Міністерство освіти і науки України

Тернопільський державний економічний університет

Курс лекцій з дисципліни «Статистика».

Частина 1. Теорія статистики.

Усі цитати, цифровий і фактичний матеріал, бібліографічні дані перевірені. Написання одиниць відповідає стандартам. Зауваження рецензентів враховано.
Автор____________ Є.І. Ткач

Затверджено на засідання кафедри статистики ТДЕУ. Протокол № 8від 28 березня 2006 року.
Зав. кафедри статистики

__________________ О.В.Кустовська

Рекомендовано до видання Вченою радою Тернопільського державного економічного університету.
Протокол № 10 від 26.05.2006 р.
Вчений секретар

__________________М.Я.Шелестовська

Вимогам, що ставляться до навчально-методичної літератури, відповідає.
До друку і в світ дозволяю.
Перший проректор

__________________Г.П.Журавель

Тернопіль – ТДЕУ – 2006

УДК 311.1
ББК 60.6я73
С-48

Рецензенти:

А.А. Григорук, *кандидат економічних наук, доцент, завідувач кафедри філософії та економічної теорії Тернопільського національного педагогічного університету ім. В. Гнатюка.*
О.І. Демчишин*, кандидат фізико-математичних наук, доцент, завідувач кафедри загальноекономічних дисциплін Інституту економіки і підприємництва.*
С.І. Шкарабан, *доктор економічних наук, професор, завідувач кафедри економічного аналізу Тернопільського державного економічного університету.*

Курс лекцій з дисципліни «Статистика». Частина 1. Теорія статистики: В.П. Сторожук, О.В. Кустовська, Є.І. Ткач, І.М. Шост та ін.; За ред. Є.І. Ткача – Тернопіль: Економічна думка, 2006 . – 224 с.

ISBN

Навальне видання представлено у вигляді курсу лекцій із статистики. В курсі лекцій детально розглянуто методику обчислення статистичних показників, необхідних для проведення комплексного аналізу явищ та процесів, діяльності підприємств, організацій і установ. Викладання матеріалу супроводжується наглядними прикладами, таблицями, графіками та розв'язанням типових завдань.

Посібник розраховано на студентів економічних спеціальностей вищих навчальних закладів.

УДК 311.1
ББК 60.6я73
С-48

Навчальне видання

Збірник задач для самостійної та індивідуальної роботи студентів з дисципліни

«Статистика»

Курс лекцій з дисциплін «СТАТИСТИКА»

Частина 1 Теорія статистики

ПЕРЕДМОВА

Інтерес до статистики постійно зростає в усьому світі. Праця економіста любої спеціалізації неминуче зв'язана із збиранням, обробкою і аналізом статистичних матеріалів. Тому вивчення і оволодіння статистичною наукою при підготовці економістів високої кваліфікації має велике значення в системі вищої економічної освіти.

В нашій країні увага до статистичної науки надзвичайно загострена у зв'язку з проведенням економічних реформ, які зачіпають інтереси всіх людей.

Для підняття статистики до сучасного наукового рівня, задоволення потреб системи управління та інших соціально-економічних суб'єктів в якісній, повній, різноманітній і своєчасній інформації, вкрай необхідна докорінна її перебудова.

Важливою умовою правильного сприйняття і практичного використання статистичної інформації, кваліфікованих висновків і обґрунтування прогнозів є завдання статистичної методології вивчення кількісної сторони соціально-економічних яви, природи масових статистичних сукупностей, пізнавальних властивостей статистичних показників, умов їх застосування в економічному дослідженні.

Одним із основних завдань статистики є оптимізація звітності, проведення об'єму інформації о сучасних потреб системи управління в умовах переходу до ринку. Потрібно впроваджувати замість суцільної звітності вибіркові обстеження, одноразові обліки чи опитування, що приведе до оперативного і поглибленого аналізу.

Забезпечення надійності і достовірності статистичної інформації можливе через підвищення наукового рівня всієї статистичної методології, наближення її до методології і стандартів світової статистичної практики.

В даний час перед статистикою стоять проблеми подальшого вдосконалення системи показників, прийомів і методів збору, обробки, зберігання і аналізу статистичної інформації. Це має важливе значення для розвитку і підвищення ефективності автоматизованих систем управління, створення автоматизованих банків даних, розподільчих банків даних, які в свою чергу могли б сприяти створенню Єдиної статистичної інформації системи (ЄСІС), що надасть можливість запровадити в практику сучасні статистичні методи аналізу, імітаційні та прогнозні методи.

В запропонованому посібнику міститься системний вклад загальних категорій, принципів і методів статистичної науки, теоретичних основ соціально-економічних методів аналізу і прогнозування із застосуванням кореляційно-регресійного, табличного і графічного методів.

В посібнику «Статистика» послідовно розглядаються питання, які виникають на стадії статистичного спостереження, зведення і групування матеріалів спостереження та їх наступної обробки.

Даний посібник написаний для студентів економічних вузів і факультетів.

ЛЕКЦІЯ 1. СТАТИСТИКА ЯК НАУКА. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

ПЛАН:

1.1. Предмет, метод та завдання статистики в Україні.

1.2. Поняття про статистичне спостереження і його завдання.

1.3. Основні організаційні форми, види і способи статистичного спостереження.

1.4. Помилки статистичного спостереження і способи контролю добутих даних.

1.1. Предмет, метод та завдання статистики в Україні

Інтерес до статистики постійно зростає в усьому світі. Праця економіста любої спеціалізації неминуче зв’язана із збиранням, обробкою і аналізом статистичних матеріалів. Тому вивчення і оволодіння статистичною наукою при підготовці економістів високої кваліфікації має велике значення в системі вищої економічної освіти.

В нашій країні увага до статистичної науки надзвичайно загострена у зв’язку з проведенням економічних реформ, які зачіпають інтерес всіх людей.

Для підняття статистики до сучасного наукового рівня, задоволення потреб систем управління та інших соціально-економічних суб’єктів в якісній, повній, різноманітній і своєчасній інформації, вкрай необхідна докорінна її перебудова.

Важливою умовою правильного сприйняття і практичного використання статистичної інформації, кваліфікованих висновків і обґрунтованих прогнозів є знання статистичної методології кількісної сторони соціально-економічних явищ, природи масових статистичних сукупностей, пізнавальних властивостей статистичних показників, умов їх застосування в економічному дослідженні.

Термін «статистика» походить від латинського слова «статус» (status), що означає суму знань про державу.

В даний час статистика має три основних значення: 1) під статистикою розуміють практичну діяльність працівників статистичних органів, які збирають, обробляють і аналізують дані про соціально-економічний розвиток країни в цілому, а також окремих її регіонів, галузей економіки, конкретних підприємств і населення; 2) статистикою вважають статистичні дані подані в звітах підприємств, організацій і установ, а також опубліковані в статистичних збірниках, довідниках і періодичній пресі; 3) статистикою називають спеціальну науку, яка займається розробкою теоретичних положень і методів її практичного використання.

Між статистичною наукою і практичною діяльністю існує тісний зв'язок і взаємозалежність. Статистична наука використовує інформацію практичної діяльності господарських організацій, узагальнює її і розробляє методи проведення статистичних досліджень. В свою чергу, підприємства, організації і установи для практичної діяльності використовують теоретичні розробки і положення статистичної науки для розв'язання конкретних управлінських завдань.

Статистика являється складною і багатогранною наукою, яка вивчає суспільні явища і процеси з їх кількісної сторони. Як навчальна дисципліна статистика включає в себе цілий ряд розділів, таких як загальна теорія статистики, соціально-економічна статистика і серію галузевих статистик (статистику промисловості, статистику сільського господарства, статистику капітальних вкладень, статистику торгівлі та інші). Таким чином, курс «Загальна теорія статистики» відкриває цикл статистичних дисциплін.

Кожний економіст вищої кваліфікації повинен вміти читати статистичні цифри, аналізувати їх і використовувати в своїй практичній роботі для обґрунтування своє позиції і своїх висновків.

Кожна наука володіє своїми суттєвими особливостями, які відрізняють її від інших наук і дають їй право на самостійне існування як осібної галузі знань. Головна особливість всякої науки заключається в предметі пізнання, в принципах і методах його вивчення, які в сукупності утворюють його методологію.

Отже, **предметом статистики** є наука, яка вивчає кількісну сторону масових явищ соціально-економічного життя у нерозривному зв'язку з їх якісним змістом в конкретних умовах місця і часу.

Наявність будь-якої властивості у окремого елемента є випадковістю, а при об'єднанні їх в єдине ціле, вплив випадковостей нівелюється.

Законом великих чисел називається принцип, відповідно до якого закономірність масових явищ може проявлятися тільки при достатньо великому числі випадків.

Для своїх досліджень статистика розробляє комплекс методів і засобів, сукупність яких формує **статистичну методологію**.

Загальною основою розробки і застосування статистичної методології є положення соціально-економічної теорії і принцип діалектичного методу пізнання явищ суспільного життя.

У відповідності з діалектичним методом пізнання статистика вивчає всі явища і процеси в їх взаємозв'язку і взаємозалежності, в русі і зміні, виділяючи їх різні типи і форми, встановлює нове і прогресивне та визначає напрямки розвитку.

Весь арсенал статистичних методів зводиться до трьох основних етапів (стадій) статистичного дослідження: 1) статистичного спостереження; 2) зведення і групування; 3) економічного аналізу.

В даний час головним обліково-статистичним центром в країні є Державний комітет статистики України (Держкомстат України). Він здійснює керівництво статистикою України у відповідності із статтею 92 п.12 Конституції України «Виключно законами України визначаються: організація і діяльність органів виконавчої влади, основи державної служби, організації державної статистики та інформатики». В його завдання входить подання офіційної статистичної інформації Президенту, Уряду, Парламенту, громадським і міжнародним організаціям, розробка науково обґрунтованої статистичної методології, координація статистичної діяльності регіональних органів виконавчої влади, аналіз економіко-статистичної інформації, складання національних рахунків і балансових розрахунків.

Система органів державної статистики утворена у відповідності з адміністративно-територіальним розподілом України. В Автономній Республіці Крим діє Державний комітет по статистиці, в областях – обласні управління статистики, а в районах і містах – відділи статистики.

Поряд із загальнодержавною статистикою існує відомча статистика, яка обслуговує підприємства, об'єднання, відомства і міністерства.

Статистичні дані використовують усі науки для обґрунтування, перевірки, ілюстрації своїх висновків в конкретних умовах місця і часу.

Виходячи з цього, можна визначити такі основні завдання статистики:

1) систематичне спостереження і економічний аналіз матеріалів які характеризують хід виконання виробничих програм і подання їх у відповідні керівні структури для прийняття конкретних рішень;

2) вивчення і розробка цільових комплексних програм з соціально-економічних проблем конкретних регіонів і держави в цілому;

3) удосконалення системи статистичних показників, які характеризують розвиток і результати соціально-економічних явищ і процесів;

4) економічний аналіз стану і розвитку галузей економіки;

5) вивчення чинників підвищення ефективності суспільного виробництва;

6) створення загальнодержавної автоматизованої системи збирання, обробки і передачі інформації;

7) аналіз демографічних процесів.

Управляти складними соціальними і економічними системами можна лише володіючи оперативною, вірогідною і повною статистичною інформацією.

1.2. Поняття про статистичне спостереження і його завдання

Для отримання повних і точних даних про стан будь-якого явища на певний момент часу, або про результати його розвитку за відповідний період часу проводять статистичне дослідження, яке складається з трьох послідовних етапів: 1) статистичного спостереження; 2) зведення і групування матеріалів статистичного спостереження; 3) економічного аналізу даних, отриманих в результаті зведення і групування.

Етапи статистичного дослідження тісно пов'язані між собою, тому успіху можна досягти лише при добре підготовленій і організованій роботі на всіх його стадіях.

Статистичне спостереження – як перша стадія статистичного дослідження, являє собою планомірну, систематизовану, науково організовану роботу по збиранню і реєстрації масових первинних даних про явища і процеси суспільного життя.

Ці дані в залежності від мети статистичного дослідження можуть бути різними за своїм змістом і способом отримання. Вони пізніше систематизуються, групуються, обробляються, аналізуються і узагальнюються.

Саме статистичне спостереження також складається з трьох етапів: а) підготовки спостереження; б) збирання матеріалу; в) контроль зібраного матеріалу.

На підготовчому етапі статистичного спостереження, відповідно до його мети і завдань, розробляється програма і організаційний план проведення спостереження. Тут вирішують питання про зміст вихідної інформації, яким способом, якими засобами і в які терміни буде проведений облік фактів, як будуть організовані збирання і контроль отриманих первинних матеріалів. Повинні бути враховані також відповідні вимоги до оформлення цих матеріалів, яких вимагає техніка подальшої їх обробки на ПЕОМ.

Від якості статистичного спостереження залежить успіх всього статистичного дослідження. Статистичне спостереження повинно бути організоване таким чином, щоб в результаті його проведення були отримані об'єктивні, вірогідні, повні дані про досліджуване явище і, по можливості, в короткий термін. Це дасть змогу зробити правильні узагальнення і висновки.

Соціально-економічні явища і процеси, які спостерігаються, повинні мати наукову і практичну цінність та виражати їх типи.

Одним з важливих завдань статистичного спостереження є ретельна і всебічна перевірка якості зібраних матеріалів для забезпечення їх вірогідності.

Наукова організація статистичного спостереження потрібна для створення найкращих умов для отримання об'єктивно правильних матеріалів, які б давали змогу передбачити майбутні ситуації і робити обґрунтовані прогнози.

Статистичне спостереження проводять за строго визначеним планом, який включає програмно-методологічні і організаційні питання.

До програмно-методологічних відносять питання, зв'язані з розробкою програми спостереження, вивченням мети, об'єкта і одиниці спостереження, проектування формулярів і текстів інструкцій, встановленням джерел і способів збирання інформації.

До організаційних відносять питання про органи спостереження, терміни і місце проведення спостереження, складання попередніх списків одиниць досліджуваної сукупності, розставлення і підготовка кадрів та деякі інші.

Програма статистичного спостереження визначається правильно встановленими і конкретно сформульованими завданнями дослідження. Тому, перш за все, потрібно чітко сформулювати мету всієї роботи, а потім вирішувати всі інші питання програми спостереження.

Мета спостереження являє собою основний результат статистичного дослідження. Чітке і конкретне формулювання мети спостереження потрібне для того, щоб не збирати зайвих, непотрібних і неповних даних.

Завдання статистичного дослідження необхідно відобразити в статистичних показниках, для чого розробляють і складають макети кінцевих статистичних таблиць, в які заносять результати всієї роботи.

При організації статистичного спостереження важливо точно визначити об'єкт спостереження.

Об'єктом статистичного спостереження називається сукупність одиниць досліджуваного явища, про які повинні бути зібрані потрібні статистичні дані.

Визначивши об'єкт статистичного спостереження потрібно вказати на його важливі ознаки і основні розпізнавальні риси, тобто встановити межі досліджуваної сукупності.

При періодичному обстеженні потрібно слідкувати, щоб досліджувана сукупність була більш менш однорідною. Для цього статистика використовує **ценз** – обмежувальну ознаку, яку повинні задовольняти всі одиниці досліджуваної сукупності.

Поряд з визначенням об'єкта статистичного спостереження визначають також одиницю сукупності і одиницю спостереження.

Одиницею спостереження називають той первинний складовий елемент об'єкта статистичного спостереження, який є носієм ознак, що підлягають реєстрації.

Одиницею сукупності називається та первинна ланка, від якої отримують необхідні статистичні відомості.

Після того як визначені об'єкт, одиниця спостереження і одиниця сукупності, потрібно розробити зміст програми спостереження, що є основним питанням статистичного спостереження.

Програмою статистичного спостереження називається перелік чітко сформульованих питань, на які намічають отримати відповідні в процесі обстеження. Від якості її розробки залежать якість і цінність зібраного статистичного матеріалу.

Статистичним формуляром називається документ особливої форми куди збирають і записують відповіді на питання програми спостереження. Обов'язковим елементом статистичного формуляра є титульна і адресна частини, які необхідні для перевірки зібраних даних і їх наступного розроблення.

В титульній частині представляють назву статистичного спостереження, назву органу, який проводить спостереження, ким і коли затверджений формуляр, присвоєний йому номер.

В адресній частині записують точний адрес одиниці або сукупності одиниць спостереження і деякі інші відомості.

В практиці застосовують два види або дві системи статистичних формулярів: індивідуальну (карткову) і спискову.

Індивідуальним називається такий статистичний формуляр, в який заносять відомості про одну одиницю спостереження (листок обліку кадрів, одне підприємство, один робітник і т.д.).

Списковим називається такий статистичний формуляр, в якому реєструються відомості по декількох одиницях спостереження (відомість на заробітну плату, екзаменаційна відомість тощо).

Відповіді які заносять до формулярів виражаються словами, цифрами або у формі альтернативних відповідей (так чи ні).

Інструкцією називають сукупність роз'яснень і вказівок до програми статистичного спостереження. Вона повинна бути написана коротко, просто, вказівки повинні бути конкретними і чіткими.

1.3. Основні організаційні форми, види і способи статистичного спостереження

З метою успішного проведення спостереження складають організаційний план.

Організаційний план – це основний документ, в якому зосереджені розв'язки важливих питань організації і проведення статистичного спостереження.

До організаційних питань статистичного спостереження належать: визначення об'єкта, місця, часу і термінів спостереження; постановка мети і завдань спостереження; визначення органів спостереження; визначення прав і обов'язків окремих установ і організацій, які беруть участь у спостереженні; підготовчі роботи проведення спостереження; добір, навчання і інструктаж масових кадрів, потрібних для проведення спостереження; розмноження і розсилки формулярів спостереження; порядок здачі і приймання матеріалів спостереження; порядок отримання і подання попередніх і остаточних підсумків спостереження та інші практичні питання.

Організаційні плани складаються статистичними органами держави починаючи з вищих і закінчуючи нижчими ланками.

Вищі статистичні органи головну увагу приділяють розв'язку загальних організаційно-методологічних питань, таких як визначення завдань спостереження, його об'єкта, одиниці і термінів проведення та ін.

Нижчі ланки статистичних органів розв'язують в основному конкретні організаційні завдання на місцях.

Статистичне спостереження в загальнодержавному масштабі організовує Державний комітет статистики України та їх місцеві органи.

Міністерства, відомства, наукові і інші установи проводять статистичні спостереження в основному локального характеру.

Місце спостереження – це місце де проводиться реєстрація фактів спостереження, які записуються у спеціальних формулярах.

Вибір місця спостереження повинен забезпечити повноту охоплення об’єкта спостереження, високу якість фіксації даних і простоту проведення спостереження.

Вся різноманітність форм статистичного спостереження здійснюється в двох основних формах: 1) у формі звітності підприємств, організацій і установ; 2) у формі спеціально організованого спостереження (перепис населення, облік багаторічних насаджень, переоцінка основних фондів і т.п.).

Звітністю називають таку організацію статистичного спостереження, за якою відомості поступають в статистичні органи від підприємств, організацій і установ у вигляді обов’язкових звітів про їх діяльність в точно встановлені терміни.

Статистичну звітність складають на підставі даних первинного обліку.

Первинним обліком в статистиці називається ведення систематичних записів у формах первинних облікових документів про різні явища і процеси, які стосуються діяльності підприємств, організацій чи установ.

Звітність подають вищим організаціям і органам Державного комітету статистики в порядку, встановленому Державним комітетом статистики України щодо кожної форми.

В нашій державі розрізняють дві основні форми звітності: а) загальнодержавну; б) внутрівідомчу.

Загальнодержавна звітність обов’язкова для всіх підприємств, організацій і установ. Вона подається уряду міністерствами і відомствами у зведеному вигляді безпосередньо, або через Державний комітет статистики України.

Внутрівідомча – це звітність розроблена міністерствами і відомствами для своїх оперативно-господарських потреб.

Звітність в даний час є одним з основних джерел статистичної інформації про соціально-економічний розвиток держави.

Поряд із звітністю в практиці різних статистичних досліджень широко використовують спеціально організовані статистичні спостереження.

Спеціально-організованим статистичним спостереженням називається таке спостереження, яке проводиться із спеціальною метою на якусь дату для отримання інформації, котру, в силу певних причин, не можна зібрати із звітів, або для перевірки і уточнення даних звітності.

Одним з основних видів спеціально організованого спостереження є переписи.

Перепис – це спеціально організоване статистичне спостереження великого масштабу, яке охоплює всю країну, або значну її частину і проводиться одночасно за єдиною програмою. Його метою є визначення чисельності, складу, стану і розміщення досліджуваного об'єкта на встановлений критичний момент.

За повнотою охоплення спостереженням досліджуваного об'єкта розрізняють два його види: а) суцільне; б) несуцільне.

Суцільним називається таке спостереження, при якому обстеженню і реєстрації підлягають всі без винятку одиниці досліджуваної сукупності.

Несуцільним називається таке спостереження, при якому обстеженню і реєстрації підлягають не всі одиниці досліджуваної сукупності, а лише певна їх частина.

Несуцільні спостереження мають ту перевагу перед суцільним, що вони вимагають значно менше затрат сил і засобів, дозволяють застосувати докладнішу програму і досконаліший спосіб обліку фактів, швидше підводити підсумки обстеження і, отже, підвищують оперативність статистичних матеріалів.

В багатьох випадках несуцільне спостереження є єдино можливим способом дослідження статистичної сукупності.

Несуцільні спостереження в статистиці суттєво доповнюють основні матеріали, отримані в результаті суцільних спостережень.

В практиці статистичної роботи застосовують наступні види несуцільного спостереження: 1) вибіркове спостереження; 2) монографічне спостереження; 3) метод основного масиву; 4) анкетне.

Вибірковим називається таке спостереження, при якому вся сукупність фактів характеризується за деякою її частиною, відібраною випадково. В його основні лежить випадковий відбір одиниць для обстеження, що гарантує незалежність результатів вибірки від волі осіб, які її проводять і не допускає тенденційних помилок.

Монографічне спостереження являє собою детальне вивчення і опис окремого об’єкта, або невеликої їх кількості за розширеною програмою. Таке спостереження проводиться з метою виявлення певних тенденцій і закономірностей розвитку явища, або для вивчення і розповсюдження передового досвіду окремих підприємств, організацій і установ. Воно також використовується для виявлення недоліків в роботі окремих підприємств з метою їх усунення і недопущення в майбутньому.

Метод основного масиву заклечається в тому, що з усієї сукупності одиниць спостереженню підлягає переважна їх частка, в яку, як правило, попадають найбільш суттєві і крупні одиниці досліджуваної сукупності. Взяті разом вони мають значну питому вагу в сукупності за однією чи декількома основними для даного дослідження ознаками.

Анкетне спостереження ґрунтується на принципі добровільного заповнення адресатами надісланих або розданих їм спеціальних анкет з метою отримання потрібної для дослідження інформації. Недоліком анкетного спостереження є те, що перевірити достовірність зібраного матеріалу досить складно або неможливо. Його застосовують у випадках, коли не вимагається висока точність інформації, а лише наближені її характеристики.

За часом проведення статистичне спостереження поділяють на: а) поточне; б) періодичне; в) одноразове.

Поточним називається таке спостереження, яке ведеться систематично при безперервній реєстрації фактів в міру їх виникнення. Наприклад, реєстрація громадських актів (народження, смерть, шлюб, розлучення), облік виходів працівників на роботу, облік виробленої продукції на підприємстві та ін.

Періодичним називається таке спостереження, яке повторюється через певні, заздалегідь установлені рівні проміжки часу. Такі спостереження, як правило, характеризують стан явища на певний момент часу. Наприклад, щорічний перепис худоби станом на 1 січня, облік чисельності працівників, товарних запасів, залишків матеріальних цінностей на 1 число кожного місяця і т.д.

Одноразовим називається таке спостереження, яке проводиться в міру потреби один раз, або час від часу, без дотримання точної періодичності (переписи виробничого устаткування, переписи багаторічних плодово-ягідних насаджень та ін.).

За способом збирання статистичних даних розрізняють: а) безпосереднє спостереження; б) документальне спостереження; в) опитування.

Безпосереднім називається таке спостереження, при якому самі реєстратори збирають потрібні дані шляхом особистих замірювань, зважувань і підрахунків одиниць об'єкта і на цій основі проводять записи у формулярі спостереження.

Документальним називається таке спостереження, при якому потрібні дані збирають і записують у формуляри на підставі використання різної документації.

Опитування – це таке спостереження, при якому відповіді на питання записують зі слів опитуваної особи. Так проводять перепис населення.

В статистичній практиці використовують наступні три способи опитування: а) усне; б) самореєстрація; в) кореспондентський спосіб.

При **усному** опитуванні спеціально виділений працівник (реєстратор) розмовляє з опитуваною особою і з її слів сам заповнює формуляр.

При **самореєстрації** опитуваній особі вручають бланк обстеження, пояснюють питання і опитувана особа сама заповнює формуляр. В назначений час обліковець збирає заповненні формуляри і перевіряє повноту і правильність їх заповнення.

Кореспондентський спосіб полягає в тому, що інформацію в органи, які проводять спостереження, надсилають добровільні кореспонденти, які попередньо отримують від статистичних органів формуляри і інструкції щодо їх заповнення.

1.4. Помилки статистичного спостереження і способи контролю добутих даних

Точність статистичного спостереження являється важливою і основною вимогою органів державної статистики. Однак, хоч як би старанно не було підготовлене статистичне спостереження, в процесі його проведення трапляються помилки, які призводять до зниження його точності.

Точністю статистичного спостереження називають ступінь відповідності значення будь-якої ознаки визначеної за допомогою статистичного спостереження її дійсному значенню. Чим ближчі значення ознак, отриманих в результаті статистичного спостереження до їх фактичних значень, тим точніше проведене спостереження.

Точність статистичного спостереження визначається як відношення даних спостереження до дійсних значень досліджуваних величин, або як різниця між ними.

Помилками спостереження називаються розходження між встановленими статистичним спостереженням і дійсними значеннями досліджуваних величин. Помилки спостереження виникають внаслідок неточностей при збиранні і реєстрації значень досліджуваних ознак.

Недопущення і попередження помилок є одним з важливих завдань організації і проведення статистичного спостереження. Невірні статистичні дані можуть призвести до прорахунків в державному управлінні економікою, серйозних помилок в науковому плануванні і прогнозуванні та інших негативних наслідків. Тому в Україні встановлена сувора відповідальність посадових осіб за навмисні викривлення статистичних даних.

В залежності від характеру, ступеня впливу на кінцеві результати, джерел і причин виникнення неточностей розрізняють наступні типи помилок статистичного спостереження: а) помилки реєстрації; б) помилки репрезентативності. Кожний з цих типів помилок ділиться на випадкові (ненавмисні) і систематичні (навмисні).

Помилки реєстрації виникають внаслідок неправильного встановлення фактів в процесі спостереження, помилкового запису їх значень, або обох причин разом.

Випадковими називаються помилки реєстрації, які можуть виникати внаслідок різних випадкових причин. Наприклад, опитувана особа може обмовитись, а реєстратор недочути чи випадково переставити місцями цифри, замість віку 23 роки записати 32 і навпаки. Такі неточності діють в протилежних напрямках і при достатньо великому числі спостережень взаємно погашаються.

Систематичні помилки реєстрації виникають внаслідок певних причин, діють в одному і тому ж напрямку і спричиняють серйозні викривлення загальних результатів статистичного спостереження. Наприклад, під час перепису населення опитувані особи часто округлюють свій вік, як правило, на цифрах, які закінчуються «5» і «0». Замість 34-36 років говорять 35, замість 49-51 говорять, що їм 50 років і т.п. Внаслідок цього виходить, що 35, 40, 45, 50 – річних громадян значено більше ніж 34, 41, 46, 51 – річних.

Статистичні помилки реєстрації можуть бути внаслідок свідомого викривлення фактів. Це навмисні приписки або приховування у звітах фактичних даних.

Помилки реєстрації виникають як при суцільному, так і при несуцільному спостереженні.

На відміну від помилок реєстрації, помилки репрезентативності властиві тільки несуцільному спостереженню.

Помилками репрезентативності називаються відхилення значень ознак відібраної і обстеженої частини сукупності від значень ознак всієї досліджуваної сукупності.

Випадкові помилки репрезентативності виникають внаслідок того, що відібрана випадковим, неупередженим способом частина досліджуваної сукупності недостатньо повно відтворює всю сукупність в цілому.

Систематичні помилки репрезентативності виникають внаслідок порушення принципів неупередженого, випадкового відбору одиниць для обстеження.

З метою отримання в процесі статистичного спостереження високоякісних матеріалів, статистичні органи здійснюють постійний контроль за ходом проведення спостереження, систематично перевіряють стан первинного обліку і звітності на підприємствах, організаціях і установах.

Після закінчення спостереження матеріали, зібрані в процесі його проведення, старанно перевіряються за повнотою охоплення об'єкта спостереження, якістю заповнення формулярів і інших документів.

Статистика використовує два способи контролю матеріалів спостереження: а) арифметичний (лічильний); б) логічний.

Арифметичний контроль полягає в лічильній перевірці підсумкових даних звітів або формулярів і погодженні тих показників, які взаємозв'язані між собою і можуть бути виведені одні з одних. Наприклад, в шаховій таблиці любого значення підсумки рядків і колонок повинні співпадати, а якщо такого співпадання не має, тоді шукають помилку в рядках чи колонках.

Логічний контроль полягає в співставленні взаємозв'язаних між собою відповідей на питання формуляра статистичного спостереження і виясненні їх логічної сумісності. Якщо виявляють логічно несумісні відповіді, шляхом подальшого співставлення з відповідями на інші питання встановлюють яка з відповідей записане невірно. Наприклад, якщо у формулярі переписного листа перепису населення записано, що опитувана особа у віці 7 років має сім'ю, вищу освіту, працює лікарем, то зрозуміло, що неправильно записаний вік.

Основною умовою успішного проведення любого статистичного дослідження на всіх його стадіях, в тому числі і на стадії спостереження є висока якість зібраного матеріалу.

ЛЕКЦІЯ 2. ЗВЕДЕННЯ І ГРУПУВАННЯ СТАТИСТИЧНИХ МАТЕРІАЛІВ. СТАТИСТИЧНІ ТАБЛИЦІ

ПЛАН:

2.1. Зміст і завдання статистичного зведення.

2.2. Зміст і завдання статистичних групувань.

2.3. Основні правила утворення груп.

2.4. Типологічні групування.

2.5. Структурні групування.

2.6. Аналітичні групування.

2.7. Вторинні групування.

2.8. Складні групування.

2.9. Необхідність створення системи групувань, та основні вимоги до них.

2.10. Статистичні таблиці.

2.1. Зміст і завдання статистичного зведення

В результаті статистичного спостереження отримують велику кількість різноманітних відомостей про кожну одиницю досліджуваної сукупності. Проте, щоб на основі цих відомостей можна було зробити певні висновки, потрібно всю масу окремих даних привести до відповідного порядку, систематизувати, обробити і на цій основі дати зведену характеристику всієї сукупності фактів за допомогою узагальнюючих статистичних показників. Цього досягають на другому етапі статистичного дослідження, який називається зведенням і групуванням статистичних матеріалів.

Отже, **статистичним зведенням** називається наукова обробка первинних даних статистичного спостереження з метою отримання узагальнюючих характеристик досліджуваного явища чи процесу за рядом суттєвих для них ознак.

Перш ніж приступити до зведення зібраного первинного матеріалу його потрібно проконтролювати і прийняти. Попередній теоретичний аналіз повинен сприяти тому, щоб під час зведення не губились основні риси досліджуваних явищ в загальних підсумках. Опрацьований матеріал необхідно перевірити за повнотою охоплення обстежуваних одиниць і якістю отриманих про них даних. Якість і повноту зібраної інформації перевіряють за допомогою логічного і лічильного контролю, виявлені дефекти виправляють. Важливою умовою своєчасного і правильного проведення статистичного зведення і суворе дотримання звітної дисципліни. І тільки після того, як весь первинний статистичний матеріал старанно проконтрольований і належним чином виправлений, можна приступати до його зведення.

Зведення може бути **просте** – як вузькотехнічна операція по підрахунку підсумків первинного статистичного матеріалу, а також **складне** – яке передбачає групування даних, розробку системи показників, підрахунок групових і загальних підсумків та виклад результатів зведення у вигляді статистичних таблиць чи графіків.

Статистичне зведення проводять за наперед розробленою програмою, яка відповідає завданням статистичного дослідження з врахуванням прийнятої форми організації зведення.

За формою організації зведення буває централізоване і децентралізоване. При централізованій формі організації зведення всі матеріали спостереження обробляють і синтезують в Державному комітеті статистики України. Суттєвою перевагою даної форми зведення є те, що вона дає можливість його автоматизації і використання єдиної методології обробки даних. При децентралізованій формі організації зведення матеріали спостереження обробляють і узагальнюють на місцях, а в центральні органи надсилають зведену інформацію по регіонах. Децентралізована форма зведення дещо дешевша і оперативніша за централізовану.

На практиці поєднують територіально-децентралізовану і централізовану форму зведення.

В залежності від завдань статистичного дослідження програма зведення встановлює групувальні ознаки, кількість груп та макети розроблювальних таблиць. Програма повинна бути складена таким чином, щоб в результаті зведення отримати матеріал, який характеризує досліджуване явище з різних його сторін.

Для успішного здійснення статистичного зведення складається план його проведення. План має містити розв'язок питань організації зведення куди входять: послідовність і терміни виконання окремих частин зведення, оформлення його результатів у вигляді таблиць, публікацій у вигляді статистичних збірників і ін.

2.2. Зміст і завдання статистичних групувань

Під час обробки статистичних матеріалів виникає потреба виділення однорідних груп, типів, а вже потім їх описання за допомогою відповідних кількісних характеристик. Застосування таких узагальнюючих показників як відносні і середні величини, індекси та інші, можливе лише після того, як статистичний матеріал розподілений на однорідні групи.

Групуванням в статистиці називається розчленування усіх одиниць досліджуваної сукупності на групи за повними істотними для них ознаками. Серед багатьох методів, які роблять статистику одним з наймодніших знарядь соціального пізнання, групування вважається найбільш ефективним. Воно є центральним моментом любого зведення, завдяки чому матеріал статистичного спостереження приймає систематизованого вигляду.

При статистичному вивченні соціально-економічних явищ і процесів групування є одним з основних методів аналізу і синтезу.

Ознаки, покладені в основу групування називаються групувальними. Групування одиниць досліджуваної сукупності за якою-небудь ознакою веде до рядів розподілу.

Групувальні ознаки можуть мати кількісний вираз (наприклад, вік працівника, стаж роботи, заробітна плата і т.п.), тому вони називаються кількісними, а ряди їх розподілу **– варіаційними рядами**.

Якщо групувальні ознаки відображають певні властивості одиниць сукупності (наприклад, стать, національність, освіту і т.п.), вони називаються якісними, а ряди розподілу – **атрибутивними**.

При групуванні одиниць сукупності за територіальною ознакою отримують **географічні** або **територіальні** ряди розподілу. Вони дають уяву про розміщення або ступінь розповсюдження тих чи інших явищ в просторі.

Особливим видом групувань в статистиці є класифікація.

Класифікацією в статистиці називається стійке і фундаментальне групування одиниць сукупності за атрибутивною ознакою на подібні і відмінні групи і підгрупи. Перелік цих груп і підгруп розглядається як своєрідний статистичний стандарт, затверджений Державний комітетом статистики України. Наприклад, класифікація галузей економіки, класифікація основних фондів, класифікація професій і т.д. Статистичні класифікації ґрунтуються на таких суттєвих ознаках, які мало змінюються і існують тривалий час.

Групування, будучи першою сходинкою статистичного аналізу, одночасно є підготовчою стадією для більш глибокого аналізу досліджуваного статистичного матеріалу.

Із багатьох завдань, які розв'язуються з допомогою статистичних групувань, можна виділити три основних: 1) розподіл всієї сукупності на якісно однорідні соціально-економічні типи; 2) вивчення структури явищ і структурних зрушень в них; 3) виявлення і характеристика взаємозв'язку між ознаками досліджуваного явища.

2.3. Основні правила утворення груп

Перед проведенням простих, а тим більше комбінованих групувань потрібно розв'язати питання про кількість груп, розмір інтервалі та ін.

При групуванні за атрибутивними ознаками число груп, на які ділять досліджувану сукупність, визначається кількістю різновидів цієї ознаки.

Окремим випадком атрибутивних групувань є **альтернативне групування**, коли є всього два варіанти атрибутивної ознаки, один з яких виключає інший. Наприклад, розподіл робітників які мають спеціальну освіту і які такої освіти не мають і т.п.

Інший характер має групування за кількісними ознаками, при якому виникає питання про кількість груп, числові межі окремих груп та інтервали групування. Наприклад, групування робітників за стажем роботи, тарифним розрядом, чи заробітною платою; групування заводів за вартістю основних виробничих фондів, випуском або реалізацією продукції і т.д.

При розв'язанні питання про те, скільки доцільно утворити груп, беруть до уваги варіацію ознаки і число спостережень. Чим інтенсивніше змінюється ознака і чим більша сукупність одиниць, тим більше число груп потрібно утворити. Однак, як загальний принцип розв'язання питання про необхідну кількість груп виступає вимога, щоб вона була оптимальною, і щоб до кожної групи потрапила достатньо велика кількість одиниць. При великій кількості груп відбудеться розпорошення одиниць досліджуваної сукупності, однорідні одиниці попадуть в різні групи. А при малій кількості груп до однієї і тієї ж групи попадуть одиниці різних типів.

При групуванні за кількісними ознаками виникає суттєве питання про вибір розміру інтервалів групування.

Інтервалом групування називається різниця між максимальним і мінімальним значеннями ознаки в кожній групі.

Інтервали в структурних і аналітичних групуваннях можуть рівними і нерівними в залежності від характеру розподілу одиниць сукупності за даною ознакою. Нерівні інтервали, в свою чергу, можуть бути прогресивно-зростаючими або прогресивно-спадаючими.

Якщо варіація досліджуваної ознаки знаходиться в порівняно вузьких межах і розподіл близький до нормального, то застосовують **рівні інтервали**.

Величину інтервалу, при групуванні з рівними інтервалами, визначають шляхом ділення розмаху варіації на число груп за формулою:

$$i = \frac{x_{\max} - x_{\min}}{n}$$

де: x_{max} — максимальна величина ознаки;

x_{min} — мінімальна величина ознаки;

i — розмір інтервалу;

n — число груп.

Розмір інтервалу залежить від числа груп і варіації досліджуваної ознаки. Чим більшою буде варіація ознаки, тим більшим буде розмір інтервалу і чим більше число груп, тим менший розмір інтервалу.

Число груп орієнтовно можна визначити за формулою американського вченого Стерджеса:

$$n \approx 1 + 3{,}321 \ell_g N \quad ,$$

де: N —число одиниць сукупності, яка підлягає групуванню.

Тоді формула для визначення розміру інтервалу матиме вигляд:

$$i = \frac{x_{\max} - x_{\min}}{1 + 3{,}321 \ell g N}.$$

Якщо, наприклад, потрібно провести групування з рівними інтервалами 100 робітників підприємства за розміром заробітної плати, максимальне значення якої 1800 грн., а мінімальне – 1200 грн., то $n \approx 1 + 3{,}321 \ell_g N = 7{,}642 \approx 8$ груп, а $i = \frac{1800 - 1200}{8} = 75$ грн.

В даному випадку оптимальним розміром інтервалу може бути величина 75 грн., а число груп – 7 або 8.

Якщо в результаті ділення отримують на кругле число і виникає потреба в закругленні, то заокруглюють в більшу сторону.

В соціально-економічній статистиці часто застосовують групування з **нерівними інтервалами**. Застосування нерівних, які прогресивно збільшуються або зменшуються, інтервалів зумовлено самою природою більшості соціально-економічних явищ, коли в нижчих групах велике значення мають навіть малі відмінності в показниках, а у вищих групах такі відмінності суттєвого значення не мають. Так, наприклад, для нижчих груп, при групуванні підприємств за чисельністю працюючих, різниця в 50 чи 100 чоловік має велике значення, а для вищих груп, в яких зосереджені великі підприємства, така різниця не суттєва.

В статистичних групуваннях часто розмежовують дві якісно відмінні групи підприємств. Наприклад, підприємства які не виконали план і ті, які виконали план на 100% і більше.

Групування, метою яких є утворення якісно однорідних груп використовують **спеціалізовані інтервали**. В таких групуваннях межа інтервалу встановлюється там, де відбувається перехід від однієї якості до іншої. Наприклад, групування дітей за віком, за характером відношення чоловічого населення до трудової діяльності і т.п.

2.4. Типологічні групування

Аналізуючи розвиток суспільних явищ і процесів в часі, потрібно виділити соціально-економічні типи, так як в зародженні, розвитку, боротьбі і відмиранні різних соціально-економічних типів заключається суть історичного процесу розвитку любого суспільства. Виділення соціально-економічних типів при дослідженні певного явища є одним з головних і вирішальних завдань методу статистичних групувань.

Типологічними називаються групування, за допомогою яких проводять розподіл досліджуваного суспільного явища на класи або соціально-економічні типи.

Типологічні групування в статистичних дослідженнях займають одне з центральних місць. На основі всебічного теоретичного аналізу досліджуваної сукупності виділяють її головні і найхарактерніші типи або групи, вивчають істотні відмінності між ними, а також спільні ознаки для всіх груп.

За допомогою типологічних групувань вивчають класовий склад населення, розподіл підприємств за формами власності, поділ економіки на сферу матеріального виробництва і невиробничу сферу та ін.

Типологічне групування дає можливість простежити тенденції зрушень в структурі досліджуваної сукупності за ряд років.

Без типологічного групування важко зрозуміти і статистично правильно охарактеризувати процеси розвитку промисловості, сільського господарства, будівництва, торгівлі, питання виробництва і розподілу валового внутрішнього продукту в любому суспільстві без врахування класової структури, виділення соціально-економічних типів явищ і однорідних в класовому відношенні груп.

Користуючись методом статистичних групувань досліджують утворення і розвиток нових економічних типів явищ.

У випадках коли статистика характеризує явища, які складаються з різних соціально-економічних типів і мають різні закони розвитку, зведені статистичні характеристики у вигляді середніх величин будуть правильно характеризувати розвиток явища тільки в тому випадку, якщо попередньо виділені за допомогою групувань якісно однорідні типи явищ.

Типологічні групування дають можливість аналізувати своєрідні особливості і розвиток окремих типів, зміну їх співвідношення в загальному обсязі певного явища. Такі групування містять цілу систему статистичних показників, які дозволяють глибоко і всебічно проаналізувати відмінності окремих типів і їх питому вагу за рядом показників в загальній сукупності.

Важливим питанням методу любих групувань є правильний вибір групувальної ознаки, від якої залежать результати групування. Типологічні групування вимагають особливого підходу до вибору групувальної ознаки. Якщо такими ознаками виступають атрибутивні ознаки, наприклад, класовий склад населення, форма власності, галузь виробництва, то утворення числа груп і їх назви визначаються самою ознакою.

Однак, часто доводиться виділяти типи на основі групувань за кількісною ознакою. Наприклад, виділяючи із підприємств великі, середні і малі, або передові і відстаючі, ми групуємо їх відповідно за вартістю основних виробничих фондів, рівнем виконання плану, тобто за кількісною ознакою. Тут важливо правильно встановити інтервал групування, щоб кількісно виділити одні типи від інших. Це питання розв’язується на основі визначення таких кількісних меж, які виділяють нову якість.

Типологічні групування використовують всюди, де потрібно охарактеризувати якісні особливості окремих груп.

2.5. Структурні групування

Виділені в результаті типологічних групувань, окремі типи явищ вивчаються за їх складом. Групування дозволяє також охарактеризувати структуру і структурні зрушення досліджуваного явища. Так, за допомогою групувань вивчають склад населення за віком, статтю, освітою, національністю, сімейними положеннями; склад працівників – за професіями, стажем роботи, віком та іншими ознаками.

