2

Прежде, чем окончательные выплаты будут произведены, проходит, как правило, не менее трех месяцев.

Если денежная сумма на счету составляет менее 1000 ф. ст., то компания продает неоплаченные счета другой компании, специализирующейся на скупке долговых обязательств, получая за это 50% исходной суммы, выплаты которой производятся в течение одного месяца, т.е. к концу четвертого месяца.

Суммы счетов, выписанных компаний за последние месяцы, распределяются следующим образом:

Сумма на счету, ф. ст.	0–200	200–500	500–1000	1000–2000	2000–5000
Вероятность	0,1	0,2	0,3	0,3	0,1

Примем предпосылки о том, что связь между суммой счета в момент его оплаты и долей общей суммы, которую удается возвратить, отсутствует, а также о том, что оплата всех счетов производится в последний день каждого месяца. Темпы роста основного капитала компании равны 1,5% в месяц.

Требуется:

1. Определить вероятность того, что выплаты по любому счету будут произведены в конце:
   а) второго месяца;
   б) третьего месяца;
   в) четвертого месяца;
   г) шестого месяца.
2. Определить, какова ожидаемая стоимость в настоящий момент нового счета, размер неоплаченной суммы которого составляет 2000 ф. ст.?

Значения ежемесячных коэффициентов дисконтирования:

Месяц	1	2	3	4	5	6
Коэффициент	0,9852	0,9707	0,9563	0,9422	0,9283	0,9145

3. Показать, используя указанные ниже случайные числа, каким образом можно построить имитационную модель системы в целом и определить текущее значение стоимости двух моделируемых счетов.

Счет 1	8	8	7	5	7
Счет 2	9	9	8	2	9

(АССА, декабрь 1986 г.).

Упражнение 14.5

Администрация корпорации "Hentor Products plc" рассматривает вопрос о покупке новой упаковочной машины с автоматическим управлением. Новая машина позволит ликвидировать две старые, которые в настоящее время используются для упаковки продукта X. Ввиду более высокой автоматизации технологического процесса новая машина также позволит сократить затраты на оплату труда, а ввиду ее более высокой производительности — увеличить объемы производства. В связи со значительным ростом спроса на продукт X, по оценкам специалистов, установка новой машины приведет к увеличению прибыли за каждый из ближайших трех лет. Однако ввиду неопределенности спроса ежегодный приток капитала (включая сумму экономии), вызванный покупкой новой машины, нельзя оценить точно, поэтому при его расчете за каждый год использовались следующие вероятностные оценки:

Ежегодный приток капитала, тыс. ф. ст.					
1-й год	*Вероятность*	*2-й год*	*Вероятность*	*3-й год*	*Вероятность*
10	0,3	10	0,1	10	0,3
15	0,4	20	0,2	20	0,5
20	0,3	30	0,4	30	0,2
		40	0,3		

Поскольку в объемах продаж продукта X также существует неопределенность, то для решения вопроса о целесообразности или нецелесообразности покупки новой упаковочной машины необходимо учитывать величину притока капитала только за ближайшие три года. Чистая стоимость новой машины за вычетом ликвидационной стоимости старого оборудования составит 42000 ф. ст. Воздействие налоговой политики можно не учитывать.

Требуется:

1. Не принимая во внимание изменение стоимости во времени, определить, какая из комбинаций ежегодных притоков капитала приведет к отрицательному значению общего притока капитала, а также вероятность появления этого события.
2. Рассчитать на основе средних значений притока капитала за каждый год текущее значение чистой стоимости новой машины, если стоимость капитала для компании равна 15%. Соответствующие коэффициенты дисконтирования имеют следующие значения:

Год	*Значение коэффициента дисконтирования*
1	0,8696
2	0,7561
3	0,6575

3. Построив имитационную модель расчета текущего значения чистой стоимости, оценить значение риска, соответствующее данной ситуации. При моделировании 5 множеств притоков капитала используйте приведенные значения случайных чисел. На основе полученных по имитационной модели результатов рассчитать

ожидаемое значение чистой стоимости на настоящий момент и вероятность того, что внедрение новой машины приведет к получению отрицательного значения чистой стоимости на настоящий момент.

	Множество 1	*Множество 2*	*Множество 3*	*Множество 4*	*Множество 5*
1-й год	4	7	6	5	0
2-й год	2	4	8	0	1
3-й год	7	9	4	0	1

(АССА, декабрь 1985 г.).

Упражнение 14.6

Мойщик машин, который в настоящее время работает на себя, установил свое оборудование на автомобильной стоянке около оживленной транспортной магистрали. На автомобильной стоянке могут находиться одновременно максимум два автомобиля, включая автомобиль, обслуживание которого производится в данный момент, причем в соответствии с местными правилами движения стоянка на дороге запрещена. Следовательно, любой потенциальный клиент, подъезжающий к стоянке в тот момент, когда все места парковки заняты, является потерянным для мойщика машин. В нижеследующей таблице приведены распределение интервалов прибытия потенциальных клиентов и распределение времени их обслуживания.

Время, мин.	*Интервалы между прибытиями, %*	*Длительность обслуживания, %*
0–2	15	10
2–4	50	25
4–6	20	30
6–8	5	25
8–10	5	10
10–12	5	0
Итого	100	100

Требуется:

а) Оценить среднее число потенциальных клиентов в час.

б) Используя приведенные ниже случайные числа, построить имитационную модель работы мойщика машин по обслуживанию 10 потенциальных клиентов и таким образом оценить число потерянных клиентов за 1 ч, пояснив при этом, каким образом генерируются интервалы между прибытиями клиентов и длительность обслуживания.

Номер клиента	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Случайные числа:										
для интервалов прибытия	87	69	19	21	62	07	31	30	60	88
для длительности обслуживания	90	67	83	32	65	18	74	8	18	38

в) Почему оценки, найденные в б), недостаточно надежны, и каковы пути улучшения их обоснованности?

г) Мойщик машин рассматривает вопрос о принятии на работу ассистента. В предположении, что ассистент будет обслуживать клиентов в соответствии с распределением времени обслуживания, приведенным выше для мойщика машин, постройте имитационную модель работы обслуживающего персонала из двух человек по оказанию услуг 10 потенциальным клиентам, используя приведенные выше случайные числа.

Оцените число потерянных клиентов в час для новой системы обслуживания.

Объясните, почему при сравнении двух имитационных моделей в расчетах по ним лучше использовать одно и то же множество случайных чисел.

д) Если математическое ожидание прибыли на одного клиента равно 50 пенсам, рассчитать размер заработной платы ассистента на основе результатов имитационного моделирования, проведенного ранее.

(АССА, июнь 1990 г.).

ОТВЕТЫ К УПРАЖНЕНИЯМ

Глава 1

Упражнение 1.1 (i) 1/16; (ii) 3/8.

Упражнение 1.2 (i) 0; (ii) 4/36.

Упражнение 1.3 (i) 21/74; (ii) 13/74; (iii) 5/74; (iv) 7/74; (v) 41/74; (vi) 0,0777; (vii) 0,1203; (viii) 0,0555; (ix) 0,056; (x) 0,09996.

Упражнение 1.4 (i) 2/15; (ii) 8.15; (iii) 0,64; (iv) 0,72.

Упражнение 1.5 0,6532.

Упражнение 1.6 (i) 0,9802; (ii) 0,0196; (iii) 0,9851 в (i), 0,0148 в (ii).

Упражнение 1.7 (i) 0,7952; (ii) 0,9853; (iii) 0,99965.

Упражнение 1.8

Время, годы	*1*	*2*	*3*	*4*	*5*
(i) Вероятность	0,02	0,13	0,25	0,40	0,20
(ii) G: 3,3 года	H: 2,5 года		Система: 3,63 года		

Упражнение 1.9. (i) 0,1327; (ii) 0,2912.

Упражнение 1.10 Вероятность высоких объемов продаж 0,273;
Вероятность низких объемов продаж 0,72.

Упражнение 1.11 156 ф. ст. в год.

Упражнение 1.12 (i) 13,800; (ii) 7,315; (iii) 100 947 000.

Упражнение 1.13 (i) 30.240 (ii) 113.400.

Упражнение 1.14 (i) 5/9; (ii) 0,4109.

Упражнение 1.15 (i) 1200 ф. ст.; (ii) 600 ф. ст.

Упражнение 1.16 1. а) 0,252; б) 0,208; в) 0,150;
2. 1500 ф. ст. 3. 0,2642. 5. а) 56 лет; б) 1266 ф. ст.

Упражнение 1.17 1. а) 1/120; б) 0,951; в) 0,4.
2. 0,083 для 1 а), 0,573 для 1 б), 2/3 для 1 в).
3. 0,744.

Глава 2

Упражнение 2.1 Ситуация 1.

Упражнение 2.2 1. 0,1139. 2. 0,1611.

Упражнение 2.3 0,90112.

Упражнение 2.4 0,18782.

Упражнение 2.6 2. 0,1847.

Упражнение 2.7 Приблизительно 39 раз в неделю.

Упражнение 2.8 1. Среднее значение = 2,15. Стандартное отклонение = 1,558.

3. *Число доступов к устаревшим файлам*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*>7*
Ожидаемая частота	11,65	25,04	26,92	19,29	10,37	4,46	1,60	0,49	0,18

Упражнение 2.9 8 упаковок.

Упражнение 2.10 0,8088.

Упражнение 2.11 1. 6,68%.
2. Переналадка станка на средний вес упаковки в 940 г.

Упражнение 2.12 12,1% 40 ч.

Упражнение 2.13 58,6%.

Упражнение 2.14 1. 4,01%. 2. 0,135%. 3. 1024,6 г либо 1025 г.

Упражнение 2.15 15570 ф. ст.

Упражнение 2.16 99,65%.

Упражнение 2.17 9,5%.

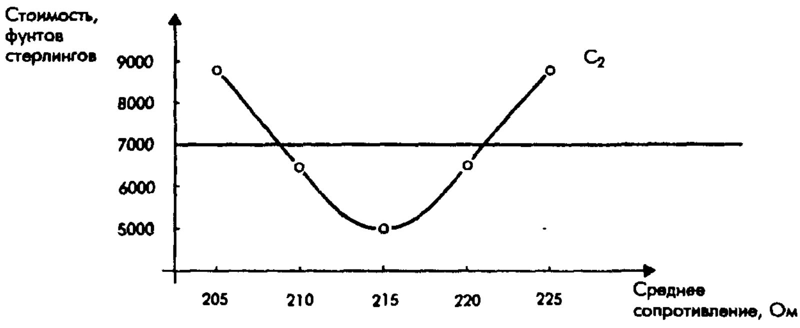
Упражнение 2.18 1. а) 0,06;
б) Предполагается, что из 500 упаковок 367 будут иметь 0 дефектов, 117 — 1 дефект и 15 — 2 дефекта.

Упражнение 2.19 а) (i) 10,6%; (ii) 427,8 ч; (iii) 0,059. б) 0,8202. в) 0,5665.

Упражнение 2.20 1. 0,2963. 2. 0,7771; 5 ф. ст.

Упражнение 2.21 1. Вероятность использования 10 машин = 0,0183;
Вероятность использования 11 машин = 0,0733;
Вероятность использования 12 машин = 0,1465;
Вероятность использования 13 машин = 0,1954;
Вероятность использования 14 машин = 1954;
Вероятность использования 15 машин = 0,3711;
Математическое ожидание ежедневно арендуемых машин = 13,6.
2. 154,62 ф. ст. в день. 3. от 3,86 ф. ст. до 6,69 ф. ст.

Упражнение 2.22 1. 1045; проверка всего запаса.



3. Замена всего запаса; рекомендации не изменятся.

Глава 3

Упражнение 3.1 1) 18000; 2) 10000; 3) 12000; 4) 14000; 5) 14000. Стоимость точной информации (Value of perfeet information) = 260 ф. ст. в день.

Упражнение 3.2 1. 1) 2400; 2) 1600; 3) 2000; 4) 2000; 5) 2000. Стоимость точной информации = 54 ф. с. за период. 2. 1) 2400 2) 1600 3) 2200 4) 2000 5) 2000. 3. Решение не изменится.

Упражнение 3.3 1. а) 7; б) 4; в) 6; г) 6; д) 6. Стоимость точной информации = 36,8 ф. ст. в неделю. 2. Все решения останутся прежними.

Упражнение 3.4 1. а) 5000; б) 2000; в) 4000; г) 4000; д) 4000. Стоимость точной информации = 4800 ф. ст. за 3 года.

Упражнение 3.5 2. Директор по финансам: 4000 книг. Директор по маркетингу: 5000 книг.

Упражнение 3.6 2) (i) 95,25%, (ii) 90,82%; 3) ожидаемая прибыль: при использовании существующих мощностей 25000 ф. ст., при введении специального станка 26000 ф. ст.; 4) 2112,5 ф. ст.

Упражнение 3.7 2. 15 т в месяц. 3. μ = 14 т, σ = 3 т, Q = 13 т в месяц.

Упражнение 3.8 1. б) А; в) 0,367. 2. Инвестиции в А; в конце 1 года, если норма отдачи составит 8%, — переход к инвестициям в В; в иных случаях — продолжение инвестиций в А.

Упражнение 3.9 б) Указание: для построенного дерева решений рассчитать чистую общую отдачу от инвестиций; в) 4%.

Глава 4

Упражнение 4.5

Выборочное распределение выборочного среднего	
$\bar{x}_4$	*Частоты*
10	1
11	1
12	2
13	2
14	3
15	2
16	2
17	1
18	1
	Итого 15

Выборочное распределение стандартного отклонения	
S	*Частоты*
4,47	3
5,92	4
6,32	2
7,21	1
7,48	2
7,68	2
8,25	1
	Итого 15

Глава 5

Упражнение 5.1 От 98,24 г до 100,77 г.

Упражнение 5.2 От 831,6 г до 848,4 г.

Упражнение 5.3 1. Вероятность того, что генеральное среднее отклоняется от выборочного среднего, равного 2920 г, не более чем на 93 г, равна 95%. Следовательно, вариация составит ± 3,2%. 2. Увеличить размер выборки до 250.

Упражнение 5.4 От 0,57 до 0,63.

Упражнение 5.5 1. От 0,19 до 0,31. 2. 451.

Упражнение 5.6 1. От 29,3% до 50,7%. 2. От 397 ф. ст. до 503 ф. ст. 3. $n \geq 369$.

Упражнение 5.7 1. $\hat{\mu} = 86{,}0$ мин; $\hat{\sigma} = 12{,}544$ мин. 2. $\hat{\mu} = 84{,}5$ мин; $\hat{\sigma} = 14{,}023$ мин; от 80,0 мин до 89,0 мин. 3. 84. 4. 19875 ф. ст.

Упражнение 5.8 1. 120000 ф. ст. 2. (120000 ± 23870) ф. ст. 3. 336.

Глава 6

Упражнение 6.1[1] 2,5.

Упражнение 6.2 −2,85.

Упражнение 6.3 2. −1,695.

Упражнение 6.4 −1,0.

[1] В данном и последующих упражнениях главы 6 на проверку гипотез в качестве ответа указаны фактические значения соответствующих критериев.

Упражнение 6.5 4,17. Указание: предварительно используйте критерий проверки равенства дисперсий.

Упражнение 6.6 – 1,59. Указание: предварительно используйте критерий проверки равенства дисперсий.

Упражнение 6.7 – 2,38.

Упражнение 6.8 2,45.

Упражнение 6.9 5,2.

Упражнение 6.10 **1**. 3,3. **2**. 0,0401. **3**. 0,1522.

Упражнение 6.11 2,60.

Упражнение 6.12 **1**. $\bar{n} = 7$, $S = 4{,}50$. **2**. $7 \pm 0{,}723$. **3**. – 2,03. Указание: введите предпосылку о распределении спроса.

Упражнение 6.13 **1**. 1,24 (без поправки на непрерывность), 1,03 (с поправкой на непрерывность). **2**. $n \geq 10$ (если поправка не непрерывность не используется), $n \geq 11$ (если поправка на непрерывность используется).

Упражнение 6.14 **2**. $t = 1{,}94$.

Упражнение 6.15 **1**. 0,56. **2**. 4,05. **3**. – 0,972. **4**. Апекс: $p < 1{,}5$ пенсов за 1 л, Бемакс: $p < 2{,}24$ пенса за 1 л.

Глава 7

Упражнение 7.1 **2**. от 24,94 до 25,06.

Упражнение 7.2 **3**. 60,35 мм; 2,275%.

Упражнение 7.3

Вид продукта	*Верхний предел, см*	*Нижний предел, см*
Ножи (станок С)	21,26	21,24
Вилки (станок А)	20,01	19,99
Ложки (станок В)	21,215	21,185

Упражнение 7.4

1. Параметры контрольной карты среднего:

Верхняя предупреждающая граница	433,94 г
Нижняя предупреждающая граница	416,06 г
Верхняя граница регулирования	438,42 г
Нижняя граница регулирования	411,58 г

   Параметры контрольной карты размахов:

Граница центра	23,26 г
Верхняя предупреждающая граница	42,00 г
Верхняя граница регулирования	54,50 г

2. Указание: нанесите рассчитанные по данным выборки параметры на соответствующие контрольные карты, построенные в 1.

Упражнение 7.5

1. Параметры контрольной карты среднего:

Верхняя предупреждающая граница	6,72 см
Нижняя предупреждающая граница	6,58 см
Верхняя граница регулирования	6,755 см
Нижняя граница регулирования	6,545 см

Параметры контрольной карты размахов:

Граница центра	0,144 см
Верхняя предупреждающая граница	0,279 см
Верхняя граница регулирования	0,371 см

3. 4,4%
4. Параметры контрольной карты среднего:

Верхняя граница регулирования	6,516 см
Нижняя граница регулирования	6,784 см

Параметры контрольной карты размахов:

Верхняя граница регулирования	0,417 см
Нижняя граница регулирования	0

Упражнение 7.6

1.

Параметры контрольной карты	*Аппроксимация*	
Схема	*Нормальным распределением*	*Распределением Пуассона*
Граница центра	0,03	0,03
Верхняя предупреждающая граница	0,058	0,057
Нижняя предупреждающая граница	0,002	0
Верхняя граница регулирования	0,072	0,077
Нижняя граница регулирования	0	0

2. Указание: по каждой из 10 выборок рассчитайте долю дефектов и нанесите их значения на соответствующую карту.

Упражнение 7.7 1. Обе схемы 2. В. 3. В. 4. А. 5. В.

6.

Схема	*Риск производителя, %*	*Риск потребителя, %*
А	9	37,6
В	8	20,6
С	14	2,8

7. 40. 8.970.

Упражнение 7.8 Риск производителя 4,8%; риск потребителя 12,2%

Упражнение 7.9

2.

Схема	*Риск производителя, %*	*Риск потребителя, %*
А	0,08	0,16
В	0,08	0,03

Глава 8

Упражнение 8.1 а) $r = -0{,}977$ $t = 11{,}2$; б) $Q = 56{,}5 - 0{,}887$ p, где Q — количество произведенной продукции (тыс.шт. в день), p — цена единицы продукции.; в) 16600 штук в день.