Отже, **структурним групуванням** в статистиці називається розчленування якісно однорідної сукупності одиниць на групи, які характеризують її структуру і структурні зрушення за відповідний період часу.

Велике значення мають структурні групування при вивченні концентрації промислових та інших підприємств.

Для вивчення процесу концентрації в промисловості, підприємства групують за вартістю основних виробничих фондів, за середньоспискової чисельністю працюючих, за обсягом виробленої чи реалізованої продукції, за рівнем механізації і автоматизації виробництва і т.д. Кожна з наведених групувальних ознак по-своєму відображають концентрацію. Для трудомістких галузей економіки концентрацію відображають через групування за числом робітників, для енергомістких галузей – з допомогою групувань за енергозатратами, а в деяких випадках групування за однією ознакою доповнюють групуваннями за іншими ознаками.

Структурні групування широко використовують в аналізі виконання плану. Групування підприємств певної галузі за процентом виконання плану дозволяє виявити які з них виконали і перевиконали план, а які його не виконали, встановити причини невиконання плану і показати шляхи ліквідації відставання від планових завдань. Це дозволить розрахувати певні резерви для виконання і перевиконання плану по галузі в цілому.

Метод структурних групувань дає можливість аналізувати статистичні сукупності за економічними і адміністративними регіонами, за галузями економіки, за географічними зонами та іншими ознаками.

В зміні структури суспільних явищ відображаються важливі закономірності їх розвитку. Наприклад, індустріалізація економіки проявляється в рості частки промислової продукції у всій продукції економіки, а також в збільшенні питомої ваги виробництва засобів виробництва в складі продукції промисловості.

Зростання культурного рівня населення простежується в підвищенні його грамотності і частки людей із середньою спеціальною і вищою освітою.

Якщо в структурних групуваннях порівняти дані в часі, то отримаємо інформацію, яка характеризуватиме закономірності в зміні структури, тобто про структурні зрушення в досліджуваному явищі. Тому структурні групування часто подаються в динамічних таблицях.

В багатьох випадках структурне групування представляється цілою системою показників, що дозволяє вивчати розподіл підприємств за рядом показників, які характеризують їх роботу.

Багатопоказникове структурне групування покажемо на прикладі групування промислових підприємств одного з економічних регіонів України за обсягом виробництва валової продукції.

Групування промислових підприємств за обсягом валової продукції в 2006 р, %

Групи підприємств за річним обсягом валової продукції, тис.грн.	Число підпри-ємств	Валова продук-ція	Середньорічна чисельність промислово-виробничого персоналу	Середньорічна вартість основних виробничих фондів	Спожи-вання електро-енергії
До 100	5,8	0,1	0,2	0,1	0,1
101-500	13,3	0,2	1,5	0,7	0,2
501-1000	12,8	0,7	2,1	1,0	0,5
1001-5000	37,0	8,6	14,0	8,9	1,6
5001-10000	13,1	8,9	12,1	8,0	7,0
10001-50000	14,2	25,3	30,0	24,5	19,0
50001-100000	2,5	18,0	14,9	15,9	15,6
100001 і більше	1,3	38,2	25,2	40,9	56,0
Розом:	100,0	100,0	100,0	100,0	100,0

За матеріалами даного групування простежується значна концентрація промисловості. Аналізуючи наведені показники бачимо, що великі підприємства мають більш високу продуктивність праці, більш високу ступінь використання основних виробничих фондів та більш високий рівень енергоозброєності праці.

Таке структурне групування можна одночасно розглядати і як типологічне (виділення малих, середніх і великих підприємств) і, як аналітичне, яке показує зв'язок між рядом показників, що характеризують окремі групи.

2.6. Аналітичні групування

Одним з основних завдань, яке розв’язується за допомогою статистичних групувань є дослідження взаємозв’язків варіаційних ознак в межах, як правило, однорідної сукупності.

Аналітичними групуваннями в статистиці називаються такі, за допомогою яких виявляють і вивчають взаємозв’язок між окремими ознаками досліджуваного соціально-економічного явища.

Всі явища суспільного життя і їх ознаки зв’язані між собою і залежать одне від одного. Взаємозв’язані ознаки діляться на факторні і результативні.

Факторною називається ознака, під впливом якої змінюється залежна від неї інша ознака.

Результативною називається ознака, яка змінюється під впливом факторної ознаки. Наприклад, внесення добрив – це фактор того, що в результаті буде вища врожайність сільськогосподарських культур. Внесення добрив впливає на урожайність, яка, в свою чергу, змінюється від кількості і якості цих добрив. Отже, внесення добрив – це факторна ознака, а вища врожайність сільськогосподарських культур – результативна ознака.

Розглянемо для прикладу аналітичне групування промислових підприємств однієї із галузей економіки за вартістю основних виробничих фондів і випуском валової продукції.

Залежність випуску валової продукції від вартості основних виробничих фондів

Групи заводів за вартістю основних виробничих фондів, млн.грн.	Кількість заводів	Вартість основних виробничих фондів в середньому на один завод, млн.грн.	Валова продукція за звітний період (у порівняльних цінах) в середньому на один завод, млн.грн.	Фондо-віддача, грн.
I. 1,0-2,5	3	1,70	1,97	1,16
II. 2,5-4,0	10	3,16	3,79	1,20
III. 4,0-5,5	7	4,60	5,78	1,25
IV. 5,5-7,0	5	6,26	10,68	1,71
Разом:	25	4,01	5,51	1,37

Із таблиці чітко простежується тенденція до збільшення випуску валової продукції із зростанням вартості основних виробничих фондів. Групування показує залежність випуску валової продукції від вартості основних виробничих фондів, тобто від розміру підприємства. Це пояснюється тим, що на великих підприємствах підсилюється роль науково-технічного прогресу, покращується організація праці, збільшується число висококваліфікованих спеціалістів та ін.

Аналітичне групування показує закономірність, яка заклечається в тому, що із ростом обсягу виробництва підвищується ефективність роботи підприємств. В нашому прикладі про це свідчить тенденція росту фондовіддачі.

Взаємозв'язок між досліджуваними ознаками проявляється також в тому, що із збільшенням факторної ознаки результативна може не тільки збільшуватись, але й зменшуватись. Наприклад, із збільшенням рівня продуктивності праці зменшується собівартість одиниці продукції, із ростом обсягу виробництва знижується фондомісткість продукції і т.д.

Аналітичне групування може проводитись як за факторною, так і за результативною ознаками в залежності від того, що є основним при статистичному дослідженні. Якщо вивчається вплив якоїсь однієї причини на різні явища – групування проводять за факторною ознакою, а якщо вивчають вплив різних причин на яке-небудь одне явище, то групують сукупність за результативною ознакою.

Характерною особливістю аналітичного групування є те, що кожна група, утворена за суттєвою факторною ознакою, характеризується середніми величинами результативних ознак.

Аналітичні групування дозволяють, при більш глибокому аналізі, знайти форму і виміряти тісноту зв'язку між варіаційними ознаками і на цій основі зробити важливі практичні висновки для планування і прогнозування.

2.7. Вторинні групування

Всі раніше наведені групування, зроблені на основі первинного статистичного матеріалу називаються **первинними групуваннями**.

Іноді, при статистичному дослідженні, раніше згрупований матеріал доводиться перегруповувати.

Вторинним групуванням в статистиці називається процес утворення нових груп на основі раніше проведеного групування первинних даних.

Метод вторинного групування використовується для утворення, на основі групування за кількісною ознакою, якісно однорідних груп або типів, приведення кількох групувань з різними інтервалами до одного виду, з метою порівняльності та утворення більш укрупнених груп, в яких чіткіше простежується характер розподілу. Його застосовують також при економічному аналізі роботи господарств для приведення до порівняльного виду їх вже згрупованих даних, але не зіставних за територією чи періодами часу.

Статистика використовує два способи утворення нових груп. **Перший**, найбільш простий і розповсюджений – це спосіб зміни (як правило укрупнення) інтервалів первинного групування. В більшості випадків тут виходять із передбачення про нормальний розподіл ознак усередині інтервалів.

Другий спосіб вторинного групування базується на закріпленні за кожною групою повної частки одиниць сукупності (спосіб дольового перегрупування).

Перший спосіб вторинного групування застосовується для зведення двох і більше групувань з неоднаковими інтервалами до одного виду, з метою зіставлення. Проілюструємо цей спосіб перегрупування на прикладі.

Розподіл робітників за виконанням норм виробітку

Базисний період		Звітний період	
Групи робітників за виконанням норм виробітку	Число робітників, %	Групи робітників за виконанням норм виробітку	Число робітників, %
I. До 80	0,8	I. До 90	2,0
II. 80-85	1,2	II. 90-100	10,0
III. 85-90	2,3	III. 100-110	71,0
IV. 90-100	2,5	IV. 110-120	11,0
V. 100-110	70,2	V. 120 і більше	6,0
VI. 110-115	16,0		
VII. 115-120	2,8		
VIII. 120-125	2,2		
IX. 125 і більше	2,0		
Разом:	100,0	Разом:	100,0

Порівняти ці два періоди між собою не можна через те, що вони мають різні інтервали. Але їх можна порівняти, провівши вторинне групування.

Узявши за основу групи робітників звітного періоду, встановлюємо, що в першу групу увійдуть всі три групи робітників базисного періоду, інтервали від 90 % до 100 % і 100 % – 110 % обох періодів співпали, в четверту групу (110-120 %) – шоста і сьома групи робітників базисного періоду, і в останню групу (120 % і більше) – восьма і дев'ята групи.

Групування робітників за виконанням норм виробітку

Групи робітників за виконанням норм виробітку	Число робітників, % до підсумку	
	Базисний період	Звітний період
I. До 90	4,3	2,0
II. 90-100	2,5	10,0
III. 100-110	70,2	71,0
IV. 110-120	18,8	11,0
V. 120 і більше	4,2	6,0
Разом:	100,0	100,0

Як видно з таблиці найбільшою є частка робітників (понад 80 %) з виконанням норм виробітку від 100 до 120 %. Вторинне групування дало можливість зіставити дані обох періодів.

Вторинне групування, для приведення двох групувань з різними розмірами інтервалів до єдиного виду, з метою порівняльності, деколи проводять двома способами одночасно. Покажемо проведення такого групування на прикладі.

Розподіл робітників двох цехів одного промислового підприємства за розміром місячної заробітної плати

ЦЕХ 1		ЦЕХ 2	
Групи робітників за розміром заробітної плати, грн.	Питома вага робітників в групі, %	Групи робітників за розміром заробітної плати, грн.	Питома вага робітників в групі, %
1200-1400	5	-	-
1400-1600	12	1300-1600	14
1600-1800	18	1600-1900	30
1800-2000	26	1900-2200	21
2000-2200	25	2200-2500	15
2200-2400	7	2500-2800	16
2400-2600	4	2800-3100	4
2600-2800	3	-	-
Разом:	100	Разом:	100

Безпосередньо порівняти розподіл робітників на двох цехах підприємства неможливо, так як цей розподіл має різні інтервали (200 і 300 грн.).

В цьому випадку при допомозі вторинного групування можна привести ці групування до порівняльного виду, перегрупувавши їх з інтервалом 400 грн.

Розподіл робітників двох цехів одного підприємства за розміром місячної заробітної плати

Групи робітників за розміром заробітної плати, грн.	Питома вага робітників, в % до підсумку	
	ЦЕХ 1	ЦЕХ 2
1200-1600	17(5+12)	14
1600-2000	44(18+26)	37(30+1/3 * 21)
2000-2400	32(25+7)	24(2/3 * 21 +2/3 * 15)
2400-2800	7(4+3)	21(1/3 *15+16)
2800-3200	-	4
Разом:	100,0	100,0

Проведене вторинне групування по обох цехах дало можливість порівняти їх дані. Вони показують, що в цеху 2 місячна заробітна плата робітників більш диференційована, ніж в цеху 1.

2.8. Складні групування

Зв’язки між явищами суспільного життя складні і різноманітні, вони залежать від багатьох причин, в яких переплетені різні, часто суперечливі тенденції. Для повнішого і глибшого дослідження процесів розвитку цих явищ, потрібно групувати дані за двома і більше ознаками. Такі групування в статистиці називаються **складними**.

Найбільш розповсюдженим видом складних групувань є комбіновані групування, при яких розчленовані на групи сукупності піддаються подальшому дрібненню на підгрупи за іншими додатковими ознаками.

Отже, **комбінованими** в статистиці називаються групування, коли утворені групи за однією ознакою, діляться потім на підгрупи за однією і більше ознаками, взятими в комбінації.

Одночасне використання декількох групувальних ознак дозволяє виявити і порівняти такі відомості і зв’язки між досліджуваними ознаками, які не можна знайти через ізольовані групування за рядом групувальних ознак. Наприклад.

Групування 40-ка підприємств за вартістю основних виробничих фондів і чисельністю промислово-виробничого персоналу

Групи підприємств за вартістю основних виробничих фондів, млн.грн	Підгрупи за чисельністю промислово-виробничого персоналу, чол.	Число підпри-ємств	Середньорічна вартість основних виробничих фондів, млн.грн.	Середньо-спискова чисельність промислово-виробничого персоналу,чол.	Товарна продукція в середньому на одне підприємство, млн.грн.
А	1	2	3	4	5
I. 10 - 160	I. до 15000	18	80,7	6540	103,5
	II. 15000 і більше	7	102,7	16258	166,4
Всього по групі		**25**	**86,9**	**9262**	**121,2**
II. 160 - 310	I. до 15000	3	184,7	12633	142,7
	II. 15000 і більше	5	230,0	20764	223,4
Всього по групі		**8**	**213,0**	**17715**	**193,1**
III. 310 - 460	I. до 15000	3	328,0	12850	359,3
	II. 15000 і більше	4	380,5	22077	513,7
Всього по групі		**7**	**358,0**	**18123**	**447,6**
Разом:		**40**	**159,5**	**12503**	**192,7**

В таблиці чітко простежується залежність результативного показника (товарної продукції) від вартості основних виробничих фондів. Підсумкові дані по групах показують, що із збільшенням вартості основних виробничих фондів зростає середня величина товарної продукції (122,2; 193,1; 447,6).

За даними таблиці також спостерігається залежність обсягу товарної продукції від чисельності промислово-виробничого персоналу. Для цього потрібно розглянути зміну результативного показника по підгрупах. В підгрупах всіх груп товарна продукція в середньому на одне підприємство зростає.

Таким чином, комбіноване групування за двома ознаками дозволяє розглянути вплив обох групувальних ознак на результативний показник.

На практиці, при проведенні комбінованого групування, обмежуються трьома-чотирма ознаками. Все це вимагає пошуків нових принципів групування, які б зняли вказані вище обмеження.

Для розв'язання цього завдання математичною статистикою розроблений **метод багатовимірних групувань – метод розпізнавання образів**. На основі розпізнавання образів віднесення до окремих груп проводять за комплексом ознак, які утворюють «ознаковий простір», де кожна ознака розглядається як координата.

Якщо в наборі сукупності є ряд одиниць з фіксованими ознаками, то кожна з цих одиниць розглядається як точка в багатовимірному просторі. Всі точки, розташовані в багатовимірному просторі, утворюють скупчення, які складають групи (таксони, кластери). Завданням виділення цих скупчень точок в просторі і займається теорія розпізнавання образів.

Геометрична близькість декількох точок в багатовимірному просторі означає подібність становища відповідних об'єктів, тобто їх кількісна однорідність за досліджуваними ознаками.

Одним з розповсюджених критеріїв міри близькості між об'єктами є евклідова віддаль:

$$d_{ij} = \sqrt{\sum_{i=1}^{n} \left(x_{ki} - x_{kj}\right)^2} \ ,$$

де: x_{ki} – значення *k*-ої ознаки в *i*-му об'єкті;

x_{kj} – значення *k*-ої ознаки в *j*-му об'єкті.

При зменшенні цієї віддалі близькість зростає, а при її збільшення – зменшується.

2.9. Необхідність створення системи групувань, та основні вимоги до них

Завдання всебічного аналізу соціально-економічних явищ і процесів не розв'язується через складання якогось одного групування, яке б характеризувало типи, структуру і взаємозв'язок даного економічного явища. Такий аналіз вимагає складання системи групувань за багатьма ознаками, яка допоможе охарактеризувати розвиток певного явища в цілому.

Утворені системи групувань повинні відповідати загальним методологічним вимогам і підпорядковуватись логічним і формальним критеріям.

Логічні критерії вимагають наступного:

а) система повинна всебічно характеризувати досліджуваний об'єкт;

б) системою групувань повинні розв'язуватись типологічні, структурні і аналітичні завдання дослідження;

в) кожне окреме групування повинно бути однією з логічних ланок в загальній системі;

г) висновки за кожним групуванням не повинні суперечити одні одним;

д) система групувань повинна бути стабільною і не піддаватись чистим змінам в часі.

Основними **формальними критеріями** є:

а) формування за якісними ознаками повинні передувати групуванням за кількісними ознаками;

б) результативні ознаки повинні бути виражені в однакових для всієї системи абсолютних, відносних і середніх показниках;

в) таблиці системи групувань повинні мати протягом тривалого часу стабільну нумерацію;

г) інтервали групувань також повинні бути стабільними і без потреби часто не змінюватись.

Дотримання вищенаведених вимог робить систему групувань більш гнучкою і одночасно стабільною, що є необхідною умовою всебічного економіко-статистичного аналізу.

2.10. Статистичні таблиці

Результати статистичного зведення і групування викладаються, як правило, у формі таблиць.

В науковій діяльності і практичній роботі досить широко застосовують таблиці. Важливу роль вони відіграють в економічній роботі, а тому кожний економіст повинен вміти правильно скласти статистичну таблицю і зробити вірні висновки.

Важливою вимогою, яка пред'являється до статистичної таблиці, – це подання досліджуваного матеріалу в наглядній для читача формі.

Основною особливістю табличного викладу матеріалу є те, що показники, які характеризуються в статистичній таблиці можна об'єднати під одним загальним заголовком.

Проте не всяка таблиця являється статистичною. На відміну від допоміжних розрахункових таблиць (логарифмів, коефіцієнтів, множення ті ін.) статистичними таблицями можуть вважатися тільки ті, що містять інформацію статистичного аналізу соціально-економічних явищ і процесів.

Отже, **статистична таблиця** – це форма раціонального, наочного та систематизованого викладання числових характеристик досліджуваних суспільних явищ і процесів.

У статистичній таблиці за аналогією з граматичним реченням є підмет і присудок.

Підметом таблиці являється статистична сукупність, тобто ті об'єкти або їх частини, які характеризуються рядом числових показників.

Присудком статистичної таблиці називаються показники, що характеризують досліджувану сукупність і її частини.

Підмет таблиці переважно розташовується зліва і складає зміст рядків. Присудок таблиці, як правило, поміщається зверху і складає зміст колонок. Підмет і присудок інколи можуть мінятись місцями.

Кожна таблиця має ряд горизонтальних рядків і вертикальних граф (стовпчиків, колонок), які при перетині утворюють клітки, котрі заповнюються статистичними даними.

Розграфлена сітка (без слів і цифр) складає скелет таблиці. Якщо записати назви підмета і присудка, то матимемо макет таблиці. Для отримання повної таблиці потрібно заповнити всі клітки відповідними статистичними даними.

Обов'язковою складовою частиною статистичної таблиці є загальний і внутрішні заголовки. Загальний заголовок розташовується над таблицею, в якому коротко зазначається, про що йде мова в таблиці, до якого місця і часу вона відноситься і в яких одиницях наведені дані.

Внутрішні заголовки розміщаються усередині таблиці (збоку і зверху).

Приступаючи до складання статистичної таблиці, потрібно перш за все розробити її макет.

МАКЕТ СТАТИСТИЧНОЇ ТАБЛИЦІ

Назва таблиці
(загальний заголовок)

<table>
<tr><td>Присудок
Підмет</td><td colspan="9">Назва граф
(верхні заголовки)</td></tr>
<tr><td>А</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td rowspan="7">Назва рядків (бокові заголовки)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Підсумок</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

← Нумерація граф

← Рядки таблиці

← Підсумковий рядок

Графи таблиці

До деяких таблиць подаються примітки, де роз’яснюється зміст окремих заголовків чи показників.

В залежності від побудови підмета розрізняють три види статистичних таблиць: прості, групові, комбінаційні.

В простих таблицях в підметі перераховуються лише одиниці сукупності, в групах – цифрові дані об’єднуються в групи, а в комбінаційних – групи ще розбиваються на підгрупи.

Простими називаються статистичні таблиці в підметі яких немає груп. Їх ще називають **інформаційними**.

Вся різноманітність простих статистичних таблиць може бути зведена до перелікових, територіальних і хронологічних, або до деяких їх сполучень.

Простими **переліковими** таблицями називаються такі, в підметі яких наводиться перелік одиниць або показників, що вивчаються.

Простими **територіальними** називаються таблиці, в підметі яких дається перелік територій (країн, областей, районів), кожна з яких характеризується відповідними показниками.

Простими **хронологічними** називаються таблиці, в підметі яких наводяться певні відрізки часу, а в присудку – один або декілька статистичних показників.

Груповими називаються такі статистичні таблиці, підмет яких утворений в результаті групування одиниць досліджуваного об’єкта за тією чи іншою ознакою.

Комбінаційними називаються статистичні таблиці, у підметі яких групи утворені за однією ознакою діляться на підгрупи за іншими ознаками.

Крім розглянутих вище таблиць, статистика використовує також типові і балансові таблиці.

Типовими називаються статистичні таблиці, підмет яких утворений в результаті типологічного групування.

Балансовими називаються таблиці, які характеризують зв'язок між поступленням і витрачанням ресурсів.

При складанні статистичних таблиць потрібно дотримуватись таких правил:

1. Таблиця повинна включати тільки ті відмінності, які необхідні для розуміння і вивчення даного явища.
2. Заголовки повинні бути сформульовані точно, коротко і ясно.
3. Рядки в підметі і графи в присудку, як правило, повинні нумеруватися.
4. Суворо дотримуватись таких умовних позначень: якщо явище відсутнє, ставиться тире (–); якщо відсутні відомості про його розмір, ставиться три крапки (…) або пишеться «нема відомостей».
5. Абсолютні дані в межах однієї графи повинні бути округлені з однаковим ступенем точності (до 0,1; 0,01 або 0,001).
6. У випадку великої різноманітності одиниць виміру виділяється спеціальна графа «одиниці виміру».
7. Таблиці, як правило, повинні бути замкнені, тобто мати підсумкові рядки.

ЛЕКЦІЯ 3. АБСОЛЮТНІ, ВІДНОСНІ ТА СЕРЕДНІ ВЕЛИЧИНИ

ПЛАН:

3.1. Поняття про абсолютні і відносні величини та їх значення.

3.2. Види відносних величин і способи їх обчислення.

3.3. Суть і значення середніх величин.

3.4. Середня арифметична і її властивості.

3.5. Середня гармонічна.

3.6. Структурні середні.

3.7. Основні правила застосування середніх в статистиці.

3.1. Поняття про абсолютні і відносні величини та їх значення

Інформація про суспільні явища і процеси створюється, передається і зберігається у вигляді **статистичних показників**. Вони є однією з основних економічних категорій, за допомогою яких відображають кількісну і якісну сторони явищ і процесів.

Адекватність або **відповідність** показника заклечається в його спроможності відобразити ті властивості явищ, які намічені програмою статистичного дослідження.

Точність оцінки залежить від структури показника, організації статистичного спостереження та обробки отриманих даних.

За **способом вирахування** розрізняють первинні і вторинні (похідні) статистичні показники. Первинні отримують в результаті зведення матеріалів статистичного спостереження у формі абсолютних величин (кількість заводів галузі, вартість основних виробничих фондів і випущеної ними продукції). На основі первинних даних обчислюють вторинні (середню вартість основних виробничих фондів і випуск продукції в середньому на один завод), а при визначенні інтенсивності використання основних виробничих фондів матимемо похідні показники другого порядку (фондовіддача, фондоємкість).

За **часовою ознакою** всі статистичні показники поділяють на інтервальні і моментні. Інтервальні показники характеризують явища за певний проміжок часу (тиждень, декада, місяць, квартал, рік). Моментні показники дають кількісну характеристику явищ на певний момент часу (поголів'я великої рогатої худоби на початок кожного року, перепис багаторічних насаджень на початок року, чисельність працівників на перше число кожного місяця і т.д.).

Абсолютними величинами в статистиці називаються первинні узагальнюючі показники, які характеризують суспільні явища і процеси в конкретних умовах місця і часу.

Абсолютні величини як узагальнюючі показники характеризують сукупність за її **чисельністю** (число працівників, кількість магазинів, лікарень) і **обсягом** (валовий випуск продукції, фонд заробітної плати, обсяг роздрібного товарообороту і т.д.).

Статистика виділяє три види абсолютних величин: індивідуальні, групові і загальні.

Індивідуальними називаються такі абсолютні величини, які виражають розміри кількісних ознак окремих одиниць досліджуваної сукупності.

Групові і **загальні** абсолютні статистичні величини виражають величину ознаки у всіх одиниць даної сукупності, або окремих її груп.

Абсолютні статистичні величини виражають розміри явищ в таких одиницях міри як: вага, об'єм, площа, довжина, вартість і ін.

Абсолютні статистичні величини завжди числа іменовані.

В статистиці використовується велике число різноманітних одиниць виміру, які можна об'єднати в три групи: натуральні, вартісні і трудові.

Натуральними називаються одиниці виміру, які виражають розміри конкретних явищ у фізичних вимірниках (тоннах, кілограмах, метрах, гектарах, літрах, кубометрах і ін.).

Натуральні одиниці виміру можуть бути простими, складними і умовно-натуральними.

Складні натуральні одиниці виміру отримують шляхом перемноження двох величин різних розмірностей. Наприклад, потужність електродвигунів вимірюється в кіловатах, а спожита ними енергія в складних одиницях – кіловат-годинах, обсяг перевезених вантажів вимірюється в тоннах, а вантажооборот – в тонно-кілометрах, верстатний парк цеху обчислюється в штуках, а робота верстатів у верстато-днях, верстато-змінах і т.д.

В ряді випадків статистика використовує умовно-натуральні одиниці виміру. Такі одиниці виміру використовуються для зведення докупи декількох різновидностей однакової споживної вартості. Одну з них приймають за еталон, а всі інші перераховують за допомогою спеціальних перевідних коефіцієнтів в одиниці виміру взятого еталону. Перерахунок в умовно-натуральні одиниці здійснюють за формулою:

$$у = е + k \cdot н,$$

де: $у$ – кількість умовно-натуральних одиниць;

$е$ – кількість еталонних одиниць;

$н$ – кількість одиниць сукупності, які відрізняються від еталонних;

k – коефіцієнт перерахунку не еталонних одиниць сукупності в еталонні.

Розглянемо приклад. Нехай маємо 150 т. мила із 40 % вмістом жирних кислот, 100 т. – із 50 % – із 60 % вмістом жирних кислот. Потрібно перерахувати все мило на 40 % -не. Спочатку визначимо коефіцієнти перерахунку: для 40 % мила $k_1 = \frac{40}{40} = 1$, для 50 % $k_2 = \frac{50}{40} = 1{,}25$ і для 60% – $k_3 = \frac{60}{40} = 1{,}5$. Перерахуємо все мило в умовно-натуральне (на 40 % - не, взяте за еталон).

$$150 \cdot 1 + 100 \cdot 1{,}25 + 50 \cdot 1{,}5 = 350 \text{ т.}$$

Отже, ми маємо 350 умовно-натуральних тонн мила.

Вартісними називаються одиниці виміру, які використовуються для характеристики в грошовому виразі багатьох різноманітних статистичних показників. Наприклад, собівартість і ціна одиниці продукції обліковується в гривнях і копійках, обсяг товарообороту продуктового магазину – в тисячах гривень, а валовий внутрішній продукт держави в мільйонах або мільярдах гривень.

Трудовими називаються одиниці виміру, які використовуються для обліку затрат робочого часу, для визначення рівня продуктивності праці, величини трудових ресурсів і раціонального їх використання та ін. Трудові вимірники виражаються в людино-годинах, людино-роках, машино-днях, верстато-днях, коне-днях.

Абсолютні величини є основою для обчислення різних видів відносних і середніх величин, індексів та інших узагальнюючих показників.

3.2. Види відносних величин і способи їх обчислення

Відносними величинами називають статистичні показники, які виражають кількісні співвідношення між соціально-економічними явищами і процесами. Їх отримують шляхом порівняння (ділення) двох однойменних, або різнойменних величин.

Величина, з якою проводять порівняння, називається основою відносної величини, базою порівняння або базисною величиною. Величина, яку порівнюють, називається поточною, порівнюваною чи звітною величиною.

Відносні величини показують, у скільки разів порівнювана величина більша (менша) за базисну, або яку частку перша займає в другій, або скільки одиниць однієї величини припадає на одиницю іншої.

В залежності від бази порівняння відносні величини можуть виражатись у формі:

а) коефіцієнтів – якщо база порівняння приймається за одиницю;

б) процентів (%) – якщо база порівняння береться за 100;

в) проміле ($\%_0$) – якщо за базу порівняння взято 1000;

г) продецеміле ($\%_{00}$) – якщо база порівняння становить 10 000;

д) просантиміле ($\%_{000}$) – якщо база порівняння прийнята за 100 000.

Абсолютна величина, взята сама по собі, не завжди дає правильну оцінку явища. В багатьох випадках тільки в порівнянні з іншою абсолютною величиною дана величина проявляє свою істинну значущість.

В залежності від змісту і пізнавального значення відносні величини, що використовуються в статистиці, поділяються на наступні основні види: відносні величини планового завдання, виконання плану, динаміки, структури, координації, порівняння і інтенсивності.

Відносні величини планового завдання характеризують відношення запланованого рівня показника до його фактично досягнутого рівня в минулому періоді.

Відносною величиною виконання плану називається статистичний показник, який характеризує ступінь виконання планового завдання, встановленого на даний період. Тобто це процентне відношення фактично досягнутого рівня до запланованого за відповідний період часу (місяць, квартал і т.д.).

Відносні величини динаміки характеризують зміну рівня того чи іншого явища в часі. В залежності від характеру бази порівняння, розрізняють відносні величини динаміки із змінною базою порівняння, або ланцюгові, і відносні величини з постійною базою порівняння, або базисні.

Відносні величини структури характеризують питому вагу окремих частин досліджуваної сукупності в загальному її обсязі. Їх обчислюють шляхом відношення частини до цілого.

Відносні величини координації характеризують співвідношення частин цілого між собою.

За допомогою відносних величин координації визначають, скільки одиниць даної частини цілого припадає на 1, на 100, на 1000, на 10 000 одиниць іншої частини, взятої за базу порівняння.

Відносні величини порівняння характеризують співвідношення ознайомлених величин, що стосуються одного й того ж періоду або моменту часу, але різних об’єктів чи територій.

Відносні величини інтенсивності характеризують ступінь поширення або розвитку даного явища в певному середовищі. Їх отримують шляхом зіставлення різнойменних абсолютних величин, пов’язаних між собою, але які не являються складовими цілого.

3.3. Суть і значення середніх величин

Для зведеної кількісної характеристики багатьох явищ і процесів суспільного життя статистика широко використовує такий розповсюджений узагальнюючий показник як середня величина (середня врожайність, середній процент виконання плану, середня частка і т.п.). Вона дає узагальнюючу характеристику однорідних елементів масових явищ, які мають різне кількісне значення (варіацію) в залежності від конкретних умов.

В середній погашаються випадкові відхилення індивідуальних значень і відображаються ті загальні умови, під впливом яких формувалась сукупність.

Середня величина – це узагальнюючий показник, який характеризує однорідну сукупність явищ за якою-небудь кількісною варіаційною ознакою в даних умовах місця і часу.

Тільки за допомогою середньої можна охарактеризувати сукупність за кількісною варіаційною ознакою.

Середні величини використовують для порівняння показників двох і більше об’єктів (порівняння урожайності окремих культур по господарствах області, порівняння цін на деякі товари на ринках певного регіону і т.п.).

Середніми величинами користуються для характеристики зміни рівнів явищ в часі.

До середніх звертаються при вивченні взаємозв’язків між явищами та їх ознаками.

Середні величини застосовують для проведення факторного аналізу явищ, з метою виявлення невикористаних резервів.

Велике значення мають середні величини в плануванні і прогнозуванні завдань для економіки в цілому і окремих його галузей.

Багатогранність суспільних явищ обумовлює виняткову важливість застосування середніх величин в економіко-статистичних дослідженнях. Вони є активним засобом управління, планування і прогнозування економіки держави.

3.4. Середня арифметична і її властивості

Найбільш поширеним видом середніх величин в статистці є середня арифметична. Вона застосовується у формі простої середньої і зваженої середньої.

Середня арифметична проста застосовується в таких випадках, коли всі варіанти зустрічаються один раз, або мають однакові частоти в досліджуваній сукупності. Її отримують шляхом додаванням окремих варіантів і діленням суми на число доданків.

Формула середньої арифметичної простої має вигляд:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum x}{n},$$

де $\bar{x}$ – середня величина ознак;

$x_{1,}\ x_{2,}\ x_{3,}\ ..., x_n$ – окремі варіанти ознаки;

$\sum$ – (велика грецька літера «сігма») – знак суми;

n – кількість варіантів.

Розглянемо приклад. Маємо такі дані про добове видобування вугілля на шахті за першу декаду червня.

Числа місяця	1	2	3	4	5	6	7	8	9	10
Добове видобування вугілля, тис.т.	5,4	5,3	5,6	5,5	5,7	5,9	6,0	5,8	6,2	6,1

$$\bar{x} = \frac{\sum x}{n} = \frac{5{,}4 + 5{,}3 + 5{,}6 + 5{,}5 + 5{,}7 + 5{,}9 + 6{,}0 + 5{,}8 + 6{,}2 + 6{,}1}{10} = \frac{57{,}5}{10} = 5{,}75 \text{ тис.т.}$$

Якщо в сукупності варіанти зустрічаються неоднакову кількість раз, то їх об'єднують в групи і, таким чином, переходять від середньої арифметичної простої до зваженої.

Середня арифметична зважена обчислюється як частка від ділення суми добутків варіантів і їх частот на суму частот. Вона визначається за формулою:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + ... + x_n f_n}{f_1 + f_2 + f_3 + ... + f_n} = \frac{\sum xf}{\sum f},$$

де f – частоти (ваги), які показують скільки раз зустрічаються значення ознаки в сукупності.

Отже, для обчислення середньої арифметичної зваженої потрібно: а) кожний варіант перемножити на його частоту; б) знайти суму їх добутків; в) суму добутку поділити на суму ваг.

Наведемо приклад. Маємо дані про заробітну плату робітників і число робітників, які отримують дану заробітну плату.

Табельний номер робітника	Заробітна плата одного робітника, грн. *(x)*	Число робітників, чол. *(f)*
1	1500	10
2	1800	17
3	2000	40
4	2100	18
5	2500	15
Разом:	x	100

Середня заробітна плата за цими даними становитиме:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{1500 \cdot 10 + 1800 \cdot 17 + 2000 \cdot 40 + 2100 \cdot 18 + 2500 \cdot 15}{10 + 17 + 40 + 18 + 15} = \frac{200900}{100} = 2009 \text{ грн.}$$

Іноді середні величини потрібно обчислити не з конкретних значень варіантів досліджуваної ознаки, а із значень величин, виражених у вигляді інтервалів. В таких випадках потрібно для кожного інтервалу знайти його середину за простою середньою між верхньою і нижньою межею кожного інтервалу і після цього проводити обчислення за формулою середньої арифметичної зваженої.

Покажемо обчислення середньої на основі інтервального ряду.

Розподіл виробів за вагою

Вага виробів, г. **(x)**	Число виробів, шт. ***(f)***	Середина інтервалу ***(x)***	Вага всіх деталей, г ***(x, f)***
До 100	5	97,5	487,5
100-105	19	102,5	1947,5
105-110	52	107,5	5590,0
110-115	18	112,5	2025,0
115 і більше	6	117,5	705,0
Разом:	100	x	10755,0

$$\bar{x} = \frac{\sum xf}{\sum f} == \frac{10755{,}0}{100} = 107{,}55\,\text{г}$$

Середня арифметична має деякі математичні властивості, що мають практичне значення для спрощеного обчислення середньої за даними варіаційного ряду.

Найважливіші з них такі:

1. Якщо всі варіанти збільшити або зменшити на одне й теж число (А), то й середня арифметична збільшиться (зменшиться) на теж число (А).

2. Якщо всі варіанти збільшити або зменшити в одне й теж число (і) раз, то й середня арифметична відповідно збільшиться (зменшиться) в (і) раз.

3. Якщо всі частоти (ваги) поділити або помножити на яке-небудь число (z), то середня арифметична від цього не зміниться.

Використання першої і другої властивостей середньої арифметичної дозволяє значно спростити її обчислення. Цей метод в статистиці називають **методом моментів**, або **метод відліку від умовного нуля**. Розглянемо спрощений спосіб обчислення середньої арифметичної методом моментів за даними попереднього прикладу.

Вага виробів, г **(x)**	Число виробів, шт. ***(f)***	Середина інтервалу ***(x)***	Скорочені варіанти		Зважені скорочені варіанти $\left(\frac{x-A}{i}\right)f$
			X-A A=107,5	$\frac{x-A}{i}$ *i=5*	
До 100	5	97,5	-10	-2	-10
100-105	19	102,5	-5	-1	-19
105-110	52	107,5	0	0	0
110-115	18	112,5	5	1	18
115 і більше	6	117,5	10	2	12
Разом:	100	х	х	х	+1

Формула для знаходження середньої арифметичної способом моментів має вигляд:

$$\bar{x} = i \cdot m_1 + A,$$

де $m_1 = \frac{\sum\left(\frac{x-A}{i}\right)f}{\sum f}$ – момент першого порядку;

x – варіант;

f – частота;

A – умовно взяте число;

i – розмір інтервалу.

Визначимо момент першого порядку:

$$m_1 = \frac{\sum\left(\frac{x-A}{i}\right)f}{\sum f} = \frac{1}{100} = 0{,}01.$$

Підставляємо значення в формулу:

$$\bar{x} = i \cdot m_1 + A = 5 \cdot 0{,}01 + 107{,}5 = 107{,}55 \text{ г.}$$

Отже, ми отримали той самий результат, що й при обчисленні за звичайною формулою середньої арифметичної зваженої.

Спосіб моментів використовується в тих випадках, коли вихідні дані наведені у вигляді інтервального ряду з рівними інтервалами.

Третю властивість середньої арифметичної використовують тоді, коли доводиться мати справу з великими частотами. В цьому випадку частоти скорочують і подальші розрахунки проводять з малими числами.

На цій же властивості середньої арифметичної ґрунтується перехід від абсолютних значень частот до відносних величин, які визначаються як процентне відношення кожної групи до їх загального підсумку.

$$f_1' = \frac{f_1}{10\,000}; \quad f_2' = \frac{f_2}{10\,000}; \quad f_3' = \frac{f_3}{10\,000} \text{ і т.д.}$$

$$\bar{x} = \frac{\sum xf'}{\sum f'}.$$

$$P_1 = \frac{f_1}{\sum f} \cdot 100; \; P_2 = \frac{f_2}{\sum f} \cdot 100; \; P_3 = \frac{f_3}{\sum f} \cdot 100 \text{ і т.д.}$$

$$\bar{x} = \frac{\sum xP}{\sum P} = \frac{\sum xP}{100}.$$

3.5. Середня гармонічна

В статистичній практиці часто зустрічаються випадки, коли середню потрібно обчислювати за формулою **середньої гармонічної**. Це відбувається тоді, коли підсумовуванню підлягають не самі варіанти, а обернені їм числа. В цьому випадку, для знаходження середнього значення варіаційної ознаки, застосовують формулу **середньої гармонічної простої**, яка має вигляд:

$$\bar{x} = \frac{n}{\sum \frac{1}{x}},$$

де n – число індивідуальних значень ознак;

$\sum\frac{1}{n}$ – сума обернених значень ознак;

Звернемось до прикладу. Три трактори при обробітку ґрунту умовної оранки одного гектара затратили часу: перший – 2,2 години, другий – 2,4 години, третій – 2,6 години. Визначимо середні затрати часу на обробіток одного гектара умовної оранки трактором.

$$\bar{x} = \frac{n}{\sum\frac{1}{x}} = \frac{3}{\frac{1}{2,2}+\frac{1}{2,4}+\frac{1}{2,6}} = \frac{3}{1,2558} = 2,38 \text{ год.}$$

Середню гармонічну зважену застосовують в тих випадках, коли є дані про індивідуальні значення ознаки в загальній сукупності і загальний обсяг сукупності, але в готовому виді немає частот.

$$\bar{x} = \frac{\sum w}{\sum\frac{w}{x}},$$

де $\sum\frac{w}{x}$ – сума добутку обернених ознак і частот, тобто $x \cdot f = w$, звідси $f = \frac{w}{x}$.

Розглянемо приклад. Маємо дані про заробітну плату робітників завод в розрізі цехів і фонд заробітної плати.

Номер цеху	Середня заробітна плата одного робітника, грн. *(x)*	Фонд заробітної плати, грн. *(w)*
1	2900	174000
2	3200	160 000
3	3500	140 000

Підставивши у формулу середньої гармонічної зваженої дані з нашого прикладу, отримаємо середню заробітну плату одного робітника по заводу в цілому.

$$\bar{x} = \frac{\sum w}{\sum\frac{w}{x}} = \frac{174\,000+160\,000+140\,000}{\frac{174\,000}{2900}+\frac{160\,000}{3200}+\frac{140\,000}{3500}} = \frac{474\,000}{150} = 3160 \text{ грн.}$$

3.6. Структурні середні

Поряд з розглянутими вище середніми, для статистичної характеристики варіаційних рядів обчислюють структурні (порядкові) середні, до яких відносять моду і медіану.

Модою називається величина ознаки (варіанта), яка найчастіше зустрічається в даній сукупності.

Знаходження моди в дискретному варіаційному ряду не представляє складності. Розглянемо приклад.

Маємо розподіл студентів за їх ростом.

Ріст студентів, см.	165	167	170	173	176	178	180	182	185	187	189	191	195
Число студентів, чол.	8	13	24	30	38	47	55	31	16	9	7	3	2

Очевидно, в цьому прикладі модою буде студент, який має ріст 180см., так як цьому значенню варіанти відповідає найбільше число студентів (55 чол.).

Медіана – це варіант, який займає середнє положення в рангованому варіаційному ряду.

Щоб знайти медіану в дискретному варіаційному ряду, потрібно спочатку розташувати всі варіанти в зростаючому або спадаючому порядку. Потім визначити номер медіани, який вкаже на її розташування в рангованому ряді за формулою:

$$№_{M_e} = \frac{n+1}{2},$$

де M_e – медіана;

n – число варіантів.

Розглянемо приклад. Маємо дані про розподіл дев’яти деталей за їх масою.

Номер деталі	1	2	3	4	5	6	7	8	9
Маса, г.	2,6	3,4	3,3	2,7	3,0	2,9	2,8	3,1	3,2

Перегрупуємо деталі за їх масою в зростаючому порядку.

Номер деталі	1	2	3	4	5	6	7	8	9
Маса, г.	2,6	2,7	2,8	2,9	3,0	3,1	3,2	3,3	3,4

Визначаємо номер медіани:

$$№_{M_e} = \frac{n+1}{2} = \frac{9+1}{2} = 5\,.$$

Тобто під п’ятим номером від початку або від кінця ряду маса деталі буде медіаною. $M_e = 3{,}0$ г.

Коли варіаційний ряд має парну кількість членів, тоді медіана буде розраховуватись як півсума двох варіантів, які займають середнє положення в ранговому ряду. Припустимо, що в нас є ще десята деталь з масою 3,5г. Номер медіани буде рівним $5{,}5\left(\frac{10+1}{2}\right)$. В даному випадку медіана буде розташована між п’ятим і шостим порядковим номером деталей.

$$M_e = \frac{3{,}0+3{,}1}{2} = 3{,}05 \text{ г.}$$

Моду і медіану із інтервальних рядів визначають розрахунковим шляхом за наступними формулами:

$$M_0 = x_{m0} + i \cdot \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)},$$

де M_0 – мода;

x_{m_o} – нижня межа модального інтервалу;

i – розмір модального інтервалу;

f_1 – передмодальна частота;

f_2 – модальна частота;

f_3 – післямодальна частота.

$$M_e = x_{me} + i \cdot \frac{\frac{\sum f}{2} - S_{me-1}}{f_{me}},$$

де M_e – медіана;

x_{me} – нижня межа медіанного інтервалу;

i – розмір медіанного інтервалу;

$\sum f$ – сума частот;

$\frac{\sum f}{2}$ – порядковий номер медіани;

S_{me-1} – сума нагромаджених частот до медіанного інтервалу;

f_{me} – частота медіанного інтервалу.

Покажемо обчислення моди і медіани для інтервального варіаційного ряду на прикладі.

Розподіл 500 робітників за заробітною платою

Заробітна плата, грн.	Число робітників, чол.	Нагромаджені частоти
1500-1600	5	5
1600-1700	10	15
1700-1800	61	76
1800-1900	105	181
1900-2000	130	311
2000-2100	109	420
2100-2200	62	482
2200-2300	11	493
2300-2400	7	500
Разом:	500	х

В нашому прикладі мода знаходиться в інтервалі від 1900 до 2000 грн., тому, що йому відповідає найбільша частота (130 чол.). Цей інтервал називається модальним.

$$M_o = 1900 + 100\frac{130-105}{(130-105)+(130-109)} = 1954{,}3 \text{ грн.}$$

Цей показник означає, що найбільше робітників було із заробітною платою 1954,3 грн.

Щоб визначити медіану інтервального варіаційного ряду спочатку, за допомогою нагромаджених частот, потрібно знайти інтервал, що містить медіану. Медіанному інтервалу відповідає перша з нагромаджених частот, яка перевищує півсуму частот всього обсягу сукупності ($\frac{\sum f}{2} = 250$). Отже, медіана знаходиться в інтервалі від 1900 до 2000 грн.

$$M_e = 1900 + 100\frac{250-181}{130} = 1953{,}1 \text{ грн.}$$

Це означає, що половинна робітників отримує заробітну плату меншу 1953,1 грн., а друга половина – більшу.

В доповнення до медіани для характеристики структури варіаційного ряду, вираховують **квартилі** і **децилі**, які відповідно ділять ряд за сумою частот на чотири і десять рівних частин.

Перший або нижній квартиль відсікає чверть сукупності знизу, другий – рівний медіані, а третій або верхній – відсікає чверть сукупності зверху.

В інтервальному варіаційному ряду квартилі у середині, визначеного за нагромадженими частотами інтервала, обчислюються за формулами:

а) перший (нижній) квартиль:

$$Q_1 = X_{Q1} + i \cdot \frac{\frac{\sum f}{4} - S_{Q1-1}}{f_{Q1}},$$

де Q_1 – перший (нижній) квартиль;

X_{Q_1} – нижня межа першого квартильного інтервалу;

i – величина інтервалу;

$\frac{\sum f}{4}$ – порядковий номер першого квартиля;

S_{Q_1-1} – нагромаджена частота перед першим квартильним інтервалом;

f_{Q_1} – частота першого квартильного інтервалу.

б) третій (верхній) квартиль:

$$Q_3 = X_{Q_3} + i \cdot \frac{3\frac{\sum f}{4} - S_{Q_3-1}}{f_{Q3}},$$

де Q_3 – третій (верхній) квартал;

X_{Q_3} – нижня межа третього квартильного інтервалу;

i – величина інтервалу;

$3\frac{\sum f}{4}$ – порядковий номер третього квартиля;

S_{Q_3-1} – нагромаджена частота перед третім квартильним інтервалом;

f_{Q_3} – частота третього квартильного інтервалу.

Обчислення цих показників у варіаційному ряді абсолютно аналогічне знаходженню медіани і починається з розрахунку порядкового номера відповідного варіанта а потім за нагромадженими частотами визначається інтервал, в якому міститься шуканий варіант.

$$Q_1 = 1800 + 100 \cdot \frac{125 - 76}{105} = 1846{,}7 \text{ грн.}$$

Це означає, що у одної четверті всіх робітників заробітна плата не перевищує 1846,7 грн., а в трьох четвертях – вона рівна або перевищує 1846,7 грн.

$$Q_3 = 2000 + 100 \cdot \frac{375 - 311}{109} = 2058{,}7 \text{ грн.}$$

Отже, заробітна плата кожного четвертого робітника перевищує 2058,7 грн.

Формули для децилів в інтервальному варіаційному ряду мають вигляд:

а) дециль перший:

$$D_1 = X_{01} + i \cdot \frac{1/10 \sum f - S_{D-1}}{f_{D_1}};$$

де D_1 – перший дециль;

X_{01} – нижня межа першого дециля;

i – розмір інтервалу;

$1/10 \sum f$ – порядковий номер першого дециля;

f_{D-1} – нагромаджена частота перед першим децильним інтервалом;

f_{D_1} – частота першого децильного інтервалу.

б) дециль другий:

$$D_2 = X_{02} + i \cdot \frac{2/10 \sum f - S_{D_2 - 1}}{f_{D_2}};$$

в) дециль третій:

$$D_3 = X_{03} + i \cdot \frac{3/10 \sum f - S_{D_3 - 1}}{f_{D_3}} \text{ і т.д.}$$

Вирахуємо децилі по тій же схемі, що медіани і квартилі.

Так, $D_1 = 1700 + 100 \dfrac{50 - 15}{61} = 1757{,}4$ грн.;

$$D_2 = 1800 + 100 \frac{100 - 76}{105} = 1822{,}8 \text{ грн.};$$

$$D_3 = 1800 + 100\frac{150-76}{105} = 1870{,}5 \text{ грн.};$$

....

$$D_{10} = 2300 + 100\frac{500-493}{7} = 2400{,}0 \text{ грн.}$$

Мода, медіана, квартилі і децилі відносяться до так званих порядкових статистик, під якими розуміють варіант, який займає певне порядкове місце в рангованому варіаційному ряду.

Їх використання в статистичному аналізі варіаційних рядів дозволяє більш глибоко дослідити і детальніше охарактеризувати сукупність, яка вивчається.

3.7. Основні правила застосування середніх в статистиці

В статистичних дослідженнях вірну характеристику сукупності за варіаційною ознакою в кожному окремому випадку дає тільки правильно визначений вид середньої. В залежності від утворення загального обсягу варіаційної ознаки визначається вид вибраної середньої.

Так, середня арифметична застосовується тоді, коли загальний обсяг варіаційної ознаки утворюється як **сума квадратів** окремих варіантів; середня гармонічна – коли загальний обсяг утворюється як **сума обернених значень** окремих варіантів; середня геометрична – коли обсяг варіаційної ознаки утворюється як **добуток окремих варіантів**.

Головна умова наукового використання середньої полягає в тому, що середні характеристики повинні вираховуватись на основі масового узагальнення фактів. Тільки тоді вони відображають суть явища, на значення якого не впливають одиничні фактори. Ця умова пов’язує статистичні середні із законом великих чисел.

Іншою важливою умовою застосування середніх в статистиці є якісна однорідність всіх одиниць сукупності. Вона заключається в тому, що не можна обчислювати середню з неоднорідної сукупності, окремі елементи якої підпорядковані різним законам розвитку по відношенню до осереднювані ознаки.

Середня величина тільки тоді відобразить типовий розмір ознаки та її загальні риси, якщо це загальне реально існує, всі елементи якого якісно однорідні і типові.

Застосування методу середніх в статистиці тісно і нерозривно зв'язане з методом групувань.

Загальні середні потрібно доповнювати груповими середніми в тих випадках, коли варіаційна ознака суттєво відрізняється в окремих групах і в порівнюваних групах існує різне співвідношення груп.

Особливого значення набуває доповнення загальної середньої груповими середніми при вивченні взаємозв'язку і взаємозалежності одних показників ознак від інших.

При використанні середніх потрібно пам'ятати, що середні величини не можуть і не повинні підміняти індивідуальні показники, а доповнюватись вивченням кращих і гірших одиниць сукупності.

ЛЕКЦІЯ 4. ПОКАЗНИКИ ВАРІАЦІЇ

ПЛАН:

4.1. Поняття про показники варіації і способи їх обчислення.

4.2. Спрощені способи розрахунку дисперсії.

4.3. Дисперсія альтернативної ознаки.

4.4. Види дисперсій і правило їх додавання.

4.1. Поняття про показники варіації і способи їх обчислення

Середні величини мають велике теоретичне і практичне значення, вони дають узагальнюючу характеристику сукупності за варіаційними ознаками, виражають типовий, для даних умов, рівень цих ознак. Проте, для характеристики досліджуваних явищ одних тільки середніх величин недостатньо, оскільки, при однакових значеннях середньої величини, різні сукупності можуть істотно відрізнятись одна від одної за характером варіації величини досліджуваної ознаки.

Середні величини не виражають індивідуальних особливостей досліджуваної сукупності, які породжують варіацією ознаки її окремих елементів, а тому, їх потрібно доповнювати показниками, що характеризують коливання значень ознаки в сукупності.

Варіацією в статистиці називаються коливання ознаки в одиниць сукупності, а показники, що характеризують ці коливання називаються показниками варіації. Вони показують як розміщуються навколо середньої окремі значення осереднюваної ознаки.

Розглянемо приклад. Маємо дані про продуктивність праці робітників-відрядників в двох бригадах:

Табельний номер робітника	Вироблено деталей за зміну, шт.	
	І бригада	ІІ бригада
1	10	120
2	20	110
3	100	100
4	180	90
5	190	80
Разом:	500	500

Середня продуктивність праці в двох бригадах однакова.

$$\overline{X}_1 = \frac{\sum X_1}{n} = \frac{500}{5} = 100 \text{ шт.} \qquad \overline{X}_2 = \frac{\sum X_2}{n} = \frac{500}{5} = 100 \text{ шт.}$$

Однак, коливання продуктивності праці відрядників–робітників в першій бригаді значно більше, ніж у другій. Отже, друга бригада працює ритмічніше, ніж перша.

Для вимірювання варіації у статистиці використовують такі показники як: розмах варіації, середнє лінійне відхилення, середній квадрат відхилення (дисперсія), середнє квадратичне відхилення і коефіцієнт варіації.

Розмах варіації являє собою різницю між найбільшим і найменшим значенням ознаки.

$$R = X_{\max} - X_{\min},$$

де R – розмах варіації;

$X_{\max}$ – максимальне значення ознаки;

$X_{\min}$ – мінімальне значення ознаки.

В нашому прикладі розмах варіації:

для першої бригади $R_1 = 190 - 10 = 180$ шт.;

для другої бригади $R_2 = 120 - 80 = 40$ шт.

Розмах варіації простий для обчислення, але він відображає лише крайні значення ознаки і не дає уяви про ступінь варіації усередині сукупності.

Середнє лінійне відхилення $(\bar{l})$ являє собою середню арифметичну з абсолютних значень відхилень окремих варіантів від середньої арифметичної.

Середнє лінійне відхилення – величина іменована і визначається за формулами:

а) середнє лінійне відхилення просте

$$\bar{l} = \frac{\sum\left|X - \overline{X}\right|}{n};$$

б) середнє лінійне відхилення зважене

$$\bar{l} = \frac{\sum\left|X - \overline{X}\right| f}{\sum f}.$$

Обчислимо середнє лінійне відхилення для нашого прикладу:

Табельний номер робітника (n)	I бригада		II бригада	
	Вироблено деталей за зміну, шт. (X_1)	$\left\|X_1 - \overline{X}_1\right\|$	Вироблено деталей за зміну, шт. (X_2)	$\left\|X_2 - \overline{X}_2\right\|$
1	10	90	120	20
2	20	80	110	10
3	100	0	100	0
4	180	80	90	10
5	190	90	80	20
Разом:	500	340	500	60

$$\bar{l}_1 \frac{\sum\left|X_1 - \overline{X}_1\right|}{n} = \frac{340}{5} = 68 \text{ шт.} \qquad \bar{l}_2 \frac{\sum\left|X_2 - \overline{X}_2\right|}{n} = \frac{60}{5} = 12 \text{ шт.}$$

Отже, кількість вроблених деталей за зміну окремими робітниками відрізняється від середньої в першій бригаді в середньому на 86 шт., а в другій бригаді – в середньому на 12 шт. Таким чином середнє лінійне відхилення по виробництву деталей за зміну в першій бригаді у 5,7 разів більше, ніж у другій.

Середній квадрат відхилення або дисперсія (σ^2) визначається як середня арифметична з квадратів відхилень окремих варіантів від їх середньої.

В залежності від вихідних даних, дисперсію обчислюють за формулами:

а) дисперсія проста:

$$\sigma^2 = \frac{\sum \left(X - \overline{X}\right)^2}{n};$$

б) дисперсія зважена:

$$\sigma^2 = \frac{\sum \left(X - \overline{X}\right)^2 f}{\sum f}.$$

Середнє квадратичне відхилення (σ), являє собою корінь квадратичний з дисперсії.

Воно визначається за формулами:

а) середнє квадратичне відхилення просте:

$$\sigma = \sqrt{\frac{\sum \left(X - \overline{X}\right)^2}{n}};$$

б) середнє квадратичне відхилення зважене:

$$\sigma = \sqrt{\frac{\sum \left(X - \overline{X}\right)^2 f}{\sum f}}.$$

Середнє квадратичне відхилення називають стандартним відхиленням. Воно як і середнє лінійне відхилення, є іменованою величиною. Середнє квадратичне відхилення використовують при оцінці тісноти зв'язку між явищами, при обчисленні помилок вибіркового спостереження, дослідженні рядів розподілу та ін.

Для нормального або близького до нормального розподілу між середнім квадратичним і лінійним відхиленнями встановлено таке співвідношення: $\sigma = 1{,}25\,\bar{l}$.

Середнє квадратичне відхилення не завжди зручне для використання, тому що воно не дозволяє порівнювати між собою середні квадратичні відхилення у варіаційних рядах, варіанти яких виражені у різних одиниць виміру.

Щоб мати можливість порівнювати середні квадратичні відхилення різних варіаційних рядів, потрібно перейти від абсолютних до відносних показників варіації.

До числа відносних показників відносять коефіцієнти варіації:

а) лінійний $V_{\bar{l}} = \frac{\bar{l}}{\bar{x}} \cdot 100;$

б) квадратичний $V_{\sigma} = \frac{\sigma}{\bar{x}} \cdot 100;$

в) осциляції $V_R = \frac{R}{\bar{x}} \cdot 100;$

г) квадратний $V_Q = \frac{1/2\, R_Q}{M_e}.$

Розглянемо обчислення вищенаведених показників варіації на прикладі. Нехай маємо дані про розподіл виробів за масою.

Розрахункова таблиця

Маса виробу, г. (*X*)	Число виробів, шт. (*f*)	X	Xf	$\lvert X-\overline{X}\rvert$	$\lvert X-\overline{X}\rvert f$	$(X-\overline{X})^2$	$(X-\overline{X})^2 f$
До 100	5	97,5	487,5	10	50	100	500
100-105	19	102,5	1947,5	5	95	25	475
105-110	53	107,5	5697,5	0	0	0	0
110-115	17	112,5	1912,5	5	85	25	425
115-120	6	117,5	705,0	10	60	100	600
Разом:	100	x	10750,0	x	290	x	2000

Розмах варіації дорівнює:

$R = X_{\max} - X_{\min} = 117{,}5 - 97{,}5 = 20$ г.

Середня вага виробу:

$\overline{X} = \frac{\sum Xf}{\sum f} = \frac{10750}{100} = 107{,}5$ г.

Середнє лінійне відхилення:

$\bar{l} = \frac{\sum \lvert X-\overline{X}\rvert f}{\sum f} = \frac{290}{100} = 2{,}9$ г.

Середній квадрат відхилення (дисперсія):

$$\sigma^2 = \frac{\sum\left(X-\overline{X}\right)^2 f}{\sum f} = \frac{2000}{100} = 20 .$$

Середнє квадратичне відхилення:

$$\sigma = \sqrt{\sigma^2} = \sqrt{20} = 4{,}472 \text{ г.}$$

Коефіцієнт варіації середньої маси виробу:

$$V_\sigma = \frac{\sigma}{\overline{\overline{X}}} \cdot 100 = \frac{4{,}472}{107{,}5} \cdot 100 = 4{,}16\% .$$

Лінійний коефіцієнт варіації:

$$V_{\bar{l}} = \frac{\bar{l}}{\overline{\overline{X}}} \cdot 100 = \frac{2{,}9}{107{,}5} \cdot 100 = 2{,}7\% .$$

Як бачимо: $\sigma > \bar{l}$ $(4{,}472 > 2{,}900)$, $V_\sigma > V_{\bar{l}}$ $(4{,}16 > 2{,}70)$.

4.2. Спрощені способи розрахунку дисперсії

Обчислення дисперсії і середнього квадратичного відхилення пов’язано з великими і складними розрахунками, які потребують значних затрат часу і праці. Однак, їх можна значно спростити, якщо використати деякі математичні властивості дисперсії.

1. Якщо всі варіанти ознаки (X) зменшити на довільну величину (A), то дисперсія від цього не зміниться.

2. Якщо всі значення варіантів (X) зменшити в (i) раз, то дисперсія зменшиться в (i^2) раз, а середнє квадратичне відхилення – в (i) раз.

3. Якщо обчислити середній квадрат відхилень від любої величини «А», яка в тій чи іншій мірі відрізняється від середньої арифметичної ($\overline{x}$), то він завжди буде більший за середній квадрат відхилень, обчислений від середньої арифметичної, на квадрат різниці між середньою і цією умовно взятою величиною, тобто на $(\overline{x}-A)^2$.

$$\sigma_A^2 = \sigma^2 + (\overline{X} - A)^2, \quad \textit{або} \quad \sigma = \sigma_A^2 - (\overline{X} - A)^2,$$

де σ^2 – середній квадрат відхилень від середньої арифметичної ($\overline{x}$).

σ_A^2 – середній квадрат відхилень від довільної величини (А) .

Розглянемо методи розрахунку (σ_A^2) на прикладі.

Розрахунок дисперсії σ_A^2 при А=107,5 г.

Маса виробу, г. (Х)	Число виробів, шт. (f)	Х-А А=107,5	$(X\text{-}A)^2$	$(X\text{-}A)^2f$
97,5	5	10	100	500
102,5	19	5	25	475
107,5	53	0	0	0
112,5	17	5	25	425
117,5	6	10	100	600
Разом:	100	х	х	2000

$$\sigma_A^2 = \frac{\sum(X-A)^2 f}{\sum f} = \frac{2000}{100} = 20, \quad або$$

$$\sigma^2 = \sigma_A^2 - (\overline{X} - A)^2 = 20(107{,}5 - 107{,}5)^2 = 20.$$

Техніку розрахунку дисперсії і середнього квадратичного відхилення можна значно спростити використавши математичні властивості дисперсії, тобто спосіб моментів або відліку від умовного нуля.

Покажемо методику цього розрахунку на прикладі.

Розрахунок дисперсії способом моментів

Маса виробу, г. (Х)	Число виробів, шт. (f)	Х	Х-А А=107,5	$\frac{X-A}{i}$ i = 5	$\left(\frac{X-A}{i}\right)f$	$\left(\frac{X-A}{i}\right)^2$	$\left(\frac{X-A}{i}\right)^2 f$
До 100	5	97,5	-10	-2	-10	4	20
100-105	19	102,5	-5	-1	-19	1	19
105-110	53	107,5	0	0	0	0	0
110-115	17	112,5	5	1	17	1	17
115-120	6	117,5	10	2	12	4	24
Разом:	100	х	х	х	0	х	80

Обчислимо моменти першого і другого порядків.

$$m_1 = \frac{\sum\left(\frac{X-A}{i}\right)f}{\sum f} = \frac{0}{100} = 0;$$

$$m_2 = \frac{\sum\left(\frac{X-A}{i}\right)^2 f}{\sum f} = \frac{80}{100} = 0{,}8.$$

Дисперсія способом моментів визначається за формулою:

$$\sigma^2 = i^2\left(m_2 - m_1^2\right) = 5^2\left(0{,}8 - 0^2\right) = 25 \cdot 0{,}8 = 20.$$

В тому випадку, коли довільна величина A=0, а i=1, дисперсію за способом відліку від умовного нуля визначають за формулою:

$$\sigma^2 = \overline{X^2} - (\overline{X})^2 = \frac{\sum X^2 f}{\sum f} - \left(\frac{\sum Xf}{\sum f}\right)^2.$$

Розрахунок дисперсії за формулою $\sigma^2 = \overline{X^2} - (\overline{X})^2$

Маса виробу, г. (X)	Число виробів, шт. (f)	Xf	X^2	X^2f
97,5 102,5 107,5 112,5 117,5	5 19 53 17 6	487,5 1947,5 5697,5 1912,5 705,0	9506,25 10506,25 11556,25 12656,25 13806,25	47531,25 199618,75 612481,25 215156,25 82837,50
Разом:	100	10750,0	х	1 157 625,00

Підставимо розраховані дані в таблиці і отримаємо:

$$\sigma^2 = \frac{\sum X^2 f}{\sum f} - \left(\frac{\sum Xf}{\sum f}\right)^2 = \frac{1157625}{100} - \left(\frac{10750}{100}\right)^2 = 11576{,}25 - 11556{,}25 = 20.$$

Як бачимо, результати обчислення дисперсії за всіма вищенаведеними способами однакові.

4.3. Дисперсія альтернативної ознаки

Поряд з вимірюванням варіації кількісних ознак, в статистичній практиці доводиться вивчати і обчислювати варіації якісних (альтернативних) ознак.

Альтернативною називається ознака, яка може набирати лише два взаємо протилежних значення. Наприклад, продукція на підприємстві може бути якісна і не якісна, товарна і нетоварна, стандартна і нестандартна і т.д.

Кількісно варіацією альтернативної ознаки виражають двома значеннями: наявність ознаки у одиниць сукупності позначають через «1», а її відсутність – через «0». Тоді, якщо частку одиниць, які володіють даною ознакою позначити через «P», а частку одиниць, які не володіють ознакою, через «q», то P + q =1, звідси P = 1 – q , а q = 1 – P.

Спочатку обчислимо середнє значення альтернативної ознаки:

$$\overline{X} = \frac{\sum Xf}{\sum f} = \frac{(1 \cdot P) + (0 \cdot q)}{P + q} = P.$$

Отже, середнє значення альтернативної ознаки дорівнює частці одиниць, що володіють даною ознакою.

Дисперсія альтернативної ознаки визначається за формулою:

$$\sigma^2 = \frac{\sum\left(X-\overline{X}\right)^2 f}{\sum f} = \frac{(1-P)^2 P + (0-q)^2 q}{P+q} = q^2 P + P^2 q = Pq(q+P) = Pq = P(1-P).$$

Таким чином, дисперсія альтернативної ознаки визначається як добуток частки одиниць, які володіють даною ознакою, на частку одиниць, які нею не володіють.

Середнє квадратичне відхилення альтернативної ознаки визначається як корінь квадратний з дисперсії цієї ознаки:

$$\sigma = \sqrt{Pq} = \sqrt{P(1-P)}.$$

Розглянемо приклад. При обстеженні 100 взірців готових виробів, відібраних у випадковому порядку, 20% виявились неякісними. Потрібно визначити середню величину, дисперсію і середнє квадратичне відхилення.

Середня величина дорівнює частці:

$$\overline{X} = P = \frac{80}{100} = 0{,}8 \quad \text{або} \;\; 80\,\%.$$

Дисперсію альтернативної ознаки обчислимо за формулою:

$$\sigma^2 = P(1-P) = 0{,}8(1-0{,}8) = 0{,}16.$$

Середнє квадратичне відхилення дорівнює:

$$\sigma = \sqrt{\sigma^2} = \sqrt{0{,}16} = 0{,}4 \quad \text{або} \;\; 40\,\%.$$

4.4. Види дисперсій і правило їх додавання

Для більш детального вивчення тієї чи іншої ознаки, статистика за допомогою правила додавання дисперсій виявляє і досліджує вплив окремих чинників і умов, які зумовили дану варіацію в цілому. Виявити частку варіації, зумовлену певними чинниками, можна розбивши всю сукупність на групи за ознакою, вплив якої досліджується.

Якщо сукупність розбита на групи за одним чинником, то для неї можна обчислити такі види дисперсій: загальну, групові (часткові), середню з групових і міжгрупову.

Загальна дисперсія вивчається як середня арифметична з квадратів відхилень кожного значення ознаки від їх загальної середньої величини. Дана дисперсія характеризує варіацію досліджуваної ознаки за рахунок впливу всіх чинників. Її вираховують за формулами:

$$\sigma_з^2 = \frac{\sum\left(X - \overline{X}_з\right)^2}{n} \text{ – проста;} \qquad \sigma_з^2 = \frac{\sum\left(X - \overline{X}_з\right)^2 f}{\sum f} \text{ – зважена.}$$

Групова (часткова) дисперсія визначається як середня арифметична з квадратів відхилень кожного значення ознаки в групі від групової середньої. Групові дисперсії визначаються за формулами:

$$\sigma_i^2 = \frac{\sum\left(X_i - \overline{X}_i\right)^2}{n_i} \text{ – проста;} \qquad \sigma_i^2 = \frac{\sum\left(X_i - \overline{X}_i\right)^2 f_i}{\sum f_i} \text{ – зважена,}$$

де: σ_i^2 – групова (часткова) дисперсія;

X_i – індивідуальні значення групових ознак;

$\overline{X}_i$ – середнє значення ознак i-ї групи;

n – число одиниць сукупності в групі;

f_i – частоти в групі.

Групова (часткова) дисперсія вимірює варіацію ознаки тільки за рахунок чинників, які діють усередині групи, тобто всіх інших чинників, крім покладеного в основу групування.

Середню внутрішньогрупову (залишкову) дисперсію визначають за формулами середньої арифметичної в групових дисперсій:

$$\overline{\sigma_i^2} = \frac{\sum \sigma^2}{n} \text{ – проста;} \qquad \overline{\sigma_i^2} = \frac{\sum \sigma_i^2 f_i}{\sum f_i} \text{ – зважена.}$$

Міжгрупова дисперсія визначається як середня арифметична з квадратів відхилень групових середніх від загальної середньої за формулами:

$$\delta^2 = \frac{\sum\left(\overline{X_i} - \overline{X_з}\right)}{n} \text{ – проста;} \qquad \delta^2 = \frac{\sum\left(\overline{X_i} - \overline{X_з}\right) f}{\sum f_i} \text{ – зважена,}$$

де: δ^2 – міжгрупова дисперсія;

$\overline{X}_i$ – групові середні;

$\overline{X}_з$ – загальна середня;

n – число одиниць сукупності;

f_i – ваги або частоти.

Міжгрупова дисперсія відображає варіацію досліджуваної ознаки під впливом чинника, покладеного в основу групування.

Математичною статистикою доведено, що між загальною дисперсією, середньою з групових дисперсій і міжгруповою дисперсією існує зв'язок:

$$\sigma_з^2 = \overline{\sigma_i^2} + \delta^2.$$

Ця рівність в статистиці називається правилом додавання дисперсій. За допомогою даного правила, знаючи два види дисперсій, завжди можна визначити невідомий третій вид:

$$\overline{\sigma_i^2} = \sigma_з^2 - \delta^2; \qquad \delta^2 = \sigma_з^2 - \overline{\sigma_i^2}.$$

Правило додавання дисперсій використовують при проведенні вибіркового спостереження, для спрощеного обчислення дисперсій громіздкого варіаційного ряду, вимірювання сили зв’язку між явищами ті ін.

Проілюструємо застосування правила додавання дисперсій на прикладі. Нехай маємо відомості про годинну заробітну плату десяти робітників, розчленованих на дві групи за рівнем фахової підготовки.

Розрахункова таблиця

Перша група				Друга група			
Табельний номер робітника (n_1)	Годинна заробітна плата робітників-практиків, грн. (x_1)	$x_1 - \overline{x_1}$	$(x_1 - \overline{x_1})^2$	Табельний номер робітника (n_2)	Годинна заробітна плата робітників з ПТУ, грн. (x_2)	$x_2 - \overline{x_2}$	$(x_2 - \overline{x_2})^2$
1	6,0	-0,52	0,2704	6	8,2	-0,42	0,1764
2	6,3	-0,22	0,0484	7	8,5	-0,12	0,0144
3	6,5	-0,02	0,0004	8	8,6	-0,02	0,0004
4	6,8	0,28	0,0784	9	8,8	0,18	0,0324
5	7,0	0,48	0,2304	10	9,0	0,38	0,1444
Разом:	32,6	х	0,6280	Разом:	43,1	х	0,3680

Обчислимо середньогодинну заробітну плату робітників для кожної групи:

$$\overline{X}_1 = \frac{\sum X_1}{n_1} = \frac{32,6}{5} = 6,52 \text{ грн.}, \qquad \overline{X}_2 = \frac{\sum X_2}{n_2} = \frac{43,1}{5} = 8,62 \text{ грн.}$$

За даними розрахункової таблиці визначимо групові дисперсії:

$$\sigma_i^2 = \frac{\sum\left(X_1 - \overline{X}_1\right)^2}{n_1} = \frac{0,6280}{5} = 0,1256\,; \sigma_2^2 = \frac{\sum\left(X_2 - \overline{X}_2\right)^2}{n_2} = \frac{0,3680}{5} = 0,0736\,.$$

Середня з групових дисперсій дорівнює:

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 f_i}{\sum f_i} = \frac{0,1256 \cdot 5 + 0,0736 \cdot 5}{10} = \frac{0,996}{10} = 0,0996\,.$$

Для знаходження міжгрупової дисперсії потрібно обчислити загальну середню годинну заробітну плату всіх робітників за формулою середньої арифметично зваженої з групових середніх:

$$\overline{X}_з = \frac{\sum \overline{X}_i f_i}{\sum f_i} = \frac{6,52 \cdot 5 + 8,62 \cdot 5}{10} = \frac{75,7}{10} = 7,57 \text{ грн.},$$

або за формулою середньої арифметичної простої:

$$\overline{X}_з = \frac{\sum X_i}{n_i} = \frac{32,6 + 43,1}{10} = \frac{75,7}{10} = 7,57 \text{ грн.}$$

Обчислимо міжгрупову дисперсію:

$$\delta^2 = \frac{\sum\left(\overline{X}_i - \overline{X}_з\right)^2 f_i}{\sum f_i} = \frac{(6,52 - 7,57)^2 \cdot 5 + (8,62 - 7,57)^2 \cdot 5}{10} = \frac{11,025}{10} = 1,1025$$

Використавши правило додавання дисперсій, визначимо загальну дисперсію:

$$\sigma_з^2 = \overline{\sigma_i^2} + \delta^2 = 0,0996 + 1,1025 = 1,2021\,.$$

Правильність наших розрахунків можна перевірити, обчисливши загальну дисперсію звичайним способом:

$$\sigma_з^2 = \frac{\sum\left(X - \overline{X}\right)^2}{n} = \frac{(6,0 - 7,57)^2 + (6,3 - 5,57)^2 + (6,5 - 7,57)^2 + (6,8 - 7,57)^2 + (7,0 - 7,57) +}{10}$$

$$\frac{+ (8,2 - 7,57)^2 + (8,5 - 7,57)^2 + (8,6 - 7,57)^2 + (8,8 - 7,57)^2 + (9,0 - 7,57)^2}{10} = \frac{12,021}{10} = 1,2021.$$

Як показала перевірка, результати обчислення загальної дисперсії за обома способами такі самі, що свідчить про правильність проведених розрахунків.

Середнє квадратичне відхилення із загальної дисперсії дорівнює:

$$\sigma_з = \sqrt{\sigma_з^2} = \sqrt{1{,}2021} = 1{,}096\,\text{грн.}$$

Таким чином ми визначили, що середня годинна заробітна плата групи робітників в кількості десяти чоловік склала 7,57 грн., при середньому квадратичному відхиленні 1,096 грн.

При цьому можемо стверджувати, що якщо загальна дисперсія склала 1,2021, то міжгрупова дисперсія в розмірі 1,1025 викликана різницями кваліфікації в групах робітників, а середня внутрішньогрупова дисперсія в розмірі 0,0996 показує частку впливу інших, крім покладеного в основу групування, чинників. Правило додавання дисперсій дозволяє визначити частку складових частин в загальній дисперсії.

Поділивши міжгрупову дисперсію на загальну дисперсію отримаємо показник, який називається **коефіцієнтом детермінації** і показує, яка частка всієї варіації ознаки, зумовлена ознакою, покладеною в основу групування. В нашому прикладі:

$$\eta^2 = \frac{\delta^2}{\sigma_з^2} = \frac{1{,}1025}{1{,}2021} = 0{,}917 \quad \textit{або} \quad 91{,}7\,\%,$$

де: η^2 – (грецька буква «ета» в квадраті) – коефіцієнт детермінації.

Значить, фактор кваліфікації робітників на 91,7 % зумовлює варіацію їхньої середньогодинної заробітної плати.

Корінь квадратний з коефіцієнта детермінації називають **емпіричним кореляційним відношенням**, яке показує тісноту (силу) зв'язку між групувальною та результативною ознаками.

$$\eta = \sqrt{\frac{\delta^2}{\sigma_з^2}} = \sqrt{0{,}917} = 0{,}958.$$

Це говорить про те, що зв'язок між кваліфікацією робітників і їхньою середньогодинною заробітною платою дуже сильний (тісний).

ЛЕКЦІЯ 5. РЯДИ РОЗПОДІЛУ

ПЛАН:

5.1. Поняття про ряди розподілу, їх види.

5.2. Форми рядів розподілу та їх характеристика.

5.3. Криві розподілу та способи перевірки гіпотез.

5.4. Графічне зображення рядів розподілу.

5.1. Поняття про ряди розподілу, їх види

В результаті статистичного групування отримують ряди цифрових показників, які характеризують розподіл одиниць досліджуваної сукупності за варіаційною ознакою. Такі ряди називаються **рядами розподілу**.

Ряд розподілу складається з двох елементів – варіантів та частот. **Варіанти** «Х» – це окремі значення групувальної ознаки, які розташовані в певній послідовності. **Частоти** «*f*» – це числа, які показують, скільки разів певне значення ознаки зустрічається у сукупності, або скільки одиниць сукупності припадає на кожну групу.

Ряди розподілу відіграють важливу роль при вивченні складу сукупності за досліджуваною ознакою, закономірностей розподілу, використовуються при визначенні середніх величин, показників варіації, взаємозв’язку та ін.

В залежності від характеру групувальної ознаки ряди розподілу поділяються на **атрибутивні** та **варіаційні (кількісні)**.

В атрибутивних рядах розподілу варіанти не мають чисельного виразу.

Варіаційні ряди, в яких варіанти мають числовий вираз, поділяються на дискретні та інтервальні. В дискретних рядах варіанти являють собою дискретні числа, в інтервальних – це інтервали групування.

В тому випадку, коли виконуються групування за двома і більше ознаками отримують комбінаційний ряд розподілу.

При побудові атрибутивних рядів розподілу варіанти потрібно розташувати в логічній послідовності. При використанні дискретних та інтервальних варіаційних рядів варіанти записують в порядку зростання. Для інтервальних рядів важливим чітке розмежування варіант.

Розрізняють ряди розподілу з абсолютними, відносними та нагромадженими частотами. Нагромаджені частоти ще називають кумулятивними. В першому випадку чистоти являють собою абсолютні числа, в другому – питому вагу або частку кожної групи.

Ряди розподілу з абсолютними частотами характеризують склад сукупності, а з відносними – структуру сукупності.

Ряди розподілу з нагромадженими (кумулятивними) частотами показують, яка чисельність або питома вага одиниць має значення ознаки менше за дане. Кумулятивні частоти знаходять шляхом їх сумування по групах.

Щільність розподілу – це кількість одиниць сукупності, що припадає на одиницю ширини інтервалу групувальної ознаки. Розрізняють **абсолютну** щільність і **відносну**:

$$f_d^a = \frac{f}{i}; \qquad f_d^в = \frac{p}{i},$$

де f – частота; p – частка (доля);

i – величина (розмір) інтервалу.

Приклад 1

Маємо наступні дані про розподіл сімей за місячним доходом:

Місячний дохід на одного члена сім'ї, грн.	Число сімей		Щільність розподілу	
	одиниць	%	число сімей	%
До 200,0	34	13,2	0,17	0,066
200,0 – 400,0	52	20,2	0,26	0,101
400,0 – 600,0	72	27,9	0,36	0,140
600,0 – 1000,0	70	27,1	0,18	0,068
1000,0 – 1500,0	30	11,6	0,06	0,023
Разом:	258	100,0	х	х

Отже, найбільшу щільність розподілу має третя група сімей з доходом 400,0 – 600,0 грн. на одного члена сім'ї.

Інтерполяція в рядах розподілу визначає, скільки одиниць сукупності (або процентів) мають значення ознаки менше за завдане. Для інтерполяції використовують як абсолютні так і відносні нагромаджені частоти.

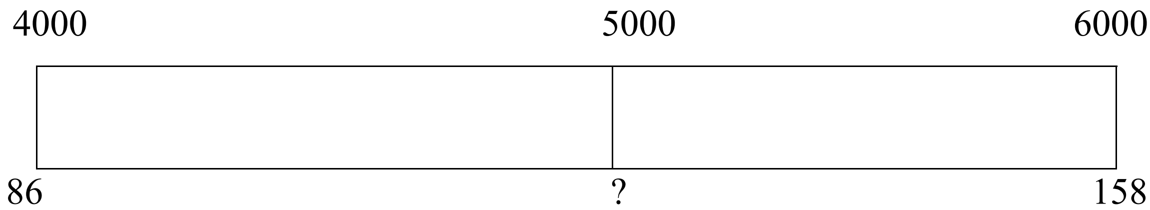
Приклад 2

Маємо наступні дані про розподіл сімей за місячним доходом:

Місячний дохід на одного члена сім'ї, грн.	Нагромаджені частоти	
	число сімей	%
До 200,0	34	13,0
2000 – 4000	86	33,4
4000 – 6000	158	61,3
6000 – 10000	228	88,4
10000 – 15000	258	100,0

Визначити: а) скільки сімей мають дохід менше 5000 грн. на одного члена сім'ї.

Розрахунок:



$f_x < 5000 = 86 + 0{,}5(158 - 86) = 122$ сім'ї, або

$f_x < 5000 = 33{,}4 + 0{,}5(61{,}3 - 33{,}4) = 47{,}4\%$.

б) скільки сімей мають дохід менше 7300 грн. на одного члені сім'ї.

$$f_x < 7300 = 158 + \frac{7300 - 6000}{4000} \cdot (228 - 158) = 181 \text{ сім'я.}$$

$$f_x < 7300 = 61{,}3 + 0{,}325 \cdot (88{,}4 - 61{,}3) = 70{,}1\% .$$

З допомогою інтерполяції визначають менше якого значення мають ознаку певна кількість (питома вага) одиниць сукупності. Наприклад, менше якого мають дохід 50 %, або 129 сімей.

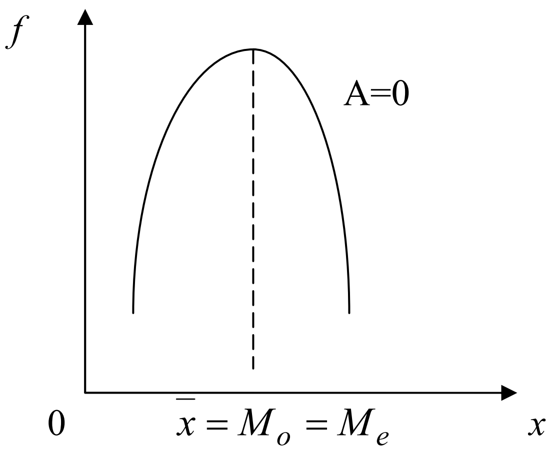
Розрахунок: $xf = 50\% = 4000 + \dfrac{50{,}0 - 33{,}4}{61{,}3 - 33{,}4} \cdot (6000 - 4000) = 5200$ грн.

Отже, половина сімей має дохід менше 5200 грн. на одного члена сім'ї.

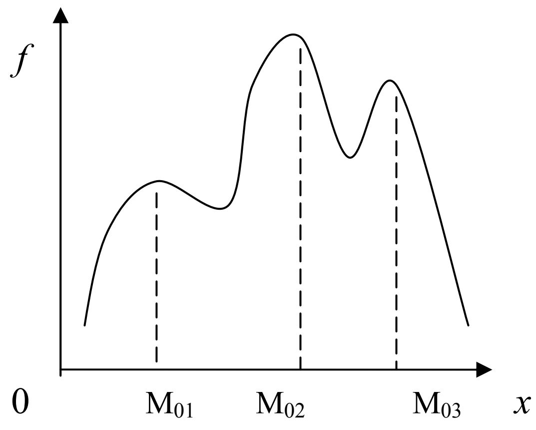
5.2. Форми рядів розподілу, та їх характеристика

Різноманітність статистичних сукупностей – передумова різних форм співвідношення частот і значень варіюючої ознаки. За своєю формою розподіли поділяють на такі види: одно-, дво і багатовершинні. Наявність двох і більше вершин свідчить про неоднорідність сукупності, про поєднання в ній груп з різними рівнями ознаки. Розподіли якісно однорідних сукупностей, як правило, одновершинні. Серед одновершинних розподілів є симетричні (скошені), гостро- і плосковершинні.

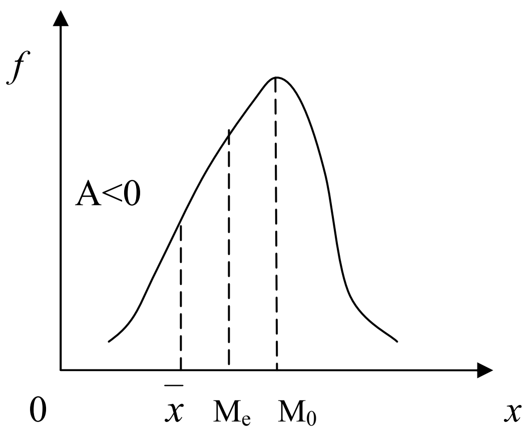
а)



б)

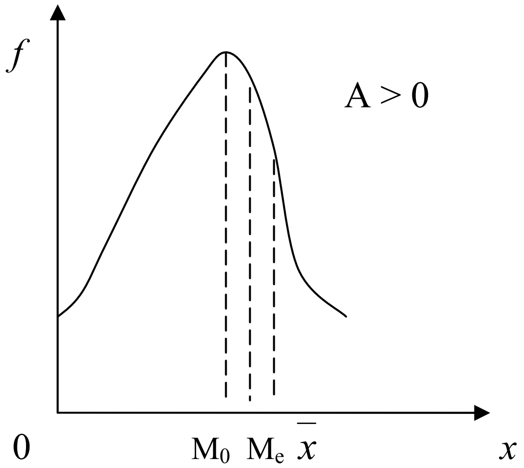


в)



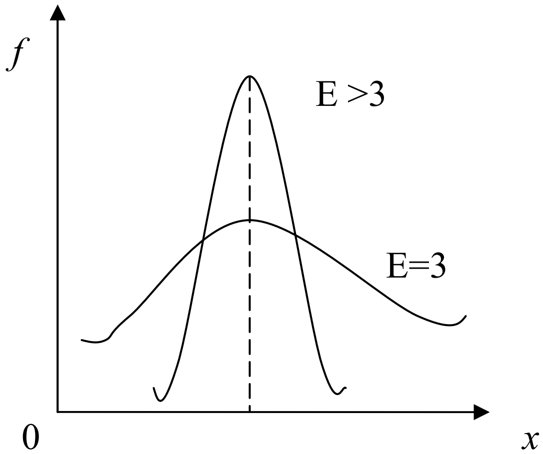
- лівостороння (асиметрія)

г)



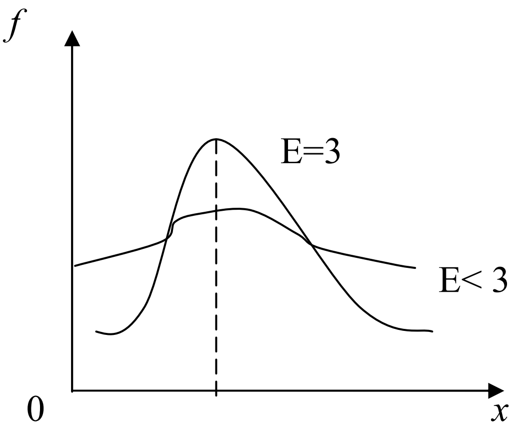
- правостороння (асиметрія)

д)



- гостровершинний (ексцес)

е)



- плосковершинний (ексцес)

Мал. 1. Різновиди форм розподілу

У **симетричному** розподілі рівновіддалені від центру значення ознаки мають однакові частоти, в **асиметричному** – вершина розподілу зміщена. Напрям асиметрії протилежний напряму зміщення вершини. Якщо вершина зміщені вліво, **то** це правостороння асиметрія, і навпаки. Асиметрія виникає внаслідок обмеженої варіації в одному напряму або під впливом домінуючої причини розвитку, яка веде до зміщення центру розподілу.