Упражнение 8.2 1. $r = 0{,}892$ $t = 5{,}23$; 2. $S = -191 + 0{,}138\ I$, где S — сбережения (ф. ст. в год), I — доход (тыс. ф. ст. в год).

Упражнение 8.3 1. Для линейной модели: $r = 0{,}82$ $t = 2{,}87$ $NG = 450 + 0{,}10 \times A$, где NG — число посетителей, A — затраты на рекламу (ф. ст.).

Упражнение 8.4 1. $r = 0{,}98t = 13{,}9C = 1{,}18 + 0{,}157Q$, где C — издержки (тыс. ф. ст. в день), Q — объем выпуска (тыс. штук в день).

Упражнение 8.5 1. $r = 0{,}997P = -14{,}7 + 0{,}338T$, где P — прибыль (тыс. ф. ст. в год), T — товарооборот (тыс. ф. ст. в год). 2. Слияние двух любых магазинов позволит увеличить прибыль на 14700 ф. ст. в год.

Упражнение 8.6 $r_s = 0{,}882$ $Z = 3{,}30$.

Упражнение 8.7 1. (i) 840 ф. ст. (ii) ($4800 \pm 760{,}2$) миль. 2. (i) $y = 16{,}91 + 1{,}11\ x$, (ii) 879 ф. ст. 3. Указание: постройте диаграмму рассеяния, рассчитайте коэффициент корреляции ($r = 0{,}882$, $t = 5{,}29$) и постройте доверительный интервал для среднего значения транспортных расходов в квартале 2 (для 95%-ного уровня значимости он составит (879 ± 63) ф. ст.).

Упражнение 8.8 3. 9,36.

Упражнение 8.9 2. Коэффициент корреляции между продолжительностью одного рейса и числом перевозимых за один рейс коробок 0,916; коэффициент корреляции между продолжительностью одного рейса и числом потребителей, охваченным за один рейс 0,679. 3. Указание: примените частный F-критерий для определения значимости введения в модель второго фактора ($F = 21{,}0$).

Упражнение 8.10 3. $\pm$ 2,27 ф. ст. за 1 т.

Упражнение 8.11

1.

Год	*Приведенные объемы продаж, x, тыс. ф. ст.*	*Количество проданных полисов, y*
1982	3	4
1983	5	3
1984	7	6
1985	10	12
1986	15	17
1987	18	22
1988	26	20

2. 0,837. 3. $y = 1{,}44 + 0{,}88\ x$. 4. 2784.

Упражнение 8.12 2. $y = 748 \times 1{,}053^x$;

Упражнение 12.6 Цех изготовления смесей (оптимальный план: производство 42,857 т "Crunchy" и 142,857 т "Chewy" в месяц; максимальный доход 17049 ф. ст. в месяц).

Упражнение 12.7 2. а) Стандартный напиток 1466 2/3, "Здоровье" – 1800 кг и диетический напиток – 1200 кг; б) 3144,33 ф. ст.; в) сахар – 110 кг; молоко – 396 2/3 кг; Спрос – объем выпуска стандартного напитка на 533 1/3 кг ниже спроса. 3. а) Стандартный напиток – 883,57 кг, "Здоровье" – 2500 кг, Диетический напиток – 1200 кг; б) увеличение на 155,87 ф. ст.; 110 кг.

Упражнение 12.9 2. Перламутровые румяна – 25000 л, жидкие румяна 30000 л и проявляющиеся румяна 25000 л; 23000 ф. ст.; 3. От 110 ф. ст. до 133 ф. ст. за 100 л; 4. 230 ч; 95 ф. ст.

Упражнение 12.10 2. 1090000 ф. ст.

Упражнение 12.11 б) Облигации государственного займа – 200000 ф. ст., облигации корпораций – 50000 ф. ст., обычные акции отраслей сферы обслуживания – 125000 ф. ст.; обычные акции – 125000 ф. ст. ; количество денег, оставшихся в банке 50000 ф. ст.; ежегодный доход – 9,09%.

Упражнение 12.12 1. а) 3357 единиц A, 2321 единиц E, 0 единиц B, 0 единиц C и 0 единиц D; максимальный доход 105791 ф. ст.; 3. Нецелесообразно (чистые убытки – 1,03 ф. ст.).

Упражнение 12.13 б) Матовый лак – 1500 галлонов, Глянцевый лак – 1250 галлонов, 13500 ф. ст. в день; в) 100 чел.-ч.; г) От 1,60 ф. ст. до 8,00 ф. ст.

Глава 13

Упражнение 13.1

Склад	*Магазин*			
	G	*H*	*I*	*J*
1	50			50
2		100	75	25

Минимальная стоимость транспортировки 1175 ф. ст.

Упражнение 13.2 19350 ф. ст.

Упражнение 13.3

Пекарня	*Магазин*			
	I	*II*	*III*	*IV*
X			350	350
H				650
Z	400	400		
Фиктивный		100		

Минимальная стоимость транспортировки 30,25 ф. ст.

Упражнение 13.4

Торговый склад	*Магазин*				*Фиктивный*
	P	Q	R	S	
A	27		13		
B		23	17		
Вариант 2		2		35	23

или

Торговый склад	*Магазин*				*Фиктивный*
	P	Q	R	S	
A	27		13		
B		25	15		
Вариант 2			2	35	23

Минимальная стоимость транспортировки 8765 ф. ст.

Упражнение 13.5

1.

Ферма	*"Cravt"*	*"Liver"*	*Фиктивный*
"Scent Hill"	2000		
"Midrow Top"	750	1750	500
"Alum Vp"		1500	

Максимальный доход 458750 ф. ст.;

2. На 7 ф. ст. за 1 т.

Упражнение 13.6 Общее время выполнения работ 83 ч.

Работник	*Работа 1*	*Работа 2*	*Работа 3*	*Работа 4*	*Работа 5*
M1				0	
M2	0				
M3		0			
M4					0
M5			0		

Упражнение 13.7 Общее время выполнения работ 81 ч.

Работник	*Работа 1*	*Работа 2*	*Работа 3*	*Работа 4*	*Работа 5*	*Фиктивная*
M1			0			
M2	0					
M3		0				
M4						0
M5					0	
M6				0		

Упражнение 13.8

Месяц 1:250 (производство) + 50 (запас) = 300 (спрос); 50 не производится;

Месяц 2: 350 (производство) = 275 (спрос) + 75 (запас, используемый в месяце 3);
Месяц 3: 325 (производство) + 75 (запас) = 400 (спрос);
Месяц 4: 300 (производство) = 300 (спрос) ; 75 не производится.

Общие издержки, не включая издержки производства начального запаса, в 50 единиц равны 122750 ф. ст.

Упражнение 13.9 Максимальный объем продаж 371000 ф. ст.

	I	*II*	*III*	*IV*	*V*	*VI*
A				0		
B		0				
C			0			
D					0	
E						0
F	0					

Упражнение 13.10

1.

	A	*B*	*C*
1	70	50	
2		150	
3	40		40
4			160
5	140		
Фиктивный	50		

Общая стоимость перевозки — 26070 ф. ст. в день.

2. Снижение общей стоимости транспортировки — на 320 ф. ст. в день;
3. Увеличение общей стоимости транспортировки — на 1680 ф. ст. в день.

Упражнение 13.11 б) Общие издержки производства 251000 ф. ст.

		Октябрь	*Ноябрь*	*Декабрь*	*Фиктивный*
Октябрь	Урочное	300	50	50	
	Сверхурочное			50	100
Ноябрь	Урочное		400		
	Сверхурочное			150	
Декабрь	Урочное			400	
	Сверхурочное			150	

Упражнение 13.12 (i) Минимальные издержки производства — 8720 ф. ст.

Станок	*Продукция*			
	Кубки	*Чашки*	*Вазы*	*Фиктивная*
A	1000		1000	
B	2400			600
C	300	2400		

(Две переналадки станков)

(ii)

Станок	*Продукция*			
	Кубки	*Чашки*	*Вазы*	*Фиктивная*
A	1000		1000	
B		2400		600
C	600			

(Одна переналадка станков)

Упражнение 13.13

(ii)

	Аббатстаун	*Бесвич*	*Карлик*	*Денстоун*
Рексфорд	625			
Сиадон	25	420	380	
Тристон		220		230
Фиктивный	200			

(iii) 111965 ф. ст.

Упражнение 13.14 Браун должен выполнять проект 3, Адамс может выполнять проекты 2 или 4, Дей – проекты 1, 2 или 4, Иванс – проекты 1 и 4. Время выполнения проекта в целом – 324 дня.

	1	*2*	*3*	*4*
Адамс				0
Браун			0	
Дей		0		
Иванс	0			

Глава 14

Упражнение 14.2

При условии использования тех же случайных чисел, что и в примере **14.3**:

Средний спрос – 502,5 аккумуляторов в неделю;

Средний запас на конец периода – 875,75 аккумуляторов в неделю;

Средний уровень нехватки запаса	— 126,75 аккумуляторов в неделю;
Среднее количество заказов в течение недели	— 0,15;
Ожидаемые издержки	— 2980,38 ф. ст. в неделю.

Упражнение 14.3 в) 30%; 2 ч 33 мин в день.

Упражнение 14.4 **1**. а) 0,14; б) 0,03; в) 0,018; г) 0,012; **2**. 1940,06 ф. ст.; **3**. Счет 1: 1456,05 ф. ст.; счет 2: 1600,38 ф. ст.

Упражнение 14.5 **1**. 0,063; **2**. 5463 ф. ст.; **3**. 2340,4 ф. ст.; 0,6.

Упражнение 14.6 а) 15; б) 6,32 потенциальных клиента за 1 ч; г) 1,58 потенциальных клиентов за 1 ч; д) не более 2,37 ф. ст. за 1 ч.

АССОЦИАЦИЯ ДИПЛОМИРОВАННЫХ АУДИТОРОВ (The Chartered Association of Certified Accountants)

Декабрь 1990 года

УРОВЕНЬ 2 — ЭКЗАМЕН ПО СПЕЦИАЛЬНОСТИ

Вариант 2.6

Количественный анализ

Отведенное время — 3 часа
Число вопросов в варианте — 7
Следует ответить ТОЛЬКО НА 5 ВОПРОСОВ
Формулы и выписка из таблиц приведены на с. 548.

1. Назовите две причины, почему нормальное распределение является важным элементом статистического анализа *(4 балла)*

Компания с ограниченной ответственностью "Broadway Plastics" специализируется на производстве сверхмощных пластиковых рессор. Компания получила заказ на 800 рессор, параметры которых в основном совпадают с параметрами стандартных рессор SX-1, за исключением результирующего веса. Руководитель производства первоначально предполагал, что издержки производства специальной партии в 800 рессор составят 5,20 ф. ст за единицу. Эти издержки включают в себя затраты на наладочные работы и другие расходы, связанные с производством небольшой партии нестандартных рессор.

Затем руководитель производства поинтересовался, действительно ли необходимо организовывать производство специальных рессор. При обычной длине вес заказанных пластиковых рессор превышает стандартный и составляет 2,9 кг. Допустимое отклонение по данной спецификации равно 0,2 кг, т.е. любая рессора, вес которой находится в пределах между 2,7 и 3,1 кг, имеющая обычную длину, была бы приемлемой. Эти требования ненамного отличаются от требований к стандартным рессорам SX-1, которые компания выпускает в большом количестве и издержки производства которых значительно ниже. Известно, что вес рессор

SX-1 имеет приблизительно нормальное распределение со средним весом 2,5 кг при обычной длине и стандартным отклонением в 0,2 кг. Цена одной рессоры типа SX-1 равна 3,5 ф. ст. Руководитель производства должен принять решение: произвести 800 рессор либо отобрать рессоры, удовлетворяющие требуемым допустимым отклонениям, из существующего запаса стандартных рессор. Стоимость отбора и замеров одной рессоры 0,20 ф. ст.

Требуется:

а) Определить долю рессор SX-1 из существующего запаса, которые удовлетворяют требуемым допустимым отклонениям. *(3 балла)*

б) Определить ожидаемое число рессор SX-1, требующих проверки в случае, если руководитель производства примет решение об отборе 800 рессор из существующего запаса. Определить ожидаемую стоимость такой проверки. *(3 балла)*

в) Найти общие издержки 800 рессор, удовлетворяющих допустимым отклонениям, для случаев, если они

(i) выбраны из существующего запаса рессор SX-1;

(ii) произведены как особая партия.

Таким образом, определить, следует или нет руководителю производства налаживать производство 800 рессор отдельной партией. *(3 балла)*

г) Какое процентное изменение первоначальных издержек производства особой партии рессор, т.е. 5,20 ф. ст. за единицу, необходимо, чтобы изменить решение, принятое в п. в)? *(3 балла)*

д) Компания также производит другой тип стандартных рессор SY-2, менее широко применяемых, цена которых несколько выше и составляет 4,25 ф. ст. за единицу. Вес этих рессор при обычной длине имеет нормальное распределение со средним значением 2,7 кг, но неизвестным стандартным отклонением σ.

В предположении, что расходы, связанные с проверкой и замерами, останутся такими же, как и в случае рессор SX-1, определить интервал значений, при котором предпочтение отдавалось бы использованию в процессе выполнения заказа рессорам SY-2, а не SX-1. *(4 балла)*

2. Администрация компаний "Gordon Electronics" рассматривает программу выпуска нового осветительного оборудования, которая рассчитана на пять лет. Если новое оборудование будет вводиться без проведения дальнейших исследований рынка, предполагается, что спрос на него будет либо высоким (с вероятностью 0,7), либо низким (с вероятностью 0,3). Однако существует возможность организовать выборочное обследование рынка, которая даст либо хороший (с вероятностью 0,6), либо плохой (с вероятностью 0,4) прогноз.

Было принято решение, что если выборочное обследование приведет к хорошим прогнозным результатам, на исследование рынка будут направлены дополнительные усилия, что приведет к увеличению годового дохода. В этом случае вероятность высокого спроса составит 0,9. Если же, однако, выборочное обследование приведет к плохим прогнозным результатам, компания не будет налаживать выпуск нового оборудования, а продолжит выпуск старых видов продукции.

Бухгалтерия фирмы предоставила следующую финансовую информацию.

Ежегодный доход в случае, если будет вводиться новое оборудование и проводиться выборочное обследование

Год	*1*	*2*	*3*	*4*	*5*
Высокий спрос	20000	40000	50000	30000	10000
Низкий спрос	10000	15000	20000	10000	5000

Увеличенный ежегодный доход в случае, если выборочное обследование даст хороший прогноз

Высокий спрос	10000 ф. ст. в год
Низкий спрос	5000 ф. ст. в год

Стоимость проведения выборочного обследования — 10000 ф. ст.

Ежегодный доход в случае, если выпуск нового оборудования не будет налажен, а выпуск старых видов продукции будет продолжаться

Год	*1*	*2*	*3*	*4*	*5*
Высокий спрос	20000	20000	15000	10000	5000

Стоимость капитала компании — 15% годовых
Предполагается, что все доходы появляются в конце года

Коэффициенты дисконтирования (при 15% годовых)	*1*	*2*	*3*	*4*	*5*
	0,8696	0,7561	0,6575	0,5718	0,497

Требуется:

а) Построить дерево решений, отражающее различные варианты действий, открывающиеся перед компанией. *(5 баллов)*

б) Рассчитать приведенный доход для каждой возможной комбинации действий и результатов. *(6 баллов)*

в) Рассчитать ожидаемую приведенную стоимость в каждой из трех следующих ситуаций:

(i) внедрение выпуска нового оборудования без проведения специального выборочного обследования;

(ii) отказ от производства нового оборудования;

(iii) проведение выборочного обследования рынка и принятие соответствующего решения. *(6 баллов)*

г) Выработать план действий для компании. *(1 балл)*

д) Если вероятность того, что выборочное обследование рынка приведет к хорошим прогнозам, изменится с 0,6 до 0,5, то изменит ли это ваши рекомендации? *(2 балла)*

(20 баллов)

3. Компания "Crown Chemical" выпускает три вида химикатов – Ростик, Сеноупур и Телтрейт, используя при этом четыре вида сырья, – Авекс, Бумакс, Сидекс и Доракс. Информация с расходе сырья на производство 1 кг каждого продукта и ежедневном запасе каждого вида сырья приведена в следующей таблице:

Продукт	*Сырье, кг*			
	Авекс	*Бумакс*	*Сидекс*	*Доракс*
Ростик	0,4	0,0	0,3	0,3
Сеноупур	0,4	0,2	0,3	0,1
Телтрейт	0,0	0,4	0,2	0,4
Ежегодный запас, кг	200	120	250	150

Себестоимость единицы каждого из четырех видов сырья приведена ниже:

Вид сырья	*Себестоимость, ф. ст. за 1 кг*
Авекс	15
Бумакс	25
Сидекс	16
Доракс	20

Цены продажи трех выпускаемых химикатов таковы:

Продукт	*Цена продажи, ф. ст. за единицу*
Ростик	35
Сеноупур	32
Телтрейт	45

Химические продукты Ростик и Сеноупур можно использовать во многих целях, поэтому рыночный спрос на них превышает производственные мощности. Телтрейт, напротив, предназначен только для специального применения, поэтому максимальный ежедневный спрос на него не превышает 500 кг.

Требуется:

а) Сформулировать модель линейного программирования, позволяющую найти ежедневные объемы производства каждого из трех химикатов таким образом, чтобы прибыль была максимальной. Кратко обоснуйте введенные вами предпосылки. *(8 баллов)*

б) Итоговая симплекс-таблица для данной задачи выглядит следующим образом:

Базис	R S T	A	B	C	D	E	Значение
S	0 1 0	1,875	2,5	0	−2,5	0	300
T	0 0 1	−0,9375	1,25	0	1,25	0	150
C	0 0 0	−0,5625	−0,25	1	−0,25	0	70
R	1 0 0	0,625	−2,5	0	2,5	0	200
E	0 0 0	0,9375	−1,25	0	−1,25	1	350
	0 0 0	15,6875	19,75	0	39,75	0	11470

Здесь R, S и T – ежедневные объемы производства химикатов Ростик, Сеноупур и Телтрейт соответственно; A, B, C и D – количество недостающего от максимального запаса сырья вида Авекс, Бумакс, Сидекс и Доракс соответственно. E – количество химиката Телтрейт, находящегося в дефиците от максимального спроса на него. Найдите по таблице ежедневный оптимальный план производства. Определите итоговый показатель прибыли и используемое количество каждого вида сырья. *(4 балла)*

в) Записать двойственную к задаче линейного программирования, сформулированную в п. а). Найти решение двойственной задачи и прокомментировать свой ответ в контексте поставленного вопроса. *(5 баллов)*

г) У компании появилась возможность производить четвертый вид продукции – химикат Упелин, который предполагается продавать по 50 ф. ст. за 1 кг. Для производства 1 кг Упелина требуется 0,5 кг Бумакса и 0,5 кг Доракса. Посоветуйте компании, стоит ли ей производить Упелин. *(3 балла)*

(20 баллов)

4. Администрация крупного гаража, открытого ежегодно в течение 50 недель, контролирует политику управления запасами в отношении одного из видов шин. Распределение еженедельного спроса на данный вид шин приведено ниже:

Спрос за неделю	Вероятность
20	0,1
30	0,6
40	0,3

Уровень повторного заказа, используемый в настоящий момент, составляет 100 шин, а время поставки заказа зафиксировано на уровне трех недель.

а) Определите:

(i) ожидаемое значение ежегодного спроса; *(2 балла)*

(ii) вероятность того, что в период поставки заказа, т.е. в течение трех недель спрос превысит 100 едини *(4 балла)*

Было оценено, что стоимость подачи заказа на шины составляет 12 ф. ст. Кроме того, согласно оценкам специалистов, ежегодные издержки хранения шин каждого вида составляют 15% их стоимости. Одна шина стоит 10 ф. ст.

б) Используйте простую формулу EOQ, чтобы доказать, что оптимальный размер заказа для данного вида шин равен 160 единицам. *(4 балла)* в) Используя приведенные ниже случайные числа, построить имитационн модель движения запасов на 10-недельный период, позволяющую оцени общие издержки хранения для данного гаража. Начиная с начальнс запаса в 150 единиц, используйте размер заказа, найденный в п. (предположив, что стоимость нехватки запаса равна 0,27 ф. ст. за кажд недостающую шину. Зафиксируйте введенные предпосылки и поясни процесс генерирования данных относительно спроса.

Случайные числа: 6 9 1 4 7 1 8 9 3 0 *(8 балло*

г) Объясните, как данный метод имитационного моделирования мож использовать для определения оптимального размера заказа и уров повторного заказа *(2 балл*

(20 баллов

5. При исследовании взаимосвязи между числом заявок на недельные ссуды текущей ставкой процента по закладным из 200 недель последних пяти ле случайным образом были выбраны 15. Ниже приведены соответствующи данные.

Номер недели	*Ставка процента по закладным, x*	*Число заявок на ссуды, y*
1	11,0	75
2	13,5	65
3	13,0	62
4	12,0	76
5	15,0	50
6	14,0	58
7	14,5	54
8	13,5	64
9	10,0	87
10	11,0	79
11	10,5	80
12	12,0	72
13	12,5	69
14	13,0	65
15	13,0	61

Ниже представлены результаты расчетов уравнения регрессии методом наименьших квадратов с помощью ППП (используется модель y = a+вx):

$\bar{x}$ = 12,5667 SD (x) = 1,4744
$\bar{y}$ = 67,8000 SD (y) = 10,3316
a = 152,3990 SE (a) = 5,7701
b = − 6,8116 SE (b) = 0,4562
r = − 0,9721 остаточная $\hat{\sigma}^2$ = 6,3349

а) Объясните менеджеру, не имеющему статистической подготовки, значение следующих терминов. В каждом случае используйте для иллюстрации своих ответов значения, полученные с помощью компьютера:

(i) коэффициенты регрессии, наклон и постоянный член *(4 балла)*

(ii) коэффициент детерминации *(3 балла)*

(iii) остаточное стандартное отклонение *(3 балла)*

б) Используя линованную бумагу, постройте по исходным данным диаграмму рассеяния и нанесите на нее линию регрессии, построенную с помощью метода наименьших квадратов *(4 балла)*

в) Рассчитайте доверительный интервал для количества заявок на недельные ссуды при 95%-ном уровне значимости, если ставка процента по закладным равна 14%. *(6 баллов)*

(20 баллов)

6. Отдел рыночных исследований планирует провести выборочное обследование в форме анкеты. В нижеследующей таблице представлены задачи, которые необходимо выполнить, непосредственно им предшествующие задачи, а для продолжительности выполнения каждой задачи, наиболее вероятная, оптимистическая и пессимистическая оценки.

Наименование задачи	*Предшествующие задачи*	*Продолжительность, дни*		
		наиболее вероятная (m)	*оптимистическая (a)*	*пессимистическая (b)*
A. Разработка анкеты	–	3	2	4
B. План выборочного обследования	–	12	10	20
C. Предварительный сбор информации	A	5	4	12
D Найм интервьюеров	B	4	2	6
E. Обучение интервьюеров	D,A	3	3	3
F. Распределение интервьюеров	B	4	3	5
G. Проведение интервью	C,E,F	10	8	18
H. Ввод данных в компьютер	G	3	2	4
I. Заключительное собрание интервьюеров	G	2	2	2
J. Анализ данных	H	5	4	6
K. Письменный отчет	J	4	2	12

Используя метод ПЕРТ, средний срок μ и стандартное отклонение σ для продолжительности выполнения каждой задачи оцениваются на основе наиболее вероятной (m),оптимистической (a) и пессимистической (в) оценок с помощью формул:

$$\mu = \frac{1}{6}(4m + a + b),$$

$$\sigma = \frac{1}{6}(b - a).$$

Требуется:

а) Найти среднюю продолжительность и стандартное отклонение для каждой задачи; *(4 балла)*

б) Построить сетевой граф данного обследования и найти критический путь в графе, опираясь на показатели средних сроков; *(8 баллов)* в) Определить среднее значение и стандартное отклонение продолжительности критического пути; *(4 балла)*

г) Какова вероятность того, что длина критического пути превысит 50 дней? Обоснуйте все предпосылки, которые вы вводите. *(4 балла)*

(20 баллов)

7. На большинстве заводов корпорации "Hoopoint", занимающейся производством кухонного оборудования, вводится новая структура заработной платы. Данные о выработке за 1 чел-ч. были получены по 8 случайно отобранным заводам до введения новой структуры заработной платы и сразу же после него по другим 8 заводам (в целом данные по 16 заводам).

Выработка за 1 чел.-ч.	
Старая структура заработной платы	*Новая структура заработной платы*
54	55
81	56
50	47
40	64
49	26
58	51
63	48
45	53

Требуется:

а) Рассчитать доверительный интервал для общей средней выработки за 1 чел.-ч. при 95%-ном уровне значимости в условиях новой структуры заработной платы. Какие предположения вы делаете при вычислении границ этого интервала? *(6 баллов)*

б) Предполагается, что изменение в структуре заработной платы фактически не привело к изменению выработки за 1 чел.-ч. в целом.

i) Поясните, почему попарный t-критерий не является подходящим инструментом при проверке данного положения *(2 балла)*

ii) Проверьте данную гипотезу, используя более подходящий t-критерий, и зафиксируйте ваши выводы. Зафиксируйте также все предпосылки, которые вы сделали; *(10 баллов)*

iii) Объясните, почему для таких данных желательно использовать тест, не предполагающий какого-либо конкретного распределения элементов совокупности. *(2 балла)*

(20 баллов)

НОВЫЕ ФОРМУЛЫ И ВЫДЕРЖКИ ИЗ ТАБЛИЦ

A. Выборочные статистики:

(i) Арифметическое среднее — $\dfrac{\sum x}{n}$

(ii) Стандартное отклонение — $\sqrt{\dfrac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\dfrac{\sum x^2 - (\sum x)^2/n}{n-1}}$

B. Распределение вероятностей:

(i) Биномиальное $P(r) = {}^nC_r\, p^r(1-p)^{n-r}$

(ii) Распределение Пуассона $P(r) = \dfrac{e^{-m}\, m^r}{r!}$

C. Стандартные ошибки:

(i) Среднего $\dfrac{\sigma}{\sqrt{n}}$

(ii) Доли $\sqrt{\dfrac{p(1-p)}{n}}$

D. Критерии проверки гипотез:

(i) $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ (одна выборка),
где σ известно;

(ii) $t = \dfrac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{n}}$ (одна выборка),

где $\hat{\sigma} = \sqrt{\dfrac{\sum (x-\bar{x})^2}{n-1}}$;

$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ (две выборки),

где $\hat{\sigma} = \sqrt{\dfrac{(n_1-1)\,\hat{\sigma}_1^2 + (n_2-1)\,\hat{\sigma}_2^2}{(n_1+n_2-2)}}$;

(iii) $$\chi^2 = \sum \frac{(0 - E)^2}{E}.$$

E. Корреляция и регрессия

(i) Смешанный момент корреляции

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}};$$

(ii) Ранговый коэффициент корреляции Спирмена

$$R = 1 - \frac{6 \sum d^2}{n(n^2 - 1)};$$

(iii) Линия регрессии по методу наименьших квадратов

$$y = a + bx,$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2},$$

$$a = \frac{\sum y}{n} - \frac{b \sum x}{n};$$

(iv) Для конкретного значениях стандартное отклонение отдельного наблюдения

$$\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x - \bar{x})^2}},$$

где $$\hat{\sigma}^2 = \frac{\sum (y - \hat{y})^2}{n - 2}.$$

F. Контроль статистических процессов (для неизвестных параметров генеральной совокупности)

(i) Средние

Контрольные границы $\bar{\bar{X}} \pm A \bar{R}$

где $\bar{\bar{X}}$ — среднее значение внутригрупповых средних,

$\bar{R}$ — среднее значение внутригрупповых размахов;

(ii) Размахи

Нижняя контрольная граница $B \bar{R}$

Верхняя контрольная граница $C \bar{R}$,

где A,B,C приведены в таблице.

(iii) Доли

Контрольные границы $p \pm 3\sqrt{\frac{p(1-p)}{n}}$,

где p — общая доля изделий, имеющих дефекты.

G. Сетевой анализ: метод ПЕРТ

Ожидаемый срок выполнения операции $\frac{1}{6}(a + 4m + b)$

с дисперсией $\frac{1}{36}(b - a)^2$.

H. Управление запасами

$$EOQ = \sqrt{\frac{2CD}{H}}.$$

Студентам разрешается при ответах на вопросы использовать таблицы нормального распределения, t – распределения и распределения χ^2.

Факторы для построения контрольных карт "3 σ" для средних и размахов

Число наблюдений в подгруппах, n	*A*	*B*	*C*
2	1,88	0	3,27
3	1,02	0	2,57
4	0,73	0	2,28
5	0,58	0	2,11
6	0,48	0	2,00
7	0,42	0,08	1,92
8	0,37	0,14	1,86
9	0,34	0,18	1,82
10	0,31	0,22	1,78
11	0,29	0,26	1,74
12	0,27	0,28	1,72
13	0,25	0,31	1,69
14	0,24	0,33	1,67
15	0,22	0,35	1,65
16	0,21	0,36	1,64
17	0,20	0,38	1,62
18	0,19	0,39	1,61
19	0,19	0,40	1,60
20	0,18	0,41	1,59

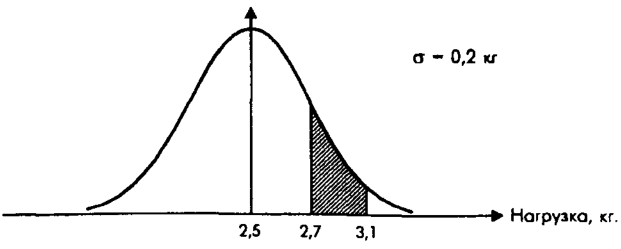
ОТВЕТЫ АВТОРОВ МОДЕЛЕЙ НА ЭКЗАМЕНАЦИОННЫЕ ВОПРОСЫ ПО ТЕМЕ "КОЛИЧЕСТВЕННЫЙ АНАЛИЗ" (декабрь 1990, АССА)

1. Две причины, почему нормальное распределение является важным элементом статистического анализа:

   (i) Нормальное распределение очень часто присутствует в естественных и производственных процессах. Например, вес чая, который сортируется по пачкам с помощью загрузочного станка, должен иметь нормальное распределение. Большинство статистических тестов было разработано в предположении, что исходные данные подчиняются нормальному закону распределения.

   (ii) Если несколько выборок определенного размера случайным образом сделаны из нормальной генеральной совокупности, выборочное распределение выборочных средних также будет нормальным, причем его среднее значение будет совпадать со средним значением генеральной совокупности. Более того, если генеральная совокупность не подчиняется нормальному закону распределения, то выборочное распределение выборочных средних приближается к нормальному (см. центральную предельную теорему). Это позволяет использовать выборочное среднее для оценки генерального среднего.

   а) Показатель нагрузки на рессоры SX-1 имеет нормальное распределение со средним значением 2,5 кг и стандартным отклонением 0,2 кг. Требуемое допустимое отклонение для нового вида рессор составляет от 2,7 кг до 3,1 кг.



Для рессор SX-1 2,7 кг составляет z стандартных отклонений от среднего значения, равного 2,5 кг, где

$$z = \frac{x - \mu}{\sigma} = \frac{2,7 - 2,5}{2,5} = 1,0.$$

Из таблиц для стандартного нормального распределения находим, что $P\ (z > 1,0) = 0,15866$.

Аналогично 3,1 кг равны 3,0 стандартным отклонениям от среднего. Из таблицы для стандартного нормального распределения $P\ (z > 3,0) = 0,00135$. Следовательно, $(0,15866 - 0,00135) = 0,15731$, или 15,731% рессор SX-1 удовлетворяет требуемому интервалу.

б) Число рессор SX-1, которое необходимо проверить, чтобы получить 800 рессор с требуемыми параметрами.

Пусть необходимо проверить X рессор. Тогда из п. а) следует, что число рессор, параметры которых удовлетворяют спецификации, будет равно:

$$X \times 0,15731 = 800,$$

т.е. $X = 5085,5$ рессор.

Следовательно, для получения 800 рессор, удовлетворяющих допустимому отклонению, необходимо провести проверку и замеры 5086 рессор. Ожидаемая стоимость такой проверки составит: $0,20 \times 5086 = 1017,20$ ф. ст.

в) Общая стоимость получения 800 рессор, удовлетворяющих допустимому отклонению:

(i) из рессор SX-1:

стоимость проверки	– 1017,20 ф. ст.,
издержки производства	– $3,50 \times 800 = 2800$ ф. ст.,
общая стоимость	– 3817,20 ф. ст. за 800 рессор.

(ii) издержки производства особой партии рессор:

$$5,20 \times 800 = 4160 \text{ ф. ст.}$$

Получить рессоры особого типа из существующего запаса рессор SX-1 по сравнению с их производством как особой партии дешевле на 342,80 ф. ст.

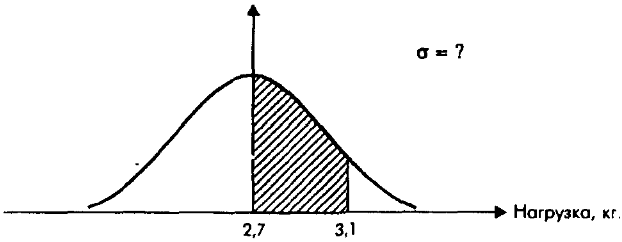
г) Процентное изменение первоначальных издержек производства особой партии рессор, необходимое для изменения решения, принятого в п. **в**):

Единичные издержки должны снизиться до $3817,20/800 = 4,77$ ф. ст. за одну рессору. По сравнению с 5,20 ф. ст. это означает снижение на 0,43 ф. ст., или на 8,27 %.

д) Издержки производства рессор SY-2 равны 4,25 ф. ст. за единицу, тогда как рессор SX-1 – 5,20 ф. ст. за единицу. Тем не менее средняя нагрузка рессор SY-2 составляет 2,7 кг для неизвестного стандартного отклонения. Для какого интервала значений стандартного отклонения было бы дешевле выбирать рессоры, требуемые в заказе, из рессор типа SY-2 или SX-1?

Общая стоимость получения требуемых рессор из рессор типа SY-2 должна быть меньше, чем 3817,20 ф. ст. Стоимость рессор SY-2 – это общая стоимость = = издержки производства + стоимость проверки = 4,25 ф. ст. × 800 + 0,20 ф. ст. × 800 / / (доля рессор SY-2, удовлетворяющих спецификации) = 3400 +160/(доля рессор SY-2, удовлетворяющих спецификации).

Следовательно, 3400 + 160/(доля SY-2, удовлетворяющих спецификации) < 3817,20, т.е. (доля SY-2, удовлетворяющих спецификации) > 160/417,20 = 0,3835. Таким образом, менее чем (0,5 − 0,3835) = 0,1165 элементов распределения должны превышать значение 3,1 кг, являющееся верхним пределом допустимого отклонения в соответствии со спецификацией.



Так как 3,1 кг составляют z' стандартных отклонений от среднего, имеем:

$$P\ (z > z') = 0{,}1165.$$

Из таблиц для стандартного нормального распределения находим, что z'= 1,19 (округление до двух десятичных знаков).

$$\text{Так как}\ z' = \frac{(x-\mu)}{\sigma},\ \text{то}\ 1{,}19 = \frac{3{,}1-2{,}7}{\sigma}.$$

Следовательно,

$$\sigma = 0{,}4/1{,}19 = 0{,}336\ \text{кг}.$$

Если стандартное отклонение для рессор типа SY-2 будет меньше, чем 0,336 кг, то выбирать рессоры, требуемые в заказе, из данного типа рессор будет дешевле.

2. а) Ниже приводится дерево решений.

б) В предположении, что все доходы рассчитываются на конец года, определить приведенный доход для каждой возможной комбинации действий и результатов, используя коэффициент дисконтирования, равный 15% годовых.

В соответствии с деревом решений необходимо сделать шесть вариантов расчетов для конечных узлов F–K.

Узел	*Исследование рынка*	*Прогноз*	*Спрос*	*Ежегодный доход, тыс. ф. ст.*					*Общий приведенный доход за пять лет при 15% годовых*
				1	*2*	*3*	*4*	*5*	
F	да	хороший	высокий	30	50	60	40	20	136,16
G	да	хороший	низкий	15	20	25	15	10	58,15
H	да	плохой	прежний	20	20	15	10	5	50,58
I	нет	–	высокий	20	40	50	30	10	102,64
J	нет	–	низкий	10	15	20	10	5	41,39
K	нет	–	прежний	20	20	15	10	5	50,58

Необходимо найти ожидаемый приведенный доход для узла А дерева решений:

$$E(A) = 0{,}6 \times (0{,}9 \times 136{,}16 + 0{,}1 \times 58{,}15) + 0{,}4 \times 50{,}58 = 97{,}25 \text{ тыс.ф. ст.}$$

Стоимость проведения исследования рынка равна 10000 ф. ст., поэтому ожидаемая чистая приведенная стоимость исследования рынка составит 97,25 − 10 = 87,25 тыс.ф. ст.

г) Рекомендуемый план действий с использованием критерия максимизации ожидаемого приведенного дохода состоит в том, чтобы предпринять исследование рынка. Это позволит получить ожидаемый приведенный доход, равный 87250 ф. ст. Если исследование рынка даст хорошие прогнозные результаты, следует внедрить выпуск нового оборудования. Если же оно даст плохие прогнозы, следует продолжить выпуск старой продукции.