Найпростішою мірою асиметрії є відхилення між середньою арифметичною і медіаною чи модою. В симетричному розподілі характеристики центру мають однакові значення $\overline{X} = M_e = M_0$; в асиметричному між ними існують певні розбіжності. При правосторонній асиметрії $\overline{X} > M_e > M_0$, при лівосторонній асиметрії, навпаки, $\overline{X} < M_e < M_0$.

Стандартизовані відхилення $A = \frac{\overline{X} - M_e}{\sigma}$**,** або $A = \frac{\overline{X} - M_0}{\sigma}$, характеризують напрям і міру скошеності розподілу. В симетричному розподілі А=0, при правосторонній асиметрії А>0, при лівосторонній А<0.

Приклад 3

Маємо такі дані про розподіл посівної площі господарств району за урожайністю гречки

Урожайність гречки, ц/га.	Посівна площа, га. (f)	x	xf	$x-\overline{x}$	$(x-\overline{x})^2$	$(x-\overline{x})^2 f$
10-12	17	11	187	-3,9	15,21	258,57
12-14	20	13	260	-1,9	3,61	72,20
14-16	27	15	405	0,1	0,01	0,27
16-18	23	17	391	2,1	4,41	101,43
18-20	13	19	247	4,1	16,81	218,53
Разом:	100	x	1490	x	x	651,00

$$\overline{x} = \frac{\sum xf}{\sum f} = \frac{1490}{100} = 14{,}9 \text{ ц/га}; \quad \sigma^2 = \frac{\sum\left(x-\overline{x}\right)^2 f}{\sum f} = \frac{651}{100} = 6{,}51;$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{6{,}51} = 2{,}55 \text{ ц/га};$$

$$M_0 = x_{m0} + i\frac{f_2 - f_1}{\left(f_2 - f_1\right) + \left(f_2 - f_3\right)} = 14 + 2\frac{27-20}{\left(27-20\right)+\left(27-23\right)} = 15{,}27 \text{ ц/га.}$$

$$A = \frac{\overline{x} - M_0}{\sigma} = \frac{14{,}9 - 15{,}27}{2{,}55} = -0{,}145.$$

Стандартизоване відхилення свідчить про дуже незначну лівосторонню асиметрію, а тому розподіл можна вважати симетричним.

Гостровершинність розподілу відображає скупченість значень ознаки навколо середньої величини і називається **ексцесом**. На практиці часто в одному розподілі поєднуються всі названі особливості: одновершинний розподіл може бути симетричним і гостровершинним або скошеним і плосковершинним.

Як узагальнюючі характеристики розподілу використовують **моменти**. За допомогою невеликого їх числа можна описати будь-який розподіл. **Момент розподілу** – це середня арифметична k-го ступеня відхилень x-A.

Залежно від величини «A» моменти поділяють на первинні A=0, центральні A=$\bar{x}$ і умовні A=const. Ступінь «k» визначає порядок моменту.

На практиці використовуються:

Початкові моменти:

– нульового порядку (k=0) $M_0 = \dfrac{\sum x^0 f}{\sum f} = 1$;

– першого порядку (k=1) $M_1 = \dfrac{\sum x^1 f}{\sum f} = \bar{x}$;

– другого порядку (k=2) $M_2 = \dfrac{\sum x^2 f}{\sum f} = \overline{x^2}$;

– третього порядку (k=3) $M_3 = \dfrac{\sum x^3 f}{\sum f} = \overline{x^3}$;

– четвертого порядку (k=4) $M_4 = \dfrac{\sum x^4 f}{\sum f} = \overline{x^4}$.

Початкові моменти відносно «x_0»(умовні):

– нульового порядку $M'_0 = \dfrac{\sum (x - x_0)^0 f}{\sum f} = 1$;

– першого порядку $M'_1 = \dfrac{\sum (x - x_0)^1 f}{\sum f} = \bar{x} - x_0$;

– другого порядку $M'_2 = \dfrac{\sum (x - x_0)^2 f}{\sum f}$;

– третього порядку $M'_3 = \dfrac{\sum (x - x_0)^3 f}{\sum f}$;

– четвертого порядку $M'_4 = \dfrac{\sum (x - x_0)^4 f}{\sum f}$.

Центральні моменти:

– нульового порядку $\mu_0 = \dfrac{\sum (x - \bar{x})^0 f}{\sum f} = 1$;

– першого порядку $\mu_1 = \dfrac{\sum (x - \bar{x}) f}{\sum f} = 0$; $\mu_1 = \dfrac{\sum xf - \bar{x} \sum f}{\sum f} = 0$;

– другого порядку $\mu_2 = \dfrac{\sum (x - \bar{x})^2 f}{\sum f} = \sigma^2$;

– третього порядку $\mu_3 = \dfrac{\sum (x - \bar{x})^3 f}{\sum f}$;

– четвертого порядку $\mu_4 = \dfrac{\sum (x - \bar{x})^4 f}{\sum f}$.

Первинний момент першого порядку – це середня арифметична «$\bar{x}$», другого – середній квадрат значень ознаки «$\overline{x^2}$». Центральний момент другого порядку характеризує варіацію $\mu_2 = \sigma^2$, третього – асиметрію, четвертого – ексцес.

Для порівняння ступеня асиметрії різних розподілів, використовують стандартизований момент: $A = \dfrac{\mu_3}{\sigma^3}$.

Вважають, що при А<0,25 асиметрія низька, якщо А не перевищує 0,5 – середня і при А>0,5 – висока.

Для вимірювання ексцесу використовують аналогічно побудований коефіцієнт, тобто стандартизований момент четвертого порядку: $E = \dfrac{\mu_4}{\sigma^4}$.

У симетричному розподілі Е=3, при гостровершинному – Е>3, плосковершинному Е<3.

Оцінка **концентрації** розподілу ґрунтується на порівнянні часток розподілу елементів сукупності «d_j» та обсягу ознаки «xd_j». Якщо розподіл рівномірний, $d_j = xd_j$. Значні відхилення $xd_j - d_j$ свідчить про певну концентрацію елементів сукупності. $K = \frac{1}{2}\left|xd_j - d_j\right|$. При рівномірному розподілі К=0, повній концентрації К=1. «К» буде тим більший, чим більший ступінь концентрації.

5.3. Криві розподілу та способи перевірки гіпотез

Поглиблюючи аналіз, можна описати закономірність співвідношення варіантів і частот певною функцією, яку називають **теоретичною кривою**. Серед безлічі кривих розподілу найпоширенішою виявилась нормальна крива. Вона застосовується як стандарт, з яким порівнюють інші розподіли, а також відіграє значну роль при вирішенні завдань вибіркового, кореляційно-регресійного, факторного та інших статистичних методів.

Нормальний розподіл близький до інших одновершинних розподілів. Його часто використовують як перше наближення при моделюванні. Деякі розподіли, які не є нормальними, приводять до такого виду перетворенням змінної «х» на «lgx». Логарифмічною нормальною кривою можна описати низку асиметричних розподілів, передусім з правосторонньою асиметрією.

Частоти, які відповідають теоретичній кривій, називають **теоретичними**. Для нормального розподілу їх визначають за формулою: $f' = \sum f \frac{i}{\sigma} f(t)$,

де $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.

Інтегральна функція розподілу:

$$F(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}}; \quad dt = \frac{1}{\sqrt{2\pi}} \int_{-x}^{+x} e^{-\frac{t^2}{2}} dt.$$

Функція «$F(x)$» табульована.

Аналітично нормальний розподіл списується рівнянням: $\hat{y}_t = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}}$,

де $\hat{y}_t$ – ордината кривої нормального розподілу (частоти);

t – нормоване відхилення, що дорівнює $\frac{x - \bar{x}}{\sigma}$;

σ – середнє квадратичне відхилення;

π – величина відношення довжини кола до діаметру, π =3,1415;

е – основа натуральних логарифмів, $e \approx 2{,}7182$.

Для перевірки відповідності фактичного розподілу нормальному, потрібно частоти фактичного розподілу порівняти з теоретичними частотами, які характерні для нормального розподілу. Для цього потрібно за фактичними даними вирахувати теоретичні частоти кривої нормального розподілу. Тобто, фактичні криву розподілу потрібно вирівняти кривою нормального розподілу.

Після розрахунку теоретичних частот виникає потреба перевірки висунутої гіпотези про відповідність або невідповідність того чи іншого теоретичного закону розподілу, прийнятого в якості математичної моделі для емпіричного розподілу. При цьому виходять з того, що якщо відхилення між емпіричними і теоретичними частотами можна вважати випадковими, тоді гіпотеза про те, що прийнятий теоретичний розподіл відповідає даному емпіричному, не відхиляється.

Математична статистика дає декілька показників, за котрими можна судити, наскільки фактичний розподіл узгоджується з нормальним. Ці показники називаються **критеріями узгодження**. Критерій узгодження виступає у вигляді деякої величини, яка оцінює досліджуване явище з певною ймовірністю.

Статистика використовує критерії узгодження К. Пірсона (χ^2), А.Н. Колмогорова (λ), П.С. Ястремського (L), В.І. Романовського (R), Р. Фішера (Ƶ), Вілконсона і ін.

Одними із основних і найбільш розповсюджених показників є критерій «χ^2» і «λ». **Критерій «χ^2»** запропонував англійський статистик К. Пірсон:

$$\chi^2 = \sum \frac{(f - f')^2}{f'},$$

де f i f' – відповідно фактичні і теоретичні частоти.

По таблицях визначаємо імовірність досліджуваного значення χ^2 в залежності від *числа ступенів вільності*. Число ступенів вільності визначається за формулою: k = m – r,

де m – число груп; r – число обмежених зв'язків.

Якщо фактичне χ^2 менше табличного $\left(\chi^2_{ф} < \chi^2_{табл.}\right)$, тоді при прийнятому рівні значимості розходження між фактичними і теоретичними частотами вважаються випадковими, гіпотеза про закон розподілу приймається.

Приклад 4

Маємо дані про урожайність озимої пшениці господарств регіону

Урожайність озимої пшениці, ц/га	Кількість господарств (f)	X	$\lvert X-\overline{X}\rvert$	$t=\frac{X-\overline{X}}{\sigma}$	$f(t)$	f'	$f-f'$	$(f-f')^2$	$\frac{(f-f')^2}{f'}$
40-42	4	41	5,72	2,22	0,0339	2	2	4	2,00
42-44	7	43	3,72	1,44	0,1415	11	-4	16	1,45
44-46	28	45	1,72	0,67	0,3187	25	3	9	0,36
46-48	35	47	0,28	0,11	0,3965	31	4	16	0,52
48-50	16	49	2,28	0,88	0,2709	21	-5	25	1,19
50-52	6	51	4,28	1,66	0,1006	8	-2	4	0,50
52-54	4	53	6,28	2,43	0,0208	2	2	4	2,00
Разом:	100	x	x	x	x	100	x	x	8,02

$\overline{X}=\frac{\sum xf}{\sum f}=\frac{4672}{100}=46{,}72$ ц/га; $\sigma=\sqrt{6{,}6414}=2{,}58$ ц/га;

$\sum f\frac{i}{\sigma}=100\frac{2}{2{,}58}=77{,}5$; $f'=\sum f\frac{i}{\sigma}\cdot f(t)=77{,}5\cdot 0{,}0339=2$;

$\left(\chi^2_{ф}<\chi^2_{табл.}\right)$; $k=m-r=7-3=4$; $\chi^2_{табл.}(p=0{,}99)=13{,}28.$

Висновок: оскільки фактичний критерій $\chi^2_{ф}$ значно менший $\chi^2_{табл.}$(8,02<13,28), то із ймовірністю 0,99 можна вважати доведеним, що тип розподілу вибраний правильно, тобто, що розподіл господарств за врожайністю озимої пшениці є нормальним.

Мал. 2. Крива розподілу господарств регіону за урожайністю озимої пшениці

Критерій узгодження А.Н. Колмогорова (λ) оцінює близькість фактичного розподілу до теоретичного шляхом знаходження величини (D), тобто максимальної різниці нагромаджених частот фактичного і теоретичного розподілів.

Критерій А.Н. Колмогорова розраховується за формулою: $\lambda = D \cdot \sqrt{n}$,
де D – абсолютна величина максимальної різниці між нагромадженими частками (долями) фактичного і теоретичного рядів розподілу;

n – чисельність одиниць сукупності.

Якщо розподіл задано в частотах, тоді формула має вигляд:

$$\lambda = \frac{D}{\sqrt{n}}.$$

Приклад 5

Розрахунок критерія «λ»

Номер груп	Нагромаджені частоти		Відхилення
	Емпіричні S_f	Теоретичні $S_{f'}$	$\left\|S_f - S_{f'}\right\|$
1	4	2	2
2	11	13	2
3	39	38	1
4	74	69	5
5	90	90	0
6	96	98	2
7	100	100	0

$$\lambda = \frac{D}{\sqrt{n}} = \frac{5}{\sqrt{100}} = \frac{5}{10} = 0{,}5.$$

По спеціальній таблиці ймовірностей для критерія узгодження «λ» знаходимо, що значенню $\lambda = 0{,}5$ відповідає ймовірність 0,9639, це означає, що з ймовірністю 0,9639 можна стверджувати, що розподіл ділянок за урожайністю є нормальним.

Критерій узгодження В.І. Романовського використовується для оцінки наближення фактичного розподілу до теоретичного.

$$R = \frac{\chi^2 - k}{\sqrt{2k}}.$$

Приклад 6

$$R = \frac{\chi^2 - k}{\sqrt{2k}} = \frac{8{,}02 - 4}{\sqrt{2 \cdot 4}} = \frac{4{,}02}{2{,}8284} = 1{,}42 < 3.$$

Якщо при дослідженні наближення фактичного розподілу до теоретичного величина цього виразу менша трьох (1,42<3), це дає підставу для ствердження про можливість прийняття теоретичного розподілу за закон даного розподілу. Тобто розподіл господарств за врожайністю озимої пшениці є нормальним.

Критерій узгодження Б.С. Ястремського (L) використовується для прямої відповіді на питання про міру розбіжності між фактичним і теоретичним розподілом:

$$L = \frac{\chi^2 - n}{\sqrt{2n + 4Q}},$$

де $Q = \frac{(f_i - f_i')^2}{f_i(1 - p_i)}$ – при кількості груп, менше 20(n<20), дорівнює 0,6.

Приклад 7

$$L = \frac{\chi^2 - n}{\sqrt{2n + 4Q}} = \frac{8{,}02 - 7}{\sqrt{2 \cdot 7 + 4 \cdot 0{,}6}} = \frac{1{,}02}{4{,}02} = 0{,}25.$$

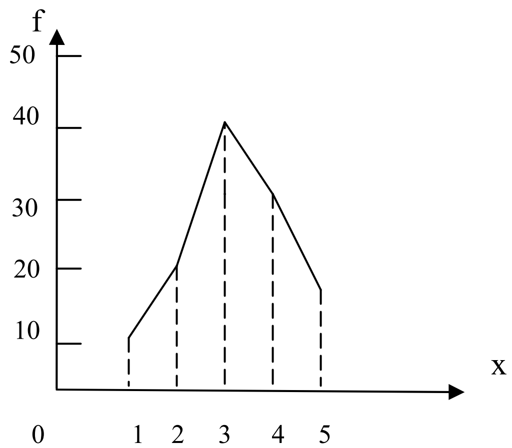
Так як величина L<3 (0,25<3), то з ймовірністю 0,997 можна стверджувати, що розподіл господарств за врожайністю озимої пшениці є нормальним.

5.4. Графічне зображення рядів розподілу

Для графічного зображення рядів розподілу використовують такі графіки як полігон, гістограма, кумулята, огіва, крива концентрації (Лоренца), показникова крива, крива Парето, антимода.

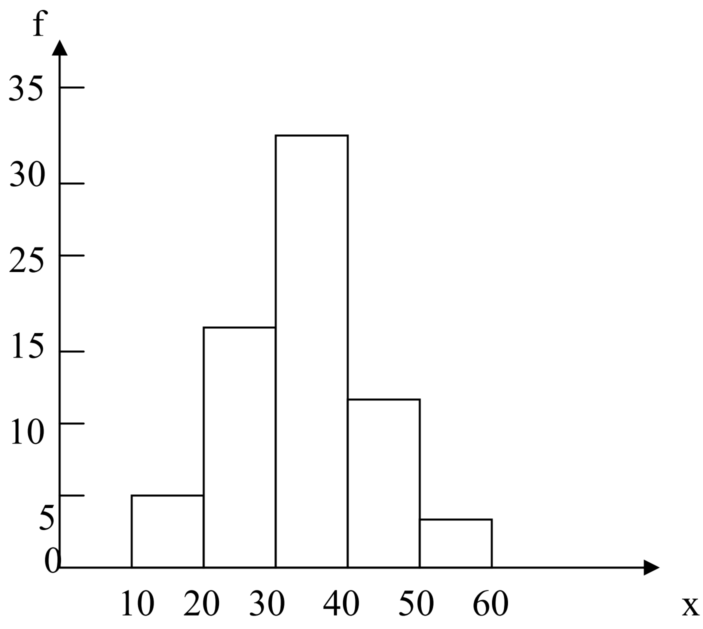
Полігон – графічне зображення варіаційного ряду в прямокутній системі координат, при котрому величина ознаки відкладається на осі абсцис, частоти або частки (щільність розподілу) – на осі ординат.

Частіше полігон застосовується для зображення дискретного варіаційного ряду, однак може використовуватись і для інтервального ряду. Графік полігона має вигляд:



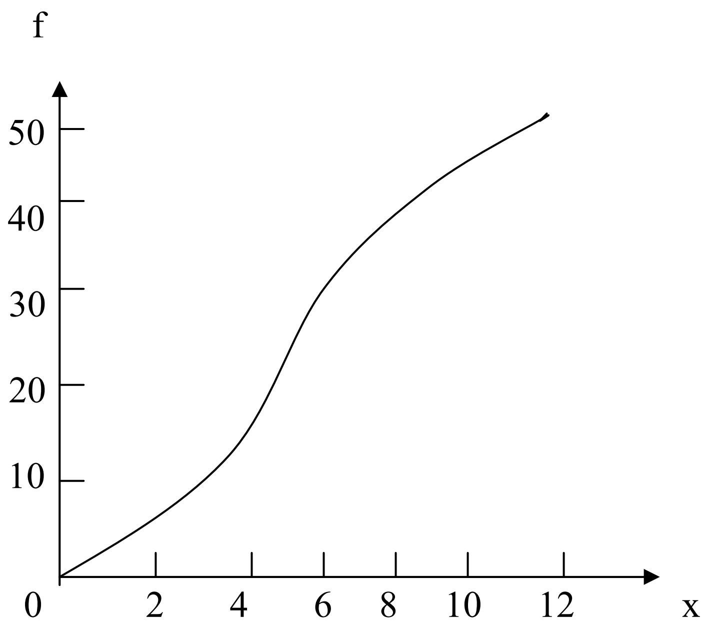
Мал. 3. Полігон

Гістограма – графічне зображення інтервального варіаційного ряду у вигляді прямокутників різної висоти, основа яких – відрізки осі абсцис, котрі відповідають інтервалам зміни ознаки. Висоти прямокутників пропорціональні при рівності інтервалів частотам або часткам інтервалів, а при нерівності – щільностям (абсолютним чи відносним). Графік гістограми має вигляд:



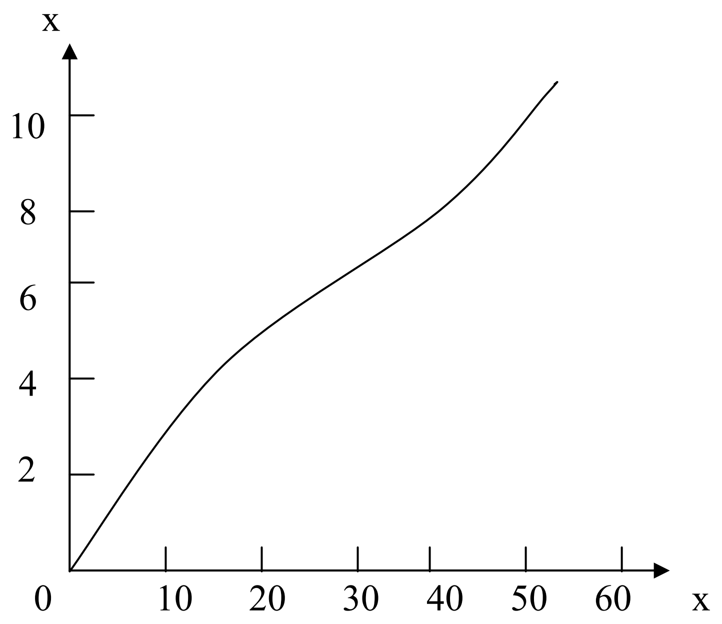
Мал. 4. Гістограма

Кумулятивна крива (кумулята) – графічне зображення варіаційного ряду, складене за нагромадженими частотами або частками.



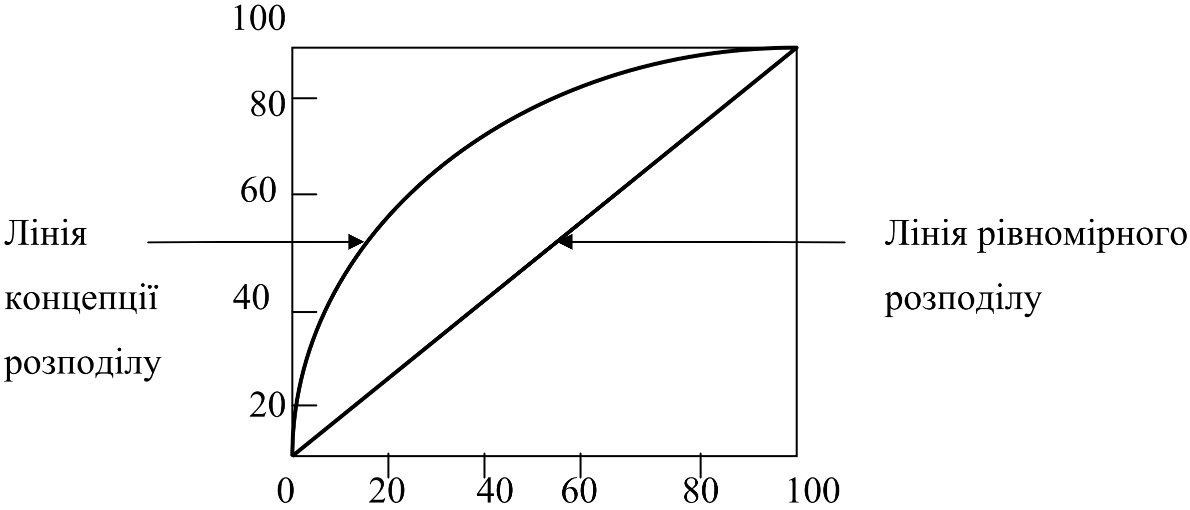
Мал. 5. Кумулята

Огіва – графічне зображення варіаційного ряду, складене аналогічно кумуляті, але на осі ординат наносяться значення ознаки, а на осі абсцис – нагромаджені частоти.



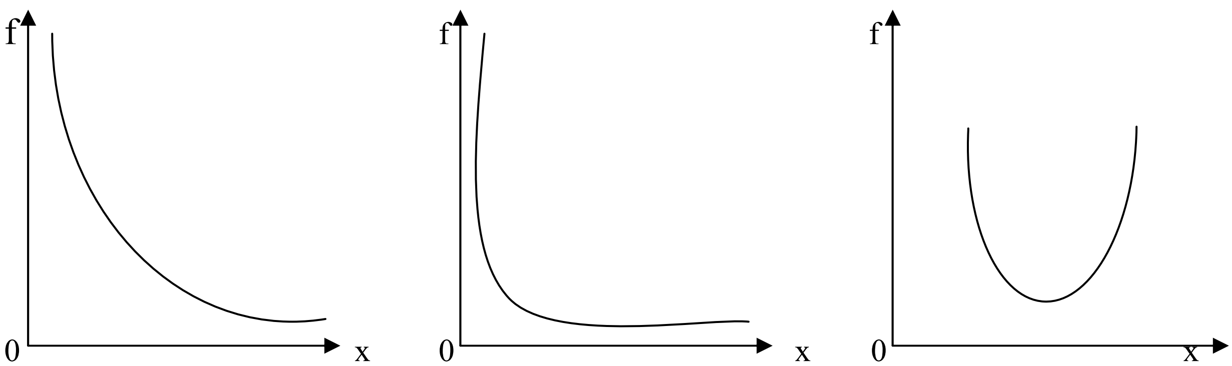
Мал. 6. Огіва

Крива концентрації Лоренца – використовується в якості показника ступеня оцінки концентрації розподілу частот. Крива Лоренца має вигляд:



Мал. 7. Крива концентрації (Лоренца)

До інших видів графічного зображення рядів розподілу відносять: показникові криву, криву Парето і антимоду. Ці графіки мають вигляд:



Мал.8. Показникові крива   Мал.9. Крива Парето   Мал. 10. Антимода

Прикладом показникового розподілу є термін служби предметів, які вибувають із-за аварійності – довговічність посуду в підприємствах громадського харчування, тривалість телефонних викликів, час між простоями верстатів і т.д.

Крива Парето показує розподіл доходів у відповідності з їх величиною.

Антимода – абсциса нижньої точки, розміщеної в центральній частині U – подібного розподілу ознаки. По обидві сторони антимоди частоти зростають.

ЛЕКЦІЯ 6. СТАТИСТИЧНІ МЕТОДИ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКІВ

ПЛАН:

6.1. Зв'язки суспільних явищ і завдання їх статистичного вивчення.

6.2. Загальні методи вивчення зв'язків.

6.3. Кореляційний і регресійний методи аналізу зв'язку.

6.4. Нелінійні залежності.

6.5. Багатофакторний кореляційно-регресійний аналіз.

6.6. Непараметричні показники тісноти зв'язку.

6.1. Зв'язки суспільних явищ і завдання їх статистичного вивчення

Одним з найбільш загальних законів об'єктивного світу є закон зв'язку і залежності між явищами суспільного життя. Ці явища найбільш складні, оскільки вони формуються під дією багаточисельних, різноманітних і взаємозв'язаних чинників.

Усі явища суспільного життя існують не ізольовано, вони органічно зв'язані між собою, залежать одні від одних і знаходяться в постійному русі і розвитку. Розкриваючи взаємозв'язки і взаємозалежності між явищами можна пізнати їх суть і закони розвитку. Тому вивчення взаємозв'язків є основним завданням всякого статистичного аналізу.

Суспільні явища або окремі їх ознаки, які впливають на інші і обумовлюють їх зміну називаються **факторними**, а суспільні явища або окремі їх ознаки, які змінюються під впливом факторних, називаються **результативними**.

За характером залежності явищ розрізняють функціональні і кореляційні зв’язки.

Функціональним називається зв'язок, при якому певному значенню факторної ознаки завжди відповідає одне значення результативної ознаки. Функціональні зв’язки характеризуються певною відповідністю між причиною і наслідком.

Кореляційним називається зв'язок, при якому кожному значенню факторної ознаки, відповідає декілька значень результативної ознаки. В кореляційних зв’язках між причиною і наслідком немає повної відповідності, а спостерігається лише певне співвідношення.

За напрямом розрізняють зв’язки прямі і обернені.

Прямий зв'язок – це такий зв'язок, коли із зростанням факторної ознаки, результативна також зростає.

При оберненому зв’язку із збільшенням факторної ознаки результативна зменшується або, навпаки, із зменшенням факторної ознаки, результативна зростає.

За формою зв'язок ділиться на прямолінійний і криволінійний.

При **прямолінійній** кореляційній залежності рівним змінним середніх значень факторної ознаки відповідають приблизно рівні зміни середніх значень результативної ознаки.

При **криволінійній** кореляційній залежності рівним змінним середніх значень факторної ознаки відповідають нерівні зміни середніх значень результативної ознаки.

Статистичне вивчення взаємозв’язків розв’язує наступні завдання:

а) визначаються форми зв’язку;

б) вимірюється тіснота (сила) зв’язку;

в) виявляється вплив окремих чинників на результативну ознаку.

6.2. Загальні методи вивчення зв’язків

Зв’язки і залежності суспільних явищ вивчаються різними методами, які дають уявлення про їх наявність і характер. До цих методів відносять:

балансовий метод, метод порівняння паралельних рядів, графічний метод, метод аналітичних групувань, індексний метод, кореляційно-регресійний аналіз та ін.

Одним з поширених методів статистичного вивчення зв'язків суспільних явищ є балансовий метод, як прийом аналізу зв'язків і пропорцій в економіці.

Статистичний баланс являє собою систему показників, яка складається із двох сум абсолютних величин, зв'язаних між собою знаком рівності.

$$а+б=в+г$$

Цю балансову ув'язку можна зобразити через балансове рівняння: залишок на початок + поступлення = видатки + залишок на кінець. Наведена балансова рівність характеризує єдиний процес руху матеріальних ресурсів і показує взаємозв'язок і пропорції окремих елементів цього процесу.

Метод порівняння паралельних рядів полягає в тому, що отримані в результаті групування і лічильної обробки матеріали статистичного спостереження рангованими паралельними рядами за факторною ознакою. Паралельно записуються значення результативної ознаки. Це дає можливість, порівнюючи їх, простежити співвідношення, виявити існування зв'язку і його напрямок.

Покажемо застосування цього методу на прикладі. Нехай маємо такі дані про роботу десяти однотипних підприємств.

Номер підприємства	1	2	3	4	5	6	7	8	9	10
Вартість основних виробничих фондів, млн. грн., (x)	5,3	6,4	7,9	8,3	9,2	10,1	12,5	13,0	14,6	15,7
Випуск продукції, млн. грн., (y)	5,8	7,6	8,7	9,1	11,9	12,3	13,8	14,0	15,2	17,6

З таблиці видно, що із збільшенням вартості основних виробничих фондів випуск продукції зростає.

На основі порівняння паралельних рядів вираховують напрямок і силу зв'язку за допомогою коефіцієнта Фехнера і кореляції рангів Спірмена.

Взаємозв'язок між вартістю основних виробничих фондів і випуском однорідної продукції по десяти підприємствах

Номер підпри-ємства	Вартість основних виробничих фондів, млн. грн. (x)	Випуск продукції, млн. грн., (y)	$x-\overline{x}$	$y-\overline{y}$	С або Н
1	5,3	5,8	-	-	С
2	6,4	7,6	-	-	С
3	7,9	8,7	-	-	С
4	8,3	9,1	-	-	С
5	9,2	11,9	-	+	Н
6	10,1	12,3	-	+	Н
7	12,5	13,8	+	+	С
8	13,0	14,0	+	+	С
9	14,6	15,2	+	+	С
10	15,7	17,6	+	+	С
Разом:	103,0	116,0	х	х	х

Коефіцієнт Фехнера оцінює силу зв'язку на основі порівняння знаків відхилень значень варіантів від їх середньої по кожній ознаці. Визначимо середні:

$$\overline{x} = \frac{\sum x}{n} = \frac{103}{10} = 10{,}3 \text{ млн.грн.} \qquad \overline{y} = \frac{\sum y}{n} = \frac{116}{10} = 11{,}6 \text{ млн.грн.}$$

Знак мінус означає, що значення ознаки менше середньої, а знак плюс – більше середньої. Співпадання знаків за обома ознаками означає узгоджену варіацію, неспівпадання – порушення такої узгодженості. За цим принципом побудований коефіцієнт Фехнера:

$$K_{\phi} = \frac{\sum C - \sum H}{\sum C + \sum H} = \frac{8-2}{8+2} = \frac{6}{10} = 0{,}6,$$

де $\sum C$ – сума знаків, які співпали по обох рядах;

$\sum H$ – сума знаків, які не співпали.

Коефіцієнт Фехнера коливається в межах від +1 до -1. При наближенні цього коефіцієнта до +1 спостерігається пряма і сильна узгодженість, при -1 будемо мати сильну але обернену узгодженість. При нулю, узгодженість між досліджуваними ознаками відсутня.

В нашому прикладі $К_ф$ показує, що між вартістю основних виробничих фондів і випуском продукції існує прямий і досить тісний зв'язок.

Більш точно оцінює силу зв'язку **коефіцієнт кореляції рангів**. Цей коефіцієнт враховує узгодженість рангів, які займають окремі одиниці сукупності по кожній із двох досліджуваних ознак.

Сукупність рангується за факторною ознакою в порядку зростання і проставляються відповідні ранги. Паралельно проставляються ранги тих же одиниць сукупності, які вони б зайняли в рангованому ряду за результативною ознакою.

Коефіцієнт кореляції рангів запропонований американським вченим К. Спірменом має вигляд: $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$,

де ρ (грецька буква «ро») – коефіцієнт кореляції рангів;

d^2 – квадрат різниці між величинами рангів в порівняльних рядах;

n – число рангів.

Існує правило, що для варіантів, які повторюються, ранг визначається як середня арифметична відповідних рангів, наприклад, ранг однакових величин, які займають 4 і 5 місця, дорівнює 4,5.

Коефіцієнт рангової кореляції може приймати значення в межах: $-1 \le \rho \le +1$.

Коли ранги факторної ознаки R_x повністю співпадають з рангами результативної ознаки R_y, тоді кожне значення R_x=R_y і $\sum d^2 = 0$. В цьому випадку можна судити про майже повний прямий зв'язок, $\rho = 1$.

Якщо ранги розташовані строго в протилежному напрямку, тоді спостерігається повна обернена залежність кореляції рангів, $\rho = -1$.

Коли $\frac{6\sum d^2}{n(n^2-1)} = 1$, кореляція рангів відсутня і $\rho = 0$.

Розглянемо приклад. Потрібно визначити силу зв'язку між вартістю основних виробничих фондів і виробітком продукції на одного робітника за такими даними:

Показники роботи десяти підприємств і розрахунок зв'язку між ними

Номер підпри-ємства (n)	Вартість основних виробничих фондів, тис. грн. (x)	Виробіток одного робітника, тис. грн. (y)	Ранги		Різниця рангів	
			R_x	R_y	$d= R_x- R_y$	d^2
1	2348	20	1	2	-1	1
2	2654	32	2	4	-2	4
3	2780	41	3	7	-4	16
4	2891	43	4	8	-4	16
5	3125	18	5	1	+4	16
6	3240	24	6	3	+3	9
7	3915	37	7	5	+2	4
8	4000	39	8	6	+2	4
9	4137	43	9	9	0	0
10	5199	45	10	10	0	0
Разом:	x	x	x	x	x	70

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \cdot 70}{10(100-1)} = 1 - 0{,}424 = 0{,}576.$$

Коефіцієнт кореляції рангів К. Спірмена вказує на помітний прямий зв'язок між вартістю основних виробничих фондів і виробітком продукції на одного робітника.

Англійський статистик Кендел для визначення тісноти зв'язку між корельованими ознаками запропонував свою формулу коефіцієнта кореляції рангів:

$$\tau = \frac{S}{\frac{1}{2}n(n-1)}, \quad або \quad \tau = \frac{2S}{n(n-1)},$$

де τ (грецька літера «тау») – коефіцієнт кореляції рангів Кендела;

S – фактична сума балів;

n – число рангів.

Величина $S = \sum S_1 - \sum S_2$,

де S_1 – число рангів, які перевищують номер рангу, записаного в рангах за результативною ознакою R_y;

S_2 – число рангів менших R_y в наступних записах.

Розрахункова таблиця для S_1 і S_2

Підприємства	R_y	S_1	S_2
1	2	8	1
2	4	6	2
3	7	3	4
4	8	2	4
5	1	5	0
6	9	4	0
7	5	3	0
8	6	2	0
9	9	1	0
10	10	0	0
Разом:	x	34	11

$$S = \sum S_1 - \sum S_2 = 34 - 11 = 23.$$

Звідси
$$\tau = \frac{S}{\frac{1}{2}n(n-1)} = \frac{23}{\frac{1}{2}10 \cdot (10-1)} = \frac{23}{45} = 0{,}511.$$

Таким чином коефіцієнт кореляції рангів Кендела оцінює зв'язок між даними ознаками більш обережно, ніж коефіцієнт Спірмена.

Між коефіцієнтами кореляції рангів існує співвідношення:

$$\rho = \frac{3}{2}\tau\ .$$

Графічний метод виявлення кореляційної залежності заключається в зображенні статистичних характеристик, отриманих в результаті зведення і обробки вихідної інформації на графіку, яке наочно покаже форму зв'язку між досліджуваними ознаками, та його напрямком.

Зв'язок між вартістю основних виробничих фондів і випуском продукції можна наочно уявити, якщо побудувати графік. Нанісши на графіку точки, які відповідають значенням «х» і «у», отримаємо кореляційне поле, де за характером розміщення точок можна судити про напрямок і силу зв'язку.

Мал. 6.1. Кореляційне поле зв'язку між вартістю основних виробничих фондів і випуском продукції

Якщо точки розташовані хаотично по всьому полю, це говорить про відсутність залежності між двома ознаками; якщо вони сконцентровані навколо осі, яка йде від нижнього лівого кута до верхнього правого – це пряма залежність між досліджуваними ознаками; якщо точки будуть сконцентровані навколо осі, яка проляже від верхнього лівого кута до нижнього правого – маємо обернену залежність.

Метод статистичних групувань, як прийом виявлення кореляційної залежності, відноситься до числа найважливіших прийомів дослідження взаємозв'язків. Для виявлення залежності між ознаками за допомогою цього методу матеріал статистичного спостереження групується за факторною ознакою, і для кожної групи вираховуються середні значення як факторної так і результативної ознаки. Порівнюючи зміни середніх значень результативної ознаки в міру зміни середніх значень факторної ознаки, виявляють характер зв'язку між ними.

Статистичні групування, проведені з метою виявлення і аналізу взаємозв'язків між ознаками, називаються **аналітичними**. Розглянемо приклад. Нехай ми провели аналітичне групування 20 робітників за стажем роботи з метою виявлення його впливу на місячну заробітну плату, утворивши за факторною ознакою п'ять груп з рівними інтервалами.

Залежність місячної заробітної плати робітників від стажу роботи

Групи робітників за стажем роботи, років	Число робітників, чол.	Середні рівні	
		стажу роботи робітників, років	місячної заробітної плати, грн.
I. 1-4	3	2,07	1546,70
II. 4-7	6	5,40	1718,30
III. 7-10	5	8,44	1910,00
IV. 10-13	4	10,92	1935,00
V. 13-16	2	15,00	2140,00
Разом:	20	7,72	1826,00

Як видно із таблиці, середній місячний заробіток робітників збільшується разом із ростом стажу їх роботи. Це свідчить про пряму залежність заробітної плати робітників від стажу їх роботи.

Групування дозволяє також виявити одночасний вплив декількох чинників на результативну ознаку. Для цього проводять комбіновані групування, дані яких викладають в комбінованих таблицях.

Аналітичні групування характеризують лише загальні риси зв'язку, його тенденцію, але не дають кількісної оцінки його сили. На основі аналітичних групувань це завдання розв'язується за допомогою розрахунку **емпіричного кореляційного відношення**.

Для кількісної оцінки зв'язку між явищами на базі матеріалів аналітичного групування вираховують коефіцієнт детермінації і емпіричне кореляційне відношення.

Коефіцієнт детермінації показує ступінь варіації ознаки під впливом чинника покладеного в основу групування.

Він визначається як відношення міжгрупової дисперсії до загальної:

$$\eta^2 = \frac{\delta^2}{\sigma_з^2},$$

де η^2 – коефіцієнт детермінації;

δ^2 – міжгрупова дисперсія;

$\sigma_з^2$ – загальна дисперсія.

Критерієм суттєвості і сили зв'язку між факторною і результативною ознаками виступає емпіричне кореляційне відношення.

$$\eta = \sqrt{\frac{\delta^2}{\sigma_з^2}}\ .$$

Для якісної оцінки сили зв'язку між досліджуваними ознаками на основі емпіричного кореляційного відношення використовують наступну шкалу.

Величина (η)	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Сила зв'язку	слабкий	помірний	помітний	сильний	дуже сильний

Розрахунок міжгрупової дисперсії

Групи робітників за стажем роботи, років	Число робітників, чол. (f_i)	Середня місячна заробітна плата одного робітника, грн. ($\bar{y}_i$)	$\bar{y}_i - \bar{y}_з$	$(\bar{y}_i - \bar{y}_з)^2$	$(\bar{y}_i - \bar{y}_з)^2 f_i$
I. 1-4	3	1546,7	-279,3	78008,49	234025,47
II. 4-7	6	1718,3	-107,7	11599,29	69595,74
III. 7-10	5	1910,0	84,0	7056,00	35280,00
IV. 10-13	4	1935,0	109,0	11881,00	47524,00
V. 13-16	2	2140,0	314,0	98596,00	197192,00
Разом:	20	$\bar{y}_з = 1826,0$	x	x	583617,21

Міжгрупова дисперсія дорівнює:

$$\delta^2 = \frac{\Sigma\left(\bar{y}_i - \bar{y}_з\right)^2 f_i}{\Sigma f_i} = \frac{583617,21}{20} = 29180,86.$$

Розрахуємо загальну дисперсію:

$$\sigma_з^2 = \overline{y^2} - (\bar{y})^2 = 3366640 - 3334276 = 32364,00.$$

Обчислимо коефіцієнт детермінації і емпіричне кореляційне відношення:

$$\eta^2 = \frac{\delta^2}{\sigma_з^2} = \frac{29180,86}{3236,00} = 0,9016\ \textit{або}\ 90,16\,\%;$$

$$\eta = \sqrt{\frac{\delta_2}{\sigma_з^2}} = \sqrt{0,9016} = 0,9495.$$

Коефіцієнт детермінації показує, що заробітна плата робітників на 90,16 % залежить від стажу їх роботи і на 9,84 % від інших чинників.

Емпіричне кореляційне відношення свідчить про те, що зв'язок між стажем роботи і середньою місячною заробітною платою робітників дуже сильний.

Емпіричне кореляційне відношення повинно мати високий рівень надійності. Для оцінки надійності кореляційних характеристик використовують критерій Фішера (F – критерій) або Стьюдента (t – критерій).

Критерій Фішера (F – критерій) визначається за формулою:

$$F_{\phi} = \frac{\delta^2}{\overline{\sigma_i^2}} \cdot \frac{k_2}{k_1}, \text{ або } F_{\phi} = \frac{\delta^2 \cdot k_2}{\overline{\sigma_i^2} \cdot k_1},$$

де δ^2 – міжгрупова дисперсія:

$\overline{\sigma_i^2}$ – середня групова (залишкова) дисперсія;

k_1, k_2 – ступені вільності для великої і малої дисперсій.

В аналітичному групуванні критерій вільності обчислюються за формулами:

$k_1=m-1; \quad k_2=n-m,$

де n – кількість елементів досліджуваної сукупності;

m – число груп.