д) Предположим, что вероятность того, что исследование рынка приведет к хорошим прогнозам, изменится с 0,6 до 0,5. Это приведет к изменению ожидаемого приведенного дохода:

$$E(A) = 0{,}5 \times (0{,}9 \times 136{,}16 + 0{,}1 \times 58{,}15) + 0{,}5 \times 50{,}58 - 10 = 79{,}47 \text{ тыс.ф. ст.}$$

Данное значение ниже, чем ожидаемый приведенный доход для узла D, равный 84,26 тыс. ф. ст.

Такое изменение значения вероятности изменило бы и наши рекомендации. Мы бы уже не рекомендовали проведение исследования рынка, однако по-прежнему придерживались бы программы выпуска нового осветительного оборудования. Этот факт говорит о чувствительности решения к небольшим изменениям значения вероятности прогноза.

3. **а)** Модель линейного программирования для производства компанией "Crown Chemical" химикатов ростик, сеноупур и телтрейт. Рассчитаем доход от производства каждого вида продукта:

Единичный доход = Цена продажи 1 кг − Единичные переменные издержки.

В данном примере нам даны только затраты на сырье. Поскольку никакие иные переменные издержки не указаны, мы предположим, что они одинаковы для всех трех видов продукции и, следовательно, не влияют на любые решения, принятые на основе критерия максимизации дохода.

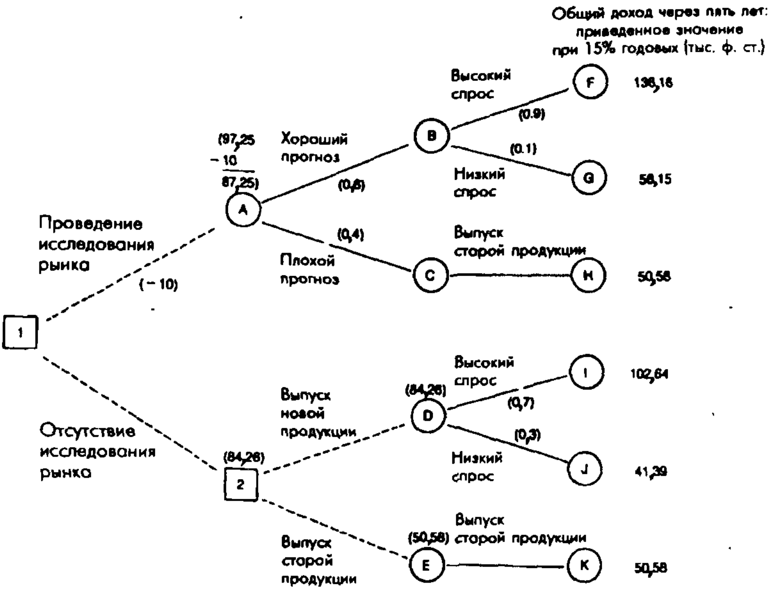
Доход от производства 1 кг:

Ростик	$35 - (0{,}4 \times 15 + 0{,}3 \times 16 + 0{,}3 \times 20) = 18{,}20$ ф.ст за 1кг.
Сеноупур:	$32 - (0{,}4 \times 15 + 0{,}2 \times 25 + 0{,}3 \times 16 + 0{,}1 \times 20) = 14{,}20$ ф.ст за 1кг.
Телтрейт:	$45 - (0{,}4 \times 25 + 0{,}2 \times 16 + 0{,}4 \times 20) = 23{,}80$ ф. ст. за 1 кг.

По условию задачи даны только две группы ограничений − на наличие сырья и максимальный спрос на телтрейт.

Переменные модели:

Пусть R кг в день − объем производства и продажи химиката ростик,
S кг в день − объем производства и продажи химиката сеноупур,
T кг в день − объем производства и продажи химиката телтрейт.



В качестве примера расчета приведенного дохода используем потоки наличности, соответствующие узлу F:

$$\text{Общий приведенный доход при 15\% годовых} = \frac{30}{1,15} + \frac{50}{1,15^2} + \frac{60}{1,15^3} + \frac{40}{1,15^4} + \frac{20}{1,15^5}.$$

Используя коэффициенты дисконтирования, данные в условии задачи, получим:

Общий приведенный доход = 30 × 8696 + 50 × 0,7561 + 60 × 0,6575 + 40 × 0,5718 + 20 × × 0,4972 = 136,16 тыс.ф. ст.

в) Рассчитайте ожидаемый приведенный доход, если компания:

(i) внедряет выпуск нового оборудования без проведения исследований рынка:

Необходимо найти ожидаемый приведенный доход E (D), соответствующий узлу D дерева решений:

E (D)= 0,7 × 102,64 + 0,3 × 41,39 = 84,265 тыс.ф.ст;

(ii) не внедряет выпуск нового оборудования:

В предположении, что в этом случае исследование рынка также не проводится, требуется найти ожидаемое приведенное значение дохода для узла E:

E (E) = 50,58 тыс.ф. ст.;

(iii) предпринимает исследование рынка:

Целевая функция:
Максимизировать ежедневный доход P, где

$$P = 18{,}2\ R + 14{,}2\ S + 23{,}8\ T \text{ (ф. ст. в день)}$$

В условиях следующих ограничений:

	R, S, T	≥ 0
Наличие авекс:	$0{,}4\ R + 0{,}4\ S$	≤ 200 кг в день,
Наличие бумакс:	$0{,}2\ S + 0{,}4\ T$	≤ 120 кг в день,
Наличие сидекс:	$0{,}3\ R + 0{,}3\ S + 0{,}2\ T$	≤ 250 кг в день,
Наличие доракс:	$0{,}3\ R + 0{,}1\ S + 0{,}4\ T$	≤ 150 кг в день,
Спрос на телтрейт:	T	≤ 500 кг в день.

б) Ниже показана итоговая симплекс-таблица:

Базис	*R*	*S*	*T*	*A*	*B*	*C*	*D*	*E*	*Значение*	
S	0	1	0	1,875	2,5	0	−2,5	0	300	=S
T	0	0	1	−0,9375	1,25	0	1,25	0	150	=T
C	0	0	0	−0,5625	−0,25	1	−0,25	0	70	
R	1	0	0	0,625	−2,5	0	2,5	0	200	=R
E	0	0	0	0,9375	−1,25	0	−1,25	1	350	
	0	0	0	15,6875	19,75	0	39,75	0	11,470	Итого

R, S, T определены выше. A, B, C и D — неиспользуемые излишки ресурсов авекс, бумакс, сидекс и дорекс. е — количество химиката телтрейт, находящегося в дефиците от максимального спроса на него, равного 500 кг в день. Оптимальный план (его значения указаны в правом столбце симплекс-таблицы):

300 кг в день химиката сеноупур;
150 кг в день химиката телтрейт;
200 кг в день химиката ростик.

Максимальный доход составляет 11470 ф. ст. в день.

Использование сырья:

Из Правого столбца симплекс-таблицы следует, что C = 70 кг, т.е. 70 кг ресурса сидекс не используется.

A = B =D=0.

Следовательно, сырье расходуется следующим образом:

200 кг ресурса авекс используется полностью;
120 кг ресурса бумакс используется полностью;
180 кг ресурса сидекс (250 − 70 кг (C));
150 кг ресурса Доракс используется полностью.

(Кроме того, объем производства химиката телтрейт составляет 350 кг (E), что ниже максимального спроса, равного 500 кг).

в) Двойственная модель:

Переменные модели:

Пусть A, B, C и D — стоимость 1 кг соответствующего сырья. Пусть E — стоимость спроса на 1 кг химиката телтрейт.

Заметим, что в данном случае те обозначения, которые использовались нами в п б), имеют другое значения, но оба определения тесно связаны друг с другом.

Целевая функция минимизирует общую стоимость используемого сырья и стоимость спроса на телтрейт V, где:

$$V = 200A + 120B + 250C + 150D + 500E \text{ ф. ст. в день.}$$

В условиях соблюдения системы ограничений:

Доход от ростик:	$0{,}4\,A + 0{,}3\,C + 0{,}3\,D$	$\geq 18{,}2$ ф. ст./кг
Доход от сеноупур:	$0{,}4\,A + 0{,}2\,B + 0{,}3\,C + 0{,}1\,D$	≥ 14.2 ф. ст./кг
Доход от телтрейт:	$0{,}4\,B + 0{,}2\,C + 0{,}4\,D + E$	$\geq 23{,}8$ ф. ст./кг
	A, B, C, D, E	≥ 0

Из нижней строки итоговой симплекс-таблицы получаем:

Стоимость 1 кг авекс,	$A = 15{,}6875$ ф.ст
Стоимость 1 кг бумакс,	$B = 19{,}75$ ф.ст
Стоимость 1 кг сидекс,	$C = 0$ ф.ст
Стоимость 1 кг доракс,	$D = 39{,}75$ ф. ст.
Стоимость спроса на телтрейт,	$E = 0$ ф. ст.

Подстановка данных значений переменных в целевую функцию дает минимальное значение стоимости/цены $V = 11470$ ф. ст. в день.

г) Следует ли налаживать выпуск четвертого вида продукции упелин?

Каждый килограмм химиката упелин продавался бы по 50 ф.ст, а его производство потребовало бы 0,5 кг бумакса и 0,5 кг доракса. Единичный доход от упелина равен:

$$50 - (0{,}5 \times 25 + 0{,}5 \times 20) = 27{,}5 \text{ ф. ст. за 1 кг.}$$

Из решения двойственной задачи (или теневой цены) стоимость используемого при производстве Упелина сырья составит:

$$0{,}5 \times 19{,}75 + 0{,}5 \times 39{,}75 = 29{,}75 \text{ ф. ст. за 1 кг.}$$

Производство 1 кг упелина требует использования сырья стоимостью 29,75 ф. ст., а отдача от него составляет 27,50 ф.ст, следовательно, выпуск этого продукта налаживать не следует. Производство 1 кг упелина влечет за собой снижение общего дохода на (29,75 − 27,5) = 2,25 ф. ст.

4. а) (i) Определите ожидаемое значение ежегодного спроса:

Ожидаемое значение спроса в неделю =
= сумма (спроса за неделю) x вероятность =
$= 20 \times 0{,}1 + 30 \times 0{,}6 + 40 \times 0{,}3 = 32$ шины в неделю.

Ожидаемое значение ежегодного спроса =
= 50 недель x ожидаемое значение спроса в неделю
$= 50 \times 32 = 1600$ шин в год.

(ii) Определите вероятность того, что в течение трех недель поставки заказа спрос превысит 100 единиц.

Существуют четыре возможных исхода, которые приводят к тому, что спрос превышает 100 единиц:

Спрос за неделю	*Итого*	*Вероятность*
1 2 3		
40+40+40	= 110	0,3 × 0,3 × 0,3 = 0,027
40+40+30	= 110	0,3 × 0,3 × 0,6 = 0,054
40+30+40	= 110	0,3 × 0,6 × 0,3 = 0,054
30+40+40	= 110	0,6 × 0,3 × 0,3 = 0,054
		Итого 0,189

Общая вероятность того, что спрос больше 100 равен 0,189.

При проведении расчетов предполагалось, что значения спроса в течение каждой недели не зависят друг от друга.

б) Использование простой формулы EOQ, чтобы доказать, что оптимальный размер заказа равен 160.

Стоимость подачи заказа = C_0 = 12 ф.ст за заказ.

Ожидаемый ежегодный спрос = D= 1600 шин в год.

Издержки хранения запаса = $C_н$=0,15 × 10 = 1,50 ф. ст. в год.

$$EOQ = \sqrt{\frac{2D\,C_o}{C_н}} = \sqrt{\frac{2 \cdot 1600 \cdot 12}{1,5}} = 160 \text{ шин в одном заказе.}$$

в) Построить имитационную модель движения запасов на 10-недельный период и оценить общие издержки хранения запаса в течение недели.

Начальный запас составляет 150 шин.

Размер заказа равен 160 шин.

Уровень повторного заказа равен 100 шин и включает в себя как находящиеся в запасе, так и уже заказанные шины.

Поскольку ежедневный спрос нам неизвестен, будем нсследовать только уровень запаса, остающийся на конец недели. После этого будет подаваться новый заказ, который будет получен через три недели, т.е. непосредственно перед началом процесса торговли в течение следующей недели, например, заказ, поданный в конце недели 4, прибудет непосредственно перед началом недели 8. Спрос, неудовлетворенный из-за отсутствия запаса, по мере получения нового заказа не удовлетворяется. Издержки, связанные с отсутствием запаса, равны 0,25 ф. ст. за шину.

Моделирование спроса осуществляется с помощью распределения нижеследующих случайных чисел в соответствии с вероятностями, указанными в п. а);

Спрос за неделю	*Вероятность*	*Случайные числа, СЧ*
20	0,1	0
30	0,6	1–6
40	0,3	7–9

В течение десяти недель мы подадим два заказа, что будет стоить 2 × 12 = 24 ф. ст. в неделю.

Средний уровень запаса на начало недели: 870/10 = 87 шин.

Средний уровень запаса на конец недели: 560/10 = 56 шин.

Следовательно, средний уровень запаса составляет приблизительно 71,5 шин в неделю.

Издержки хранения запаса в течение недели:

$$71,5 \times 10 \text{ ф. ст.} \times (0,15/50) = 2,145 \text{ ф. ст.}$$

Издержки хранения запаса в течение 10 недель =

$$= 2,145 \text{ ф.ст} \times 10 = 21,45 \text{ ф. ст. за 10 недель.}$$

Стоимость отсутствия запаса:

$$20 \times 0,25 = 5,00 \text{ ф. ст. за 10 недель.}$$

Общая стоимость функционирования системы запасов в течение 10 недель:

$$(24 + 21,45 + 5,0) = 50,45 \text{ ф. ст. за 10 недель.}$$

Следовательно, оценка общей недельной стоимости движения запасов составляет 5,05 ф.ст в неделю.

Номер недели	*Прибытие заказа*	*Запас на начало недели*	*Спрос*		*Запас на конец недели*	*Дефицит шин из-за отсутствия запаса*	*Комментарии*
			СЧ	*Число шин*			
1	–	150	6	30	120	–	
2	–	120	9	40	80	–	Уровень повторного заказа 160 для недели 6
3	–	80	1	30	50	–	
4	–	50	4	30	20	–	Необходимо 40, в наличии 20
5	–	20	7	40	0	20	Уровень повторного заказа 160 для недели 11
6	160	160	1	30	130	–	
7	–	130	8	40	90	–	
8	–	90	9	40	50	–	
9	–	50	3	30	20	–	
10	–	20	0	20	0	–	
Всего		870		330	560	20	

г) Каким образом данный метод имитационного моделирования может использоваться для определения оптимального размера заказа и уровня повторного заказа?

Необходимо разработать модель для более длительного периода, скажем, от 50 до 100 недель. Следует повторить процесс моделирования для ряда различных комбинаций размера заказа и уровня повторного заказа, используя при этом один и тот же набор случайных чисел. После этого появится возможность определить лучшие значения этих параметров и рассмотреть их более детально. Для выбранных значений параметров можно повторить метод имитационного моделирования, используя различные наборы случайных чисел Повторяемые процессы имитационного моделирования дадут также возможность изучить чувствительность стоимости по отношению к размеру заказа и т.д. Данный метод предполагает, что модель спроса остается неизменной в течение всего периода моделирования.

5. а) Объяснение сущности следующих терминов:

(i) Коэффициенты регрессии, наклон и постоянный член:

Для оценки линейной связи между числом поданных заявок и ставкой процента по закладным использовался пакет прикладных программ. Данная взаимосвязь имеет вид:

Число заявок, поданных в течение недели:

153,4 — 6,81 × (% по закладным).

153,4 и −6,81 — это коэффициенты регрессии.

Эмпирические значения ставки процента по закладным в течение 15 недель изменяются от 10 до 15%. В рамках данного интервала модель позволяет рассчитать оценку среднего количества заявок, которое можно ожидать в течение недели. Например, если бы ставка процента по закладным составляла 12%, оценка была бы равна:

количество заявок за неделю: 153,4 − 6,81 × 12 = 71,7, или 72.

Наклон линии регрессии равен −6,81 и означает, что при выходе за пределы указанного выше интервала увеличение ставки на 1% приведет к уменьшению числа подаваемых заявок в среднем на 7 (6,81) единиц. Постоянный член 153,7 не имеет определенного значения сам по себе. Его можно интерпретировать как основное, или базисное значение, из которого мы вычитаем воздействие ставки процента по закладным при выходе за определенный интервал ее значений.

(ii) Коэффициент детерминации.

Мы пытаемся объяснить, почему число заявок в неделю варьирует от 50 за 5-ю неделю до 87 за 9-ю неделю. Для объяснения подобной вариации мы применяем показатель ставки процента по закладным. Коэффициент детерминации является мерой, позволяющей установить, насколько хорошо данная вариация объясняется с помощью регрессионной модели. Если коэффициент детерминации r^2 равен 1, нам удалось объяснить всю вариацию количества заявок на ссуды через взаимосвязь со ставкой процента по закладным. В этом случае линейная модель является идеальной. Если r^2 равен 0, вариация вообще не была объяснена, и модель является непригодной к использованию.

Коэффициент детерминации: $r^2 = (-0{,}9721)^2 = 0{,}945$.

Данное значение достаточно близко к 1. Линейная модель хорошо объясняет вариацию числа заявок на ссуды. Мы объяснили 94,5% вариации и не объяснили оставшееся 5,5%.

(iii) Остаточное стандартное отклонение.

Мы уже установили, что построенная модель не является совершенной. Фактически именно это и следовало предполагать, поскольку нам известно, что кроме ставки процента по закладным спрос на них подвержен воздействию и других факторов. Данные, собранные за 15 недель, наглядно демонстрируют это положение. Так, для 3-й, 14-й и 15-й недель ставка по закладным составила 13%, однако число заявок было равно 62,65 и 61 соответственно. Если данная модель используется для прогнозирования числа поданных заявок при ставке по закладным в 13%, полученное прогнозное значение окажется единственным и составит:

153,4 − 6,81 × 13) = 64,9 заявок.

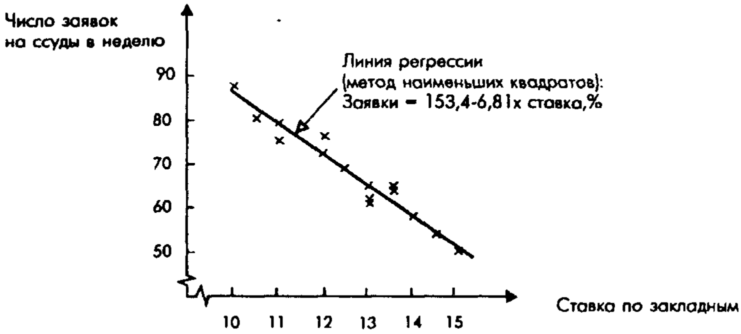
Остаток — это разница между полученным по модели прогнозным и эмпирическим значениями. В целом, остаточное стандартное отклонение представляет собой оценку среднего отклонения между прогнозными и фактическими значениями. Оно равно:

$$\hat{\sigma} = \sqrt{6{,}3349} = 2{,}517.$$

В среднем предполагается, что оценки содержат ошибку в две или три заявки.

Примечание: одна из предпосылок регрессионного анализа состоит в том, что в генеральной совокупности значения у, соответствующие конкретным значениям х, имеют нормальное распределение. Предполагается, что для всех х в этих нормальных распределениях существует постоянное стандартное отклонение σ_e. Именно это стандартное отклонение оценивается с помощью остаточного стандартного отклонения σ (см. гл.8).

б) График зависимости числа поданных заявок от ставки процента по закладным:



в) Рассчитаем доверительный интервал для числа заявок на ссуды в неделю при 95%-ном уровне значимости, если ставка процента по закладным равна 14%.

Мы можем построить доверительный интервал для генерального среднего у при заданных х и $\mu_{y/x}$. И наоборот, можно построить доверительный интервал для отдельного значения у. На основе известных нам формул предположим, что требуется найти последний из указанных интервалов. Доверительный интервал для отдельного значения числа заявок на ссуды при 95%-ном уровне значимости составит:

$$\hat{y} \pm t_{0,025\,,\,(n-2)} \cdot SE_y.$$

В списке формул (раздел Е, (iv)) приведена формула расчета стандартной ошибки для отдельного значения у (в отличие от среднего значения) при заданном значении х.

$$\text{SE (отдельного значения y)} = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x - \bar{x})^2}}\,.$$

Здесь $x_o = 14\%$

$n = 15$

$\hat{\sigma} = 2{,}517$ (рассчитано в п. а))

$\bar{x} = 12{,}5667$ (дано по условию)
$(x_o - \bar{x})^2 = (14 - 12{,}5667)^2 = 1{,}4333^2 = 2{,}0543$
SD (x) = 1,4744 (дано), следовательно,

$$\text{Дисперсия (x)} = 1{,}4744^2 = 2{,}1739 = \frac{\sum(x - \bar{x})^2}{(n-1)}\text{, следовательно,}$$

$$\sum(x - \bar{x})^2 = \text{Дисперсия (x)} \cdot (n-1) = 2{,}1739 \cdot 14 = 30{,}434.$$

Следовательно,

$$\text{SE (отдельного значения y)} = 2{,}517\sqrt{1 + \frac{1}{15} + \frac{2{,}0543}{30{,}434}} = 2{,}517 \times 1{,}065 = 2{,}68.$$

Прогнозное значение числа заявок на ссуды при ставке по закладным в 14% составит:

$$\hat{y} = 153{,}4 - 6{,}81 \cdot 14 = 58{,}06 \text{ заявок в неделю.}$$

Доверительный интервал для отдельного значения числа заявок на ссуды при 95%-ном уровне значимости равен:

$58{,}06 \pm t_{0,025,\ 14} \cdot$ SE (отдельного значения y),
т.е. $58{,}06 \pm 2{,}145 \cdot 2{,}68 = 58{,}06 \pm 5{,}75$,
т.е. от 52,3 до 63,8 заявок в неделю.

Данный результат можно сравнить с числом заявок на ссуды, поданными в течение 6-й недели и равными 58, при ставке по закладным в 14%.

6. а) Найдем среднее значение и стандартное отклонение продолжительности выполнения каждой задачи. Используя указанные формулы

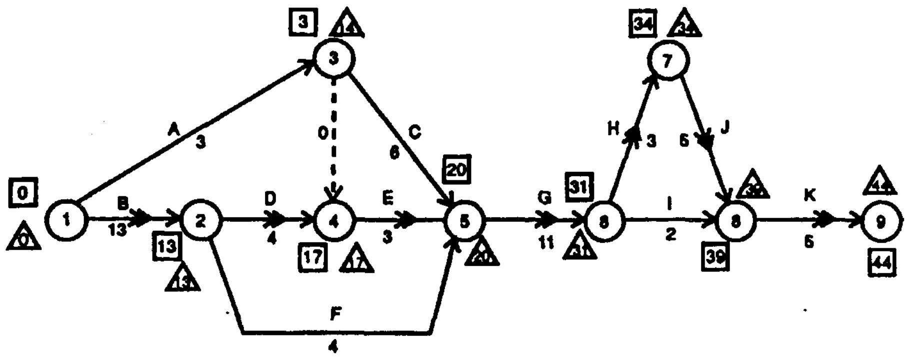
$$\mu = \frac{(4m + a + b)}{b}, \quad \sigma = \frac{(b - a)}{6},$$

получим:

Задача	*Предшествующая задача*	*Продолжительность*				
		m	*a*	*b*	μ	σ
A	–	3	2	4	3	2/6
B	–	12	10	20	13	10/6
C	A	5	4	12	6	8/6
D	B	4	2	6	4	4/6
E	D,A	3	3	3	3	0
F	B	4	3	5	4	2/6
G	C,E,F	10	8	18	11	10/6
H	G	3	2	4	3	2/6
I	G	2	2	2	2	0
J	H	5	4	6	5	2/6
K	I, J	4	2	12	5	10/6

б) Построим сетевой граф и определим критический путь (см. с. 563)

Сетевой граф приведен ниже. Наиболее ранние и наиболее поздние сроки наступления событий указаны около соответствующих узлов. Задача является критической, если $LET_{\text{окончания}} - EET_{\text{начала}} -$ продолжительность = 0.



□ – наиболее ранний срок наступления события,

△ – наиболее поздний срок наступления события
средняя продолжительность, дней.

Как следует из графа, критический путь, построенный с помощью средних сроков выполнения задач, выглядит следующим образом: B,D,E,G,H,J,K.

в) Определим среднюю продолжительность и стандартное отклонение для критического пути.

Из сетевого графа следует, что средняя продолжительность критического пути составляет 44 дня. Чтобы найти стандартное отклонение, определим сначала величину дисперсии.

$$\text{Дисперсия продолжительности критического пути} = \sum \left(\text{дисперсии всех задач, принадлежащих критическому пути} \right)$$

Следовательно, воспользовавшись известной продолжительностью критического пути, получим:

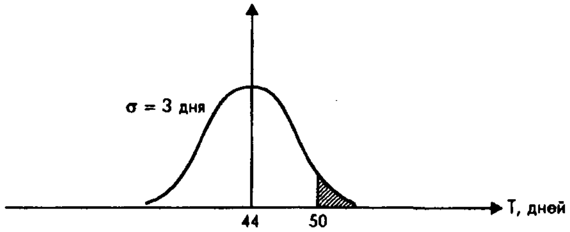
$$\sigma_T^2 = \sigma_B^2 + \sigma_D^2 + \sigma_E^2 + \sigma_G^2 + \sigma_H^2 + \sigma_J^2 + \sigma_K^2\ ;$$

$$\sigma_T^2 = (100 + 16 + 0 + 100 + 4 + 4 + 100)/36 = 324/36 = 9\ .$$

Таким образом, стандартное отклонение для продолжительности критического пути $= \sqrt{\sigma_T^2} = \sqrt{9} = 3$ дня.

г) Какова вероятность того, что продолжительность критического пути превысит 50 дней?

Предположим, что общая продолжительность выполнения проекта T имеет нормальное распределение со средним значением 44 дня и стандартным отклонением 3 дня (см. п. в)). Этот факт, в свою очередь, предполагает, что продолжительность каждой задачи, принадлежащей критическому пути, имеет β–распределение, причем продолжительности выполнения каждой из задач независимы друг от друга.



50 дней составляют z стандартных отклонений от среднего значения, равного 44 дням, где:

$$z = \frac{50 - 44}{3} = 2{,}0.$$

По таблице стандартного нормального распределения находим, что $P(z > 2{,}0) = 0{,}02275$. Следовательно, вероятность того, что продолжительность критического пути превысит 50 дней, равна 0,02275.

7. **а)** Рассчитаем доверительный интервал для выработки за 1 чел.-ч. в целом при 95%-ном уровне значимости в условиях новой структуры заработной платы.

Предпосылки:

1. Для каждого завода можно вычислить соответствующий показатель нормы выработки, т.е. в течение рассматриваемого временного промежутка каждый завод выпускает только один продукт или по крайней мере имеет постоянный ассортимент продукции.
2. Показатели нормы выработки для различных заводов сопоставимы, т.е. все заводы производят сходные виды продукции.
3. Показатели нормы выработки по всем заводам подчиняются нормальному закону распределения и имеют одинаковые среднее значение и стандартное отклонение.

Для новой структуры заработной платы имеем:

$$n = 8,\ \bar{x} = 50,\ s = 10{,}344.$$

Доверительный интервал для нормы выработки как генерального среднего при 95%-ном уровне значимости равен:

$$\bar{x} \pm t_{0.025,7}\, s/\sqrt{(n-1)} = 50 \pm 2{,}365 \times 10{,}344/\sqrt{7} = 50 \pm 9{,}246.$$

Мы можем быть на 95% уверены, что норма выработки как генеральное среднее лежит в пределах между 40,754 и 59,246.

б) (i) Почему попарный t-критерий являлся бы неподходящим инструментом при проверке предположения о том, что изменение в структуре заработной платы не привело к изменениям в норме выработки в целом?

Две выборки не являются попарно сопоставимыми. Они получены на основе различных заводов до и после изменения структуры заработной платы. Если бы проводилась выборка 8 заводов, по которым показатели выработки фиксировались бы до и после изменения структуры заработной платы, попарный t-критер можно было бы использовать.

(ii) Будем использовать t-критерий для средних значений двух выборок.

Примечание. Данной процедуре должен предшествовать F-критерий для ген ральных дисперсий. Предположим, что обе выборки были получены из генеральн совокупностей, имеющих равные дисперсии. Если сравнить приведенные ни стандартные отклонения двух выборок, сделанное нами предположение мож считать вполне допустимым (и фактически F-критерий это подтверждает). На также предполагается, что лежащие в основе выборок генеральные совокупнос распределены нормально.

H_0: $\mu_o = \mu_N$, различие в нормах выработки при старой и новой структур заработной платы отсутствует.

H_1: $\mu_o \neq \mu_N$, различие в нормах выработки существует.

Формулировка гипотезы H_1 предполагает использование двухвершинно критерия. (Если бы утверждалось, что новая структура была введена в целя увеличения выработки, то в этом случае гипотеза H_1 формулировалась бы след ющим образом: H_1: $\mu_o < \mu_N$, и можно было бы применять одновершинный критери Из постановки проблемы не ясно, действительно ли преследовалась именно э цель, однако обе выборки идут в размер с данной идеей, поскольку средняя норм выборки снизилась с 55 до 50).

Проведем проверку гипотезы H_0, используя двухсторонний t-критерий, н 5%-ном уровне значимости $(n_1 + n_2 - 2) = (8 + 8 - 2) = 14$ степеней свободы.

По таблицам стандартного t-распределения находим, что

$$t_{0,025,14} = 2,145.$$

Необходимо знать выборочные средние и стандартные отклонения.

Для старой структуры заработной платы: $n_o = 8$, $\bar{x}_o = 55$ и $s_o = 11,916$.

Для новой структуры заработной платы: $n_N = 8$, $\bar{x}_N = 50$ и $s_N = 11,916$.

Значение t-критерия равно:

$$t = \frac{\bar{x}_O - \bar{x}_N}{\hat{SE}}.$$

Следовательно,

$$\frac{\bar{x}_O - \bar{x}_N}{\sqrt{\frac{(n_O s_O^2 + n_N s_N^2)}{(n_O + n_N - 2)}\left(\frac{1}{n_O} + \frac{1}{n_N}\right)}} = \frac{55 - 50}{\sqrt{\frac{(8 \cdot 11,916^2 + 8 \cdot 10,344^2)}{(8 + 8 - 2)}\left(\frac{1}{8} + \frac{1}{8}\right)}} =$$

$$= \frac{5}{\sqrt{\frac{1991,92}{14 \cdot 4}}} = \frac{5}{\sqrt{35,57}} = 0,838.$$

Так как $0,838 < t_{0,025,14} = 2,145$ результат не является значимым на 5%-ном уровне. Выборка подтверждает гипотезу H_o, поэтому H_o принимается. Мы не находим доказательства в поддержку того, что изменения в выработке действительно имели место.

(iii) Почему для таких данных желательно использовать тест, не предполагающий какого-либо конкретного распределения элементов совокупности?

Одна из четырех введенных нами предпосылок заключалось в том, что генеральные совокупности, лежащие в основе выборок, распределены нормально. В соответствии с формулировкой вопроса применение любого критерия, не предполагающего конкретного распределения элементов совокупности, не требует принятия данной предпосылки.

Каждая из выборок содержит одно значение, которое является нетипичным. Так, в выборке показателей норм выработки при старой структуре заработной платы имеется значение 81, которое почти на 20% превышает все остальные. В выборке показателей норм выработки при новой структуре заработной платы есть значение 26, которое почти на 20% ниже остальных. На основе этого факта можно предположить, что исходные генеральные совокупности не подчиняются нормальному закону распределения. Указанные значения оказывают большое влияние на выборочные средние, а через стандартные отклонения отрицательно воздействуют и на чувствительность t-критерия.

МАТЕМАТИЧЕСКИЕ ФОРМУЛЫ

Глава 1

Правило сложения вероятностей

$$P(\text{А или В или их совместного появления}) = P(A) + P(B) - P(\text{А и В}).$$

Правило сложения вероятностей для взаимоисключающих событий

$$P(\text{А или В}) = P(A) + P(B).$$

Правило умножения вероятностей

$$P(\text{А и В}) = P(A)P(\text{В при условии наступления А}).$$

Правило умножения вероятностей для независимых событий

$$P(\text{А и В}) = P(A)P(B).$$

Формула Байеса

$$P(\text{А при условии наступления В}) = \frac{P(\text{А и В})}{P(B).}$$

Математическое ожидание

$$E(r) = \sum pr.$$

Число перестановок

$${}^{n}P_{r} = \frac{n!}{(n-1)!}, \text{ где } n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$$

причем $0! = 1$.

$${}^{n}C_{r} = \frac{n!}{r!\,(n-r)!}.$$

Глава 2

Математическое ожидание дискретного распределения вероятностей

$$E(r) = \sum r\,P(r).$$

Дисперсия дискретного распределения вероятностей

$$\sigma^2 = \sum r^2 P(r) - (E(r))^2.$$

Стандартное отклонение для дискретного распределения вероятностей

$$\sigma = \sqrt{\sum r^2 P(r) - (E(r))^2}.$$

Биномиальное распределение вероятностей

$$P(r \text{ успехов в } n \text{ испытаниях}) = {}^nC_r\, p^r q^{n-r}, \quad r = 0, 1, 2, \ldots, n$$

где

$$^nC_r = \frac{n!}{r!\,(n-r)!}.$$

Математическое ожидание для биномиального распределения

Математическое ожидание числа успехов — np,

Математическое ожидание доли успехов — p.

Стандартное отклонение для биномиального распределения

Стандартное отклонение числа успехов $= \sqrt{npq}$.

Стандартное отклонение доли успехов $= \sqrt{\frac{pq}{n}}$.

Распределение Пуассона

$$P(r \text{ успехов / единичный интервал}) = \frac{m^r}{r!} e^{-m}, \quad r = 0, 1, 2, \ldots,$$

где m — среднее число успехов, приходящееся на единичный интервал.

Математическое ожидание и стандартное отклонение распределения Пуассона

Математическое ожидание $E(r) = m$;

Стандартное отклонение $\sigma = \sqrt{m}$.

Нормальное распределение

Значения x нормальной случайной величины имеют среднее значение μ и стандартное отклонение σ.

Значения z стандартной нормальной случайной величины имеют среднее значение 0 и стандартное отклонение 1.

$$z = \frac{x - \mu}{\sigma}.$$

Сочетания независимых нормальных случайных величин

Пусть x и y — независимые случайные величины, имеющие нормальное распределение. Если $z = x \pm y$, то z также нормально распределена с параметрами

$$\mu_z = \mu_x \pm \mu_y \quad \text{и} \quad \sigma_z^2 = \sigma_x^2 \pm \sigma_y^2.$$

Если

$z = ax$, где a — константа, то

$\mu_z = a\,\mu_x$ и $\sigma_z^2 = a^2\,\sigma_x^2$.

Глава 4

Выборочное распределение выборочного среднего

$E(\bar{x}) = \mu$,

$\bar{x} = \hat{\mu}$ — несмещенная оценка,

$$SE_{\bar{x}} = \sqrt{\frac{(N-n)\,\sigma^2}{(N-1)\,n}}$$

$SE_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ при больших значениях N.

Выборочное распределение выборочной дисперсии

$$E(s^2) = \frac{N}{(N-1)}\,\frac{(n-1)}{n}\,\sigma^2;$$

$\hat{\sigma}^2 = \dfrac{(N-1)}{N} \cdot \dfrac{n}{(n-1)}\, s^2$ — несмещенная оценка,

где

$$s^2 = \frac{\sum(x-\bar{x})^2}{2};$$

$\hat{\sigma}^2 = \dfrac{n}{(n-1)}\, s^2$ при больших значениях N;

$$\hat{SE} = \frac{s}{\sqrt{(n-1)}} = \frac{\hat{\sigma}}{\sqrt{n}}.$$

Стандартное нормальное распределение выборочного среднего

σ^2 известно, генеральная совокупность имеет нормальное распределение

$$z = \frac{\bar{x}-\mu}{SE_{\bar{x}}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}},$$

t — распределение выборочного среднего

σ^2 неизвестно, генеральная совокупность нормально распределена

$t = \dfrac{\bar{x}-\mu}{\hat{SE}_{\bar{x}}} = \dfrac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{n}} = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$ с $(n-1)$ степенями свободы.