За даними нашого прикладу:

$$\overline{\sigma_i^2} = \sigma_з^2 - \delta^2 = 32364{,}00 - 29180{,}86 = 3183{,}14.$$

Тоді $F_{\phi} = \dfrac{\delta^2 \cdot k_2}{\overline{\sigma_i^2} \cdot k_1} = \dfrac{29180{,}86 \cdot 15}{3183{,}14 \cdot 4} = 34{,}38.$

Для оцінки отриманого відношення його порівнюють з табличним (F_T) при певній ймовірності.

Знаходимо F_T при ймовірності 0,95 і даних ступенях вільності (k_1=m-1= =5-1=4, k_2=n-m=20-5=15) в математичній таблиці. Воно становить $F_{T(0,95)}$=3,06.

Таким чином $F_ф > F_T$(34,38 > 3,06), що свідчить про суттєвість впливу стажу роботи на середню місячну заробітну плату робітників.

До аналітичного висновку можна прийти за оцінкою надійності кореляційного відношення за критерієм Стьюдента (t – критерій), який визначається за формулою: $t_\eta = \dfrac{\eta}{\mu_\eta}$, де μ_η – середня помилка кореляційного відношення.

Вона визначається за формулою: $\mu_\eta = \dfrac{1-\eta^2}{\sqrt{\eta}}$.

Якщо критерій Стьюдента дорівнює або більший 3 ($t_\eta \geq 3$) показник кореляційного відношення вважають вірогідним (тобто зв'язок між досліджуваними явищами є доведеним).

Для нашого прикладу: $\mu_\eta = \dfrac{1-\eta^2}{\sqrt{\eta}} = \dfrac{1-0{,}90}{\sqrt{0{,}95}} = 0{,}1026$,

$$t_\eta = \frac{\eta}{\mu_\eta} = \frac{0{,}9495}{0{,}1026} = 9{,}25.$$

Так як критерій Стьюдента значно більший за 3, то кореляційне відношення вважається вірогідним, а зв'язок між стажем роботи і середньою місячною заробітною платною робітників доведеним.

6.3. Кореляційний і регресійний методи аналізу зв’язку

Основне завдання кореляційного і регресійного аналізу статистичних даних є виявлення залежності між досліджуваними ознаками у вигляді певної математичної формули і встановлення за допомогою коефіцієнта кореляції порівняльної ознаки тісноти взаємозв’язку.

Кореляційний і регресійний методи аналізу розв’язують два основних завдання:

1) визначають з допомогою рівняння регресії аналітичну форму зв’язку між варіацією ознак «х» і «у»;

2) встановлюють міру тісноти зв’язку між ознаками.

В практиці економіко-статистичних досліджень часто доводиться мати справу з прямолінійною формою зв’язку, яка виражається за допомогою рівняння регресії.

Рівняння регресії характеризує зміну середнього рівняння результативної ознаки «у» в залежності від зміни факторної ознаки «х».

У випадку лінійної форми зв’язку рівняння регресії має вигляд:

$$\hat{y} = a_0 + a_1 x,$$

де $\hat{y}$ – вирівняне середнє значення результативної ознаки;

x – значення факторної ознаки;

a_0 *і* a_1 – параметри рівняння;

a_0 – значення «у» при х=0;

a_1 – коефіцієнт регресії.

Коефіцієнт регресії «a_1» показує наскільки зміниться результативна ознака «у» при зміні факторної ознаки «х» на одиницю.

Якщо «a_1» має позитивний знак, то зв'язок прямий, якщо від’ємний – зв'язок обернений.

Параметри рівняння зв’язку визначаються способом найменших квадратів складеної і розв’язаної системи двох рівнянь з двома невідомими.

$$\begin{cases} \sum y = na_0 + a_1 \sum x; \\ \sum xy = a_0 \sum x + a_1 \sum x^2, \end{cases}$$

де n – число членів у кожному з двох порівнюваних рядів;

$\sum x$ – сума значень факторної ознаки;

$\sum y$ – сума значень результативної ознаки;

$\sum xy$ – сума добутків значень факторної ознаки на значення результативної ознаки.

Розв’язавши дану систему рівнянь, отримаємо такі значення параметрів:

$$a_0 = \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - \sum x \sum x}; \qquad a_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \sum x \sum x}.$$

Вирахувавши за фактичними даними всі записані вище суми і підставивши їх у наведені формули, знайдемо параметри шуканої прямої.

Розрахунок параметрів лінійного рівняння зв’язку і лінійного коефіцієнта кореляції між вартістю основних виробничих фондів і випуском продукції.

Номер заводу (n)	Вартість основних виробничих фондів, млн.грн. (x)	Випуск продукції, млн.грн. (y)	x^2	xy	y^2	$\hat{y} = a_0 + a_1 x$
1	12	5,6	144	67,2	31,36	5,2
2	8	4,0	64	32,0	16,00	3,5
3	10	4,0	100	40,0	16,00	4,4
4	6	2,4	36	14,4	5,76	2,7
5	9	3,6	81	32,4	12,96	4,0
6	15	5,0	225	75,0	25,00	6,5
7	11	4,6	121	50,6	21,16	4,8
8	13	6,5	169	84,5	42,25	5,6
9	14	7,0	196	98,0	49,00	6,1
10	10	4,5	100	45,0	20,25	4,4
Разом:	108	47,2	1236	539,1	239,74	47,2
В середньому на 1 завод	10,8	4,72	132,6	53,91	23,974	4,72

За способом найменших квадратів визначаємо параметри:

$$a_0 = \frac{1236 \cdot 47{,}2 - 108 \cdot 596{,}1}{10 \cdot 1236 - 108 \cdot 108} = \frac{58339{,}2 - 58222{,}8}{12360 - 11664} = \frac{116{,}4}{696{,}0} = 0{,}167;$$

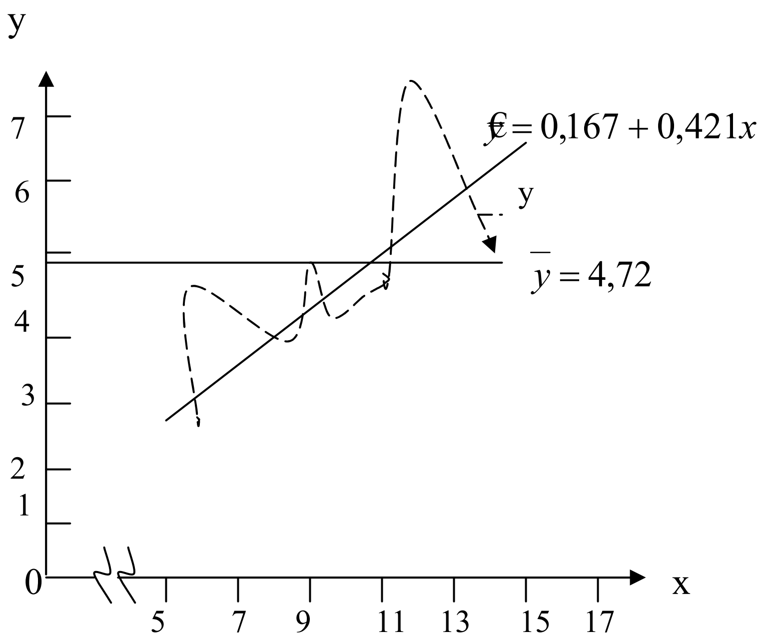
$$a_1 = \frac{10 \cdot 539{,}1 - 108 \cdot 47{,}2}{696{,}0} = \frac{5391{,}0 - 5097{,}6}{696{,}0} = \frac{293{,}4}{696{,}0} = 0{,}421.$$

Лінійне рівняння регресії між вартістю основних виробничих фондів і випуском продукції має вигляд: $\hat{y} = 0{,}167 + 0{,}421x.$

Таким чином, при збільшенні вартості основних виробничих фондів на 1 млн.грн. випуск продукції зросте на 0,421 млн.грн.

Підставляючи в дане рівняння послідовно значення факторної ознаки «х», отримаємо вирівняні значення результативної ознаки «у». Якщо параметри рівняння визначені правильно, то $\sum y = \sum \hat{y} = 47{,}2.$

Побудуємо графік, який покаже вирівнювання емпіричних даних рівнянням прямої.



Мал. 6.2. Емпіричні і вирівняні рівні ряду

Для економічної інтерпретації лінійних і нелінійних зв'язків між двома досліджуваними явищами часто використовують розраховані на основі рівнянь регресії коефіцієнти еластичності.

Коефіцієнт еластичності показує, на скільки відсотків змінюється в середньому результативна ознака «у» при зміні факторної ознаки «х» на 1%.

Для лінійної залежності коефіцієнт еластичності визначається за формулою:

$$\varepsilon = a_1 \cdot \frac{x}{\hat{y}}, \quad \varepsilon = a_1 \cdot \frac{\bar{x}}{\bar{y}},$$

де ε – коефіцієнт еластичності.

В нашому прикладі коефіцієнт еластичності на першому підприємстві при (х=12) буде дорівнювати: $\varepsilon_1 = a_1 \cdot \frac{x}{\hat{y}} = 0{,}421 \frac{12}{5{,}2} = 0{,}97\,\%.$

Отже, на 1 % приросту вартості основних виробничих фондів, випуск продукції зросте на 0,97 %. На п'ятому підприємстві при (х=9) $\varepsilon_5 = 0{,}421 \frac{9}{4} = 0{,}95\,\%,$ на десятому – при (х=10) $\varepsilon_{10} = 0{,}421 \frac{10}{4{,}4} = 0{,}96\,\%.$

Для всіх підприємств разом коефіцієнт еластичності становить:

$$\varepsilon = a_1 \cdot \frac{\bar{x}}{\bar{y}} = 0{,}421 \cdot \frac{10{,}8}{4{,}72} = 0{,}963\,\%.$$

Це означає, що при збільшенні середньої вартості основних виробничих фондів на 1 % випуск продукції зросте в середньому на 0,963 %.

Якщо залежність між ознаки параболічна, то коефіцієнт еластичності визначається за формулою:

$$\varepsilon = (a_1 + a_2 x)\frac{\bar{x}}{\bar{y}}.$$

Визначення тісноти зв'язку в кореляційно-регресійному аналізі ґрунтується на правилі складання дисперсій, але для оцінки лінії регресії використовують теоретичні значення результативної ознаки.

Різниця між загальною і залишковою дисперсіями дає нам теоретичну (факторну) дисперсію, яка вимірює варіацію, зумовлену фактором «х». На порівнянні цієї різниці із загальною дисперсією побудований **індекс кореляції**, або **теоретичне кореляційне відношення**, які обчислюються за формулами:

$$R = \sqrt{\frac{\sigma_з^2 - \sigma_e^2}{\sigma_з^2}} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_з^2}}, \quad або \quad R = \sqrt{\frac{\delta^2}{\sigma_з^2}},$$

де R – індекс кореляції (теоретичне кореляційне відношення);

$\sigma_з^2$ – загальна дисперсія;

σ_e^2 – залишкова дисперсія;

δ^2 – факторна (теоретична) дисперсія.

Факторну дисперсію з теоретичних значень обчислюють за формулою:

$$\delta^2 = \frac{\sum(\hat{y} - \bar{y})^2}{n},$$

або за формулою без теоретичних значень:

$$\delta^2 = \frac{(a_0 \sum y + a_1 \sum xy) - (\bar{y})^2}{n}.$$

Залишкову дисперсію визначають за формулою:

$$\sigma_e^2 = \frac{\sum(y - \hat{y})^2}{n},$$

або за правилом складання дисперсій: $\sigma_e^2 = \sigma_з^2 - \delta^2$.

В нашому прикладі факторна дисперсія дорівнює:

$$\delta^2 = \frac{(0{,}167 \cdot 47{,}2 + 0{,}421 \cdot 539{,}1) - 4{,}72^2}{10} = 1{,}206.$$

Загальна дисперсія становить:

$\sigma_з^2 = \overline{y^2} - \left(\overline{y}\right)^2 = 23{,}974 - 22{,}278 = 1{,}696.$

Залишкову дисперсію визначаємо як різницю між загальною і факторною дисперсіями: $\sigma_e^2 = \sigma_з^2 - \delta^2 = 1{,}699 - 1{,}206 = 0{,}490.$

Таким чином індекс кореляції за вищенаведеними формулами буде дорівнювати:

$$R = \sqrt{\frac{\sigma_з^2 - \sigma_e^2}{\sigma_з^2}} = \sqrt{\frac{1{,}696 - 0{,}490}{1{,}696}} = 0{,}843,$$

або $$R = \sqrt{1 - \frac{\sigma_e^2}{\sigma_з^2}} = \sqrt{1 - \frac{0{,}490}{1{,}96}} = 0{,}843,$$

або $$R = \sqrt{\frac{\delta^2}{\sigma_з^2}} = \sqrt{\frac{1{,}206}{1{,}696}} = \sqrt{0{,}711} = 0{,}843.$$

Індекс кореляції показує тісну залежність випуску продукції від вартості основних фондів.

Коефіцієнт детермінації (R^2) характеризує ту частину варіації результативної ознаки «у», яка відповідає лінійному рівнянню регресії:

$$R^2 = \frac{\delta^2}{\sigma_з^2} = \frac{1{,}206}{1{,}696} = 0{,}711.$$

Отже, в обстеженій сукупності заводів 71,1 % варіації випуску продукції пояснюється різними рівнями оснащеності заводів основними фондами.

Індекс кореляції приймає значення від «0» до «1». Коли R=0, то зв’язку між варіацією ознак «у» і «х» немає. Залишкова дисперсія дорівнює загальній ($\sigma_e^2 = \sigma_з^2$), а теоретична дисперсія дорівнює нулю ($\delta^2 = 0$).

При R=1 теоретична дисперсія дорівнює загальній ($\delta^2 = \sigma_з^2$), а залишкова – $\sigma_e^2 = 0$.

Для вимірювання тісноти зв’язку і визначення його напрямку при лінійній залежності використовують **лінійний коефіцієнт кореляції**, який визначається за формулою:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}.$$

Значення «r» коливається в межах від -1 до +1. Додатнє значення «r» означає прямий зв'язок між ознаками, а від’ємне – зворотній.

Оцінка тісноти зв’язку проводиться за схемою:

Сила зв’язку	Величина лінійного коефіцієнта кореляції при наявності:	
	прямого зв’язку	оберненого зв’язку
Слабка	0,1 – 0,30	(-0,1) – (-0,30)
Середня	0,3 – 0,70	(-0,3) – (-0,70)
Тісна	0,7 – 0,99	(-0,7) – (-0,99)

За даними нашого прикладу обчислимо лінійний коефіцієнт кореляції:

$$\sigma_x = \sqrt{\overline{x^2} - \left(\bar{x}\right)^2} = \sqrt{123{,}6 - 10{,}8^2} = \sqrt{6{,}96} = 2{,}638;$$

$$\sigma_y = \sqrt{\overline{y^2} - \left(\bar{y}\right)^2} = \sqrt{23{,}974 - 4{,}72^2} = 1{,}302;$$

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{53{,}91 - 10{,}8 \cdot 4{,}72}{2{,}638 \cdot 1{,}302} = \frac{2{,}9340}{3{,}4349} = 0{,}854.$$

Це означає, що зв'язок між вартістю основних виробничих фондів і випуском продукції сильний (тісний) і прямий.

Істотність зв’язку коефіцієнта детермінації R^2 перевіряють за допомогою таблиці критерію F для 5 %-го рівня значимості. Так, при k_2=n-m=10-2=8. Фактичне значення F- критерія для нашого прикладу визначають за формулою:

$$F_{ф} = \frac{R^2}{1 - R^2} \cdot \frac{k_2}{k_1} = \frac{0{,}711}{1 - 0{,}711} \cdot \frac{8}{1} = 19{,}68.$$

Критичне значення $F_{т\ (0,95)}$ значно менше від фактичного $F_{т\ (0,95)} < F_{ф}$ (5,32 < 19,68), то підтверджує істотність кореляційного зв’язку між досліджуваними ознаками.

Для встановлення достовірності обчисленого лінійного коефіцієнта кореляції використовують критерій Стьюдента (t- критерій): $t_r = \frac{|r|}{\mu_r}$,

де μ_r – середня помилка коефіцієнта кореляції, яку визначають за формулою:

$$\mu_r = \frac{1 - r^2}{\sqrt{n - 1}}.$$

При достатньо великому числі спостережень (n > 50) коефіцієнт кореляції можна вважати достовірним, якщо він перевищує свою помилку в 3 і більше раз, а якщо він менше 3, то зв'язок між досліджуваними ознаками «у» і «х» не доведений.

В нашому прикладі середня помилка коефіцієнта кореляції дорівнює:

$$\mu_r = \frac{1-0{,}853^2}{\sqrt{9}} = \frac{1-0{,}723}{3} = \frac{0{,}277}{3} = 0{,}092.$$

Відношення коефіцієнта кореляції до його середньої помилки становить:

$$t_r = \frac{0{,}853}{0{,}092} = 9{,}27.$$

Це дає нам право вважати, що обчислений лінійний коефіцієнт кореляції достатньо точно характеризує силу зв’язку між досліджуваними ознаками.

6.4. Нелінійні залежності

В практиці економічного аналізу найбільш часто використовують наступні нелінійні функції залежності: гіперболічну, параболічну другого порядку, напівлогарифмічну та деякі інші.

Якщо результативні ознака із збільшенням факторної ознаки зростає або спадає не безкінечно, а прямує до кінцевої мети, то для її аналізу застосовують **рівняння гіперболи**: $\tilde{y} = a_0 + a_1 \frac{1}{x}$.

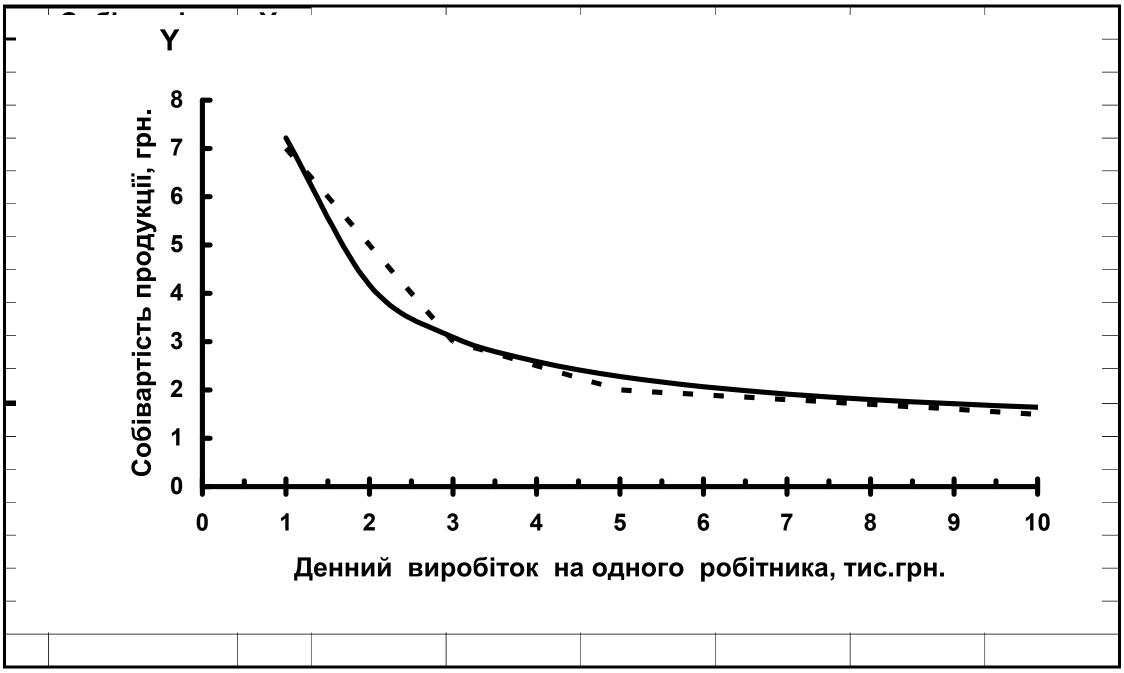
Для знаходження параметрів цього рівняння способом найменших квадратів використовується система нормальних рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum \frac{1}{x}; \\ \sum y \frac{1}{x} = a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2}. \end{cases}$$

За способом найменших квадратів параметри гіперболи визначають за формулами:

$$a_0 = \frac{\sum \frac{1}{x^2} \cdot \sum y - \sum \frac{1}{x} \cdot \sum \frac{y}{x}}{n \sum \frac{1}{x^2} - \sum \frac{1}{x} \cdot \sum \frac{1}{x}}; \qquad a_1 = \frac{n \sum \frac{x}{y} - \sum \frac{1}{x} \cdot \sum y}{n \sum \frac{1}{x^2} - \sum \frac{1}{x} \cdot \sum \frac{1}{x}}.$$

Графік гіперболи має вигляд:



Мал. 6.3. Графік кореляційної залежності собівартості одиниці продукції від денного виробітку на одного робітника

Для визначення тісноти зв'язку між результативною і факторною ознаками обчислюємо кореляційне відношення за формулою:

$$\eta_{\frac{y}{x}} = \sqrt{1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}}.$$

Парабола другого порядку застосовується в тих випадках, коли із зростанням факторної ознаки відбувається нерівномірне зростання або спадання результативної ознаки. Рівняння параболи другого порядку визначається за формулою:

$$\hat{y} = a_0 + a_1 x + a_2 x^2.$$

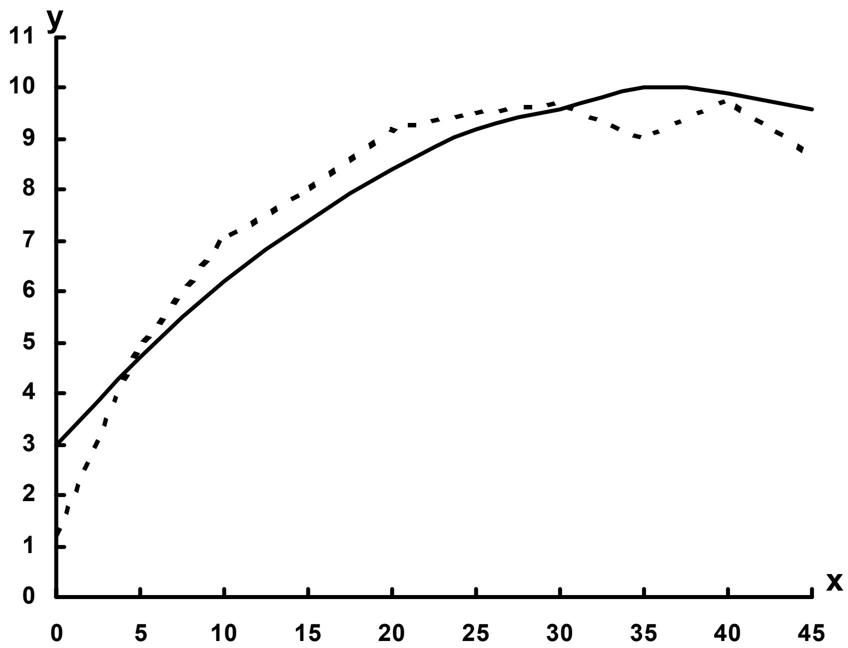
Параметри цього рівняння знаходять способом найменших квадратів шляхом складання і розв'язку системи нормальних рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum x + a_2 \sum x^2; \\ \sum xy = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3; \\ \sum x^2 y = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4. \end{cases}$$

З метою оцінки тісноти зв'язку визначають кореляційне відношення:

$$\eta_{\frac{y}{x}} = \sqrt{\frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2}}.$$

Графік параболи другого порядку має вигляд:



Мал. 6.4. Графік залежності між глибиною зрошення і урожайністю насіння багаторічних трав

Вирівнювання за напівлогарифмічною кривою проводять в тих випадках, коли із зростанням факторної ознаки, середня результативна ознака спочатку до певних меж зростає досить швидко, але пізніше темпи її зростання поступово сповільнюються.

Напівлогарифмічна функція має вигляд:

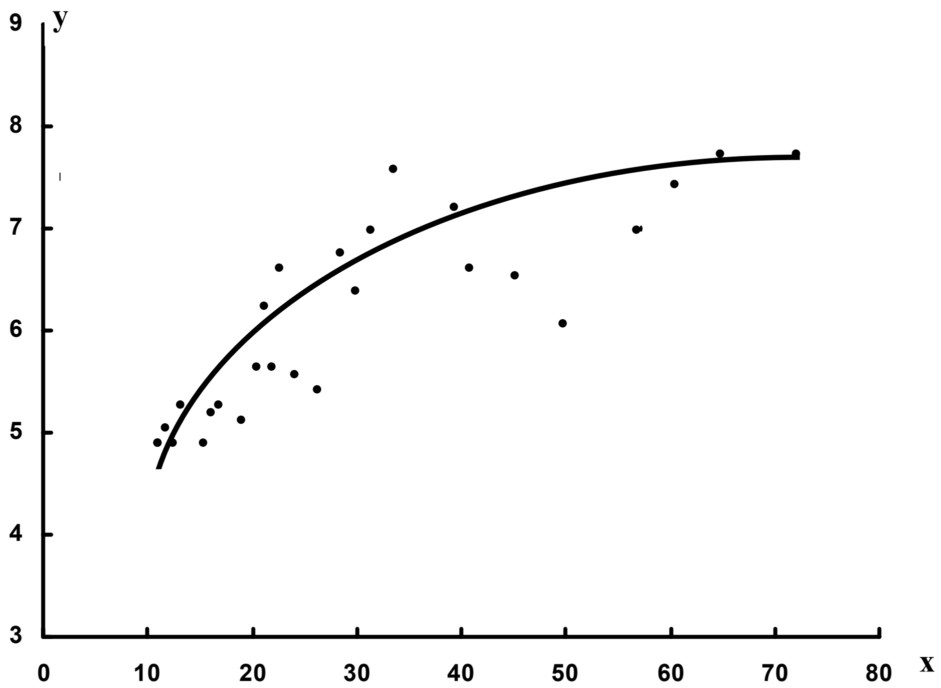
$$\hat{y} = a_0 + a_1 \ell og\, x.$$

Для знаходження параметрів напівлогарифмічної функції способом найменших квадратів, розв’язують систему двох рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum \ell og\, x; \\ \sum y\, \ell og\, x = a_0 \sum \ell og\, x + a_1 \sum (\ell og\, x)^2. \end{cases}$$

Графік напівлогарифмічної кривої.

$$\hat{Y}_X = 1{,}51 + 3{,}479 loqx.$$



Розмір товарообороту магазину, млн. грн.

Мал. 6.5. Графік кореляційної залежності між розміром товарообороту магазину і виробітком одного продавця

6.5. Багатофакторний кореляційно-регресійний аналіз

В багатьох випадках на результативну ознаку впливає не один, а декілька чинників. Між ними існують складні взаємозв'язки, тому їх вплив на результативну ознаку комплексний і його не можна розглядати як просту суму ізольованих впливів.

Багатофакторний кореляційно-регресійний аналіз дозволяє оцінити міру впливу на досліджуваний результативний показник кожного із введених в модель чинників при зафіксованому на середньому рівні інших чинників.

Форму зв'язку можна визначити шляхом перебору функцій різних типів, але це зв'язане з великою кількістю зайвих розрахунків. Однак, беручи до уваги, що любу функцію багатьох змінних шляхом логарифмування або заміни змінних можна звести до лінійного виду:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + ... + a_n x_n .$$

Параметри рівняння знаходять за способом найменших квадратів.

Так, для розрахунку параметрів рівняння лінійної двофакторної регресії:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2,$$

де $\hat{y}$ – розраховані значення результативної ознаки-функції;

x_1, x_2 – факторні ознаки;

a_0, a_1, a_2 – параметри рівняння.

Система нормальних рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum x_1 + a_2 \sum x_2; \\ \sum yx_1 = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2; \\ \sum yx_2 = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2. \end{cases}$$

Кожний коефіцієнт рівняння показує ступінь впливу відповідного чинника на результативний показник при фіксованому положенні решти чинників, тобто, як із зміною окремого чинника на одиницю змінюється результативний показник.

На основі коефіцієнтів регресії не можна судити, яка із факторних ознак найбільше впливає на результативну ознаку, так як коефіцієнти регресії між собою не порівняльні, оскільки вони володіють різними одиницями виміру.

З метою виявлення порівняльної сили впливу окремих чинників і резервів, які закладені в них, статистика вираховує часткові коефіцієнти еластичності «ε_i», а також бета-коефіцієнти «β_i» за формулами:

$$\varepsilon_i = a_i \frac{\bar{x}_i}{\bar{y}}; \qquad \beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y},$$

де a_i – коефіцієнт регресії при і-му факторі;

$\bar{x}_i$ – середнє значення і-го факторі;

$\bar{y}$ – середнє значення результативної ознаки;

σ_{x_1} – середнє квадратичне відхилення і-го фактора;

σ_y – середнє квадратичне відхилення результативної ознаки.

Часткові коефіцієнти еластичності показують, на скільки відсотків в середньому зміниться результативна ознака із зміною на 1 % кожного чинника при фіксованому положенні інших чинників.

Для визначення чинників, в розвитку котрих закладені найбільші резерви покращення досліджуваної ознаки, з врахуванням ступеня варіації чинників рівняння множинної регресії, вираховують **часткові β-коефіцієнти**, які показують на яку частину середнього квадратичного відхилення змінюється результативна ознака із зміною відповідної факторної ознаки на величину її середнього квадратичного відхилення.

Для характеристики ступеня тісноти зв'язку в множинній прямолінійній кореляції використовують **множинний коефіцієнт кореляції**.

$$R_{yx1x2} = \sqrt{\frac{r_{yx1}^2 + r_{yx2}^2 - 2r_{yx1} \cdot r_{yx2} \cdot r_{x1x2}}{1 - r_{x1x2}^2}},$$

де $R_{yx_1x_2}$ – множинний коефіцієнт кореляції;

r_{yx_1}, r_{yx_2}, $r_{x_1x_2}$ – парні коефіцієнти лінійної регресії, які визначаються за формулами: $r_{yx1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\sigma_y \cdot \sigma_{x1}}$; $r_{yx2} = \frac{\overline{yx_2} - \overline{y} \cdot \overline{x_2}}{\sigma_y \cdot \sigma_{x2}}$; $r_{yx3} = \frac{\overline{x_1x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_{x1} \cdot \sigma_{x2}}$.

Множинний коефіцієнт кореляції показує, яку частину загальної кореляції складають коливання, під впливом чинників $x_{1,}\ x_{2,\dots,}x_n$ – закладених в багатофакторну модель для дослідження.

Множинний коефіцієнт кореляції коливається в межах від «0» до «± 1». При R=0 зв'язок між досліджуваними ознаками відсутній, при R=1 – функціональний.

6.6. Непараметричні показники тісноти зв'язку

Поряд з вивченням кореляційної залежності між кількісними показниками статистика встановлює також зв'язок і між якісними ознаками.

Для вимірювання тісноти зв'язку між двома ознаками, які мають альтернативний вираз, застосування **коефіцієнт асоціації**, запропонований статистиком Юлом.

З метою розрахунку коефіцієнта асоціації використовують таблицю, яка має вигляд:

Ознака	А	не А	$\sum$В
В	а	в	а+в

не В	c	d	c+d
$\sum$А	a+c	b+d	a+в+c+d

Коефіцієнт асоціації визначається за формулою:

$$Q = \frac{ad - cb}{ad + cb}.$$

Розглянемо конкретний приклад.

Залежність урожайності озимої пшениці від внесення мінеральних добрив в ґрунт

Ділянки за урожайністю, га	Внесені добрива	Не внесені добрива	Всього
Підвищили урожайність	110	10	120
Не підвищили урожайність	10	70	80
Всього	120	80	200

Коефіцієнт асоціації дорівнює:

$$Q = \frac{ad - cb}{ad + cb} = \frac{110 \cdot 70 - 10 \cdot 10}{110 \cdot 70 + 10 \cdot 10} = 0{,}974.$$

Отже, спостерігається досить тісний зв'язок між удобренням ділянок і урожайністю озимої пшениці.

Для дослідження кореляції альтернативних ознак Юл запропонував **коефіцієнт колігації**:

$$W = \frac{\sqrt{ad} - \sqrt{cb}}{\sqrt{ad} + \sqrt{cb}} = \frac{\sqrt{7700} - \sqrt{100}}{\sqrt{7700} + \sqrt{100}} = \frac{77{,}7}{97{,}7} = 0{,}795.$$

Коефіцієнт контингенції К. Пірсона визначається за формулою:

$$K = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{7600}{\sqrt{120 \cdot 80 \cdot 120 \cdot 80}} = \frac{7600}{9600} = 0{,}792.$$

Коефіцієнт колігації і контингенції оцінюють зв'язок між внесенням мінеральних добрив в ґрунт і урожайністю озимої пшениці більш обережно, однак показують достатньо сильний зв'язок між цими ознаками.

В тому випадку, коли обидві взаємозв’язані ознаки розбиті більш ніж на дві групи, для вимірювання тісноти зв’язку застосовують показники взаємного спряження, запропоновані К. Пірсоном і А. Чупровим.

Коефіцієнт взаємного спряження К. Пірсона вираховується за формулою:

$$C_1 = \sqrt{\frac{\varphi^2}{\varphi^2 + 1}},$$

де φ – (грецька буква «фі») – сума квадратів частот кожної стрічки, розділеної на суму частот по графах, і в свою чергу, на суму частот по стрічці без одиниці.

Коефіцієнт взаємного спряження А. Чупрова визначається за формулою:

$$C_2 = \sqrt{\frac{\varphi^2}{(K_1 - 1)\cdot(K_2 - 1)}},$$

де K_1 – число груп по графах;

K_2 – число груп по стрічках.

Якщо одна із взаємозв'язаних ознак має кількісний вираз, а друга – альтернативний, то показником тісноти зв'язку виступає **бісеріальний коефіцієнт кореляції**:

$$r = \frac{\overline{y}_1 - \overline{y}_2}{\sigma_y} \cdot \frac{p \cdot q}{z},$$

де $\overline{y}_1$ – середня ознака по першій альтернативній групі;

$\overline{y}_2$ – середня ознака по другій альтернативній групі;

σ_y – середнє квадратичне відхилення по обох групах;

p – частка першої групи;

q – частка другої групи;

z – ордината нормальної кривої, яка ділить її площу у відношенні $p:q$ (значення «z» для різних «p» беруться із спеціальної таблиці).

Оцінка емпіричних мір тісноти зв'язку для всіх непараметричних показників здійснюється через t-критерій Стьюдента.

ЛЕКЦІЯ 7. РЯДИ ДИНАМІКИ

ПЛАН

7.1. Поняття про ряди динаміки, їх види та правила побудови.

7.2. Основні характеристики рядів динаміки.

7.3. Середні показники динаміки.

7.4. Виявлення тенденцій розвитку явищ.

7.5. Вимірювання сезонних коливань.

7.6. Особливості вимірювання взаємозв'язків в рядах динаміки.

7.1. Поняття про ряди динаміки, їх види та правила побудови

В статистичній практиці доводиться мати справу з великою кількістю чисел, що характеризують розвиток явищ в часі. Для кращого розуміння і аналізу досліджуваних статистичних даних, їх потрібно систематизувати, побудувавши хронологічні ряди, які називаються **рядами динаміки.** Отже, **рядами динаміки** в статистиці називаються ряди чисел, що характеризують закономірності і особливості зміни суспільних явищ і процесів в часі.

Кожний ряд динаміки складається з двох елементів:

1) періодів або моментів часу, до яких відносяться рівні ряду (t);

2) статистичних показників, які характеризують рівні ряду (y).

В залежності від характеру рівнів ряду розрізняють два види рядів динаміки: моментні і інтервальні (періодичні).

Моментним називається ряд динаміки, величини якого характеризують стан явищ на певний момент часу.

Рівні моментного ряду сумувати не має змісту.

Інтервальним називається такий ряд динаміки, величини якого характеризують розміри суспільних явищ за певні періоди часу (день, місяць, квартал і т.д.). Сума рівнів інтервального ряду динаміки характеризує рівень даних явища за більш тривалий проміжок часу.

Ряди динаміки бувають одномірні і багатомірні.

Одномірні ряди динаміки характеризують зміну одного показника (валовий збір картоплі).

Багатомірні ряди динаміки характеризують зміну двох, трьох і більше показників.

В свою чергу, багатомірні динамічні ряди поділяються на паралельні і ряди взаємозв'язаних показників.

Паралельні ряди динаміки відображають зміну або одного і того самого показника щодо різних об'єктів (чисельність населення по країнах), або різних показників щодо одного і того самого об'єкта (валовий збір пшениці, цукрових буряків і картоплі в районі).

Ряди взаємозв'язаних показників характеризують залежність одного явища від іншого (залежність заробітної плати робітників від їхнього тарифного розряду).

За повнотою часу динамічні ряди поділяються на повні і неповні.

В **повних** динамічних рядах дати або періоди ідуть один за одним з рівними інтервалами.

В **неповних** динамічних рядах в послідовності часу спостерігаються нерівні інтервали.

За способом вираження рівнів динамічного ряду вони поділяються на ряди **абсолютних, середніх і відносних** величин.

При формуванні динамічних рядів для наукового дослідження розвитку суспільних явищ в часі потрібно дотримуватись правил їх побудови. Важливим правилом побудови динамічних рядів є вимога порівняльності всіх рівнів ряду між собою. Показники ряду динаміки повинні бути порівняльні за територією, колом охоплюваних об'єктів, способами розрахунків, періодами часу, одиницями виміру.

Важливою вимогою любих динамічних порівнянь є вимога **порівняльності території**, до котрої відносяться рівні динамічного ряду. Межі територіальних одиниць держав, областей, районів на протязі досліджуваного періоду змінюються внаслідок приєднання до них нових територій, або відокремлення певних частин їх територій. В кожному окремому випадку питання порівняльності розв'язується в залежності від мети дослідження. Для приведення даних динамічного ряду до порівняльного виду проводиться перерахунок попередніх даних з врахуванням нових меж (кордонів).

Статистичні дані, які необхідні для побудови ряду динаміки повинні бути порівняльні за **колом охоплюваних об'єктів**. Непорівняльність може виникнути внаслідок переходу деяких об'єктів із одного підпорядкування в інше.

Порівняльність за колом охоплюваних об'єктів забезпечується **зімкненням динамічних рядів** шляхом заміни абсолютних рівнів відносними.

В моментних рядах динаміки виникає непорівняльність **за критичним моментом реєстрації** рівнів явищ, які піддаються сезонним коливанням (чисельність худоби зимою і літом).

Рівні динамічного ряду повинні бути порівняльні за **методикою їх розрахунку**. Наприклад, за попередні роки чисельність робітників заводу була визначена на початок кожного місяця, тобто на певну дату, а в наступні роки - як середньомісячна чисельність.

Статистичні дані динамічного ряду можуть бути непорівняльними за **різними періодами або тривалістю часу**. Наприклад, обсяг виробництва молока за різні роки потрібно порівнювати тільки таким чином: січень з січнем, квітень з квітнем і т.д. Інтервали часу, за які наведені дані динамічного ряду, повинні бути рівні (місяць, квартал, півріччя і т.д.).

Непорівняльність **через різні одиниці виміру** виникає внаслідок того, що ряд явищ обліковується паралельно в двох одиницях виміру. Наприклад, сталеві труби обліковуються в тоннах і метрах, електромотори - в штуках і кіловатах потужності і т.д. Порівняльність за одиницями виміру вимагає, щоб рівні динамічного ряду завжди були виражені в одних і тих самих одиницях виміру.

Непорівняльність рядів динаміки через одиниці виміру виникає і внаслідок непорівняльності грошової оцінки (міняється грошова одиниця, інфляція, змінюється курс валюти та ін.). Для приведення до порівняльного виду таких рядів динаміки всі попередні рівні досліджуваних ознак перераховуються за діючою грошовою оцінкою.

Непорівняльність статистичних показників динаміки може бути зумовлена також **різною структурою сукупності** за ряд років. Для приведення даних таких рядів до порівняльного виду використовують так звану стандартизацію структури (стандартизовані коефіцієнти народжуваності, смертності, природного приросту і т.д.).

7.2. Основні характеристики рядів динаміки

Завдання статистики полягає в тому, щоб шляхом аналізу рядів динаміки розкрити і охарактеризувати закономірності, що проявляються на різних етапах розвитку того чи іншого явища, виявити тенденції розвитку та їх особливості.

В процесі аналізу динаміки розраховують і використовують наступні аналітичні показники динаміки: абсолютний приріст, темп росту, темп приросту і абсолютне значення одного відсотка приросту.

Розрахунок цих показників ґрунтується на абсолютному або відносному порівнянні між собою рівнів ряду динаміки. При цьому порівнюваний рівень називається **поточним**, а рівень, з яким роблять порівняння - **базисним**. За базу порівняння часто приймають або попередній рівень, або початковий (перший) рівень ряду динаміки.

Якщо кожний рівень порівнюється з попереднім, то отримують **ланцюгові показники динаміки**, а якщо кожний рівень порівнюють з одним і тим же рівнем, взятим за базу порівняння, то такі показники називаються **базисними**.

Абсолютний приріст (Δ_y) обчислюється як різниця між поточним та базисним рівнями і показує, на скільки одиниць підвищився або зменшився рівень порівняно з базисним, за певний період часу.

$$\Delta_y^{б} = y_i - y_1, \ \ \textit{або} \quad \Delta_y^{л} = y_i - y_{i-1},$$

де Δ_y - абсолютний приріст;

y_i - поточний рівень ряду динаміки;

y_1 - початковий (перший) рівень ряду динаміки;

y_{i-1} - попередній рівень ряду динаміки.

Темп росту (T_P) вираховується як відношення порівнюваного рівня до базисного і показує, в скільки разів (відсотків) порівнюваний рівень більший або менший за базисний.

$$T_P^{б} = \frac{y_i}{y_1}, \qquad або \qquad T_P^{л} = \frac{y_i}{y_{i-1}}.$$

Між ланцюговими і базисними темпами росту існує певний взаємозв'язок. Добуток кількох послідовних ланцюгових темпів росту дорівнює базисному темпу росту за відповідний період і, навпаки, поділивши наступний базисний темп росту на попередній, отримаємо відповідний ланцюговий темп росту.

Темп приросту ($T_{ПР}$) визначається як відношення абсолютного приросту до абсолютного попереднього або початкового рівня і показує на скільки відсотків порівнюваний рівень більший або менший рівня, прийнятого за базу порівняння.

$$T_{ПР}^{б} = \frac{\Delta_y^{б}}{y_1}, \qquad або \qquad T_{ПР}^{л} = \frac{\Delta_y^{л}}{y_{i-1}}.$$

Між темпами росту і приросту існує безпосередній взаємозв’язок:

$$T_{ПР}^{б} = T_p^{б} - 1, \; або \qquad T_{ПР}^{б}(\%) = T_P^{б}(\%) - 100\%;$$

$$T_{ПР}^{л} = T_p^{л} - 1, \; або \qquad T_{ПР}^{л}(\%) = T_P^{л}(\%) - 100\%.$$

Абсолютне значення одного відсотка приросту (А) визначається шляхом ділення абсолютного приросту на темп приросту за один і той самий період. Абсолютне значення одного відсотка приросту можна вирахувати технічно більш легким шляхом, діленням початкового рівня на 100.