Распределение χ^2 выборочной дисперсии

Генеральная совокупность нормально распределена:

$$\chi^2 = \frac{ns^2}{\sigma^2} \text{ с } (n-1) \text{ степенями свободы.}$$

Распределение Фишера двух выборочных дисперсий
Обе генеральные совокупности распределены нормально

$$F = \frac{n_1 s_1^2}{(n_1-1)\sigma_1^2} \bigg/ \frac{n_2 s_2^2}{(n_2-1)\sigma_2^2} \quad \text{с } (n_1-1) \text{ и } (n_2-1) \text{ степенями свободы.}$$

Глава 5

Доверительный интервал для генерального среднего μ и вероятности $(1-\alpha)\times 100\%$

$$\bar{x} \pm z_{\alpha/2}\, SE_{\bar{x}}\,.$$

Если значение s известно, доверительный интервал равен

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\,.$$

Если значение σ неизвестно, доверительный интервал равен

$$\bar{x} \pm t_{\alpha/2,\,(n-1)} \frac{s}{\sqrt{n-1}}\,.$$

Доверительный интервал для генеральной доли p и вероятности $(1-\alpha)\times 100\%$

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\,(1-\hat{p}\,)}{n}}.$$

Доверительный интервал для генеральной дисперсии σ^2 и вероятности $(1-\alpha)\times 100\%$

$$\text{от } \frac{ns^2}{\chi^2_{\alpha/2\,,\,(n-1)}} \quad \text{до } \frac{ns^2}{\chi^2_{(1-\alpha)/2\,,\,(n-1)}}$$

Глава 6

Критерий проверки гипотезы о выборочном среднем
Единственная выборка σ^2 известно, либо $n \geq 30$

$$z = \frac{\bar{x}-\mu}{SE_{\bar{x}}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}},$$

σ^2 неизвестно

$$t = \frac{\bar{x}-\mu}{\widehat{SE}_{\bar{x}}} = \frac{\bar{x}-\mu}{\hat{\sigma}\,/\sqrt{n}} = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} \quad \text{с } (n-1) \text{ степенями свободы.}$$

Две выборки σ_1^2, σ_2^2 известны

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE_{\bar{x}_1 - \bar{x}_2}},$$

где

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

σ_1^2, σ_2^2 неизвестны, однако,предполагается, что $\sigma_1^2 = \sigma_2^2$,

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{SE}_{\bar{x}_1 - \bar{x}_2}}$$ с $(n_1 + n_2 - 2)$ степенями свободы,

где $$SE_{\bar{x}_1 - \bar{x}_2} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

и $$\hat{\sigma} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{2}}.$$

F-критерий отношения дисперсий

$$F = \frac{\text{большая } \hat{\sigma}^2}{\text{меньшая } \hat{\sigma}^2}$$

где

большая $$\hat{\sigma}^2 = \frac{n_1 s_1^2}{(n_1 - 1)},$$

меньшая $\hat{\sigma}^2 = \dfrac{n_2 s_2^2}{(n_2 - 1)}$, с $(n_1 - 1)$ и $(n_2 - 2)$ степенями свободы.

Критерий проверки гипотез о выборочной доле

$$n \geq 30.$$

Единственная выборка

$$z = \frac{\hat{p} - p}{SE_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}.$$

Две выборки

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{SE_{\hat{p}_1 - \hat{p}_2}},$$

где $SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ если p_1 и p_2 заданы,

либо

$$\widehat{SE}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\,(1 - \bar{p})\,(\frac{1}{n_1} + \frac{1}{n_2})},$$ если p_1 и p_2 неизвестны,

но предполагаемое равенство $p_1 = p_2 = p$ оценивается через $\hat{p}_1$ и $\hat{p}_2$.

Критерий χ^2

$$\chi^2 = \sum\left\{\frac{(f_E - f_0)^2}{f_E}\right\}.$$

Число степеней свободы
Таблицы сопряженности $(r - 1)(c - 1)$.

Глава 7

Контрольная карта для среднего значения

μ и σ известны	предупреждающие границы	$\mu \pm \frac{2\,\sigma}{\sqrt{n}}$,
	границы регулирования	$\mu \pm \frac{3\,\sigma}{\sqrt{n}}$,
μ и σ неизвестны	предупреждающие границы	$\bar{\bar{x}} \pm \frac{2\,\bar{R}\,/d_n}{\sqrt{n}}$,
	границы регулирования или	$\bar{\bar{x}} \pm \frac{3\,\bar{R}\,/d_n}{\sqrt{n}}$ или
		$\bar{\bar{x}} \pm A\,\bar{R}$.

Контрольные карты для доли

предупреждающие границы	$p \pm 2\sqrt{\frac{pq}{n}}$
границы регулирования	$p \pm 3\sqrt{\frac{pq}{n}}$

Контрольные карты размаха

σ известно	граница центра	$d_n\,\sigma$
	верхняя предупреждающая граница	$r_W\,\sigma$
	верхняя граница регулирования	$r_A\,\sigma$
σ неизвестно	граница центра	$\bar{R}$
	верхняя предупреждающая граница	$\frac{r_W\,\bar{R}}{d_n}$
	верхняя граница регулирования	$\frac{r_A\,\bar{R}}{d_n}$ или $C\bar{R}$
	нижняя граница регулирования	$B\bar{R}$

Глава 8

Простая линейная регрессия

$\hat{y} = a + bx$ является оценкой уравнения $y = \alpha + \beta x + \varepsilon$.

В методе наименьших квадратов параметры уравнения регрессии и коэффициент корреляции вычисляются по формулам

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2},$$

$$a = \frac{\sum y}{n} - \frac{b\sum x}{n},$$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}.$$

Статистическая оценка значимости в регрессивном анализе – в случае модели с двумя переменными

Критерий оценки значимости коэффициента корреляции

$$t = \sqrt{\frac{r^2(n-2)}{(1-r^2)}} \quad \text{с } (n-2) \text{ степенями свободы.}$$

Критерий оценки значимости коэффициента регрессии

$$t = \frac{b}{SE_b} \text{ с } (n-2) \text{ степенями свободы,}$$

где

$$SE_b = \frac{\hat{\sigma}_e}{\sqrt{\sum(x-\bar{x})^2}},$$

$$\hat{\sigma}_e^2 = \frac{\sum(y-\hat{y})^2}{(n-2)} = \frac{(\sum y^2 - a\sum y - b\sum xy)}{n-2}.$$

Доверительный интервал для среднего значения $\mu_{y/x}$ переменной y при заданном значении x_0 переменной x

$$\hat{y} \pm t_{p/2,(n-2)}\hat{\sigma}_e\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum(x-\bar{x})^2}}.$$

Доверительный интервал для отдельных значений y при заданном значении x_0 переменной x

$$\hat{y} \pm t_{p/2,(n-2)}\hat{\sigma}_e\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum(x-\bar{x})^2}}.$$

Статистическая оценка значимости в регрессионном анализе — в случае модели со многими переменными

Критерий оценки значимости модели в целом

$$F = \frac{\text{Сумма квадратов, объясненная регрессией}}{\text{Остаточная сумма квадратов}} = \frac{\sum (\hat{y} - \bar{y})^2 / d\,f_{рег}}{\sum (y - \hat{y})^2 / d\,f_{ост}}$$

где $df_{рег}$ — число независимых переменных в модели k, $df_{ост}$ — n − 1 − k. F-статистике соответствует k и (n − 1 − k) степеней свободы.

Частный F-критерий

$$\frac{F = (r^2_{больший} - r^2_{меньший})\,(k_{большее} - k_{меньшее})}{(1 - r^2_{больший})/(n - 1 - k)}.$$

Ранговая корреляция

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}.$$

Критерий значимости

$$\frac{(r_s - 0)}{1/\sqrt{(n - 1)}}, \quad n \geq 10.$$

Глава 9

Временные ряды

A = T + S + E аддитивная модель;

A = T × S × E мультипликативная модель;

либо $$MAD = \frac{\sum |\text{Факт} - \text{Прогноз}|}{n} = \sum \frac{|E_t|}{n}$$

$$MSE = \frac{\sum (E_t^2)}{n}.$$

Глава 10

Сетевой анализ

ПЕРТ

Ожидаемая продолжительность операции $t = \dfrac{a + 4m + b}{6}$;

Дисперсия $\sigma_t^2 = (\dfrac{b - a}{6})^2$.

Глава 11

Управление запасами

Основная модель

Общая годовая переменная стоимость

хранения запасов $TC = \frac{C_0 D}{q} + \frac{C_h q}{2}$ ф. ст. в год.

Интервал повторного заказа – q / D;

$$EOQ = \sqrt{\frac{2C_0 D}{C_h}};$$

$$\text{Простой } EOQ = \sqrt{\frac{2C_s D}{C_h}}.$$

Модель производственной партии — продукция используется в течение процесса производства

$$EOQ = \sqrt{\frac{2C_s D}{C_h} - \frac{P}{(P - D)}},$$

$$TC = \frac{2C_s D}{q} + \frac{C_h(P - D)q}{2P} \text{ (ф. ст.)};$$

Модель планирования запасов

1. Выполнение заказов производится из нового запаса

$$TC = \frac{2C_0 D}{q} + \frac{C_h(q - s)^2 q}{2q} + \frac{C_b s^2 q}{2q} \text{ (ф. ст.)};$$

$$\text{Оптимальный размер заказа } q_o = \sqrt{\frac{2C_0 D}{2} \frac{C_h + C_b}{C_b}} = EOQ \sqrt{\frac{C_h + C_b}{C_b}},$$

$$S = \sqrt{\frac{2C_0 D}{C_b} \frac{C_h}{C_h + C_b}}.$$

2. Выполнение заказов не производится из нового запаса

$$TC = \frac{C_0 D}{(q + s)} + \frac{C_h q^2}{2(q + s)} + \frac{C_b S^2}{2(q + s)} \text{ (ф. ст.)};$$

$$q_0 = \sqrt{\frac{2C_0 D}{C_h} \frac{C_b}{C_h + C_b}} = EOQ \sqrt{\frac{C_b}{C_h + C_b}},$$

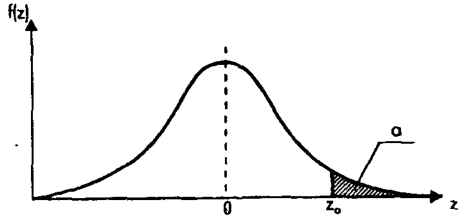
$$S = \sqrt{\frac{2C_0 D}{C_b} \frac{C_h}{C_h + C_b}},$$

Модель интервала повторного заказа

Оптимальный интервал $T = \sqrt{\frac{2C_0}{ChD}}$.

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Нормальное распределение (площади)



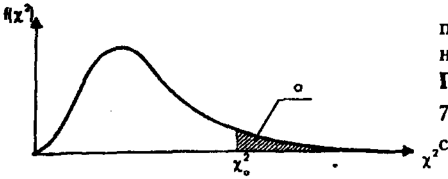
Значения площади α края стандартизированной нормальной кривой N(0,1) для различных значений z.

Пример: площадь справа от z = 1,96 (или слева от z = −1,96) составляет α = 0,02500. Для нормальной кривой с параметрами μ=10 и σ=2 площадь справа от x=12, например, равна площади справа от $z = \frac{x-\mu}{\sigma} = \frac{12-10}{2} = 1$, т.е. α = 0, 15866.

z→ ↓	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
1.0	.15866	.15625	.15386	.15150	.14917	.14686	.14457	.14231	.14007	.13786
1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07214	.07078	.06944	.06811
1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
2.2	.01390	.01355	.01321	.01287	.01254	.01222	.01191	.01160	.01130	.01101
2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00509	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00403	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00263
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.1	.00097	.00094	.00090	.00087	.00085	.00082	.00079	.00076	.00074	.00071
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00009	.00008	.00008	.00008
3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00004	.00003
4.0	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002	.00002

α	0.4	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.001
z_α	.2533	.6745	.8416	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

Распределение χ^2 (хи-квадрат)



Значения χ^2_α, определяющие площадь α в правой части кривой распределения для различного числа степеней свободы ν.

Пример: для $\nu = 15$ площадь справа от $\chi^2_{0,95} = 7{,}261$ составит 0,950, а справа от $\chi^2_{0,10} = 22{,}307$ составляет 0,100.

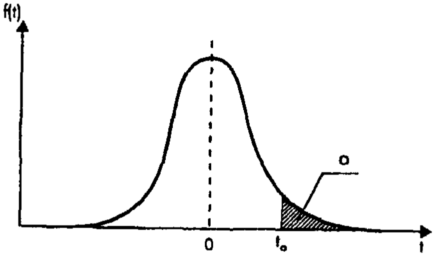
ν \ α	0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	$0.0^{4}3927^{*}$	$0.0^{3}1571^{*}$	$0.0^{3}9821^{*}$	$0.0^{2}3932^{*}$	0.01579	0.1015	0.4549	1.323	2.706	3.841	5.024	6.635	7.879
2	0.01003	0.02010	0.05065	0.1026	0.2107	0.5754	1.386	2.773	4.605	5.991	7.378	9.210	10.597
3	0.07172	0.1148	0.2158	0.3518	0.5844	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838
4	0.2070	0.2971	0.4844	0.7107	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860
5	0.4117	0.5543	0.8312	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.833	15.086	16.750
6	0.6757	0.8721	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548
7	0.9893	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.335	20.090	21.935
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.036	14.339	18.245	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.143	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.670	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.688	13.090	14.848	18.137	22.337	27.141	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	20.843	25.336	30.434	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	21.749	26.336	31.528	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	22.657	27.336	32.620	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	23.567	28.336	33.711	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	24.478	29.336	34.800	40.256	43.773	46.979	50.892	53.672
35	17.192	18.509	20.569	22.465	24.797	29.054	34.336	40.223	46.059	49.802	53.203	57.342	60.275
40	20.707	22.164	24.433	26.509	29.050	33.660	39.335	45.616	51.805	55.758	59.342	63.691	66.766
45	24.311	25.901	28.366	30.612	33.350	38.291	44.335	50.985	57.505	61.656	65.410	69.957	73.166
50	27.991	29.707	32.357	34.764	37.689	42.942	49.335	56.334	63.167	67.505	71.420	76.154	79.490
55	31.735	33.571	36.398	38.958	42.060	47.611	54.335	61.665	68.796	73.311	77.381	82.292	85.749
60	35.535	37.485	40.482	43.188	46.459	52.294	59.335	66.981	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	61.698	69.334	77.577	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	71.144	79.334	88.130	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	80.625	89.334	98.650	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	90.133	99.334	109.141	118.498	124.342	129.561	135.807	140.169
120	83.829	86.909	91.568	95.705	100.627	109.224	119.335	130.051	140.228	146.565	152.214	158.963	163.670
150	109.122	112.655	117.980	122.692	126.278	137.987	149.334	161.288	172.577	179.579	185.803	193.219	198.380
200	152.224	156.421	162.724	168.279	174.828	186.175	199.334	213.099	226.018	233.993	241.060	249.453	253.281
250	196.145	200.929	208.095	214.392	221.809	234.580	249.334	264.694	279.947	287.889	295.691	304.948	311:361

z_α	−2.5758	−2.3263	−1.9600	−1.6449	−1.2816	−0.6745	0.0000	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

* т.е. $0.0^{4}3927 = 0.00003927$

Интерполяция: При $\nu > 100$, $\chi^2_\alpha = \frac{1}{2}\left(z_\alpha + \sqrt{2\nu - 1}\right)^2$, где z_α – стандартизированная нормальная переменная, значения которой приведены под под таблицей отдельной строкой.

t-распределение



Критические значения t_α для различных уровней значимости α и числа степеней свободы ν.

Пример: при $\nu = 19$ $P(t>2,0930)=0,025$, а $P(|t|>2,0930)=0,05$.

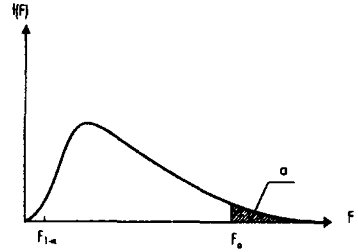
ν \ α	0.4	0.25	0.15	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.3249	1.0000	1.9626	3.0777	6.3138	12.7062	31.8205	63.6567	318.3087	636.6189
2	0.2887	0.8165	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	0.2767	0.7649	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	0.2707	0.7407	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.7267	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.7176	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.7111	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.7064	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.7027	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.6998	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.6974	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.6955	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.6938	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.6924	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.6912	1.0735	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.2576	0.6901	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.2573	0.6892	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.2571	0.6884	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.2569	0.6876	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.2567	0.6870	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.2566	0.6864	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.2564	0.6858	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.2563	0.6853	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.2562	0.6848	1.0593	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.2561	0.6844	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.2560	0.6840	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.2559	0.6837	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.2558	0.6834	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.2557	0.6830	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.6828	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
35	0.2553	0.6816	1.0520	1.3062	1.6896	2.0301	2.4377	2.7238	3.3400	3.5911
40	0.2550	0.6807	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
45	0.2549	0.6800	1.0485	1.3006	1.6794	2.0141	2.4121	2.6896	3.2815	3.5203
50	0.2547	0.6794	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960
60	0.2545	0.6786	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
70	0.2543	0.6780	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.4350
80	0.2542	0.6776	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953	3.4163
90	0.2541	0.6772	1.0424	1.2910	1.6620	1.9867	2.3685	2.6316	3.1833	3.4019
100	0.2540	0.6770	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737	3.3905
120	0.2539	0.6765	1.0409	1.2886	1.6577	1.9799	2.3578	2.6174	3.1595	3.3735
150	0.2538	0.6761	1.0400	1.2872	1.6551	1.9759	2.3515	2.6090	3.1455	3.3566
200	0.2537	0.6757	1.0391	1.2858	1.6525	1.9719	2.3451	2.6006	3.1315	3.3398
300	0.2536	0.6753	1.0382	1.2844	1.6499	1.9679	2.3388	2.5923	3.1176	3.3233
∞	0.2533	0.6745	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

Факторы построения контрольных карт средних и размахов "3σ"

Число наблюдений в подгруппах, n	*A*	*B*	*C*
2	1,88	0	3,27
3	1,02	0	2,57
4	0,73	0	2,28
5	0,58	0	2,11
6	0,48	0	2,00
7	0,42	0,08	1,92
8	0,37	0,14	1,86
9	0,34	0,18	1,82
10	0,31	0,22	1,78
11	0,29	0,26	1,74
12	0,27	0,28	1,72
13	0,25	0,31	1,69
14	0,24	0,33	1,67
15	0,22	0,35	1,65
16	0,21	0,36	1,64
17	0,20	0,38	1,62
18	0,19	0,39	1,61
19	0,19	0,40	1,60
20	0,18	0,41	1,59

Примечание: по использованию этих факторов см. Гл. 7 и Математические формулы.

F-распределение



Значения F_α(α = 0,1; 0,05; 0,025; 0,01 и 0,005) для различных комбинаций степеней свободы в числителе (ν_1) и знаменателе (ν_2).

Пример:

При ν_1=10 и ν_2=2 площадь α справа от $F_{0,05}$=19,40 равна 0,05. Чтобы найти $F_{1-\alpha}$, определяющее площадь α в оставшейся части кривой, воспользуемся соотношением $F_{1-\alpha}(\nu_1,\nu_2)=1/F_\alpha(\nu_1,\nu_2)$.

Пример:

$F_{0,95}(2,10)=1/19,40=0,05155$.

ν_2	α \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	50	100	∞
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.05	62.26	62.69	63.01	63.33
	0.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.3	250.1	251.8	253.0	254.3
	0.025	647.8	799.5	864.2	899.0	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	998.1	1001	1008	1013	1018
	0.010	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6240	6261	6303	6334	6366
	0.005	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24960	25044	25211	25337	25464
2	0.100	8.526	9.000	9.162	9.243	9.293	9.362	9.349	9.367	9.381	9.392	9.406	9.425	9.441	9.451	9.458	9.471	9.481	9.491
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.48	19.49	19.50
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.48	39.49	39.50
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.48	99.49	99.50
	0.005	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5
3	0.100	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.216	5.200	5.184	5.175	5.168	5.155	5.144	5.134
	0.050	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.660	8.634	8.617	8.581	8.554	8.526
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.01	13.96	13.90
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.58	26.50	26.35	26.24	26.13
	0.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.59	42.47	42.21	42.02	41.83
4	0.100	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.896	3.870	3.844	3.828	3.817	3.795	3.778	3.761
	0.050	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.803	5.769	5.746	5.699	5.664	5.628
	0.025	12.22	10.65	9.979	9.605	9.364	9.197	9.074	8.980	8.905	8.844	8.751	8.657	8.560	8.501	8.461	8.381	8.319	8.257
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.91	13.84	13.69	13.58	13.46
	0.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.00	19.89	19.67	19.50	19.32
5	0.100	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.268	3.238	3.207	3.187	3.174	3.147	3.126	3.105
	0.050	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.558	4.521	4.496	4.444	4.405	4.365
	0.025	10.01	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525	6.428	6.329	6.268	6.227	6.144	6.080	6.015
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.722	9.553	9.449	9.379	9.238	9.130	9.020
	0.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.76	12.66	12.45	12.30	12.14
6	0.100	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.905	2.871	2.836	2.815	2.800	2.770	2.746	2.722
	0.050	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874	3.835	3.808	3.754	3.712	3.669
	0.025	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.366	5.269	5.168	5.107	5.065	4.980	4.915	4.849
	0.010	13.75	10.92	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.296	7.229	7.091	6.987	6.880
	0.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.814	9.589	9.451	9.358	9.170	9.026	8.879
7	0.100	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.751	2.725	2.703	2.668	2.632	2.595	2.571	2.555	2.523	2.497	2.471
	0.050	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.445	3.404	3.376	3.319	3.275	3.230
	0.025	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666	4.568	4.467	4.405	4.362	4.276	4.210	4.142
	0.010	12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.058	5.992	5.858	5.755	5.650
	0.005	16,24	12.40	10.88	10.05	9.522	9.155	8.885	8.678	8.514	8.380	8.176	7.968	7.754	7.623	7.534	7.354	7.217	7.076
8	0.100	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.502	2.464	2.425	2.400	2.383	2.348	2.321	2.293
	0.050	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.150	3.108	3.079	3.020	2.975	2.928
	0.025	7.571	6.059	5.416	5.055	4.817	4.652	4.529	4.433	4.357	4.295	4.200	4.101	3.999	3.937	3.894	3.807	3.739	3.670
	0.010	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.263	5.198	5.065	4.963	4.859
	0.005	14.69	11.04	9.596	8.805	8.302	7.952	7.694	7.496	7.339	7.211	7.015	6.814	6.608	6.482	6.396	6.222	6.088	5.951
9	0.100	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.379	2.340	2.298	2.272	2.255	2.218	2.189	2.159
	0.050	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.936	2.893	2.864	2.803	2.756	2.707
	0.025	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.868	3.769	3.667	3.604	3.560	3.472	3.403	3.333
	0.010	10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962	4.808	4.713	4.649	4.517	4.415	4.311
	0.005	13.61	10.11	8.717	7.956	7.471	7.134	6.885	6.693	6.541	6.417	6.227	6.032	5.832	5.708	5.625	5.454	5.322	5.188
10	0.100	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.284	2.244	2.201	2.174	2.155	2.117	2.087	2.055
	0.050	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845	2.774	2.730	2.700	2.637	2.588	2.538
	0.025	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.621	3.522	3.419	3.355	3.311	3.221	3.152	3.080
	0.010	10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.311	4.247	4.115	4.014	3.909
	0.005	12.83	9.427	8.081	7.343	6.872	6.545	6.302	6.116	5.968	5.847	5.661	5.471	5.274	5.153	5.071	4.902	4.772	4.639
11	0.100	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.209	2.167	2.123	2.095	2.076	2.036	2.005	1.972
	0.050	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646	2.601	2.570	2.507	2.457	2.404
	0.025	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.430	3.330	3.226	3.162	3.118	3.027	2.956	2.883
	0.010	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.251	4.099	4.005	3.941	3.810	3.708	3.602
	0.005	12.23	8.912	7.600	6.881	6.422	6.102	5.865	5.682	5.537	5.418	5.236	5.049	4.855	4.736	4.654	4.486	4.359	4.226

F-распределение (продолжение)

ν_2	α \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	50	100	∞
12	0.100	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.147	2.105	2.060	2.031	2.011	1.970	1.938	1.904
	0.050	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544	2.498	2.466	2.401	2.350	2.296
	0.025	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.277	3.177	3.073	3.008	2.963	2.871	2.800	2.725
	0.010	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010	3.858	3.765	3.701	3.569	3.467	3.361
	0.005	11.75	8.510	7.226	6.521	6.071	5.757	5.525	5.345	5.202	5.085	4.906	4.721	4.530	4.412	4.331	4.165	4.037	3.904
13	0.100	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.097	2.053	2.007	1.978	1.958	1.915	1.882	1.846
	0.050	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459	2.412	2.380	2.314	2.261	2.206
	0.025	6.414	4.965	4.327	3.996	3.707	3.604	3.483	3.388	3.312	3.250	3.153	3.053	2.948	2.882	2.837	2.744	2.671	2.595
	0.010	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.665	3.571	3.507	3.375	3.272	3.165
	0.005	11.37	8.186	6.926	6.233	5.791	5.482	5.253	5.076	4.935	4.820	4.643	4.460	4.270	4.153	4.073	3.908	3.780	3.647
14	0.100	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.054	2.010	1.962	1.933	1.912	1.869	1.834	1.797
	0.050	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388	2.341	2.308	2.241	2.187	2.131
	0.025	6.298	4.857	4.222	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.050	2.949	2.844	2.778	2.732	2.638	2.565	2.487
	0.010	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.412	3.348	3.215	3.112	3.004
	0.005	11.06	7.922	6.680	5.998	5.562	5.257	5.031	4.857	4.717	4.603	4.428	4.247	4.059	3.942	3.862	3.698	3.569	3.436
15	0.100	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.017	1.972	1.924	1.894	1.873	1.828	1.793	1.755
	0.050	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.328	2.280	2.247	2.178	2.123	2.066
	0.025	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.963	2.862	2.756	2.689	2.644	2.540	2.474	2.395
	0.010	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.522	3.372	3.278	3.214	3.081	2.977	2.868
	0.005	10.80	7.701	6.476	5.805	5.372	5.071	4.847	4.674	4.536	4.424	4.250	4.070	3.883	3.766	3.687	3.523	3.394	3.260
16	0.100	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	1.985	1.940	1.891	1.860	1.839	1.793	1.757	1.718
	0.050	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.276	2.227	2.194	2.124	2.068	2.010
	0.025	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.889	2.788	2.681	2.614	2.568	2.472	2.396	2.316
	0.010	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.165	3.101	2.967	2.863	2.753
	0.005	10.58	7.514	6.303	5.638	5.212	4.913	4.692	4.521	4.384	4.272	4.099	3.920	3.734	3.618	3.539	3.375	3.246	3.112
17	0.100	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.958	1.912	1.862	1.831	1.809	1.763	1.726	1.686
	0.050	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.308	2.230	2.181	2.148	2.077	2.020	1.960
	0.025	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.825	2.723	2.616	2.548	2.502	2.405	2.329	2.247
	0.010	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.312	3.162	3.068	3.003	2.869	2.764	2.653
	0.005	10.38	7.354	6.156	5.497	5.075	4.779	4.559	4.389	4.254	4.142	3.971	3.793	3.607	3.492	3.412	3.248	3.119	2.984
18	0.100	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.997	1.933	1.887	1.837	1.805	1.783	1.736	1.698	1.657
	0.050	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191	2.141	2.107	2.035	1.978	1.917
	0.025	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.769	2.667	2.559	2.491	2.445	2.347	2.269	2.187
	0.010	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.983	2.919	2.784	2.678	2.566
	0.005	10.22	7.215	6.028	5.375	4.956	4.663	4.445	4.276	4.141	4.030	3.860	3.683	3.498	3.382	3.303	3.130	3.009	2.873
19	0.100	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.912	1.865	1.814	1.782	1.759	1.711	1.673	1.631
	0.050	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.155	2.106	2.071	1.999	1.940	1.878
	0.025	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817	2.720	2.617	2.509	2.441	2.394	2.295	2.217	2.133
	0.010	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.153	3.003	2.909	2.844	2.709	2.602	2.489
	0.005.	10.07	7.093	5.916	5.268	4.853	4.561	4.345	4.177	4.043	3.933	3.763	3.587	3.402	3.287	3.208	3.043	2.913	2.776
20	0.100	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.892	1.845	1.794	1.761	1.738	1.690	1.650	1.607
	0.050	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124	2.074	2.039	1.966	1.907	1.843
	0.025	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.676	2.573	2.464	2.396	2.349	2.249	2.170	2.085
	0.010	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.843	2.778	2.643	2.535	2.421
	0.005	9.944	6.986	5.818	5.174	4.762	4.472	4.257	4.090	3.956	3.847	3.678	3.502	3.318	3.203	3.123	2.950	2.828	2.690
21	0.100	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920	1.875	1.827	1.776	1.742	1.719	1.670	1.630	1.586
	0.050	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.250	2.176	2.096	2.045	2.010	1.936	1.876	1.812
	0.025	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735	2.637	2.534	2.425	2.356	2.308	2.106	2.128	2.042
	0.010	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.785	2.720	2.584	2.475	2.360
	0.005	9.830	6.891	5.730	5.091	4.681	4.393	4.179	4.013	3.880	3.771	3.602	3.427	3.243	3.128	3.049	2.884	2.753	2.614
22	0.100	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933	1.904	1.859	1.811	1.759	1.726	1.702	1.652	1.611	1.567
	0.050	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.226	2.151	2.071	2.020	1.984	1.909	1.849	1.783
	0.025	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700	2.602	2.498	2.389	2.320	2.272	2.171	2.090	2.003
	0.010	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.733	2.667	2.531	2.422	2.305
	0.005	9.727	6.806	5.652	5.017	4.609	4.322	4.109	3.944	3.812	3.703	3.535	3.360	3.176	3.061	2.982	2.817	2.685	2.545
23	0.100	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919	1.890	1.845	1.796	1.744	1.710	1.686	1.636	1.594	1.549
	0.050	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.204	2.128	2.048	1.996	1.961	1.885	1.823	1.757
	0.025	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731	2.668	2.570	2.466	2.357	2.287	2.239	2.137	2.056	1.968
	0.010	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.781	2.686	2.620	2.483	2.373	2.256
	0.005	9.635	6.730	5.582	4.950	4.544	4.259	4.047	3.882	3.750	3.642	3.475	3.300	3.116	3.001	2.922	2.756	2.624	2.484
24	0.100	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.832	1.783	1.730	1.696	1.672	1.621	1.579	1.533
	0.050	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183	2.108	2.027	1.975	1.939	1.863	1.800	1.733
	0.025	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640	2.541	2.437	2.327	2.257	2.209	2.107	2.024	1.935
	0.010	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.643	2.577	2.440	2.329	2.211
	0.005	9.551	6.661	5.519	4.890	4.486	4.202	3.991	3.826	3.695	3.587	3.420	3.246	3.062	2.947	2.868	2.702	2.569	2.428

F-распределение (продолжение)

ν_2	α \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	50	100	∞
25	0.100	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.820	1.771	1.718	1.683	1.659	1.607	1.565	1.518
	0.050	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.165	2.089	2.007	1.955	1.919	1.842	1.779	1.711
	0.025	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.515	2.411	2.300	2.230	2.182	2.070	1.996	1.906
	0.010	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.604	2.538	2.400	2.289	2.169
	0.005	9.475	6.598	5.462	4.835	4.433	4.150	3.939	3.776	3.645	3.537	3.370	3.196	3.013	2.898	2.819	2.652	2.519	2.377
26	0.100	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855	1.809	1.760	1.706	1.671	1.647	1.594	1.551	1.504
	0.050	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.148	2.072	1.990	1.938	1.901	1.823	1.760	1.691
	0.025	5.659	4.265	5.670	3.529	3.105	2.945	2.824	2.729	2.655	2.590	2.491	2.387	2.276	2.205	2.157	2.053	1.969	1.878
	0.010	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.815	2.664	2.569	2.503	2.364	2.252	2.131
	0.005	9.406	6.541	5.409	4.785	4.384	4.103	3.893	3.730	3.599	3.492	3.325	3.151	2.968	2.853	2.774	2.607	2.473	2.330
27	0.100	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874	1.845	1.799	1.749	1.695	1.660	1.636	1.583	1.539	1.491
	0.050	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.132	2.056	1.974	1.921	1.884	1.806	1.742	1.672
	0.025	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631	2.568	2.469	2.364	2.253	2.183	2.133	2.028	1.945	1.853
	0.010	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.783	2.632	2.536	2.470	2.330	2.218	2.097
	0.005	9.342	6.489	5.361	4.740	4.340	4.059	3.850	3.687	3.557	3.450	3.284	3.110	2.928	2.812	2.733	2.565	2.431	2.287
28	0.100	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836	1.790	1.740	1.685	1.650	1.625	1.572	1.528	1.478
	0.050	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.118	2.041	1.959	1.906	1.869	1.790	1.725	1.654
	0.025	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611	2.547	2.448	2.344	2.232	2.161	2.112	2.007	1.922	1.829
	0.010	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.753	2.602	2.506	2.440	2.300	2.187	2.064
	0.005	9.284	6.440	5.317	4.698	4.300	4.020	3.811	3.649	3.519	3.412	3.246	3.073	2.890	2.775	2.695	2.527	2.392	2.247
29	0.100	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827	1.781	1.731	1.676	1.640	1.616	1.562	1.517	1.467
	0.050	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.104	2.027	1.945	1.891	1.854	1.775	1.710	1.638
	0.025	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.669	2.592	2.529	2.430	2.325	2.213	2.142	2.092	1.987	1.901	1.807
	0.010	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.726	2.574	2.478	2.412	2.271	2.158	2.034
	0.005	9.230	6.396	5.276	4.659	4.262	3.983	3.775	3.613	3.483	3.377	3.211	3.038	2.855	2.740	2.660	2.492	2.357	2.210
30	0.100	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.773	1.722	1.667	1.632	1.606	1.552	1.507	1.456
	0.050	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.015	1.932	1.878	1.841	1.761	1.695	1.622
	0.025	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.412	2.307	2.195	2.124	2.074	1.968	1.882	1.787
	0.010	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.700	2.549	2.453	2.386	2.245	2.131	2.006
	0.005	9.180	6.555	5.239	4.623	4.228	3.949	3.742	3.580	3.450	3.344	3.179	3.006	2.823	2.708	2.628	2.459	2.323	2.176
35	0.100	2.855	2.461	2.247	2.113	2.019	1.950	1.896	1.852	1.817	1.787	1.739	1.688	1.632	1.595	1.569	1.513	1.465	1.411
	0.050	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114	2.041	1.963	1.878	1.824	1.786	1.703	1.635	1.558
	0.025	5.485	4.106	3.517	3.179	2.956	2.796	2.676	2.581	2.504	2.440	2.341	2.235	2.122	2.049	1.999	1.890	1.801	1.702
	0.010	7.419	5.268	4.396	3.908	3.592	3.368	3.200	3.069	2.963	2.876	2.740	2.597	2.445	2.348	2.281	2.137	2.020	1.891
	0.005	8.976	6.188	5.086	4.479	4.088	3.812	3.607	3.447	3.318	3.212	3.048	2.876	2.693	2.577	2.497	2.327	2.188	2.036
40	0.100	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.715	1.662	1.605	1.568	1.541	1.483	1.434	1.377
	0.050	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.003	1.924	1.839	1.783	1.744	1.660	1.589	1.509
	0.025	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.288	2.182	2.068	1.994	1.943	1.832	1.741	1.637
	0.010	7.314	5.170	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.522	2.369	2.271	2.203	2.058	1.938	1.805
	0.005	8.828	6.066	4.976	4.374	3.986	3.713	3.509	3.350	3.222	3.117	2.953	2.781	2.598	2.482	2.401	2.230	2.088	1.932
45	0.100	2.820	2.425	2.210	2.074	1.980	1.909	1.855	1.811	1.774	1.744	1.695	1.643	1.585	1.546	1.519	1.460	1.409	1.349
	0.050	4.057	3.204	2.812	2.579	2.422	2.308	2.221	2.152	2.096	2.049	1.974	1.895	1.808	1.752	1.713	1.626	1.554	1.470
	0.025	5.377	4.009	3.422	3.086	2.864	2.705	2.584	2.489	2.412	2.348	2.248	2.141	2.026	1.952	1.900	1.788	1.694	1.586
	0.010	7.234	5.110	4.249	3.767	3.454	3.232	3.066	2.935	2.830	2.743	2.608	2.464	2.311	2.213	2.144	1.997	1.875	1.737
	0.005	8.715	5.974	4.892	4.294	3.909	3.638	3.435	3.276	3.149	3.044	2.881	2.709	2.527	2.410	2.329	2.155	2.012	1.851
50	0.100	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729	1.680	1.627	1.568	1.529	1.502	1.441	1.388	1.327
	0.050	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.952	1.871	1.784	1.727	1.687	1.599	1.525	1.438
	0.025	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317	2.216	2.109	1.993	1.919	1.866	1.752	1.656	1.545
	0.010	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.562	2.419	2.265	2.167	2.098	1.949	1.825	1.683
	0.005	8.626	5.902	4.826	4.232	3.849	3.579	3.376	3.219	3.092	2.988	2.825	2.653	2.470	2.353	2.272	2.097	1.951	1.786
60	0.100	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.657	1.603	1.543	1.504	1.476	1.413	1.358	1.291
	0.50	4.001	3.130	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.917	1.836	1.748	1.690	1.649	1.559	1.481	1.389
	0.025	3.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.169	2.061	1.944	1.869	1.815	1.690	1.599	1.482
	0.010	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.352	2.198	2.098	2.028	1.877	1.749	1.601
	0.005	8.495	5.795	4.729	4.140	3.760	3.492	3.291	3.134	3.008	2.904	2.742	2.570	2.387	2.270	2.187	2.010	1.861	1.689
70	0.100	2.779	2.380	2.164	2.027	1.931	1.860	1.804	1.760	1.723	1.691	1.641	1.587	1.526	1.486	1.457	1.392	1.335	1.265
	0.050	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969	1.893	1.812	1.722	1.664	1.622	1.530	1.450	1.353
	0.025	5.247	3.890	3.309	2.975	2.754	2.595	2.474	2.379	2.302	2.237	2.136	2.028	1.910	1.833	1.779	1.660	1.558	1.436
	0.010	7.011	4.922	4.074	3.600	3.291	3.071	2.906	2.777	2.672	2.585	2.450	2.306	2.150	2.050	1.980	1.826	1.695	1.540
	0.005	8.403	5.720	4.661	4.076	3.698	3.431	3.232	3.076	2.950	2.846	2.684	2.513	2.329	2.211	2.128	1.940	1.797	1.618
80	0.100	2.769	2.370	2.154	2.016	1.921	1.849	1.795	1.748	1.711	1.680	1.629	1.574	1.513	1.472	1.443	1.377	1.318	1.245
	0.050	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.875	1.793	1.703	1.644	1.602	1.508	1.426	1.325
	0.025	5.218	3.864	3.284	2.950	2.730	2.571	2.450	2.355	2.277	2.213	2.111	2.003	1.884	1.807	1.752	1.632	1.527	1.400
	[illegible]	[illegible]	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551	2.415	2.271	2.115	2.015	1.944	1.788	[illegible]	[illegible]