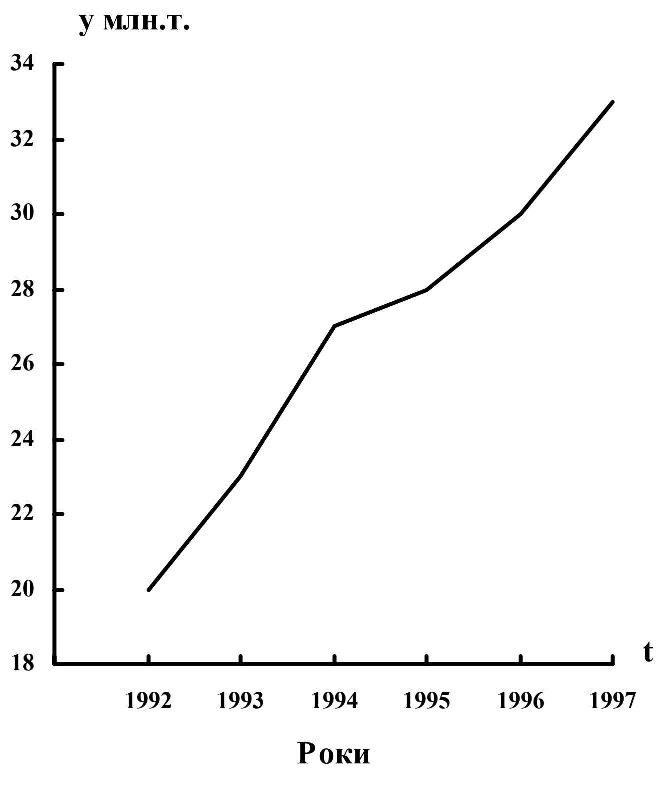
$$A = \frac{\Delta_y}{T_{ПР}(\%)}, \qquad або \qquad A = \frac{y_б}{100}.$$

Подамо всі вищенаведені показники в таблиці.

Розрахунок показників динаміки валового збору картоплі в області за 2000-20005 рр.

Роки	Валовий збір картоплі, млн. т. (у)	Абсолютний приріст, млн. грн.		Темп росту		Темп приросту		Абсолютне значення 1% приросту, млн.т. (А)
		Базисний ($\Delta_y^б$)	Ланцюговий ($\Delta_y^л$)	Базисний ($T_P^б$)	Ланцюговий ($T_P^л$)	Базисний ($T_{ПР}^б$)	Ланцюговий ($T_{ПР}^л$)	
2000	20	-	-	1,000	-	-	-	-
2001	23	3	3	1,150	1,150	0,150	0,150	0,20
2002	27	7	4	1,350	1,174	0,350	0,174	0,23
2003	28	8	1	1,400	1,037	0,400	0,037	0,27
2004	30	10	2	1,500	1,074	0,500	0,074	0,28
2005	33	13	3	1,650	1,100	0,650	0,100	0,30

Ряди динаміки, як правило, подаються не тільки в таблицях, але й показуються на графіках. При графічному зображенні динамічного ряду на осі абсцис відкладається шкала часу, а на осі ординат - шкала рівнів ряду.



Мал. 7.1. Динаміка валового збору картоплі в області за 2000 - 2005 рр.

7.3. Середні показники динаміки

Динамічні ряди складаються з багатьох варіаційних рівнів, а тому, як люба статистична сукупність, вони потребують деяких узагальнюючих характеристик.

Для цього вираховують середні показники: середні рівні ряду, середні абсолютні прирости, середні темпи росту і приросту.

В інтервальному ряду з рівними інтервалами середній рівень ряду вираховується за формулою **середньої арифметичної простої**:

$$\bar{y} = \frac{\sum y}{n},$$

де $\bar{y}$ - середній рівень ряду;

$\sum y$ - сума рівнів ряду;

n - число рівнів.

$$\bar{y} = \frac{\sum y}{n} = \frac{20+23+27+28+30+33}{6} = \frac{161}{6} = 26{,}8 \text{ млн.т.}$$

Якщо окремі періоди інтервального ряду динаміки мають різну довжину, то для визначення середнього рівня використовують **середню арифметичну зважену**:

$$\bar{y} = \frac{\sum yt}{\sum t},$$

де y - рівні ряду;

t - проміжки часу.

Розглянемо приклад.

Чисельність працівників підприємства в січні 2006 р., чол.

На 0,01	На 6,01	На 15,01	На 21,01	На 29,01	На 1,02
1210	1243	1236	1248	1238	1238

Середньоспискова чисельність працівників підприємства в січні місяці становитиме:

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{1210\cdot 5 + 1243\cdot 9 + 1236\cdot 6 + 1248\cdot 8 + 1238\cdot 3}{31} = \frac{38351}{31} = 1237 \text{ чол.}$$

Для визначення середнього рівня в моментному динамічному ряду з рівними інтервалами між сусідніми датами застосовують формулу **середньої хронологічної.**

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + y_3 ... + y_{n-1} + \frac{1}{2}y_n}{n-1},$$

Розглянемо приклад.

Парк тракторів в сільських спілках району.

Дані на початок місяця	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.01
Число тракторів, шт.	622	640	643	640	664	670	682	733	753	768	800	826	888

Визначимо середнє число тракторів за кожний квартал, перше і друге півріччя і за рік в цілому.

$$\bar{y}_{Iк.} = \frac{\frac{622}{2} + 640 + 643 + \frac{640}{2}}{3} = \frac{1914}{3} = 638\,шт.;$$

$$\bar{y}_{IIк.} = \frac{\frac{640}{2} + 664 + 670 + \frac{668}{2}}{3} = \frac{1995}{3} = 665\,шт.;$$

$$\bar{y}_{IIIк.} = \frac{\frac{682}{2}2 + 733 + 753 + \frac{768}{2}}{3} = \frac{2211}{3} = 737\,шт.;$$

$$\bar{y}_{IVк.} = \frac{\frac{768}{2} + 800 + 826 + \frac{888}{2}}{3} = \frac{2454}{3} = 188\,шт.;$$

$$\bar{y}_{I\,п.} = \frac{\bar{y}_{I\,к.} + \bar{y}_{II\,к.}}{2} = \frac{638 + 665}{2} = 651{,}5\,шт;$$

$$\bar{y}_{II\,п.} = \frac{\bar{y}_{III\,к.} + \bar{y}_{IV\,к.}}{2} = \frac{737 + 818}{2} = 777{,}5\,шт;$$

$$\bar{y}_{р} = \frac{\bar{y}_{I\,п.} + \bar{y}_{II\,п.}}{2} = \frac{651{,}5 + 777{,}5}{2} = 714{,}5\,шт.$$

Середній абсолютний приріст визначається як середня арифметична проста з ланцюгових абсолютних приростів за певні періоди і показує на скільки одиниць в середньому змінився рівень у порівнянні з попереднім.

$$\bar{\Delta}_y = \frac{\sum \Delta_y^л}{n},$$

де $\bar{\Delta}_y$ - середній абсолютний приріст;

n - кількість приростів.

Середній абсолютний приріст визначається і за іншою формулою:

$$\overline{\Delta}_y = \frac{y_n - y_1}{n-1}.$$

З попереднього нашого прикладу середньорічний абсолютний приріст валового збору картоплі складе:

$$\overline{\Delta}_y = \frac{\sum \Delta_y^л}{n} = \frac{3+4+1+2+3}{5} = \frac{13}{5} = 2{,}6\ \textit{млн.т., або}$$

$$\overline{\Delta}_y = \frac{y_n - y_1}{n-1} = \frac{33-20}{6-1} = \frac{13}{5} = 2{,}6\ \textit{млн.т.}$$

Середній темп росту вираховується за формулою **середньої геометричної**:

$$\overline{T}_P = \sqrt[n]{T_{P_1}^л \cdot T_{P_2}^л \cdot T_{P_3}^л \cdot \ldots \cdot T_{P_n}^л},$$

де $\overline{T}_P$ – середній темп росту;

$T_{P_1}^л, T_{P_2}^л, T_{P_3}^л, \ldots, T_{P_n}^л$ – ланцюгові темпи росту;

n – число темпів.

Для обчислення середнього темпу росту використовують також іншу формулу: $\overline{T}_P = \sqrt[n-1]{\frac{y_n}{y_1}}$,

де y_n – кінцевий рівень ряду;

y_1 – початковий рівень ряду;

n – кількість рівнів динамічного ряду.

$$\overline{T}_P = \sqrt[5]{1{,}150 \cdot 1{,}174 \cdot 1{,}037 \cdot 1{,}071 \cdot 1{,}100} = \sqrt[5]{1{,}650} = 1{,}105, \qquad \textit{або}$$

$$\overline{T}_P = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5]{\frac{33}{20}} = \sqrt{1{,}650} = 1{,}105.$$

Середні темпи росту обчислюють також за формулою **середньої геометричної зваженої**: $\overline{T}_P = \sqrt[\sum t]{T_{P_1}^{t1} \cdot T_{P_2}^{t2} \cdot T_{P_3}^{t3} \cdot \ldots \cdot T_{P_n}^{tn}}$,

де t - інтервал часу, на протязі якого зберігається даний темп росту;

$\sum t$ - сума відрізків часу періоду.

Середній темп приросту визначається як різниця між середнім темпом росту і одиницею (якщо середній темп росту у вигляді коефіцієнта), або 100 (якщо він у відсотках):

$$\overline{T}_{ПР} = \overline{T}_{P} - 1 = 1{,}105 - 1 = 0{,}105, \quad або$$
$$\overline{T}_{ПР} = \overline{T}_{P} - 100 = 110{,}5 - 100 = 10{,}5\,\%$$

Подальший аналіз рядів динаміки соціально-економічних показників зв'язаний з більш складними узагальненнями, з визначенням основної тенденції, вивченням сезонних коливань рівнів і дослідженням зв'язку між рядами.

7.4. Виявлення тенденцій розвитку явищ

Виявлення основної тенденції (тренду) ряду, є одним з головних методів аналізу і узагальнення динамічних рядів. Зображена на графіку лінія тренду динамічного ряду покаже плавну зміну досліджуваного явища в часі, яке звільнене від короткочасних відхилень, викликаних різними причинами. В статистичній практиці виявлення основної тенденції розвитку явищ в часі проводиться методами укрупнення інтервалів, рухомої середньої і аналітичним вирівнюванням.

Одним з найпростіших способів обробки ряду з метою виявлення закономірності зміни його рівнів є **укрупнення інтервалів** (періодів) часу. Суть цього методу полягає в тому, що дані динамічного ряду об'єднуються в групи по періодах і розраховується середній показник на період - трирічччя, п'ятиріччя і т.д.

Укрупнення інтервалів проілюструємо за даними наступного прикладу.

Динаміка врожайності озимої пшениці в селянській спілці "Колос".

Роки	Урожайність озимої пшениці, ц/га	Сумарна врожайність, ц/га (за триріччя)	Середня врожайність, ц/га (за триріччя)
1991	15,6		
1992	16,0	50,5	16,8
1993	18,9		
1994	15,7		
1995	20,0	55,3	18,4
1996	19,6		
1997	19,8		
1998	21,5	61,3	20,4
1999	20,0		
2000	27,3		
2001	24,4	79,9	26,6
2002	28,2		
2003	27,9		
2004	33,1	93,7	31,2
2005	32,7		

В результаті проведеного укрупнення інтервалів, взявши дані за триріччя, ми отримали нові ряди динаміки сумарної і середньорічної урожайності за три роки, які показують тенденцію її росту.

Важливим способом виявлення загальної тенденції ряду динаміки є **згладжування за допомогою рухомої середньої**. Тут також вдаються до укрупнення періодів, але воно проводиться шляхом послідовних зміщень на одну дату при збереженні постійного інтервалу періоду.

Порядок розрахунку рухомої середньої покажемо за даними попереднього прикладу.

Розрахунок трирічної рухомої середньої врожайності озимої пшениці.

Роки	Урожайність озимої пшениці, ц/га	Сумарна врожайність, ц/га (за трирічча)	Середня врожайність, ц/га (за трирічча)
1991	15,6	-	–
1992	16,0	50,5	16,8
1993	18,9	50,6	16,9
1994	15,7	54,6	18,2
1995	20,0	55,3	18,4
1996	19,6	59,4	19,8
1997	19,8	60,9	20,3
1998	21,5	61,3	20,4
1999	20,0	68,8	22,9
2000	27,3	71,7	23,9
2001	24,4	79,9	26,6
2002	28,2	80,5	26,8
2003	27,9	89,2	29,7
2004	33,1	93,7	31,2
2005	32,7	–	–

Як видно із таблиці, згладжений ряд, який складається з рухомих середніх показує більш плавне підвищення урожайності рухомої середньої.

Найбільш ефективним способом виявлення основної тенденції є **аналітичне вирівнювання**.

На практиці найбільш поширеними формулами, які виражають тенденцію розвитку (тренд) явищ є: пряма, гіпербола, парабола другого порядку, показникова функція, ряди Фур'є, логістична функція, експонента та інші.

Вирівнювання за прямою використовується в тих випадках, коли абсолютні прирости більш-менш постійні, тобто коли рівні динамічного ряду змінюються в арифметичній прогресії, або близькі до неї.

Рівняння прямої має вигляд: $\hat{y}_t = a_0 + a_1 t$,

де $\hat{y}_t$ - вирівняні значення динамічного ряду;

a_0, a_1 - параметри шуканої прямої (початковий рівень і щорічний приріст);

t - умовне позначення часу.

Для знаходження параметрів «a_0» і «a_1» потрібно розв'язати за способом найменших квадратів систему нормальних рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum t; \\ \sum ty = a_0 \sum t + a_1 \sum t^2, \end{cases}$$

де y - фактичні рівні динамічного ряду;

n - число членів ряду динаміки.

При відліку часу від середини ряду коли $\sum t = 0$, тоді система рівнянь для знаходження параметрів «a_0» і «a_1» матиме вигляд: $\sum y = na_0$; $\sum ty = a_0 \sum t^2$,

звідки: $a_0 = \frac{\sum y}{n}$, $a_1 = \frac{\sum yt}{\sum t^2}$.

Методику вирівнювання врожайності озимої пшениці за рівнянням прямої покажемо на прикладі.

Розрахункова таблиця для вирівнювання ряду динаміки за прямою

Роки	Урожайність озимої пшениці, ц/га (y)	Умовне позначення часу (t)	t^2	Yt	Вирівняна урожайність $\hat{y}_t = a_0 + a_1 t$
1991	15,6	-7	49	-109,2	14,1
1992	16,0	-6	36	-96,0	15,3
1993	18,9	-5	25	-94,5	16,5
1994	15,7	-4	16	-62,8	17,8
1995	20,0	-3	9	-60,0	19,0
1996	19,6	-2	4	-39,2	20,2
1997	19,8	-1	1	-19,8	21,5
1998	21,5	0	0	0	22,7
1999	20,0	1	1	20,0	23,9
2000	27,3	2	4	54,6	25,2
2001	24,4	3	9	73,2	26,4
2002	28,2	4	16	112,8	27,7
2003	27,9	5	25	139,5	28,9
2004	33,1	6	36	198,6	30,1
2005	32,7	7	49	228,9	31,4
n=15	340,7	0	280	346,1	340,7

Використовуючи розрахункові підсумки, отримуємо:

$$a_0 = \frac{\sum y}{n} = \frac{340{,}7}{15} = 22{,}713; \quad a_1 = \frac{\sum yt}{\sum t^2} = \frac{346{,}1}{280{,}0} = 1{,}236.$$

Звідси рівняння прямої буде мати наступний вигляд: $\hat{y}_t = 22{,}713 + 1{,}236t$.

Коефіцієнт регресії (a_1=1,236) характеризує середній приріст урожайності озимої пшениці за рік. Величина 22,713 буде показувати теоретичну врожайність 1998 р., для якого ми взяли «0» номер року. Підставляючи у рівняння $\hat{y}_t = 22{,}713 + 1{,}236t$ послідовно значення t (-7, -6, -5 і т.д.), отримаємо вирівняний (теоретичний) ряд динаміки урожайності озимої пшениці

Результати проведеного аналітичного вирівнювання ряду динаміки урожайності озимої пшениці за 1991-2005 рр. і фактичні дані покажемо на графіку (мал. 7.2.).

Вирівнювання за гіперболою проводиться тоді, коли із плином часу ряд динаміки зростає або спадає до певної межі.

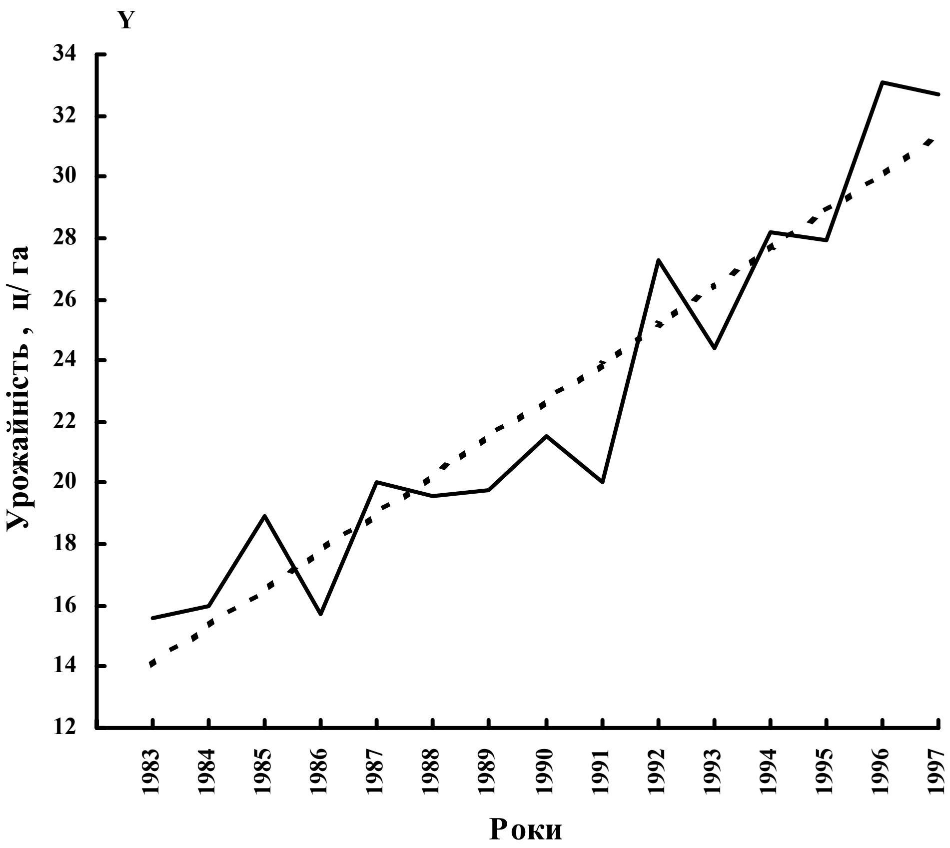
Рівняння гіперболи визначається за формулою: $\hat{y}_t = a_0 + a_1 \frac{1}{t}$.

Для знаходження параметрів «a_0» і «a_1» в даному рівнянні способом найменших квадратів застосовують систему нормальних рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum \frac{1}{t}; \\ \sum \frac{y}{t} = a_0 \sum \frac{1}{t} + a_1 \sum \frac{1}{t^2}. \end{cases}$$

Якщо добитись, щоб $\sum t = 0$, тоді параметри «a_0» і «a_1» знаходять за новою системою рівнянь: $\begin{cases} \sum y = na_0; \\ \sum \frac{y}{t} = a_1 \sum \frac{1}{t^2}. \end{cases}$

Перетворивши цю систему рівнянь в логарифмічну, будемо мати: $\begin{cases} \sum \ell gy = n\ell ga_0; \\ \sum t\ell gy = \ell ga_1 \sum t^2, \end{cases}$ звідси: $\ell g\, a_0 = \frac{\sum \ell gy}{n}$; $\quad \ell g\, a_1 = \frac{\sum t\ell gy}{\sum t^2}$.



---------- вирівняні дані;

———— фактичні дані.

Мал. 7.2 Урожайність озимої пшениці в селянській спілці "Колос" за 1991-2005 рр.

При **вирівнюванні за параболою другого порядку** $\hat{y}_t = a_0 + a_1 t + a_2 t^2$ параметри «a_0», «a_1», і «a_2» визначаються способом найменших квадратів, для чого складають і розв'язують систему нормальних рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum t + a_2 \sum t^2; \\ \sum ty = a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3; \\ \sum yt^2 = a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4. \end{cases}$$

Якщо добитись щоб $\sum t = 0$, тоді і $\sum t^3 = 0$, а отже, система рівняння спроститься:

$$\begin{cases} \sum y = na_0 + a_2 \sum t^2; \\ \sum ty = a_1 \sum t^2; \\ \sum yt^2 = a_0 \sum t^2 + a_2 \sum t^4. \end{cases}$$

Із цієї системи $a_1 = \frac{\sum yt}{\sum t^2}$, а параметри «$a_0$» і «$a_2$» визначаються розв'язком системи двох рівнянь з двома невідомими.

Параметри рівняння параболи другого порядку потрібно інтерпретувати наступним чином: «a_0» - це величина, яка виражає середні умови утворення рівнів ряду; «a_1» - швидкість розвитку даних динамічного ряду; «a_2»- характеризує прискорення цього розвитку.

Вирівнювання за показниковою функцією проводиться в тих випадках, коли динамічний ряд розвивається в геометричній прогресії, тобто тоді, коли ланцюгові темпи росту більш-менш постійні.

Показникова функція описується рівнянням: $\tilde{y}_t = a_0 + a_1^{\,t}$.

Для визначення параметрів «a_0» і «a_1» цього рівняння методом найменших квадратів попередньо логарифмують рівні, тоді логарифми показникової функції описують лінійною функцією: $\ell g\tilde{y}_t = \ell g a_0 + t\ell g a_1$.

Система нормальних рівнянь має вигляд:

$$\begin{cases} \sum \ell gy = n\ell ga_0 + \ell ga_1 \sum t; \\ \sum t\ell gy = \ell ga_0 \sum t + \ell ga_1 \sum t^2, \end{cases}$$

коли добитись, що $\sum t = 0$, тоді $\begin{cases} \sum \ell gy = n\ell ga_0; \\ \sum t\ell gy = \ell ga_1 \sum t^2, \end{cases}$

звідки $\ell ga_0 = \frac{\sum \ell gy}{n}$ і $\ell ga_1 = \frac{\sum t\ell gy}{\sum t^2}$.

Коефіцієнт «a_1», в показниковій функції характеризує середній темп росту досліджуваної ознаки.

Особливе місце в аналітичному вирівнюванні рядів динаміки займає **вирівнювання за допомогою ряду Фур'є**, який описується рівнянням:

$$\tilde{y}_t = a_0 + \sum_{n=1}^{m} \left(a_n coskt + a_{n-1} \sin kt\right),$$

де k - ступінь точності гармонік (найчастіше від 1 до 4);

t - час, виражений в радіанній мірі або градусах.

При вирівнюванні по ряду Фур'є періодичні коливання рівнів динамічного ряду виступають у вигляді суми декількох гармонік, нашарованих одна на одну. Так, наприклад, при k = 1 рівняння Фур'є матиме вигляд:

$$\hat{y}_t = a_0 + a_1 \cos t + a_2 \sin t.$$

Параметри рівняння теоретичних рівнянь визначаються за способом найменших квадратів.

Знайшовши часткові похідні функції ряду Фур'є і прирівнявши їх до нуля, отримаємо систему нормальних рівнянь, за якими можна вирахувати параметри:

$$a_0 = \frac{\sum y}{n},\ a_1 = \frac{2\sum y\cos t}{n}, \quad a_2 = \frac{2\sum y \sin t}{n}.$$

Вирівнювання радів динаміки використовують також для знаходження відсутніх членів раду за допомогою інтерполяції і екстраполяції.

Інтерполяцією називається в статистиці знаходження відсутнього показника усередині раду. Наприклад, на початок 2005 року в селянській спілці числилось 10 тис. голів овець. Потрібно визначити ймовірну чисельність овець на 1.04.2005 р.

Визначимо річний абсолютний приріст овець:

$$\Delta_y = y_i - y_1 = 12{,}4 - 10{,}0 = 2{,}4\, тис.голів.$$

Знаходимо середньомісячний абсолютний приріст:

$$\overline{\Delta}_y = \frac{\sum \Delta y}{n} = \frac{2{,}4}{12{,}0} = 0{,}2\, тис.голів.$$

Якщо припустити, що кожного місяця абсолютний приріст овець був приблизно однаковий, тоді на 1.04.2005 р. в спілці числилось 10,6 тис. голів:

$$\hat{y}_{t(1.04.2005 р.)} = y_1 + \overline{\Delta}_y \cdot t = 10 + 0.2 \cdot 3 = 10{,}6\, тис.голів.$$

Екстраполяцією в статистиці називається знаходження невідомих рівнів в кінці або на початку динамічного ряду. Звернемось до прикладу. Нехай в місті на 1.01.2005 р. проживало 200 тис. чол. Середньорічний темп приросту за попередні п'ять років склав 2 *%.* Потрібно визначити ймовірну чисельність населення міста на 1.01.2010 р.

Для знаходження перспективної чисельності населення станом на 1.01.2010р. використовуємо формули:

$$\hat{y}_t = y_1 \cdot \overline{T}^t, \text{ або } \hat{y}_t = y_1 + \overline{\Delta}_y \cdot t;$$

$$\hat{y}_{t(1.01.2010\,р.)} = 200 \cdot 1{,}02^5 = 220\,\textit{тис.чол.},$$

$$\text{або } \hat{y}_{t(1.01.2010\,р.)} = 200 + 4 \cdot 5 = 220\,\textit{тис.чол.}$$

Як інтерполяція, так і екстраполяція ґрунтуються на припущенні, що наявні величини цілком достатньо визначають темп розвитку досліджуваного явища і, отже, його можна поширювати на відсутні рівні динамічного ряду.

Коефіцієнт випередження являє собою відношення більшого середньорічного темпу приросту до меншого.

7.5. Вимірювання сезонних коливань

Сезонними коливаннями називаються більш-менш стійкі внутрішньорічні коливання в рядах динаміки, обумовлені специфічними умовами виробництва чи споживання певного виду продукції.

Для дослідження внутрішньорічних коливань можна використати цілий ряд методів (простої середньої, Персонса, рухомої середньої, аналітичного вирівнювання, рядів Фур'є), які забезпечують їх оцінку з різною точністю, надійністю і трудоємкістю.

Сезонні коливання характеризуються спеціальним показником, який називається індексом сезонності (I_s). В сукупності ці індекси утворюють сезонну хвилю.

Індекс сезонності – це процентне відношення однойменних місячних (квартальних) фактичних рівнів динамічних рядів до їх середньорічних або вирівняних рівнів.

Наочну уяву про зміну попиту населення на товари культурно-побутового призначення в окремі періоди року дають графіки. (мал. 7.3.).

Індекс сезонності (сезонну хвилю) реалізації побутових холодильників торговими підприємствами споживчої кооперації за 2003-2005 рр. розрахуємо **методом простих середніх**.

Індекс сезонності за методом простої середньої визначається за формулою:

$$I_s = \frac{\bar{y}_i}{\bar{y}_з} \cdot 100,$$

де I_s - індекс сезонності;

$\bar{y}_i$ – середні місячні або квартальні рівні;

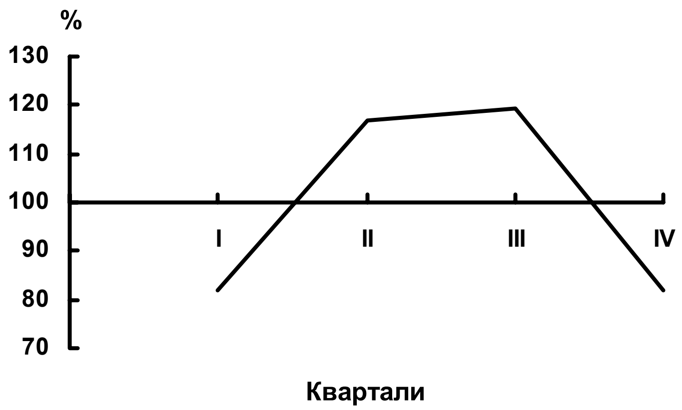
$\bar{y}_з$ – загальна середня (місячна або квартальна).

Розрахунок сезонної хвилі реалізації побутових холодильників торговими підприємствами споживчої кооперації області за 2003-2005 рр., шт.

Квартал	Роки			Разом	В середньому ($\bar{y}_i$)	Сезонна хвиля, % $I_s = \frac{\bar{y}_i}{\bar{y}_з} \cdot 100$
	2003	2004	2005			
I	1942	2126	2505	6573	2191,00	82,1
II	2957	2704	3704	9365	3121,67	117,0
III	2504	3291	3834	9629	3209,67	120,3
IV	2194	1745	2513	6452	2150,67	80,6
Разом	9597	9866	12556	32019	$\bar{y}_з$=2668,25	400,0

$$I_s^I = \frac{\bar{y}_1}{\bar{y}_з} \cdot 100 = \frac{2191{,}00}{2668{,}25} \cdot 100 = 82{,}1\%;$$

$$I_s^{II} = \frac{\bar{y}_2}{\bar{y}_з} \cdot 100 = \frac{3121{,}67}{2668{,}25} \cdot 100 = 117{,}0\%, \textit{і т.д.}$$



Мал. 7.3. Сезонна хвиля реалізації побутових холодильників системою споживчої кооперації області за 2003-2005 рр.

Судячи з таблиці і графіка (мал. 7.3.), реалізація побутових холодильників суттєво падає в першому і четвертому кварталах і різко зростає в другому і третьому кварталах року.

Більшість динамічних рядів досліджуваних явищ мають тенденцію росту, тому для більш точного визначення сезонної хвилі в таких рядах необхідна нейтралізація еволюції тренду. З цією метою використовують **метод ланцюгових індексів (метод Персонса)**.

Для вивчення сезонності часто доводиться вираховувати **рухому середню** з парним числом членів ряду, тому що характер динамічного ряду визначає тривалість періоду рухомої середньої, який повинен співпадати з періодом коливання, або бути кратним йому.

Згладжування за парним числом членів ряду незручне тим, що середня мусить бути віднесена тільки до середини між двома датами, тобто проходить зсув періоду, до якого відноситься рівень. Усунення зсуву періоду проводять способами перетворення рівнів і центруванням.

Для розрахунку сезонних коливань використовують також метод аналітичного вирівнювання за рівнянням прямої.

Рік, квартал		Реалізація мотоциклів і велосипедів, млн. грн. (y)	t	t^2	yt	$\hat{y}_t$	Сезонна хвиля, $I_s = \frac{\bar{y}_i}{\bar{y}_з} \cdot 100$
2001	I	79,4	- 19	361	-1508,6	81,2	97,8
	II	102,4	- 17	289	- 1740,8	82,1	124,2
	III	89,5	- 15	225	- 1342,5	82,9	108,0
	IV	58,9	- 13	169	- 765,7	83,8	70,3
2002	I	77,6	- 11	121	- 853,6	84,7	91,6
	II	105,6	- 9	81	- 950,4	85,6	123,4
	III	87,8	- 7	49	- 614,6	86,5	101,4
	IV	63,9	- 5	25	- 319,5	87,3	73,2
2003	I	79,1	- 3	9	- 237,3	88,2	89,7
	II	89,1	- 1	1	- 89,1	89,1	100,0
	III	93,3	1	1	93,3	90,0	103,7
	IV	87,2	3	9	261,6	90,8	96,0
2004	I	89,1	5	25	445,5	91,7	97,2
	II	120,8	7	49	845,6	92,6	130,4
	III	100,5	9	81	904,5	93,5	107,5
	IV	64,1	11	121	705,1	94,4	67,9
2005	I	95,1	13	169	1236,3	95,2	99,9
	II	130,9	15	225	1936,5	96,1	136,2
	III	107,8	17	289	1832,6	97,0	111,1

	IV	68,6	19	361	1303,4	97,8	70,1
Разом:		1790,7	x	2660	1169,3	1790,6	x

Параметри «a_0» і «a_1» для рівняння прямої $\hat{y}_t = a_0 + a_1 t$ знаходимо методом найменших квадратів розв'язавши систему нормальних рівнянь:

$$\begin{cases} \sum y = na_0 + a_1 \sum t; \\ \sum yt = a_0 \sum t + a_1 \sum t^2, \end{cases}$$

$$a_0 = \frac{\sum y}{n} = \frac{1790{,}7}{20} = 89{,}535; \quad a_1 = \frac{\sum yt}{\sum t^2} = \frac{1169{,}3}{2660{,}0} = 0{,}439,$$

звідси $\hat{y}_t = 89{,}535 + 0{,}439t.$

Підставляючи в дане рівняння послідовно значення (t) матимемо вирівняний ряд: $\hat{y}_t^I = 89{,}535 + 0{,}439 \cdot (-19) = 81{,}2;$

$$\hat{y}_t^{II} = 89{,}535 + 0{,}439 \cdot (-17) = 82{,}1, \text{ і т.д.}$$

Індекси сезонності (сезонну хвилю) реалізації мотоциклів і велосипедів визначаємо як процентне відношення емпіричних рівнів до теоретичних:

$$I_s^I = \frac{y_1}{\hat{y}_{t_1}} \cdot 100 = \frac{79{,}4}{81{,}2} \cdot 100 = 97{,}8\,\%; \quad I_s^{II} = \frac{y_2}{\hat{y}_{t_2}} \cdot 100 = \frac{102{,}4}{82{,}1} \cdot 100 = 124{,}2\,\%; \text{ і т.д.}$$

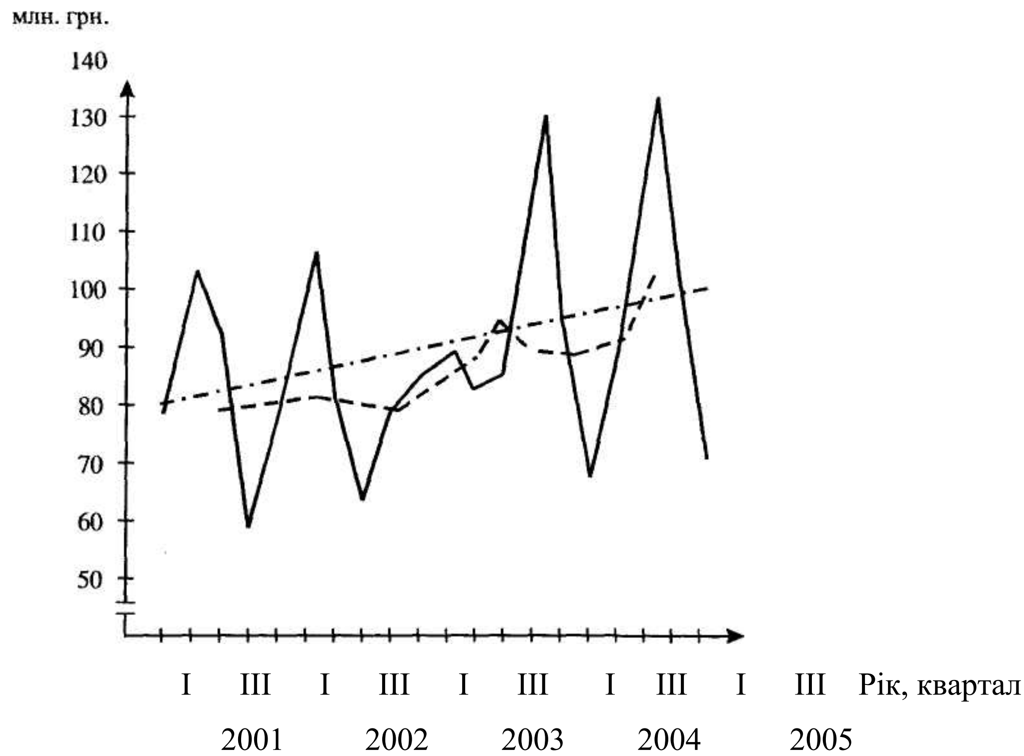
Результати згладжування внутрішньорічних коливань за методом рухомої середньої і визначення загальної тенденції реалізації мотоциклів і велосипедів населенню України за рівнянням прямої наочно показані на графіку (мал. 7.4).

Судячи з графіка, пряма лінія цілком об'єктивно відображає тенденцію розвитку досліджуваного явища.

Моделювання сезонних коливань різних явищ можна проводити і з допомогою *ряду Фур'є,* аналітичний вираз якого стосовно динаміки має наступний вигляд:

$\hat{y}_t = a_0 + a_1 \cos t + a_2 \sin t.$

Таке моделювання сезонної хвилі покажемо на прикладі реалізації холодильників побутових.



-------------- емпіричний ряд;

—————— рівняння прямої.

Мал. 7.4. Згладжування внутрішньорічних коливань реалізованого попиту населення України на мотоцикли і велосипеди рівнянням прямої.

Розрахунок сезонної хвилі реалізації холодильників побутових за допомогою ряду Фур'є ($\bar{y}_t = 2668{,}25 + 64{,}78\cos t - 381{,}41\sin t$) і середнього квадратичного відхилення індексів сезонності.

Рік, квартал		Реалізація холодильників, щт. (y)	Вирівнюваний ряд динаміки ($\bar{y}_t$)	Сезонна хвиля, $I_s = \frac{\bar{y}_i}{\bar{y}_з} \cdot 100$	$I_s - 100$	$(I_s - 100)^2$
2003	I	1942	2733,03	71,0	- 29,0	841,00
	II	2957	2533,65	116,7	16,7	278,89
	III	2504	2370,34	105,6	5,6	31,36
	IV	2194	2286,84	95,9	- 4,1	16,81
2004	I	2126	2305,56	92,2	- 7,8	60,84
	II	2704	2421,45	111,7	11,7	136,89
	III	3291	2603,47	126,4	26,4	696,96
	IV	1745	2802,85	62,2	- 37,8	1429,84
2005	I	2505	2966,16	84,4	- 15,6	243,36
	II	3704	3049,66	121,4	21,4	457,96
	III	3834	3030,94	126,5	26,5	702,25
	IV	2513	2915,05	86,2	- 13,8	190,44
Разом:		32019	32019,00	-	-	5085,60

Показник сили коливання динамічного ряду із-за сезонного характеру реалізації холодильників визначається за формулою середнього квадратичного відхилення індексів сезонності:

$$\sigma = \sqrt{\frac{\sum(I_s - 100)^2}{4}} = \sqrt{\frac{5085{,}6}{4}} = \sqrt{1271{,}4} = 35{,}6\,\%.$$

Цей показник свідчить про достатньо великий вплив на реалізацію холодильників побутових сезонного чинника.

Як показують вище наведені дані, розраховані індекси сезонності реалізованого попиту на товари культурно-побутового призначення достатньо надійні і точні. Це дозволяє використати їх для екстраполяції реалізованого попиту на дані товари.

7.6 Особливості вимірювання взаємозв'язків в рядах динаміки

При вивченні кореляційних зв'язків в багатомірних рядах динаміки спостерігається певна залежність рівнів даного періоду від попереднього, внаслідок чого виникають певні методологічні особливості. В таких динамічних рядах фактором зміни рівнів виступає, крім інших, також час.

Вплив даного рівня динамічного ряду на зміну наступного з плином часу приводить до **автокореляції**.

В практиці статистичного аналізу рядів динаміки застосовують різні способи усунення автокореляції, такі як спосіб різницевих перетворень (при лінійному тренді), спосіб відхилень тенденції (при нелінійній залежності), або введення змінної величини «t» в рівняння регресії $\tilde{y}_t = f(x_1, x_2, x_3, ..., t)$, де вона відіграє роль чинника часу.

При застосуванні методу **регресії** для дослідження динамічних рядів виникає особливість, яка заключається в тому, що в рівнях динамічних рядів присутня **авторегресія**, яка проявляється так же, як і автокореляція.

Авторегресія виражає залежність величини рівня динамічного ряду від попередніх значень рівня в певні моменти часу.

Методику побудови рівня регресії з введенням фактора часу розглянемо на прикладі двох взаємозв'язаних рядів динаміки: глибини зрошення багаторічних трав з добавленням органічних компонентів, та урожайності насіння цих трав.

Розрахункова таблиця для обчислення коефіцієнтів регресії.

Роки	Глибина зрошення з добавленням органічних компонентів, см. (x)	Урожайність насіння багаторічних трав, ц/га (y)	t	t^2	x^2	xy	xt	yt	$\hat{y}_t$
1995	12	5,6	-5	25	144	67,2	-60	-28,0	4,9
1996	8	4,0	-4	16	64	32,0	-32	-16,0	3,4
1997	10	4,0	-3	9	100	40,0	-30	-12,0	4,3
1998	6	2,4	-2	4	36	14,0	-12	-4,8	2,8
1999	9	3,6	-1	1	81	32,4	-9	-3,6	4,0
2000	15	5,0	0	0	225	75,0	0	0	6,4
2001	11	4,6	1	1	121	50,6	11	4,6	4,9
2002	13	6,5	2	4	169	84,5	26	13,0	5,8
2003	14	7,0	3	9	196	98,0	42	21,0	6,3
2004	10	4,5	4	16	100	45,0	40	18,0	4,8
2005	12	6,0	5	25	144	72,0	60	30,0	5,6
n=15	120	53,2	0	110	1380	611,1	36	22,2	53,2

Зв'язок між глибиною зрошення багаторічних трав з добавлення органічних компонентів та врожайності насіння цих трав можна відобразити лінійною функцією:

$$\hat{y}_t = a_0 + a_1 x + a_2 t,$$

де a_1 - параметр, який характеризує середній приріст результативної ознаки на одиницю приросту факторної ознаки (x);

a_2 - середній щорічний приріст (y) під впливом зміни комплексу чинників крім (x);

t - час (роки).

Параметри цього рівняння визначаються методом найменших квадратів, склавши і розв'язавши систему нормальних рівнянь з трьома невідомими:

$$\begin{cases} \sum y = na_0 + a_1 \sum t + a_2 \sum t; \\ \sum xy = a_0 \sum x + a_1 \sum x^2 + a_2 \sum xt; \\ \sum xt = a_0 \sum t + a_1 \sum xt + a_2 \sum t^2. \end{cases}$$

Якщо добитись що $\sum t = 0$, то система нормальних рівнянь матиме вигляд:

Розв'язавши цю систему, отримаємо параметрів: a_0=0,549; a_1=0,393; a_2=0,073.

Лінійне рівняння зв'язку буде мати вигляд: $\hat{y}_t = 0{,}549 + 0{,}393x + 0{,}073t$.

Параметри рівняння регресії потрібно тлумачити так: якщо при інших рівних умовах глибина зрошення багаторічних трав з добавленням органічних компонентів збільшиться на 1 см., то врожайність насіння цих трав зросте на 0,393 ц/га. За рахунок впливу інших факторів, які рівномірно змінюються протягом часу, урожайність насіння багаторічних трав щорічно зростатиме в середньому на 0,073 ц/га.

Підставляючи в отримане рівняння регресії значення ***t*** і ***x*** визначимо теоретичні рівні врожайності багаторічних трав:

$\hat{y}$1995 р. = 0>^{54}9 + °'393 • 12 + 0,073 • (- 5) = 4,9;

$\hat{y}$1996 р. = 0>549 + 0,393 • 8 + 0,073 • (- 4) = 3,4 ;

$\hat{y}$ 1997р. = °-^{54}9 + 0,393 • 10 + 0,073 • (- 3) = 4,3 ; і т.д.

Приведене рівняння регресії повинно виключити авторегресію. Для переконання в цьому знайдемо автокореляцію різниць між фактичними даними і вирівняними даними за цим рівнянням, тобто кореляцію величин $y - \hat{y} = \varepsilon_t$.

Коефіцієнт автокореляції відхилень приймає значення в межах від -1 до +1 і визначається за формулою:

$$r_a = \frac{\sum \varepsilon_t \cdot \varepsilon_{t+1}}{\sum \varepsilon_t^2}.$$

Розрахунок коефіцієнта автокореляції.

Роки	y	$\hat{y}_t$	$\varepsilon_t = (y - \hat{y}_t)$	ε_{t+1}	$\varepsilon_t \cdot \varepsilon_{t+1}$	$\varepsilon_t^2 = (y - \hat{y}_t)^2$
1995	5,6	4,9	0,7	0,6	0,42	0,49
1996	4,0	3,4	0,6	-0,3	-0,18	0,36
1997	4,0	4,3	-0,3	-0,4	0,12	0,09
1998	2,4	2,8	-0,4	-0,4	0,16	0,16
1999	3,6	4,0	-0,4	-1,4	0,56	0,16
2000	5,0	6,4	-1,4	-0,3	0,42	1,96
2001	4,6	4,9	-0,3	0,7	-0,21	0,09
2002	6,5	5,8	0,7	0,7	0,49	0,49
2003	7,0	6,3	0,7	-0,3	-0,21	0,49
2004	4,5	4,8	-0,3	0,4	-0,12	0,09
2005	6,0	5,6	0,4	0,7	0,28	0,16
Разом:	53,2	53,2	0	0	1,73	4,54

Коефіцієнт автокореляції з часовим зсувам-лагом (p=1) дорівнює:

$$r_a = \frac{\sum \varepsilon_t \cdot \varepsilon_{t+1}}{\sum \varepsilon_t^2} = \frac{1,73}{4,54} = 0,381.$$

Значення коефіцієнта додатнє ($r_a = 0,381$) свідчить про незначний ступінь кореляції залишкових величин.