F-распределение (продолжение)

ν_2	α \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	50	100	∞
90	0.100	2.762	2.363	2.146	2.008	1.912	1.841	1.785	1.739	1.702	1.670	1.620	1.564	1.503	1.461	1.432	1.365	1.304	1.228
	0.050	3.947	3.098	2.706	2.473	2.316	2.201	2.113	2.043	1.986	1.938	1.861	1.779	1.688	1.629	1.586	1.491	1.407	1.302
	0.025	5.196	3.844	3.265	2.932	2.711	2.552	2.432	2.336	2.259	2.194	2.092	1.983	1.864	1.787	1.731	1.610	1.503	1.371
	0.010	6.925	4.849	4.007	3.535	3.228	3.009	2.845	2.715	2.611	2.524	2.389	2.244	2.088	1.987	1.916	1.759	1.623	1.457
	0.005	8.282	5.623	4.573	3.992	3.617	3.352	3.154	2.999	2.873	2.770	2.608	2.437	2.253	2.134	2.051	1.868	1.711	1.520
100	0.100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663	1.612	1.557	1.494	1.453	1.423	1.355	1.293	1.214
	0.050	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.850	1.768	1.676	1.616	1.573	1.477	1.392	1.283
	0.025	5.179	3.828	3.250	2.917	2.696	2.537	2.417	2.321	2.244	2.179	2.077	1.968	1.849	1.770	1.715	1.592	1.483	1.347
	0.010	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.368	2.223	2.067	1.965	1.893	1.735	1.598	1.427
	0.005	8.241	5.589	4.542	3.963	3.589	3.325	3.127	2.972	2.847	2.744	2.583	2.411	2.227	2.108	2.024	1.840	1.681	1.485
120	0.100	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652	1.601	1.545	1.482	1.440	1.409	1.340	1.277	1.193
	0.050	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.834	1.750	1.659	1.598	1.554	1.457	1.369	1.254
	0.025	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157	2.055	1.945	1.825	1.746	1.690	1.565	1.454	1.310
	0.010	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.192	2.035	1.932	1.860	1.700	1.559	1.381
	0.005	8.179	5.539	4.497	3.921	3.548	3.285	3.087	2.933	2.808	2.705	2.544	2.373	2.188	2.069	1.984	1.798	1.636	1.431
150	0.100	2.739	2.338	2.121	1.983	1.886	1.814	1.757	1.712	1.674	1.642	1.590	1.533	1.470	1.427	1.396	1.325	1.259	1.169
	0.050	3.904	3.056	2.665	2.432	2.274	2.160	2.071	2.001	1.943	1.894	1.817	1.734	1.641	1.580	1.535	1.436	1.345	1.223
	0.025	5.126	3.781	3.204	2.87˙	2.652	2.494	2.373	2.278	2.200	2.135	2.032	1.922	1.801	1.722	1.665	1.538	1.423	1.271
	0.010	6.807	4.749	3.915	3.447	3.142	2.924	2.761	2.632	2.528	2.441	2.305	2.160	2.003	1.900	1.827	1.665	1.520	1.331
	0.005	8.118	5.490	4.453	3.878	3.508	3.245	3.048	2.894	2.770	2.667	2.506	2.335	2.150	2.030	1.944	1.756	1.590	1.374
200	0.100	2.731	2.329	2.111	1.973	1.876	1.804	1.747	1.701	1.663	1.631	1.579	1.522	1.458	1.414	1.383	1.310	1.242	1.144
	0.050	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878	1.801	1.717	1.623	1.561	1.516	1.415	1.321	1.189
	0.025	5.100	3.758	3.182	2.850	2.630	2.472	2.351	2.256	2.178	2.113	2.010	1.900	1.778	1.698	1.640	1.511	1.393	1.229
	0.010	6.763	4.713	3.881	3.414	3.110	2.893	2.730	2.601	2.497	2.411	2.275	2.129	1.971	1.868	1.794	1.629	1.481	1.279
	0.005	8.057	5.441	4.408	3.837	3.467	3.206	3.010	2.856	2.732	2.629	2.468	2.297	2.112	1.991	1.905	1.715	1.544	1.314
500	0.100	2.716	2.313	2.095	1.956	1.859	1.786	1.729	1.683	1.644	1.612	1.559	1.501	1.435	1.391	1.358	1.282	1.209	1.087
	0.050	3.860	3.014	2.623	2.390	2.232	2.117	2.028	1.957	1.899	1.850	1.772	1.686	1.592	1.528	1.482	1.376	1.275	1.113
	0.025	5.054	3.716	3.142	2.811	2.592	2.434	2.313	2.217	2.139	2.074	1.971	1.859	1.736	1.655	1.596	1.462	1.336	1.137
	0.010	6.686	4.648	3.821	3.357	3.054	2.838	2.675	2.547	2.443	2.356	2.220	2.075	1.915	1.810	1.735	1.566	1.408	1.164
	0.005	7.950	5.355	4.330	3.763	3.396	3.137	2.941	2.789	2.665	2.562	2.402	2.230	2.044	1.922	1.835	1.640	1.460	1.184
1000	0.100	2.711	2.308	2.089	1.950	1.853	1.780	1.723	1.676	1.638	1.605	1.552	1.494	1.428	1.383	1.350	1.273	1.197	1.060
	0.050	3.851	3.005	2.614	2.381	2.223	2.108	2.019	1.948	1.889	1.840	1.762	1.676	1.581	1.517	1.471	1.363	1.260	1.078
	0.025	5.039	3.703	3.129	2.799	2.579	2.421	2.300	2.204	2.126	2.061	1.958	1.846	1.722	1.640	1.581	1.445	1.316	1.094
	0.010	6.660	4.625	3.801	3.338	3.036	2.820	2.657	2.529	2.425	2.339	2.203	2.056	1.897	1.791	1.716	1.544	1.383	1.112
	0.005	7.915	5.326	4.305	3.739	3.373	3.114	2.919	2.766	2.643	2.541	2.380	2.208	2.022	1.900	1.812	1.615	1.431	1.125
∞	0.100	2.706	2.303	2.084	1.945	1.847	1.774	1.717	1.670	1.632	1.599	1.546	1.487	1.421	1.375	1.342	1.263	1.185	1.000
	0.050	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.752	1.666	1.571	1.506	1.459	1.350	1.243	1.000
	0.025	5.024	3.689	3.116	2.786	2.567	2.408	2.288	2.192	2.114	2.048	1.945	1.833	1.708	1.626	1.566	1.428	1.296	1.000
	0.010	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.039	1.878	1.773	1.696	1.523	1.358	1.000
	0.005	7.879	5.298	4.279	3.715	3.350	3.091	2.897	2.744	2.621	2.519	2.358	2.187	2.000	1.877	1.789	1.590	1.402	1.000

Случайные числа

Случайные числа от 0 до 9, которые для удобства сгруппированы в "блоки", могут использоваться любым систематическим образом: если, например, необходимо осуществить случайную выборку 5 из 83 элементов популяции, для идентификации этих элементов можно использовать два первых столбца, т.е. числа от 01 до 83, таким образом, выбираются числа 29, 12, 02, 69 и 11. Числа больше 83 можно не учитывать. Если числа двух первых столбцов уже исчерпаны, можно использовать столбцы 3, 4 и т.д.

29	32	95	99	57	98	08	36	97	08	65	30	47	22	00	38	60	10	01	10
12	11	80	16	17	01	03	97	59	73	74	98	73	65	85	59	74	66	37	58
87	58	22	25	55	35	72	79	28	15	69	17	42	98	72	05	47	12	40	99
02	92	42	87	57	53	53	34	55	75	83	64	09	10	19	33	29	57	62	98
69	28	63	73	98	45	61	10	43	20	08	10	43	16	81	17	62	99	09	16
11	95	68	77	86	91	76	11	63	34	15	08	35	39	37	12	74	15	00	10
06	43	41	02	13	65	23	94	48	88	88	87	03	90	77	68	98	09	17	22
68	55	98	08	39	59	85	46	66	13	42	90	86	13	29	12	38	48	27	54
41	01	06	65	10	29	29	91	86	24	45	59	04	88	17	68	31	01	91	13
46	75	71	76	88	04	42	94	41	42	39	79	14	46	13	49	37	18	28	08
80	14	13	43	24	47	61	47	42	24	24	82	12	23	54	81	33	18	96	89
30	56	60	77	80	33	67	68	31	67	73	23	45	30	55	81	51	87	68	58
53	50	41	02	98	49	97	32	43	55	75	33	51	20	99	64	76	20	80	98
84	14	75	87	37	58	51	94	06	73	27	94	23	76	77	81	72	90	45	41
08	27	89	33	87	52	24	57	50	22	22	76	60	05	79	86	58	83	88	41
97	08	50	16	41	67	40	56	13	12	68	67	36	22	08	55	76	86	45	67
97	08	37	42	48	95	90	48	34	88	19	66	38	94	64	95	07	78	23	86
70	15	04	10	34	95	57	63	75	82	88	74	28	24	66	99	52	65	36	98
06	38	31	17	38	24	98	52	67	04	95	54	89	79	45	28	05	18	60	17
63	87	79	25	86	56	74	17	45	32	53	62	09	04	86	65	87	48	82	02
17	00	56	31	14	18	56	97	91	78	85	82	06	24	88	49	17	68	51	50
17	76	35	38	19	24	47	21	09	43	09	72	02	64	66	06	78	21	70	41
57	77	32	13	60	37	68	66	11	23	30	62	97	71	02	20	13	22	00	40
35	86	97	84	91	77	73	03	37	77	50	24	54	51	40	20	66	16	34	84
72	68	64	77	89	72	77	67	45	72	25	56	78	69	72	63	86	52	07	43
91	01	78	50	50	91	99	15	36	02	74	42	55	33	19	88	35	17	58	37
70	37	55	94	53	05	78	53	23	29	15	57	70	30	88	63	20	12	64	38
11	06	17	48	24	57	50	76	81	77	30	12	92	27	19	32	63	70	97	80
60	37	89	98	61	05	51	89	47	28	34	83	98	44	66	96	84	64	64	92
37	41	11	09	04	84	38	51	91	49	23	78	53	95	40	17	73	23	04	70
28	97	38	27	97	54	95	94	54	79	93	88	00	82	39	61	93	78	07	88
14	29	17	18	84	03	10	62	15	70	01	15	06	30	97	79	55	98	79	39
81	70	53	83	20	25	26	56	55	56	33	58	74	21	76	94	24	80	12	50
08	20	90	25	43	22	81	74	51	76	53	39	59	35	34	46	55	54	73	50
61	95	25	85	66	34	76	39	98	88	45	57	64	11	17	06	43	35	27	09
64	58	31	05	45	77	25	20	02	09	36	87	63	01	10	08	01	19	19	06
75	49	97	87	79	31	66	57	89	56	56	97	71	43	65	62	36	77	50	87
66	95	10	78	42	24	91	82	74	29	00	53	44	70	18	23	48	09	90	99
85	37	61	48	07	99	13	01	16	94	37	31	28	96	59	77	62	24	95	84
06	87	15	09	48	31	18	66	87	11	19	71	67	20	93	92	02	96	15	65
11	15	95	59	69	81	75	75	88	69	95	12	75	69	18	10	60	35	31	47
03	64	44	33	46	16	02	28	14	33	61	57	28	33	96	47	49	86	85	83
68	89	57	51	94	84	09	80	37	90	52	99	85	52	49	66	63	69	11	31
43	13	09	12	00	65	69	54	11	00	20	94	22	93	90	16	82	64	27	46
42	68	71	56	74	17	71	63	80	81	02	41	49	27	92	44	44	13	45	21
12	55	09	80	30	50	34	96	31	71	19	21	79	42	17	57	04	04	19	00
88	84	87	74	01	39	99	02	75	76	61	88	97	89	06	97	15	70	26	27
49	27	92	08	87	65	12	32	27	96	11	26	30	88	48	89	29	73	50	47
46	51	54	92	06	44	85	83	14	78	68	83	33	17	03	10	99	10	17	34
34	96	78	90	18	41	44	69	10	30	48	98	32	76	12	81	29	83	02	87

Случайные числа - продолжение

80	07	15	41	15	37	42	39	24	45	48	73	61	15	44	74	40	27	26	47
39	08	51	67	63	03	76	76	86	09	39	32	62	77	60	85	37	14	69	76
51	32	57	06	49	13	01	25	98	83	44	96	92	78	37	24	49	35	54	52
84	46	17	46	71	53	88	78	30	71	53	85	55	10	93	40	05	66	72	38
04	88	20	78	89	94	31	36	83	74	51	25	28	43	54	76	57	08	21	23
21	45	86	26	12	21	28	37	56	47	86	18	38	39	18	89	99	62	81	98
71	38	27	31	40	52	36	03	51	54	83	14	51	17	86	77	66	84	50	84
78	50	39	32	55	17	25	06	90	90	69	48	70	68	22	07	85	07	95	84
22	76	93	40	26	30	77	61	71	74	81	13	73	21	99	00	47	52	43	18
25	21	70	62	69	05	05	58	75	92	85	60	50	87	81	35	80	83	42	16
96	79	06	87	51	04	17	61	42	12	64	77	45	06	55	68	19	39	17	22
97	76	01	89	33	70	46	23	44	83	99	55	95	03	41	89	33	49	89	86
78	03	18	58	00	47	18	01	33	49	99	55	54	70	65	34	76	58	86	20
09	63	31	80	30	17	11	75	34	81	25	45	91	80	50	25	64	70	05	48
61	33	89	72	78	98	26	56	88	66	51	69	71	48	13	71	40	57	31	22
64	83	61	76	37	68	22	25	09	82	53	59	78	66	81	66	45	56	64	78
18	93	65	67	39	81	96	44	68	46	96	50	08	71	70	81	23	32	89	61
86	84	70	40	22	89	25	42	62	69	95	98	59	26	69	55	33	62	91	88
96	57	56	48	81	92	77	95	43	50	29	89	07	58	10	83	66	04	15	74
54	35	65	28	09	99	04	41	86	60	69	54	82	74	49	86	82	25	07	29
18	79	09	01	55	60	31	19	19	48	01	89	54	63	96	70	99	15	71	84
19	78	77	63	36	52	38	88	16	92	23	42	49	79	27	15	09	94	49	35
55	71	79	75	30	29	13	32	60	07	33	73	61	89	63	64	17	15	21	39
38	58	83	62	94	73	84	48	95	17	79	74	78	38	09	37	35	75	74	70
78	29	66	85	65	45	79	70	88	92	73	24	71	71	63	70	47	56	70	28
87	55	81	22	04	62	21	45	81	82	43	96	17	70	61	80	59	10	59	00
06	98	70	24	03	20	67	45	67	65	04	61	76	89	25	13	73	06	41	16
33	08	62	21	90	70	72	16	01	23	26	05	10	33	23	23	03	07	46	08
54	03	25	45	50	40	58	15	41	07	16	24	16	63	46	64	27	85	27	47
68	90	88	08	25	70	23	82	53	40	51	91	84	67	84	08	09	76	19	19
90	18	00	18	76	88	55	07	52	00	30	04	83	72	04	74	87	56	90	80
70	07	33	78	52	59	92	46	58	33	61	42	31	47	58	89	32	02	55	36
19	13	05	69	12	74	49	85	21	49	18	11	60	96	94	04	74	26	23	44
95	70	86	00	19	44	74	51	22	34	63	14	11	30	48	54	71	78	97	12
65	12	41	20	32	33	72	70	71	24	51	39	43	28	90	51	14	46	17	40
15	53	57	75	61	54	95	63	75	51	28	43	39	55	90	58	01	50	31	88
60	27	72	94	00	25	71	09	76	19	66	69	44	09	39	12	60	43	02	52
57	91	58	68	24	78	33	54	25	46	08	87	72	85	28	98	89	67	68	92
40	15	42	80	71	35	81	75	95	40	04	85	70	88	19	44	75	50	63	41
23	97	89	48	74	96	60	10	40	24	33	88	86	93	30	79	96	32	25	34
48	25	55	19	87	97	39	79	66	73	50	78	72	75	08	78	66	69	13	35
24	58	57	51	61	90	39	52	91	33	77	67	76	78	40	42	05	70	73	08
60	22	38	11	98	95	66	00	95	19	32	99	90	77	55	50	86	94	41	83
84	89	06	96	10	47	83	22	11	81	19	13	48	21	71	99	16	81	88	56
30	80	70	60	93	09	74	04	99	72	67	91	91	75	20	36	08	45	28	35
23	95	78	32	20	71	90	24	20	66	09	27	14	97	94	78	67	45	20	62
48	52	58	73	69	63	54	77	76	89	09	15	50	05	85	91	12	10	12	29
33	69	72	87	15	96	24	09	14	84	41	57	16	17	78	18	46	46	23	04
71	71	53	72	84	65	86	16	70	43	62	10	33	15	61	60	80	73	18	21
29	53	27	21	49	53	31	68	21	10	17	47	35	74	84	18	58	07	17	32
17	70	60	84	24	50	82	33	67	40	15	88	50	22	54	28	39	46	14	28
98	37	60	93	52	27	20	93	10	62	90	69	27	96	44	54	01	13	81	14
36	39	86	14	17	56	74	44	76	20	77	74	52	52	56	06	99	78	52	67
53	17	93	61	99	15	08	47	04	09	46	95	53	02	57	60	02	02	99	83
05	38	06	80	55	75	49	12	95	96	98	63	46	51	49	74	97	71	95	88

ОГЛАВЛЕНИЕ

Переводное издание

М. Эддоус, Р. Стэнсфилд

МЕТОДЫ ПРИНЯТИЯ РЕШЕНИЙ

Редактор *О.Л. Борисова*
Корректор *Л.И. Ганина*
Оформление художника *И.А. Науменко*

Оригинал-макет выполнен *Л.Н. Гвоздевой*

Лицензия № 061071 от 17.04.92
Подписано в печать 13.01.97. Формат 70x100 1/16
Усл. печ. л. 47,73. Тираж 10 000 экз. (1-й завод - 5 000)
Заказ 64

Издательство "Аудит"
Генеральный директор *В.Н. Закаидзе*

123298, Москва, Тепличный пер., 6
Тел.: (095) 194-00-15, 131-97-01. Тел./Факс: (095) 194-00-14

Отпечатано в типографии издательства "Дом печати"
432601, г. Ульяновск, ул. Гончарова, 14