Для нашого прикладу: $r_{a_ф} = 0,381$, при n = 10 і 5 %-ному рівні ймовірності $r_{a_T} = 0,353$ і . Тобто $r_{a_ф} > r_{a_T}$ на 0,028 пункти, що і засвідчує про незначну автокореляцію.

В багатьох економіко-статистичних дослідженнях доводиться вивчати паралельно декілька динамічних рядів, в яких коливання рівнів взаємообумовлені, наприклад, динаміка цін на які-небудь овочі на ринку в значній мірі зв'язана з їх урожайністю; в свою чергу динаміка урожайності або валовий збір залежать від динаміки кількості опадів, агрохімобробітку; попит населення на певні товари народного споживання залежить від пропозиції, тобто від об'єму їх виробництва і т.д.

Для вимірювання залежності між такими рядами динаміки використовують методи кореляції, тобто розраховують різні коефіцієнти кореляції.

Якщо використовують лінійний коефіцієнт кореляції: $r = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y}$, то для знаходження коефіцієнтів автокореляції "$r_{a_1}; r_{a_2}$ *і т.д.*" вираховують наступні необхідні величини:

$$\overline{x_t \cdot x_{t+1}} = \frac{\sum x_t \cdot x_{t+1}}{n}; \qquad \overline{x_t} = \frac{\sum x_t}{n}; \qquad \sigma_{x_t}^2 = \frac{\sum x_t^2}{n} - \overline{x}_t^2;$$

$$r_{a_1} = \frac{\overline{x_t \cdot x_{t+1}} - \left(\overline{x_t}\right)^2}{\sigma_{x_t}^2}; \qquad r_{a_2} = \frac{\overline{y_t \cdot y_{t+1}} - \left(\overline{y_t}\right)^2}{\sigma_{y_t}^2} \qquad і\ т.д.$$

Для того, щоб зробити висновок про існування автокореляції порівнюють отримані коефіцієнти автокореляції з табличними за певною чисельністю спостережень «*n*» і рівні значимості $\rho = 0,05$ (5 % -ний рівень).

Одним із способів усунення автокореляції є **корелювання відхилень фактичних рівнів від вирівняних**, які відображають тренд. Для цього потрібно:

а) провести аналітичне вирівнювання порівнювальних рядів; б) визначити величини відхилень кожного фактичного рівні динаміки від їх вирівняних значень; в) провести корелювання отриманих відхилень.

Коефіцієнт кореляції відхилень розраховують за формулою:

$$r = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}},$$

де r – коефіцієнт кореляції відхилень фактичних рядів від вирівняних;

$d_x = x - \hat{x}_t$; $d_y = y - \hat{y}_t$ – відхилення фактичних рівнів динаміки від вирівняних обох рядах.

До аналогічних результатів можна прийти, якщо знайти *кореляцію різниць між наступними і попередніми рівнями* обох рядів. При заміні рівнів динамічних рядів різницями між ними усувається вплив автокореляції в кожному динамічному ряду.

Коефіцієнт кореляції перших різниць визначається за формулою:

$$r = \frac{\sum \Delta_x \cdot \Delta_y}{\sqrt{\sum \Delta_x^2 \cdot \sum \Delta_y^2}},$$

де r – коефіцієнт кореляції перших різниць;

$\Delta_x = x_i - x_{i-1}$; $\Delta_y = y_i - y_{i-1}$ – різниці між наступними і попередніми рівнями обох рядів.

ЛЕКЦІЯ 8. ІНДЕКСИ

ПЛАН

8.1. Поняття про індекси, їх види.

8.2. Агрегатні індекси як вихідна форма індексів.

8.3. Середньозважені індекси.

8.4. Базисні і ланцюгові індекси з постійними і змінними вагами.

8.5. Індекси змінного, постійного складу і структурних зрушень.

8.6. Територіальні індекси.

8.7. Використання системи взаємозв’язаних індексів в аналізі чинників динаміки.

8.1. Поняття про індекси, їх види

Для характеристики соціально-економічних явищ і процесів статистика широко використовує узагальнюючі показники у вигляді середніх, відносних величин та коефіцієнтів. Одним з таких узагальнюючих показників і є індекси. В широкому розумінні слово "Index" у перекладі з латинської означає "показник".

Індексом у статистиці називається відносний показник, що характеризує зміну рівня соціально-економічного явища в часі, порівняно з планом, базисним періодом або в просторі.

В статистичних дослідженнях складних соціально-економічних явищ і процесів виділяють три великі сфери застосування економічних індексів.

До першої сфери застосування індексів відносять порівняльну характеристику несумарних сукупностей в часі. Сюди входять синтетичні індекси динаміки, виконання плану і територіальні індекси.

Індекси динаміки показують зміну якого-небудь складного явища в звітному періоді порівняно з базисним.

Індекс виконання плану використовують для порівняння досягнутого рівня з плановими завданнями.

Теротиріальні індекси застосовують для просторового порівняння рівнів урожайності, цін, продуктивності праці і т.п., в різних регіонах.

Друга сфера застосування індексів заключається і їх використанні для факторного аналізу складного явища через систему взаємозв'язаних індексів. До таких складних явищ можуть бути віднесені вартість виробленої чи реалізованої продукції, фонд заробітної плати, валовий збір зерна та ін.

Так, вартість виробленої продукції дорівнює добутку цін на кількість продукції, валовий збір зерна – добутку урожайності на посівну площу, фонд заробітної плати – добутку заробітної плати одного працівника на їх чисельність і т.д.

За допомогою **третьої сфери застосування індексів** проводять аналіз динаміки середніх величин, зміна яких піддається впливу структурних зрушень в середині досліджуваної сукупності. В зв'язку з цим, велике значення має вивчення впливу структурних зрушень на динаміку середніх показників через застосування системи взаємозв'язаних індексів змінного складу, постійного (фіксованого) складу і структурних зрушень.

Всі економічні індекси статистика класифікує за трьома основними ознаками:

а) за характером досліджуваних об'єктів;

б) за ступенем охоплення елементів сукупності;

в) за методикою розрахунку загальних індексів.

За характером досліджуваних об'єктів індекси ділять на індекси об'ємних (кількісних) і якісних показників.

До першої групи відносяться індекси фізичного обсягу продукції промисловості, сільського господарства, будівництва та ін.

До другої групи якісних показників відносять індексів цін, собівартості, урожайності і ряд інших.

За ступеня охоплення елементів сукупності індекси ділять на:

а) індивідуальні;

б) загальні;

в) групові.

Індивідуальні індекси характеризують зміну окремих елементів складного явища. В теорії індексів показник, зміну якого характеризує індекс, називається **індексованою величиною**.

Індивідуальні індекси позначають малою латинською буквою «*i*», продукцію в натуральному виразі – через «*q*», ціну одиниці товару – через «*p*», собівартість одиниці продукції – через «*z*» і т.д. Індивідуальні індекси цих ознак визначаються за формулами:

а) фізичного обсягу: $i_q = \frac{q_1}{q_0}$;

б) ціни одиниці товару: $i_p = \frac{p_1}{p_0}$;

в) собівартості одиниці продукції: $i_z = \frac{z_1}{z_0}$;

де i_q, i_p, i_z – індивідуальні індекси фізичного обсягу, ціни і собівартості одиниці продукції;

q_1, q_0; p_1, p_0; z_1, z_0 – фізичний обсяг, ціна, собівартість у звітному і базисному періодах.

Загальні індекси характеризують зміну сукупності в цілому і являють собою відносні числа, що визначають зміни в часі порівняно з плановим, базисним періодами або в просторі складного явища, яке складається з несумірних елементів.

Груповими або **субіндексами** називаються такі індекси, які охоплюють не всі елементи сукупності, а тільки яку-небудь частину або їх групу.

В залежності від методології обчислення, загальні і групові індекси діляться на агрегатні і середні з індивідуальних індексів.

Агрегатні індекси є основною формою економічних індексів, а середні із індивідуальних індексів – похідними, отриманими в результаті перетворення агрегатних індексів.

Базисні і ланцюгові індекси обчислюють в тих випадках, коли доводиться вивчати яке-небудь явище суспільного життя за ряд послідовних років.

8.2. Агрегатні індекси як вихідна форма індексів

Агрегатний індекс являється основною формою економічного індекса. Його назва пішла від латинського слав «aqqreqo» – приєдную. Чисельник і знаменник цього індекса являє собою агрегат, набір різнорідних елементів.

Отже, **агрегатним індексом** в статистиці називається загальний індекс, який є відношенням сум добутків індексованих (зіставлюваних) величин порівнюваних періодів на ваги (співвимірники, за допомогою яких сумуються різнорідні елементи).

При побудові формул агрегатних індексів використовують наступне правило: **«якщо індексована величина – якісний показник, який знаходять шляхом ділення (ціна, собівартість, урожайність і т.д.) ваги беруться звітного періоду, а якщо індексована величина – кількісний показник, який можна**

підсумувати (фізичний обсяг продукції, чисельність працівників, посівна плаща) ваги беруться базисного періоду».

Покажемо застосування цього правила при побудові формул агрегатних індексів.

Загальний індекс цін визначається за формулою:

$$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1};$$

де *Ip* – загальний індекс цін;

p – індексована величина;

q – вага.

Цей індекс показує, як змінилися ціни на всі досліджувані товари в звітному періоді порівняно з базисним.

Загальний індекс фізичного обсягу визначається за формулою:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0},$$

де I_q – загальний індекс фізичного обсягу продукції;

q – індексована величина;

p – вага.

Даний індекс показує зміну кількості виробленої або реалізованої продукції в звітному періоді порівняно з базисним.

Загальний індекс обсягу товарообороту показує зміну виробництва або реалізації продукції в звітному періоді порівняно з базисним у фактичних цінах:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}.$$

Ці індекси взаємозв'язані:

$$I_p \cdot I_q = I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0},$$

звідси $I_p = \frac{I_{pq}}{I_q}$, *а* $I_q = \frac{I_{pq}}{I_p}$.

Абсолютна сума економії або перевитрат від зміни цін визначається як різниця між чисельником і знаменником загального індекса цін:

$$\Delta pq(p) = \sum p_1 q_1 - \sum p_0 q_1.$$

Розглянемо приклад. Маємо дані про реалізацію деяких продуктових товарів на ринках міста в січні місяці.

Назва товару	Продано товарів, тис. кг.		Середня ціна одного кілограма, грн.	
	2004р.	2005р.	2004р.	2005р.
Морква, кг	15,3	16,2	0,22	0,20
Яблука, кг	49,8	51,6	0,90	0,85

Обчислимо індивідуальні індекси цін і фізичного обсягу:

$$i_p = \frac{p_1}{p_0} = \frac{0,20}{0,22} = 0,909; \qquad i_p = \frac{p_1}{p_0} = \frac{0,85}{0,90} = 0,944;$$

$$i_q = \frac{q_1}{q_0} = \frac{16,2}{15,3} = 1,059; \qquad i_q = \frac{q_1}{q_0} = \frac{51,6}{49,8} = 1,036.$$

Ціна кілограма моркви в січні 2005р. порівняно з січнем 2004р. знизилась на 9,1 % (100 – 90,9), яблук – на 5,6 % (100 – 94,4), а кількість реалізованої моркви за цей же період збільшилась в 1,059 рази, або на 5,9 % (105,9 – 100), яблук – в 1,036 рази або на 3,6 %.

Визначимо загальний індекс цін на дані продукти:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{0,20 \cdot 16,2 + 0,85 \cdot 51,6}{0,22 \cdot 16,2 + 0,90 \cdot 51,6} = \frac{47,100}{50,004} = 0,942.$$

Абсолютна сума виграшу населення від зниження цін:

$$\Delta pq(p) = \sum p_1 q_1 - \sum p_0 q_1 = 47,100 - 50,004 = -2,904 \text{ тис. грн.}$$

Індекси фізичного обсягу продуктів дорівнює:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{16,2 \cdot 0,22 + 51,6 \cdot 0,90}{15,3 \cdot 0,22 + 49,8 \cdot 0,90} = \frac{50,004}{48,196} = 1,037.$$

Загальний індекс обсягу товарообороту у фактичних цінах:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{47,100}{48,196} = 0,977.$$

Абсолютна зміна товарообороту у фактичних цінах становить:

$$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0 = 47,100 - 48,196 = -1,096 \text{ тис. грн.,}$$

в тому числі за рахунок зміни кількості проданих товарів:

$$\Delta pq(q) = \sum q_1 p_0 - \sum q_0 p_0 = 50,004 - 48,196 = 1,0808 \text{ тис. грн.}$$

Перевіримо правильність наших розрахунків:

$$I_{pq} = I_p \cdot I_q = 0{,}977 = 0{,}942 \cdot 1{,}037;$$

$$\Delta pq = \Delta pq(p) + \Delta pq(q) = -1{,}096 = -2{,}904 + 1{,}808.$$

Отже, ціни на продукти на ринку в січні 2005р. порівняно з січнем 2004р. знизились на 5,8 %, внаслідок чого населення зекономило 2,904 тис. грн., кількість реалізованих продуктів за цей же період збільшилась в 1,037 рази або на 3,7 %, а товарооборот у фактичних цінах – зменшився на 2,3 % або на 1,096 тис. грн., в тому числі за рахунок зниження цін на 2,904 тис. грн., а за рахунок збільшення кількості проданих товарів він зріс на 1,808 тис. грн.

Аналогічно розраховують систему індексів, зв’язаних із собівартість і кількістю виготовленої продукції.

Агрегатний індекс собівартості продукції: $I_z = \dfrac{\sum z_1 q_1}{\sum z_0 q_1}$.

Загальний індекс фізичного обсягу продукції: $I_q = \dfrac{\sum q_1 z_0}{\sum q_0 z_0}$.

Загальний індекс обсягу затрат на виробництво продукції: $I_{zq} = \dfrac{\sum z_1 q_1}{\sum z_0 q_0}$.

Ці індекси взаємозв’язані: $I_z \cdot I_q = I_{zq} = \dfrac{\sum z_1 q_1}{\sum z_0 q_1} \cdot \dfrac{\sum q_1 z_0}{\sum q_0 z_0} = \dfrac{\sum z_1 q_1}{\sum z_0 q_0}$,

звідси $I_z = \dfrac{I_{zq}}{I_q}$, $I_q = \dfrac{I_{zq}}{I_z}$.

Загальна зміна затрат на виробництво продукції в звітному періоді порівняно з базисним дорівнює: $\Delta zq = \Delta zq(z) + \Delta zq(q) = \sum z_1 q_1 - \sum z_0 q_0$, в тому числі за рахунок:

а) зміни собівартості одиниці продукції: $\Delta zq(z) = \sum z_1 q_1 - \sum z_0 q_1$;

б) зміни кількості виробленої продукції: $\Delta zq(q) = \sum q_1 z_0 - \sum q_0 z_0$.

8.3. Середньозважені індекси

В деяких випадках загальні індекси обчислюють як середні перетворені з відповідних агрегатних індексів.

Перетворюють агрегатний індекс в середній з індивідуальних індексів, підставляючи у його чисельник або знаменник замість індексованого показника його вираз, виведений з формули індивідуального індекса. Якщо таку заміну роблять у чисельнику, то агрегатний індекс перетворюється у середній арифметичний, а якщо у знаменнику – в середній гармонічний.

Перетворимо агрегатний індекс фізичного обсягу в середній арифметичний.

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \qquad i_q = \frac{q_1}{q_0}, \quad \textit{звідси} \qquad q_1 = i_q \cdot q_0.$$

Замінивши в формулі агрегатного індекса фізичного обсягу продукції індексовану величину «q_1» на «$i_q \cdot q_0$», отримаємо формулу **середнього арифметичного індекса** фізичного обсягу продукції:

$$\bar{I}_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}.$$

Наведемо приклад. Нехай маємо дані про продаж товарів в універмазі міста.

Товарні групи	Продано в 2004р., тис. грн. ($q_0 p_0$)	Індекси кількості проданих товарів (i_q)
Трикотажні вироби	150	0,98
Тканини	200	1,05
Галантерея	30	1,20

Середній арифметичний індекс фізичного обсягу реалізованих товарів дорівнює: $\bar{I}_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{0{,}98 \cdot 150 + 1{,}05 \cdot 200 + 1{,}20 \cdot 30}{150 + 200 + 30} = \frac{393}{380} = 1{,}034.$

Перетворимо агрегатний індекс цін у **середній гармонічний.**

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}; i_p = \frac{p_1}{p_0}, \text{ звідси } p_0 = \frac{p_1}{i_p}, \qquad \text{а } \bar{I}_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}.$$

Покажемо визначення такого індекса на прикладі. Маємо дані про продаж товарів продуктовими магазинами за два квартали.

Товари	Товарооборот в діючих цінах, тис. грн.		Зміна середніх цін в II кварталі порівняно з I кварталом, %
	I квартал	II квартал	
Овочі	60	64	- 20
М'ясо	42	44	+ 10
Зерно	35	38	без змін

Спочатку визначимо індивідуальні індекси цін:

$$i_p = \frac{p_1}{p_0} = \frac{100-20}{100} = 0{,}8; \quad i_p = \frac{p_1}{p_0} = \frac{100+10}{100} = 1{,}1; \quad i_p = \frac{p_1}{p_0} = \frac{100}{100} = 1{,}0.$$

Обчислимо середній гармонічний індекс цін:

$$\bar{I}_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} = \frac{64+44+38}{\frac{64}{0{,}8}+\frac{44}{1{,}1}+\frac{38}{1{,}0}} = \frac{146}{158} = 0{,}924.$$

Аналогічно перетворюють агрегатний індекс собівартості в середній гармонічний індекс:

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}; \qquad i_z = \frac{z_1}{z_0}; \qquad z_0 = \frac{z_1}{i_z}; \qquad \bar{I}_z = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}}.$$

Середні арифметичні і гармонічні індекси повинні співпадати за своєю величиною з відповідними агрегатними індексами.

Вибір форми індекса залежить від поставленого завдання дослідження і від наявності даних, необхідних для обчислення того чи іншого індекса.

8.4. Базисні і ланцюгові індекси з постійними і змінними вагами

В ряді випадків доводиться аналізувати явища суспільного життя не за два, а за три і більше послідовних періодів. В такому разі, в залежності від бази порівняння, обчислюють індекси з постійною базою порівняння (базисні) і змінною базою порівняння (ланцюгові).

Базисними називаються індекси, які вираховуються шляхом порівняння даних кожного періоду з даними будь-якого одного періоду, прийнятого за базу порівняння, Наприклад:

$$i_{q1} = \frac{q_1}{q_0}; \quad i_{q2} = \frac{q_2}{q_0}; \; \ldots \; i_{qn} = \frac{q_n}{q_0}.$$

Ланцюговими називаються індекси, обчислені шляхом порівняння даних кожного періоду з даними попереднього періоду, Наприклад:

$$i_{q1} = \frac{q_1}{q_0}; \quad i_{q2} = \frac{q_2}{q_1}; \quad i_{q3} = \frac{q_3}{q_2}; \; \ldots \; i_{qn} = \frac{q_n}{q_{n-1}}.$$

Між базисними і ланцюговими індексами існує взаємозв'язок, що дозволяє переходити від одного виду індексів до іншого.

Базисні індекси можна визначити через ланцюгові, послідовно перемноживши останні:

$$i_q = \frac{q_1}{q_0} \cdot \frac{q_2}{q_1} \cdot \frac{q_3}{q_2} = \frac{q_3}{q_0}.$$

Ланцюгові індекси визначають через базисні шляхом ділення відповідного базисного індекса до попереднього базисного індекса:

$$i_q \frac{q_2}{q_0} : \frac{q_1}{q_0} = \frac{q_2 q_0}{q_0 q_1} = \frac{q_2}{q_1}.$$

Базисні і ланцюгові індекси є індивідуальні і загальні.

Якщо порівнюваних періодів три і більше, то загальні (базисні і ланцюгові) індекси обчислюють з постійними і змінними вагами.

Якщо для всього індексованого ряду беруть ваги якогось одного періоду, то отримують базисні і ланцюгові індекси з **постійними вагами**. Якщо ваги змінюються від одного індекса до іншого, то матимемо базисні і ланцюгові індекси із **змінними вагами**.

Базисні індекси цін з **постійними вагами**:

$$I_{p1} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \qquad I_{p2} = \frac{\sum p_2 q_0}{\sum p_0 q_0}; \qquad I_{p3} = \frac{\sum p_3 q_0}{\sum p_0 q_0} \text{ і т.д.}$$

Ланцюгові індекси з **постійними вагами**:

$$I_{p1} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \qquad I_{p2} = \frac{\sum p_2 q_0}{\sum p_1 q_0}; \qquad I_{p3} = \frac{\sum p_3 q_0}{\sum p_2 q_0} \text{ і т.д.}$$

Базисні індекси із **змінними вагами**:

$$I_{p1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \qquad I_{p2} = \frac{\sum p_2 q_2}{\sum p_0 q_2}; \qquad I_{p3} = \frac{\sum p_3 q_3}{\sum p_0 q_3} \text{ і т.д.}$$

Ланцюгові індекси із **змінними вагами**:

$$I_{p1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \qquad I_{p2} = \frac{\sum p_2 q_2}{\sum p_1 q_2}; \qquad I_{p3} = \frac{\sum p_3 q_3}{\sum p_2 q_3} \text{ і т.д.}$$

Між базисними і ланцюговими індексами з постійними вагами існує співвідношення:

$$\frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_2 q_0}{\sum p_1 q_0} = \frac{\sum p_2 q_0}{\sum p_0 q_0};$$

$$\frac{\sum p_2 q_0}{\sum p_0 q_0} : \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{\sum p_2 q_0}{\sum p_1 q_0}.$$

Для ланцюгових індексів із змінними вагами такого взаємозв'язку не існує.

В статистичній практиці з постійними вагами обчислюють, в основному, індекси фізичного обсягу продукції, а із змінними вагами – індекси цін, собівартості, урожайності та інших якісних показників.

Для визначення індексів якісних показників за тривалий період обчислюють, як правило, ланцюгові індекси із змінними вагами, а для індекса фізичного обсягу ефективніше вираховувати базисні індекси з постійними вагами.

8.5. Індекси змінного, постійного складу і структурних зрушень

Для якісних показників, таких як середня ціна, собівартість, урожайність та інших по однойменній продукції, але віднесеної до різних об'єктів, обчислюють загальні індекси змінного, постійного (фіксованого) складу і структурних зрушень.

Індекси, який характеризує спільний вплив обох чинників, називається **індексом змінного складу** і визначається за формулою:

$$I_{\bar{z}} = \frac{\overline{Z}_1}{\overline{Z}_0} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0},$$

де $I_{\bar{z}}$ – загальний індекс собівартості продукції змінного складу;

z_1, z_0 – собівартість одиниці продукції в звітному і базисному періодах;

q_1, q_0 – кількість виробленої продукції в натуральному виразі в звітному і базисному періодах;

$\bar{z}_1, \bar{z}_0$ – осереднені ознаки.

На величину індекса собівартості змінного складу впливають зміни рівнів собівартості і зміни в структурі (її складі). Щоб виявити роль кожного чинника в загальній динаміці середньої, потрібно індекс змінного складу розкласти на два індекси-співмножники, кожний з яких відображає вплив тільки одного чинника.

Перший індекс, який характеризує вплив тільки індексованої величини (в якому змінюється лише собівартість), називається **індексом постійного (фіксованого) складу**. Він обчислюється за формулою:

$$I_z = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_1}{\sum q_1} = \frac{\sum z_1 q_1}{\sum z_0 q_1}.$$

Другий індекс показує, як змінюється середній рівень (середня собіварттість) тільки за рахунок зміни структури явища (структури продукції). Він називається **індексом структурних зрушень** і визначається за формулою:

$$I_{zdq} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{\sum z_0 q_1}{\bar{z}_0 \sum q_1}.$$

Ці індекси взаємозв'язані.

$$I_{\bar{z}} = I_{zdq} \cdot I_z; \qquad I_z = \frac{I_{\bar{z}}}{I_{zdq}}; \quad I_{zdq} = \frac{I_{\bar{z}}}{I_z}.$$

Розглянемо приклад. Маємо дані про середню собівартість продукції «А» на двох заводах.

Номер заводу	Базисний період			Звітний період		
	Вироблено продукції, тис.шт. (q_0)	Собівартість одиниці, грн. (z_0)	Частка продукції, (d_0)	Вироблено продукції, тис.шт. (q_1)	Собівартість одиниці, грн. (z_1)	Частка продукції, (d_1)
1	70	25	50	80	22	40
2	70	23	50	120	20	60
Разом:	140	x	100	200	x	100

Обчислимо індекс собівартості продукції змінного складу:

$$I_{\bar{z}} = \frac{\bar{Z}_1}{\bar{Z}_0} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{22 \cdot 80 + 20 \cdot 120}{200} : \frac{25 \cdot 70 + 23 \cdot 70}{140} =$$

$$= \frac{4160}{200} : \frac{3360}{140} = \frac{20{,}8}{24{,}0} = 0{,}867, \text{ } або \text{ } 86{,}7\%.$$

Отже, середня собівартість знизилась на 13,3 % (100,0 – 86,7 %).

Економія на одиницю продукції становить – 3,2 грн. (20,8 – 24,0), а на весь обсяг продукції звітного періоду – [-640 тис. грн. ($-3{,}2 \cdot 200$)].

Зниження середньої собівартості одиниці продукції зумовлене зміною собівартості продукції на кожному заводі і зміною структури продукції. Обчислимо індекси собівартості:

а) постійного (фіксованого) складу:

$$I_{\bar{z}} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_1}{\sum q_1} = \frac{22 \cdot 80 + 20 \cdot 120}{200} : \frac{25 \cdot 80 + 23 \cdot 120}{200} = \frac{4160}{200} : \frac{4760}{200} =$$

$$= \frac{20,8}{23,8} = 0,874, \qquad \textit{або} \qquad 87,4\,\%.$$

Собівартість продукції по двох заводах разом в середньому знизилась на 12,6 % (100,0 – 87,4). Економія затрат на виробництво продукції в звітному періоді складає 600 тис. грн. [(20,8 – 23,8)$\cdot$ 200];

б) структурних зрушень:

$$I_{zdq} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{23,8}{24,0} = 0,992, \qquad \textit{або} \qquad 99,2\,\%.$$

Це означає, що середня собівартість вибору «А» в звітному періоді додатково знизилась на 0,8 % (100,0 – 99,2) за рахунок зміни структури, тобто за рахунок збільшення частки продукції другого заводу з 50 % до 60 %, на якому рівень собівартості був дещо нижчим в порівнянні з першим заводом. За рахунок цієї зміни економія затрат виробництва досягла в звітному періоді 40 тис. грн. [(23,8 – 24,0)$\cdot$ 200].

Проведемо перевірку наших розрахунків:

$$I_{\bar{z}} = I_z \cdot I_{zdq} = 0,867 = 0,874 \cdot 0,992;$$

$$\Delta_{zq} = \Delta_{zq}(z) + \Delta_{zq}(dq) = 640 = 600 + 40.$$

Отже, всі індекси обчислені правильно.

Вираховані нами індекси можна обчислити іншим способом, за частками продукції заводів, вираженими в коефіцієнтах:

а) Індекс собівартості змінного складу:

$$I_{\bar{z}} = \frac{\sum z_1 d_1}{\sum z_0 d_0} = \frac{22 \cdot 0,4 + 20 \cdot 0,6}{25 \cdot 0,5 + 23 \cdot 0,5} = \frac{20,8}{24,0} = 0,867, \qquad \textit{або} \qquad 86,7\,\%;$$

де d_1, d_0 – коефіцієнти частки продукції заводів в звітному і базисному періодах;

б) Індекс собівартості продукції постійного складу:

$$I_z = \frac{\sum z_1 d_1}{\sum z_0 d_1} = \frac{22 \cdot 0,4 + 20 \cdot 0,6}{25 \cdot 0,4 + 23 \cdot 0,6} = \frac{20,8}{23,8} = 0,874, \qquad \textit{або} \qquad 87,4\,\%;$$

в) Індекс струтурних зрушень:

$$I_{zdq} = \frac{\sum z_0 d_1}{\sum z_0 d_0} = \frac{25 \cdot 0{,}4 + 23 \cdot 0{,}6}{25 \cdot 0{,}5 + 23 \cdot 0{,}5} = \frac{23{,}8}{24{,}0} = 0{,}992, \quad \textit{або} \quad 99{,}2\,\%,$$

$$\textit{або} \quad I_{zdq} = \frac{I_{\bar{z}}}{I_z} = \frac{0{,}867}{0{,}874} = 0{,}992.$$

8.6. Територіальні індекси

В практиці статистичних досліджень часто виникає потреба зіставлення рівнів економічних явищ в просторі, для чого використовують територіальні індекси.

Територіальні індекси – це узагальнюючі відносні величини, що дають порівняльну характеристику в розрізі територій або об’єктів.

При побудові територіальних індексів якісних показників вагами можуть вступати:

а) кількісний (екстенсивний) показник тієї території, на якій якісний (інтенсивний) показник найбільш економічно кращий;

б) кількісний показник однієї з двох порівнюваних територій (об'єктів);

в) середній кількісний показник з багатьох порівнюваних територій (об'єктів);

г) об'ємний кількісний показник (сума екстенсивних показників декількох територій або об'єктів);

д) кількісний показник, прийнятий за стандарт.

Проілюструємо обчислення територіальних індексів цими способами на прикладі:

Культура	Середня урожайність, ц/га (y)		Посівна площа, га ($П$)			
	по району «А»	по району «В»	по району «А»	по району «В»	по області	
					в га	в %
Пшениця	36	42	190	210	2850	50
Жито	19	23	80	100	1200	21
Ячмінь	25	30	110	150	1650	29

Обчислимо територіальні індекси урожайності зернових:

а) з вагами району «В»:

$$I_y = \frac{\sum У_A П_B}{\sum У_B П_B} = \frac{36 \cdot 210 + 19 \cdot 100 + 25 \cdot 150}{42 \cdot 210 + 23 \cdot 100 + 30 \cdot 150} = \frac{13210}{15620} = 0{,}846;$$

б) з вагами району «А»:

$$I_y = \frac{\sum У_A П_A}{\sum У_B П_A} = \frac{36 \cdot 190 + 19 \cdot 80 + 25 \cdot 110}{42 \cdot 190 + 23 \cdot 80 + 30 \cdot 110} = \frac{11110}{13120} = 0{,}847;$$

в) з вагами бази порівняння:

$$I_y = \frac{\sum У_B П_A}{\sum У_A П_A} = \frac{42 \cdot 190 + 23 \cdot 80 + 30 \cdot 110}{36 \cdot 190 + 19 \cdot 80 + 25 \cdot 110} = \frac{13120}{11110} = 0{,}181;$$

г) з вагами району «В»:

$$I_y = \frac{\sum У_B П_B}{\sum У_A П_B} = \frac{42 \cdot 210 + 23 \cdot 100 + 30 \cdot 150}{36 \cdot 210 + 19 \cdot 100 + 25 \cdot 150} = \frac{15620}{13210} = 0{,}182.$$

Визначимо територіальні індекси урожайності зернових з обємними вагами:

а) для району «А» порівняно з районом «В»:

$$I_y = \frac{\sum У_A (П_A + П_B)}{\sum У_B (П_A + П_B)} = \frac{36 \cdot (190 + 210) + 19 \cdot (80 + 100) + 25 \cdot (110 + 150)}{42 \cdot (190 + 210) + 23 \cdot (80 + 100) + 30 \cdot (110 + 150)} = 0{,}846;$$ б) для району «В» порівняно з районом «А»:

$$I_y = \frac{\sum У_B (П_A + П_B)}{\sum У_A (П_A + П_B)} = \frac{28740}{24320} = 1{,}182.$$

Аналогічні результати отримаємо, використавши в якості ваг середні посівні площі для окремих культур зернових:

а) для району «А» порівняно з районом «В»:

$$I_y = \frac{\sum У_A \cdot \overline{П}_{(A+B)}}{\sum У_B \cdot \overline{П}_{(A+B)}} = \frac{36 \cdot 200{,}8 + 19 \cdot 90{,}9 + 25 \cdot 131{,}8}{42 \cdot 200{,}8 + 23 \cdot 90{,}9 + 30 \cdot 131{,}8} = \frac{12250{,}9}{14478{,}3} 0{,}846;$$

а) для району «В» порівняно з районом «А»:

$$I_y = \frac{\sum У_B \cdot \overline{П}_{(A+B)}}{\sum У_A \cdot \overline{П}_{(A+B)}} = \frac{14478{,}3}{12250{,}9} 1{,}182.$$

Розрахуємо територіальні індекси урожайності зернових із стандартними вагами:

а) для району «А» порівняно з районом «В»:

$$I_y = \frac{\sum У_A \cdot П_{ст.}}{\sum У_B \cdot П_{ст.}} = \frac{36 \cdot 50 + 19 \cdot 21 + 25 \cdot 29}{42 \cdot 50 + 23 \cdot 21 + 30 \cdot 29} = \frac{2924}{3453} = 0{,}847;$$

а) для району «В» порівняно з районом «А»:

$$I_y = \frac{\sum У_B \cdot П_{ст.}}{\sum У_A \cdot П_{ст.}} = \frac{3453}{2924} = 1{,}181.$$

Отже, урожайність зернових в районі «А» нижча ніж у районі «В» на 15,3 % (100,0 – 84,7), а в районі «В» порівняно з районом «А» вона вища в 1,181 рази, або на 18,1 %.

Будуючи територіальні індекси для кількісних показників, вагами виступає середній якісний показник по території, для якої здійснюється порівняння, або використовуються стандартні ваги.

Наведемо приклад. Маємо дані про реалізацію овочів на ринках двох районів.

Товари	Район «А»		Район «В»	
	ціна за 1 кг, грн.	продано, т	ціна за 1 кг, грн.	продано, т
Картопля	0,20	700	0,25	900
Капуста	0,25	300	0,20	500
Помідори	0,40	100	0,35	200

Визначимо середні ціни на овочі:

$$\overline{P_1} = \frac{\sum pq}{\sum q} = \frac{0{,}20 \cdot 700 + 0{,}25 \cdot 900}{700 + 900} = \frac{365}{1600} = 0{,}23\, грн.;$$

$$\overline{P_2} = \frac{\sum pq}{\sum q} = \frac{0{,}25 \cdot 300 + 0{,}20 \cdot 500}{300 + 500} = \frac{175}{800} = 0{,}22\, грн.;$$

$$\overline{P_3} = \frac{\sum pq}{\sum q} = \frac{0{,}40 \cdot 100 + 0{,}35 \cdot 200}{100 + 200} = \frac{110}{300} = 0{,}37\, грн.$$

Обчислимо територіальні індекси реалізації овочів:

а) для району «А» порівняно з районом «В»:

$$I_q = \frac{\sum q_A \cdot \overline{p}}{\sum q_B \cdot \overline{p}} = \frac{700 \cdot 0{,}23 + 300 \cdot 0{,}22 + 100 \cdot 0{,}37}{900 \cdot 0{,}23 + 500 \cdot 0{,}22 + 200 \cdot 0{,}37} = \frac{264}{391} = 0{,}675;$$

б) для району «В» порівняно з районом «А»:

$$I_q = \frac{\sum q_B \cdot \overline{p}}{\sum q_A \cdot \overline{p}} = \frac{900 \cdot 0{,}23 + 500 \cdot 0{,}22 + 200 \cdot 0{,}37}{700 \cdot 0{,}23 + 300 \cdot 0{,}22 + 100 \cdot 0{,}37} = \frac{391}{264} = 1{,}481.$$

Отже, кількість проданих овочів в районі «А» нижча, ніж в районі «В» на 32,5 % (100,0 – 67,5), відповідно в районі «В» реалізація вища за район «А» в 1,481 рази, або на 48,1 % (148,1 – 100,0).

8.7. Використання системи взаємозв'язаних індексів в аналізі чинників динаміки

Соціально-економічні явища і процеси взаємозв'язані між собою, що виражається у взаємозв'язку відповідних показників. Одна з форм взаємозв'язку між економічними показниками заключається в тому, що їх можна виразити як добуток кількох інших показників. Так, фонд заробітної плати може бути поданий у вигляді добутку заробітної плати одного працівника на загальне число працівників.

$$F = f \cdot T,$$

де F – фонд заробітної плати;

T – число працівників;

f – заробітна плата одного працівника $\left(f = \frac{F}{T} \right)$.

Товарооборот у фактичних цінах можна виразити як добуто цін на кількість проданих товарів: $p \cdot q = pq$;

валовий збір любої культури – як добуток врожайності на посівну площу: $У \cdot П = УП$, *і т.д.*

Показники-співмножники виступають тут як чинники, від величини яких залежить результативна ознака. При економічному аналізі динаміки потрібно виявити і оцінити роль кожного окремого чинника в змінні результативного показника.

Індекс результативного показника завжди виступає як добуток індекса якісного показника на індекс об'ємного показника.

Так, наприклад, загальний індекс товарообороту може бути виражений як добуток індекса цін на індекс фізичного обсягу продукції:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum q_1 p_0}{\sum q_0 p_0} = I_p \cdot I_q.$$

Аналогічно можна виразити взаємозв'язок між індексами:

а) собівартості одиниці продукції, обсягом продукції і затратами на її виробництво:

$$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0} = \frac{\sum z_1 q_1}{\sum z_0 q_1} \cdot \frac{\sum q_1 z_0}{\sum q_0 z_0} = I_z \cdot I_q;$$

б) агрегатним індексом продуктивності праці, затрат часу і фізичним обсягом продукції:

$$I_{\bar{w}} \cdot I_T = I_q = \left(\frac{\sum q_1 p_0}{\sum T_1} : \frac{\sum q_0 p_0}{\sum T_0} \right) \cdot \frac{\sum T_1}{\sum T_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0};$$

в) урожайності, посівної площі і валового збору:

$$I_{\bar{y}} \cdot I_{\Pi} = I_{y\Pi} = \left(\frac{\sum y_1 \Pi_1}{\sum \Pi_1} : \frac{\sum y_0 \Pi_0}{\sum \Pi_0} \right) \cdot \frac{\sum \Pi_1}{\sum \Pi_0} = \frac{\sum y_1 \Pi_1}{\sum y_0 \Pi_0}, i\ т.д.$$

Статистика взаємозв'язаних індексів використовується для виявлення і оцінки ролі кожного із взаємозв'язаних чинників в абсолютному вираженні.

Розглянемо приклад. Нехай маємо дані про реалізацію продуктових товарів на ринку в базисному і звітному періодах.

Назва продуктів	Базисний період			Звітний період			Вартість проданого товару у звітному році за цінами базисного року, тис.грн. ($p_0 q_1$)
	ціна одиниці продукції, грн. (p_0)	кількість проданого товару, тис.од. (q_0)	вартість проданого товару, тис.грн. ($p_0 q_0$)	ціна одиниці продукції, грн. (p_1)	кількість проданого товару, тис.од. (q_1)	вартість проданого товару, тис.грн. ($p_1 q_1$)	
Картопля, кг	0,35	45	15,75	0,30	53	15,90	18,55
Молоко, л	0,45	13	5,85	0,40	20	8,00	9,00
М'ясо, кг	4,50	10	45,00	4,00	14	56,00	63,00
Разом	x	x	66,60	x	x	79,90	90,55

Визначимо індекс товарообороту у фактичних цінах:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{79{,}90}{66{,}60} = 1{,}2.$$

Різниця між чисельником і знаменником даного індекса дасть нам абсолютний приріст товарообороту в звітному році порівняно з базисним за рахунок зміни двох чинників – цін одиниці товару і обсягу проданих товарів кожного виду:

$$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0 = 79{,}90 - 66{,}60 = 13{,}3\ тис.\ грн.$$

Загальний абсолютний приріст товарообороту розкладемо за чинниками у відносних і абсолютних показниках за рахунок зміни:

а) цін: $I_p = \dfrac{\sum p_1 q_1}{\sum p_0 q_1} = \dfrac{79{,}90}{90{,}55} = 0{,}882;$

$\Delta pq(p) = \sum p_1 q_1 - \sum p_0 q_1 = 79{,}90 - 90{,}55 = -10{,}65\,\textit{тисгрн.};$

б) фізичного обсягу продукції: $I_q = \dfrac{\sum q_1 p_0}{\sum q_0 p_0} = \dfrac{90{,}55}{66{,}60} = 1{,}360;$

$\Delta pq(q) = \sum q_1 p_0 - \sum q_0 p_0 = 90{,}55 - 66{,}60 = 23{,}95\,\textit{тисгрн.}$

Перевіримо через взаємозв'язок індексів проведені розрахунки:

$$I_{pq} = I_p \cdot I_q = 0{,}882 \cdot 1{,}360 = 1{,}200;$$

$$\Delta pq = \Delta pq(p) + \Delta pq(q) = -10{,}65 + 23{,}95 = 13{,}3\,\textit{тис.грн.}$$

Таким чином, в цілому по всіх видах товарів загальний обсяг товарообороту за рахунок зниження цін на 11,8 % (100,0 – 88,2) зменшився на 10,65 тис. грн., а за рахунок збільшення обсягу реалізації віх видів товарів в 1,36 рази, або на 36 % (136,0 – 100,0) він зріс на 23,95 тис. грн.

На практиці зустрічається дуже багато результативних показників, які залежать від трьох і більше чинників. Наприклад. Зміна чистої продукції залежить від таких чинників: а) відпрацьованого часу, б) продуктивності праці, в) частки чистої продукції у валовій продукції; вартість матеріальних затрат, зв'язаних з виробництвом любої продукції на підприємстві, залежить від: а) кількість виробленої продукції, б) питомих витрат сировини, в) цін на цю сировину і т.п.

ЛЕКЦІЯ 9. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

ПЛАН

9.1. Поняття про вибіркове спостереження та його основні завдання.

9.2. Основні умови наукової організації вибіркового спостереження.

9.3. Методи і способи відбору одиниць у вибіркову сукупність.

9.4. Знаходження середньої і граничної помилок та необхідної чисельності для різних видів вибірок.

9.5. Способи поширення даних вибіркового спостереження на генеральну сукупність.

9.1. Поняття про вибіркове спостереження та його основні завдання

Із всіх видів несуцільного спостереження в практиці статистичних досліджень найбільше визначення і застосування отримало вибіркове спостереження

Вибірковим спостереженням називається такий вид несуцільного спостереження, за характеристикою відібраної частини одиниць якого судять про всю сукупність.

Розрізняють генеральну і вибіркову сукупності:

Генеральною сукупністю називається така маса одиниць, з якої проводиться відбір для дослідження.

Вибірковою сукупністю називається частина генеральної сукупності відібрана для обстеження.

Обсяг генеральної сукупності позначають через «N», а вибіркової – через «n».

Узагальнюючими показниками генеральної сукупності є: середній розмір ознаки «$\bar{x}$», частка «p», генеральна дисперсія «σ^2».

Узагальнюючими показниками вибіркової сукупності є: середня вибіркова «$\tilde{x}$», вибіркова частка «w», дисперсія «$\sigma_в^2$».

До вибіркового спостереження статистика вдається у випадках, коли потрібно зекономити сили і засоби при проведенні дослідження, тобто, коли недоцільно або неможливо проводити суцільне спостереження.

Вибіркове спостереження застосовують також у поєднанні із суцільним для поглиблення дослідження, або уточнення і контролю результатів суцільного спостереження.

Вибіркове спостереження складається з таких етапів:

1) постановка мети спостереження;

2) складання програми спостереження і розробка відповідних даних;

3) вирішення організаційних питань проведення спостереження;

4) визначення відсотка і способу відбору одиниць;

5) проведення відбору;

6) реєстрація відповідних ознак у відібраних для дослідження одиниць;

7) узагальнення даних спостереження та розрахунок їх вибіркових характеристик;

8) знаходження помилок вибірки;

9) перерахунок характеристик вибіркового спостереження на генеральну сукупність.

Вибіркове спостереження проводиться для вирішення наступних основних завдань:

1) визначення середнього розміру досліджуваної ознаки;

2) визначення питомої ваги (частки);

3) визначення середньої і граничної помилки вибірки;

4) знаходження меж для середньої і частки при повторному і безповторному відборі;

5) визначення потрібної чисельності вибірки;

6) поширення даних вибіркового спостереження на всю сукупність.

9.2. Основні умови наукової організації вибіркового спостереження

Науковим обґрунтуванням можливості застосування вибіркового спостереження виступає діалектична єдність одиничного, особливого і загального, згідно якої в кожному одиничному є риси особливого і загального, а загальне володіє рисами одиничного і особливого. Ця єдність дозволяє за одиничним і особливим судити про загальне, за частиною – про ціле, якщо правильно встановлений зв'язок між ними.

Особливістю вибіркового спостереження в порівнянні з іншими видами несуцільного спостереження є те, що при відборі одиниць у вибіркову сукупність забезпечується рівна можливість попадання кожної одиниці у вибірку. Це досягається шляхом неупередженого строгого випадкового відбору за схемою, розробленою математичною статистикою.

Відповідь на питання про те, яка за розміром різниця між генеральними і вибірковими узагальнюючими показниками, з якою ймовірністю можна судити про цю різницю, дає теорія вибіркового методу, на основі закону великих чисел.

За допомогою цього закону розв'язують два взаємозв'язаних завдання: 1) розраховують, із заданою ймовірністю, межі можливих відхилень вибіркового від відповідного показника в генеральній сукупності; 2) визначають ймовірність того, що розмір можливих відхилень вибіркового показника від генерального не перевищить встановленої межі.

При масовому спостереженні, розподіл емпіричних частот більшості явищ підпорядковується закону нормального розподілу.

Доведено, що за нормальним розподілом більша частина величин зосереджена навколо генеральної середньої. Біля 68,3 % чисельність вибіркової середньої буде знаходитись в межах $\pm\sigma$ генеральної середньої; 95,4 % цієї чисельності знаходиться в межах $\pm 2\sigma$ і 99,7 % – не вийде за межі $\pm 3\sigma$. Нормальний розподіл показує частоту виникнення помилок даного розміру середньої.

Принцип строгої випадковості, який покладений в основу вибірки, забезпечує його об'єктивність, дозволяє встановити межі можливих помилок і отримати практично достовірні дані для характеристики всієї сукупності явищ. Така вибіркова сукупність називається **представницькою** або **репрезентативною.** В її склад входять представники всіх груп, з яких складається генеральна сукупність.

Точність результатів вибіркового спостереження, в кінцевому підсумку, буде залежати від способу відбору одиниць, ступеня коливання ознаки в сукупності та від числа одиниць, що їх спостерігатимуть.

9.3. Методи і способи відбору одиниць у вибіркову сукупність

Способом відбору називається система організації відбору одиниць з генеральної сукупності.

Розрізняють два методи відбору одиниць у вибіркову сукупність: повторний і безповторний.

Повторним називається такий метод відбору, при якому кожна раніше відібрана одиниця повертається в генеральну сукупність і може знову брати участь у вибірці.

Безповторним називається такий метод відбору, при якому кожна раніше відібрана одиниця не повертається в генеральну сукупність і в подальшій вибірці участі не бере.

Оскільки безповторний відбір охоплює постійно нові одиниці сукупності, а повторний – одну і ту ж сукупність, тому безповторний відбір дає більш точні результати.

Повторний і безповторний методи відбору, в залежності від характеру одиниці відбору, застосовується в поєднанні з іншими видами відбору. В практиці статистичного дослідження використовуються три види відбору:

1) індивідуальний – відбір окремих одиниць сукупності;

2) груповий (серійний) – відбір груп (серій) одиниць;

3) комбінований – комбінація індивідуального і групового.

За способом відбору одиниць для обстеження розрізняють такі види вибіркового спостереження:

1) власне випадкова вибірка;

2) механічна вибірка;

3) типова (районована) вибірка;

4) серійна (гніздова) вибірка;

5) комбінована вибірка;

6) одноступінчаста і багатоступінчаста вибірка;

7) однофазна і багатофазна вибірка;

8) інші види вибірки.

Власне випадковою називається така вибірка, при якій відбір одиниць з генеральної сукупності є випадковим. Часто для цього застосовують жеребкування або таблицю випадкових чисел.

Механічна вибірка – це послідовний відбір одиниць через рівні проміжки в порядку визначеного розположення їх в генеральній сукупності, або в якому-небудь переліку. Інтервали відбору визначаються у відповідності з часткою відбору одиниць (кожна п’ята, десята, сота і т.д.).

При типовому відборі генеральну сукупність поділяють на однорідні групи за певною ознакою, райони, зони. Потім з кожної групи випадковим або механічним способом відбирають певну кількість одиниць, пропорційно частці групи в загальній сукупності.

При серійній (гніздовій) вибірці відбір одиниць проводять цілими групами (серіями, гніздами) сукупності в межах яких обстежують всі одиниці без винятку. Серії для спостереження відбирають випадково, частіше безповторним способом механічної вибірки.

Комбінованою називається така вибірка, коли комбінують два або кілька видів вибірок. Перш за все, комбінують суцільне і вибіркове спостереження. В даному випадку, за основною програмою обстежується генеральна сукупність, а за додатковою – вибіркова.

Одноступінчастою називається вибірка, коли із досліджуваної сукупності зразу відбираються одиниці або серії одиниць для безпосереднього обстеження.

Багатоступінчаста вибірка передбачає поступове вилучення із генеральної сукупності спочатку укрупнення груп одиниць, потім груп менших за обсягом, і так до тих пір, поки не відберуть відповідні групи або одиниці, які будуть досліджуватись. Вибірка може бути двох-, трьох і більше ступінчастою.

Якщо необхідні дані можна отримати на основі вивчення всіх первинно відібраних одиниць, застосовують **однофазну вибірку**, а якщо тільки на основі деякої її частини, відібраної так, що вона складає підвибірку із початково проведеної вибірки – **багатофазну**.

Багатофазною називається така вибірка, коли одні відомості збираються від всіх одиниць відбору, потім відбираються ще деякі одиниці і обстежуються за більш широкою програмою. При багатофазній вибірці на кожній фазі зберігається одна і таж одиниця відбору.

Розрахунок помилок репрезентативності багатоступінчастої і багатофазної вибірок проводиться для кожної ступені і фази окремо.

Бувають випадки, коли необхідно застосувати інші види відбору, такі як взаємопроникаючі і квантильні вибірки, направлений відбір, моментні спостереження, або скористатись малою вибіркою.

Взаємопроникаючою називається така вибірка, коли із однієї генеральної сукупності проводять одним і тим же способом декілька незалежних вибірок.

Взаємопроникаючі вибірки завжди проводять різні, незалежні один від одного дослідники, що дозволяє порівнювати підсумки по всіх частинах і забезпечити взаємну перевірку їх роботи. Взаємопроникаючі вибірки дають незалежні одна від одної оцінки значень досліджуваної сукупності, і, якщо результати різних вибірок близькі між собою, то такі оцінки дуже переконливі.

Помилки взаємопроникаючих вибірок визначаються за формулами типової пропорційної вибірки.

Квантильні вибірки застосовують тоді, коли виникає потреба дослідження даних суцільного спостереження за додатковою програмою.

Для проведення квантильної вибірки рангують потрібну варіаційну ознаку і за її нагромадженими частотами будують огіву. За огівою механічним способом відбирають потрібну частину одиниць для дослідження цієї ж ознаки. Якщо огіва вибіркової сукупності добре відтворює огіву генеральної сукупності, то помилка репрезентативності буде мінімальною.

Направлений відбір використовують тоді, коли за відомим середнім значенням ознаки в генеральній сукупності вибіркова сукупність повинна характеризувати її структуру за іншими ознаками.

Направлений відбір передбачає проведення відбору таким чином, щоб середній розмір відібраних одиниць дорівнював середньому розміру одиниць всієї сукупності. В тому випадку, коли заміна однієї одиниці іншою призводить до наближеної рівності середніх генеральної і вибіркової сукупностей, вибірку вважають врівноваженою і репрезентативною за всіма іншими ознаками сукупності. Таким чином, **направленим відбором** називається врівноваження за однією ознакою для вибіркового дослідження інших ознак.

Помилку вибірки направленого відбору визначають в залежності від способу проведення відбору одиниць до врівноваження.

Моментне спостереження використовується для вивчення використання робочого часу робітниками або часу роботи устаткування. В кожний момент спостереження фіксують, чи знаходився робітник або верстат в роботі, а якщо ні , то з яких причин. Вибіркове моментне спостереження вважають через те, що охоплює не весь час роботи цеху, а лише визначені моменти часу.

Малою вибіркою називається вибіркова сукупність, яка складається з порівняно невеликої кількості одиниць (20-30). На практиці іноді доводиться обмежуватись малою кількістю спостережень (при перевірці якості продукції, зв'язаної із знищенням продукції, яку перевіряють). Математичною статистикою доведено, що і при малих вибірках характеристики вибіркової сукупності можна поширити на генеральну.

При малих вибірках дисперсію обчислюють з врахуванням кількості ступенів вільності варіації.

Англійський вчений Стьюдент винайшов закон розподілу відхилень вибіркових середніх від генеральної середньої для малих вибірок. Опираючись на цей закон, він склав спеціальні таблиці, в яких наводяться значення критерію «t» для малих вибірок.

9.4. Знаходження середньої і граничної помилок та необхідної чисельності для різних видів вибірок

Для вибіркового спостереження властиві помилки реєстрації і помилки репрезентативності.

Помилки репрезентативності становлять різницю між середніми і відносними показниками вибіркової сукупності та відповідними показниками генеральної сукупності. Вони поділяються на систематичні та випадкові.

Систематичні помилки репрезентативності зумовлені внаслідок порушення принципів проведення вибіркового спостереження.

Випадкові помилки репрезентативності зумовлені тим, що вибіркова сукупність не відображає точно середні і відносні показники генеральної сукупності.

Визначення величини випадкових помилок репрезентативності є одним з головних завдань теорій вибіркового методу.

Для узагальнюючої характеристики помилки вибірки вираховують середню помилку репрезентативності, яку позначають через грецьку букву «мю» (μ) і називають ще стандартом.

Для визначення середньої помилки репрезентативності власне випадкової і механічної вибірки застосовують чотири формули для повторного і без повторного відбору.

Спосіб відбору	При визначенні середньої	При визначенні частки
Повторний	$\mu = \sqrt{\frac{\sigma^2}{n}}$	$\mu = \sqrt{\frac{w(1-w)}{n}}$
Безповторний	$\mu = \sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$	$\mu = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$

де μ – середня помилка репрезентативності;

σ^2 – середній квадрат відхилень у вибірці;

n – чисельність вибіркової сукупності;

N – чисельність генеральної сукупності;

$\left(1-\frac{n}{N}\right)$ – необстежена частка генеральної сукупності;

$\frac{n}{N}$ – частка обстеженої частини вибіркової сукупності;

w – частка даної ознаки у вибірці;

$(1-w)$ – частка протилежної ознаки у вибірці.

На практиці частіше використовують безповторний відбір, який гарантує більш точні результати.

Для узагальнюючої характеристики помилки вибірки поряд із середньою розраховують ще і граничну помилку вибірки.

При вибірковому спостереженні розмір граничної помилки репрезентативності «Δ» може бути більший, дорівнювати або менший від середньої помилки репрезентативності «μ». Тому величину граничної помилки репрезентативності обчислюють з певною ймовірністю «р», якій відповідає t-разове значення «μ». З введенням показника кратності помилки «t», формула граничної помилки репрезентативності матиме вигляд: $\Delta = t\mu$; $t = \frac{\Delta}{\mu}$,

де μ – середня помилка вибірки;

t – коефіцієнт довір'я, який залежить від ймовірності визначення граничної помилки.

Ймовірність відхилень вибіркової середньої від генеральної середньої при достатньо великому обсязі вибірки і обмеженій дисперсії генеральної сукупності підпорядковується закону нормального розподілу. Ймовірність цих відхилень при різних значеннях «t» визначається за формулою:

$$F(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt.$$

Значення цього інтеграла при різних значеннях «t» табульовані і приводяться в спеціальних таблицях, наприклад:

для $t = 1$ $p(\Delta \le \mu) = 0{,}683$; для $t = 2$ $p(\Delta \le \mu) = 0{,}954$;

для $t = 3$ $p(\Delta \le \mu) = 0{,}997$; для $t = 4$ $p(\Delta \le \mu) = 0{,}999$.

Гранична помилка вибірки дає можливість встановити, в яких межах знаходиться величина генеральної середньої або частки.

Із теореми Чебишева знаходять, що:

$$\bar{x} - \tilde{x} = \pm\Delta_x \quad i \qquad \tilde{x} - \Delta_x \le \bar{x} \le \tilde{x} + \Delta_x.$$

Додаючи граничну помилку вибірки до вибіркової частки і віднімаючи її від неї, знаходять межі генеральної частки:

$$p - w = \pm\Delta_p \quad i \qquad w - \Delta_p \le p \le w + \Delta_p.$$

На основі формул граничної помилки вибірки розв'язують наступні завдання:

1) визначають довірчі межі генеральної середньої і частки з прийнятою ймовірністю;

2) визначають ймовірність того, що відхилення між вибірковими і генеральними характеристиками не перевищать визначену величину;

3) визначають необхідну чисельність вибірки, яка із заданою ймовірністю забезпечить прийняту точність вибіркових показників.

Розглянемо приклад. При 2% власне випадковому відборі у відібраних для обстеження 100 деталей встановлено, що середня вага однієї деталі 2500 г; дисперсія 900, із 100 деталей 10 виявились бракованими. З ймовірністю 0,954 встановити межі середньої ваги однієї деталі в генеральній сукупності, а з ймовірністю 0,997 – межі частки якісних деталей в генеральній сукупності.

Граничну помилку визначаємо $\Delta_x = t\mu$ за формулою безповторного відбору, так як чисельність генеральної сукупності можна знайти:

$$N = \frac{100 \cdot 100}{2} = 5000 \text{ шт.}$$

В спеціальній таблиці знаходимо, що для ймовірності 0,954 $t = 2$, а для ймовірності 0,997 $t = 3$.

Таким чином отримаємо:

$$\Delta_x = t\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} = 2\sqrt{\frac{900}{100}\left(1 - \frac{100}{5000}\right)} = 2\sqrt{9 \cdot 0{,}98} = 2 \cdot 3 = 6\, г.$$

Звідси довірчі межі генеральної середньої будуть наступні:

$$\tilde{x} - \Delta_x \le \bar{x} \le \tilde{x} + \Delta_x;$$
$$2500 - 6 \le \bar{x} \le 2500 + 6;$$
$$2494 \le \bar{x} \le 2506.$$

Тобто, з ймовірністю 0,954 можна стверджувати, що середня вага однієї деталі в генеральній сукупності знаходиться в межах від 2494 г. до 2506 г.

Аналогічно вирішується завдання і при визначенні інтервалів (меж) для генеральної частки. Вибіркова частка якісних деталей:

$$w = \frac{m}{n} = \frac{90}{100} = 0{,}9,$$

де m – кількість якісних деталей у вибірковій сукупності;

n – кількість відібраних деталей.

$$\Delta_p = t\sqrt{\frac{w - (1 - w)}{n}\left(1 - \frac{n}{N}\right)} = 3\sqrt{\frac{0{,}9(1 - 0{,}9)}{100}\left(1 - \frac{100}{5000}\right)} = 3 \cdot 0{,}03 = 0{,}09,$$

$$w - \Delta_p \le p \le w + \Delta_p,$$
$$0{,}9 - 0{,}09 \le p \le 0{,}9 + 0{,}09,$$
$$0{,}81 \le p \le 0{,}99.$$

Таким чином, з ймовірністю 0,997 можна стверджувати, що частка якісних деталей в генеральній сукупності знаходиться в межах від 81% до 99%.

При розрахунках вибіркових характеристик інколи ставиться завдання визначення ймовірності допуску певної помилки, тобто відхилення від відповідних характеристик генеральної сукупності не більше ніж на певну завдану величину, яку знаходять за формулою граничної помилки.

Використавши розрахунки попередньої задачі знаходимо для середньої:

$$t = \frac{\Delta_x}{\mu} = \frac{6}{3} = 2.$$

Для даного значення ($t = 2$) відповідає ймовірність ($p = 0{,}954$). Це дає право стверджувати, що при визначенні за вибірковими даними середньої ваги деталей ($\tilde{x} = 2500\,г$) допущена помилка, яка не перевищує 6 г.

Аналогічно розраховують ймовірність допуску помилки для частки:

$$t = \frac{\Delta_p}{\mu} = \frac{0{,}09}{0{,}03} = 3.$$

Для цього значення ($t = 3$) відповідає ймовірність ($p = 0{,}997$). Таким чином, майже достовірно можна стверджувати, що при визначенні за вибірковими даними частки якісних деталей ($w = 0{,}9$) допущена помилка, яка не перевищує 9 %.

При організації проведення вибіркового спостереження важливе значення має правильне визначення необхідної чисельності вибірки, яка з відповідною ймовірністю забезпечить встановлену точність результатів спостереження.

Чисельність вибірки залежить від наступних чинників:

1) від варіації досліджуваної ознаки. Чим більша варіація, тим більшою повинна бути чисельність вибірки, і навпаки;

2) від розміру можливої граничної помилки вибірки. Чим менший розмір можливої помилки, тим більша повинна бути чисельність вибірки. Існує правило, якщо помилку потрібно зменшити в три рази, то чисельність вибірки збільшують в дев’ять раз;

3) від розміру ймовірності, з якою гарантуватимуть результати вибірки. Чим більша ймовірність, тим більша повинна бути чисельність вибірки;

4) від способу відбору одиниць у вибіркову сукупність для обстеження.

Основні формули для знаходження необхідної чисельності вибірки для власне випадкової і механічної вибірки.

Способи відбору	Чисельність вибірки	
	при визначенні середньої	при визначенні частки
Повторний	$n=\frac{t^2\sigma^2}{\Delta_x^2}$	$n=\frac{t^2w(1-w)}{\Delta_p^2}$
Безповторний	$n=\frac{t^2\sigma^2N}{\Delta_x^2N+t^2\sigma^2}$	$n=\frac{t^2w(1-w)N}{\Delta_p^2N+t^2w(1-w)}$

Покажемо застосування цих формул на прикладі. Припустимо, що для району, в якому є 8000 корів, необхідно організувати вибіркове спостереження з метою встановлення річної удійності корів. Якою повинна бути чисельність вибірки?

Якщо орієнтуватись на повторний вибір , то при граничній помилці в 30 кг. з ймовірністю ($p=0{,}954$) і при середньому квадратичному відхиленні 300 кг., визначеному за результатами аналогічних обстежень, необхідна чисельність вибірки повинна бути:

$$n=\frac{t^2\sigma^2}{\Delta_x^2}=\frac{2^2\cdot 300^2}{30^2}=\frac{4\cdot 90000}{90}=400 \text{ корів.}$$

При безповторному відборі необхідна чисельність вибірки буде дорівнювати:

$$n=\frac{t^2\sigma^2N}{\Delta_x^2N+t^2\sigma^2}=\frac{4\cdot 90000\cdot 8000}{900\cdot 8000+4\cdot 90000}=380 \text{ корів.}$$

Цей розрахунок підтверджує, що при інших рівних умовах, обсяг вибірки при безповторному відборі завжди буде менший, ніж при повторному.

Для визначення частки породних корів з помилкою не більше 5 %, при ймовірності 0,954 і дисперсії альтернативної ознаки 0,25, знаходимо при повторному відборі: $n=\frac{t^2w(1-w)}{\Delta_p^2}=\frac{4\cdot 0{,}25}{0{,}0025}=400$ корів,

і при безповторному відборі:

$$n=\frac{t^2w(1-w)N}{\Delta_p^2N+t^2w(1-w)}=\frac{8000}{0{,}0025\cdot 8000+1}=\frac{8000}{21}=380 \text{ корів.}$$

Таким чином, для забезпечення прийнятої точності при повторному відборі необхідно обстежити 400 корів, а при безповторному відборі – 380.

Розрахунок граничної помилки вибірки при **типовому відборі** проводиться за допомогою формул:

Спосіб відбору	Гранична помилка вибірки	
	для середньої	для частки
Повторний	$\Delta_x = t\sqrt{\dfrac{\overline{\sigma^2}}{n}}$	$\Delta_x = t\sqrt{\dfrac{\overline{\sigma^2}}{n}}$
Безповторний	$\Delta_x = t\sqrt{\dfrac{\overline{\sigma^2}}{n} - \left(1 - \dfrac{n}{N}\right)}$	$\Delta_p = t\sqrt{\dfrac{\overline{w(1-w)}}{n}\left(1 - \dfrac{n}{N}\right)}$

де $\overline{\sigma_i^2}$ – середня групова дисперсія;

$\overline{w(1-w)}$ – середня дисперсія альтернативної ознаки для частки.

Необхідна чисельність типової вибірки визначається за формулою:

$$n = \frac{t^2\overline{\sigma^2}N}{\Delta_x^2 N + t^2\overline{\sigma^2}}.$$

Гранична помилка **серійної вибірки** визначається за формулами:

Спосіб відбору	Гранична помилка вибірки	
	для середньої	для частки
Повторний	$\Delta = t\sqrt{\dfrac{\delta_x^2}{s}}$	$\Delta = t\sqrt{\dfrac{\delta_p^2}{s}}$
Безповторний	$\Delta = t\sqrt{\dfrac{\delta_x^2}{s}(1 - \dfrac{s}{S})}$	$\Delta = t\sqrt{\dfrac{\delta_p^2}{s}(1 - \dfrac{s}{S})}$

де δ^2 – міжсерійна (міжгрупова) дисперсія середніх;

S – число серій в генеральній сукупності;

s – число відібраних серій.

Необхідна чисельність вибірки при серійному відборі визначається за формулами:

а) для повторного відбору: $s = \dfrac{t^2\delta^2}{\Delta_x^2}$;

б) для безповторного відбору: $s = \dfrac{t^2\delta^2 S}{\Delta_x^2 S + t^2\delta^2}$.

В статистичній практиці вибіркове спостереження із великих масивів генеральної сукупності часто проводиться у вигляді **комбінованої, ступінчастої або багатофазної вибірки.**

Загальна помилка при комбінованій вибірці складається із помилок, які можливі на кожній ступені, і визначається як корінь квадратний із квадратів помилок відповідних вибірок. Якщо серійну вибірку скомбінувати з власне випадковою або механічною, то гранична помилка вибірки визначається за формулою:

$$\Delta_x = t\sqrt{\frac{\delta^2}{s} - \left(1 - \frac{s}{S}\right) + \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}.$$

При **моментному методі спостереження** гранична помилка частки визначається як при звичайній повторній власне випадковій вибірці.

Для визначення необхідної чисельності моментів спостереження застосовується формула:

$$n = \frac{0{,}25t^2}{\Delta^2}.$$

При **малих вибірках** розподіл вибіркових середніх і помилок вибірки відрізняється від нормального. Тому для оцінки результатів малої вибірки використовують дещо видозмінені формули. Середня помилка малої вибірки розраховується за формулою:

$$\mu_{м.в.} = \sqrt{\frac{S^2_{м.в.}}{n-1}} = \frac{S_{м.в.}}{\sqrt{n-1}},$$

де $S_{м.в.} = \sqrt{\frac{\sum(x - \overline{x})^2}{n-1}}$;

n-1 – число ступенів вільності варіацій, які вказують на кількість різних можливих значень варіантів і їх середньою арифметичною.

9.5.Способи поширення даних вибіркового спостереження на генеральну сукупність

Кінцевою практичною метою всякого вибіркового спостереження є поширення його характеристик генеральну сукупність.

Існують два способи розповсюдження даних вибіркового спостереження: 1) спосіб прямого перерахування; 2) спосіб коефіцієнтів.

Спосіб прямого перерахування застосовують в тому випадку, коли на основі вибірки розраховують об'ємні показники генеральної сукупності, використовуючи для цього вибіркові середню або частку. В першому випадку середній розмір ознаки, визначений в результаті вибіркового спостереження, множиться на кількість одиниць генеральної сукупності.

Встановивши, наприклад, в результаті вибіркового спостереження продуктивності тваринництва, яке перебуває в особистій власності, що середній річний надій молока на одну корову в області становить 2500 кг., і знаючи, що всього в області в особистій власності є 20000 корів, можемо отримати величину валового надою молока: $2500 \cdot 20\,000 = 50\,000\,000$ кг. або 50 тис. т. Якщо при цьому відомо, що середня помилка вибірки з певною ймовірністю дорівнює ± 20 кг., а, отже, генеральна середня з тією ж ймовірністю коливається в межах від 2480 до 2520 кг., то загальний валовий надій молока з врахуванням помилки вибірки буде коливатися від 49,6 тис. т. до 50,4 тис. т.

Спосіб поправочних коефіцієнтів застосовується в тих випадках, коли вибіркове спостереження проводиться з метою перевірки і уточнення результатів суцільного спостереження. В даному випадку, співставляючи дані вибіркового спостереження із суцільним, вираховують поправочний коефіцієнт, який використовують для внесення поправок в матеріали суцільних спостережень. Так, по закінченні переписів худоби проводять 10%-не контрольне вибіркове обстеження худоби, яка знаходиться в особистій власності населення, і якщо в результаті контрольного обходу виявляється недооблік, тоді дані перепису коригуються на відсоток недообліку.

Припустимо, що в результаті перепису худоби, яка знаходиться в особистій власності населення району є 15000 свиней, в тому числі в населених пунктах, де проводиться 10%-на вибірка – 1200 свиней. В результаті контрольних обходів в цих же населених пунктах було обчислено 1215 свиней. Тобто, при переписі було недообліковано 15 голів свиней, що складає 1,25 % $\left(\frac{15}{1200} \cdot 100\right)$. За допомогою цього поправочного коефіцієнта уточнюють матеріали перепису суцільного спостереження. В даному випадку по району всього недообліковано 187 голів свиней $\left(\frac{1{,}25 \cdot 15000}{100}\right)$.

Отже, загальна кількість свиней, які знаходяться в особистій власності населення району з поправкою на недооблік, складає 1387 голів $(1200 + 187)$.

ЛЕКЦІЯ 10. СТАТИСТИЧНІ ГРАФІКИ

ПЛАН:

10.1. Поняття про статистичні графіки і правила їх побудови.

10.2. Графіки порівняння статистичних величин.

10.3. Наочне зображення структури і структурних зрушень.

10.4. Графічне зображення динаміки статистичних показників.

10.5. Контрольно-планові графіки.

10.6. Графіки просторового розміщення і просторового розповсюдження.

10.1. Поняття про статистичні графіки і правила їх побудови

В результаті опрацювання даних різного ряду спостережень отримують велику кількість цифрового матеріалу, який розміщують у таблицях. Застосування табличного методу значно полегшує орієнтацію в зібраному і згрупованому матеріалі. Проте в багатьох випадках статистичних досліджень не можна обмежуватись одними таблицями.

Таблична форма викладу цифрового матеріалу не завжди дозволяє достатньо наглядно і чітко відобразити загальну картину стану або розвитку якого-небудь явища, розкрити закономірності зв’язку статистичних показників між собою, або їх розподілу. А тому для розв’язку цих та інших завдань поряд із статистичними таблицями широко застосовується графічний спосіб зображення статистичних величин.

Статистичний графік – це особливий спосіб наочного зображення і узагальнення статистичних даних про соціально-економічні явища і процеси за допомогою геометричних образів, малюнків або схематичних географічних карт і пояснень до них.

Графіки застосовуються, головним чином, для характеристики (порівняння) розвитку показників в часі і просторі, вивчення структури і структурних зрушень, контролю за виконанням планових завдань, характеристики просторового розміщення і просторового розповсюдження явищ. Графіки застосовуються також для аналізу зв'язків і залежностей між різними показниками або між значеннями варіаційної ознаки і частотами чи частками.

При побудові статистичного графіка потрібно знати, з якою метою складається графік, вивчити вихідний матеріал і володіти методикою графічних зображень.

Основними елементами графіка є: поле графіка, графічні образи, масштабні орієнтири і експлікація графіка. Кожний елемент має своє призначення і виконує відповідну роль в побудові і інтерпретації графіка.

Поле графіка – це простір, на якому розміщуються геометричні та інші знаки, які створюють графік. Цей простір обмежується або аркушем чистого паперу, або географічною чи контурною картою.

Розмір поля залежить від призначення графіка. В статистичних дослідженнях найбільш часто зустрічаються графіки у вигляді прямокутників з нерівними сторонами по вертикалі і горизонталі, хоча також застосовуються графіки у вигляді квадратів. В практиці співвідношення нерівних сторін полів графіка береться від 1:1,33 до 1:1,50, якщо вертикальну сторону прийняти за 1.

Просторові орієнтири задаються у вигляді прямокутної системи координат, тобто координатної сітки. В картограмах засобами просторової орієнтації виступають географічні карти.

Графічний образ – це сукупність різноманітних геометричних та графічних знаків, за допомогою яких відображаються статистичні величини. В статистичних графіках використовуються такі геометричні знаки як, крапки, відрізки прямих ліній, квадрати, прямокутники, кола, півкола, сектори, а також негеометричні знаки – символи у вигляді силуетів або малюнків. Це і є основою графіка, його мовою.

Масштабні орієнтири статистичних графіків – це масштаб, масштабні шкали і масштабні знаки, які використовуються для визначення розмірів геометричних та інших графічних знаків.

Масштаб – це умовна міра переводу числової величини статистичного явища в графічну і навпаки. Тобто, це довжина відрізка шкали, прийнята за числову одиницю. Наприклад, 1 см. на графіку відповідає 1000 одиницям виробленої продукції, або 1 $см^2$ дорівнює 100 $км^2$ на досліджуваній території.

При побудові графіка масштаб повинен бути таким, щоб ясно і чітко проявлялися відмінності зображення статистичних величин і разом з цим їх легко можна було б порівнювати між собою. Найбільш розповсюдженою при побудові статистичних графіків є система прямокутних координат. При цьому найкраще співвідношення масштабу на осі абсцис і ординат 1,41:1, яке відоме під назвою «золотого перетину». На осі ординат графіка повинна бути нульова точка. У випадках, коли мінімальне значення ознаки значно вище нуля, доцільно робити розрив вертикальної шкали.

Масштабна шкала – це лінія, поділена на відрізки точками відповідно до прийнятого масштабу. Носієм шкали можуть виступати пряма або крива лінії. Залежно від цього масштабні шкали поділяють на прямолінійні і кругові.

Довжину відрізків між сусідніми поділками шкали називають графічним інтервалом, а різницю між числовими значеннями цих поділок – числовим інтервалом. Обидва інтервали можуть бути рівними і нерівними.

Шкалу, в якій рівним графічним інтервалом відповідають рівні числові інтервали називають рівномірною, або арифметичною.

Якщо рівним графічним інтервалом відповідають нерівні числові інтервали, шкалу називають нерівномірною, або функціональною. Для побудови статистичних графіків з функціональною шкалою найчастіше застосовують логарифмічну функцію « $y = \ell qx$ ».

Масштабні знаки – це еталони, які зображають на графіку статистичні величини у вигляді квадратів, кругів, силуетів тощо. Ними користуються для визначення розмірів і співвідношень статистичних величин, зображених на графіку, тобто для порівняння графічних знаків із знаком-еталоном.

Експлікація графіка – це словесні пояснення, які розкривають його зміст і основні елементи: заголовок графіка, одиниці виміру, умовні позначення.

Загальний заголовок повинен ясно, чітко і коротко розкрити основний його зміст і відповісти на три питання – що, де, коли ?

На кожній масштабній шкалі графіка вказуються розміщені на них статистичні величини і одиниці їх вимірювання.

Пояснювальні надписи до окремих елементів графічного образу можуть знаходитись на полі графіка, або у формі умовних позначень виноситись за його межі.

Класифікація графіків дає можливість визначити їх загальні риси, аналітичні можливості та техніку побудови. Графіки класифікуються за функціонально-цільовим призначенням, виданим, формами і типами основних елементів.

За загальним призначенням графіки ділять на аналітичні, ілюстративні та інформаційні.

За функціонально-цільовим призначенням розрізняють графіки групувань і рядів розподілу, динаміки, взаємозв'язку і порівняння.

За формою графічних образів графіки поділяють на крапкові, лінійні, площинні, просторові і фігурні.

За типом системи координат розрізняють графіки у прямокутній і полярній системі координат, а за масштабними шкалами – графіки з рівномірними, функціональними і змішаними шкалами.

Класифікація графіків за видом їх поля дає змогу виділити дві великі групи графіків: а) діаграми; б) статистичні карти.

З точки зору розв'язуваних завдань статистичні графіки поділяють на: 1) графіки порівняння статистичних величин; 2) графіки структури і структурних зрушень; 3) графіки зображення динаміки статистичних показників; 4) графіки контролю виконання плану; 5) графіки просторового розміщення і розповсюдження; 6) графіки варіаційних рядів; 7) графіки взаємозв'язку і взаємозалежності.

Графіки, які застосовуються для зображення статистичних даних надзвичайно різноманітні.

10.2. Графіки порівняння статистичних величин

В статистичній практиці для графічного порівняння величин статистичного показника, які характеризують його зміну в просторі , застосовують діаграми.

Діаграми – це такий вид графіків, в якому цифрові дані зображаються з допомогою різних геометричних фігур і ліній. Діаграми є стовпчикові, стрічкові, секторні, лінійні та інші.

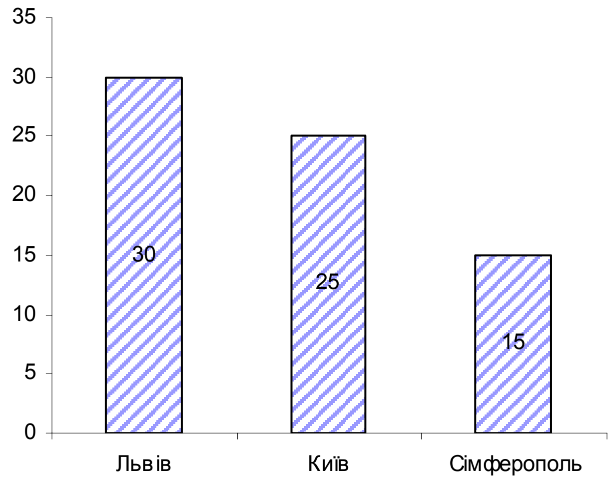
Стовпчикові діаграми являють собою найбільш простий, наочний і широко розповсюджений вид графіків в одному вимірі. В них статистичні дані зображають у вигляді стовпчиків-прямокутників однакової ширини розміщених вертикально на осі абсцис і однакової або різної висоти. Кожний окремий стовпчик характеризує окремий об'єкт. Загальне число стовпчиків дорівнює числу порівнюваних об'єктів. Віддаль між стовпчиками береться однакова, ще інколи стовпчики розташовують упритул один до одного.

Покажемо побудову стовпчикової діаграми на прикладі.

Виробництво телевізорів кольорового зображення заводами України в 2006р.

Заводи	Вироблено телевізорів, тис. шт.	Середня ціна одного телевізора, грн.	Вартість вироблених телевізорів, тис. грн.
Львівський	30	650	19500
Київський	25	450	11250
Сімферопольський	15	670	10050

Для побудови діаграми на осі абсцис на однаковій віддалі один від одного відкладемо три відрізки рівної довжини – основи стовпчиків. Заводи розмістимо на графіку ранжировано: в порядку зменшення чисельності виробництва телевізорів кольорового зображення. Масштаб на осі ординат – 5 тис. шт. телевізорів на 1 см.



Мал. 10.1. Виробництво телевізорів кольорового зображення заводами України в 2006 р.

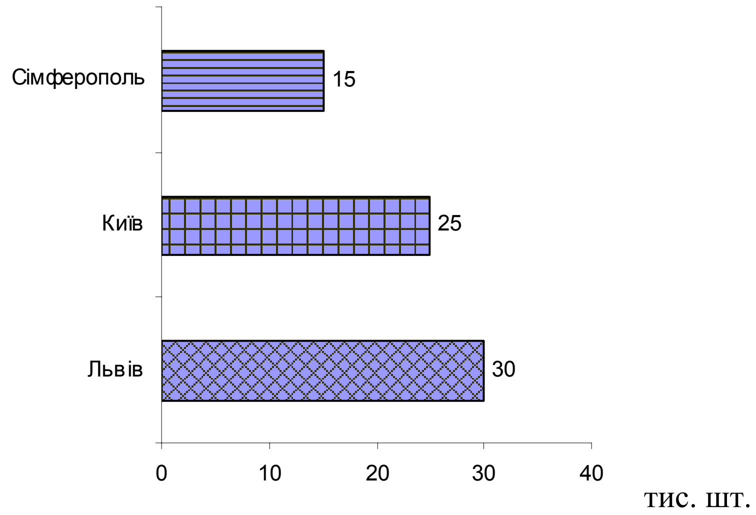
Для наочності стовпчики заштриховують або замальовують. Наочність даної діаграми досягається шляхом порівняння висоти стовпчиків, котра відповідає чисельності телевізорів кольорового зображення. Внизу під стовпчиками вказують назви об’єктів порівняння, в даному випадку міст, де знаходяться телевізорні заводи.

Якщо стовпчики-прямокутники, які зображають числа, розташувати не по вертикалі, а по горизонталі, тоді таку діаграму називають **стрічковою**.

Стовпчикові і стрічкові діаграми взаємозамінні, так як в обох випадках використовується один вимір – висота стовпчика або довжина стрічки.

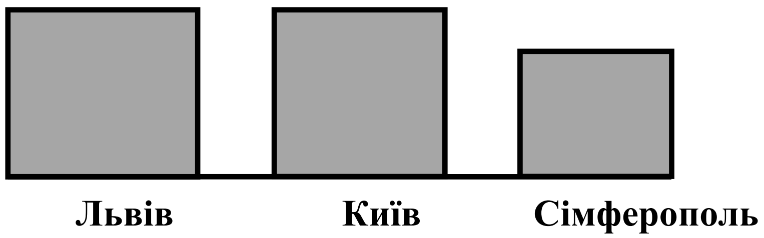
Зображення діаграм у вигляді стрічок краще ніж у вигляді стовпчиків, так як при цьому вигідніше кожному прямокутнику дати відповідну горизонтальну назву.

Проілюструємо побудову стрічкової діаграми за попередніми даними.



Мал. 10.2. Виробництво телевізорів кольорового зображення заводами України в 2006 р.

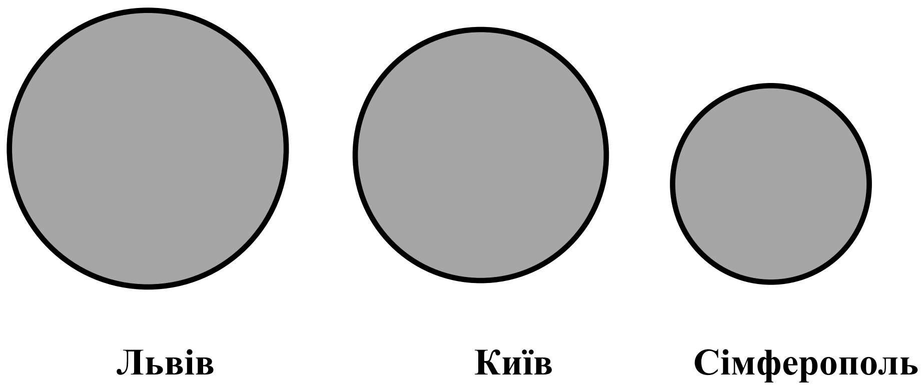
Для порівняння декількох абсолютних величин між собою використовують також **квадратні діаграми**. Для визначення сторони квадрату потрібно добути корінь квадратний із абсолютної величини явища, в даному випадку обсягу виробництва телевізорів кольорового зображення. Для Львівського заводу корінь квадратний із 30 дорівнює приблизно 5,5; Київського – корінь квадратний із 25 дорівнює 5 і Сімферопольського – корінь квадратний із 15 дорівнює 3,9. При виборі масштабу орієнтуються на найбільше число. Прийнявши масштаб 2=1 см., визначаємо сторони квадратів: для Львова – 5,5:2=2,75 см.; Києва – 5,0:2=2,5 см.; Сімферополя – 3,9:2=1,95 см. Визначивши сторони квадратів будують діаграму:



Мал. 10.3. Виробництво телевізорів заводами України в 2006 р.

Кругові діаграми основані на використанні площ кругів для порівняння однорідних абсолютних величин між собою.

При побудові кругової діаграми, потрібно прийняти до уваги, що площі кругів відносяться між собою як квадрати їх радіусів. Отже, щоб знайти радіус, потрібно добути корінь квадратний із абсолютних величин і на цій основі визначити радіуси. Довжина радіуса кола при масштабі 3=1 см. буде становити: для Львова – 1,83 (5,5:3); Києва – 1,66 (5,0:3) і Сімферополя – 1,30 (3,9:3). Після того як визначені довжини радіусів описують кола і отримують наступну діаграму:

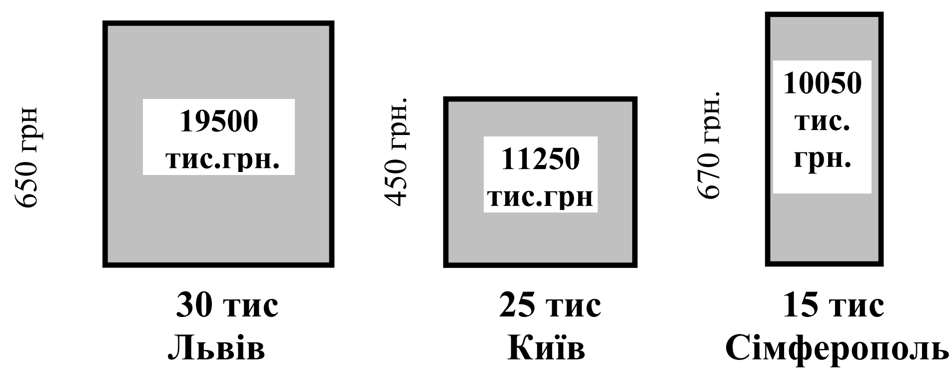


Мал. 10.4. Виробництво телевізорів заводами України в 2006 р.

Прямокутні діаграми застосовують в тих випадках, коли потрібно порівняти величини, які являють собою добуток двох співмножників і показати роль кожного з них у формуванні цієї величини. Ці діаграми вперше запропонував російський статистик В.Є Варзар (1851-1940 рр.), а тому прямокутні діаграми називають ще знаки Варзара.

При побудові прямокутних діаграм встановлюють два масштаби: один для множника, який приймають за основу, а другий для множника, який приймають за висоту. В нашому п

рикладі основа прямокутника буде характеризувати кількість телевізорів, висота – середню ціну одного телевізора, площа прямокутника – вартість всіх виготовлених телевізорів. Приймаємо масштаб для основи прямокутника (10 тис. шт.=1 см) і висоти (200 грн.=1 см), тоді діаграма матиме вигляд:



Мал. 10.5. Порівняння трьох заводі за кількістю виготовлених телевізорів, середньою ціною і вартістю

Для більшої наочності зображення статистичних явищ, можна замінити абстрактні геометричні фігури малюнками. Такого виду діаграми називаються **картинними або фігурними**.

Картинні діаграми будують двома способами:

- перший, коли малюють фігури розмір яких пропорційний величині зображуваного явища;
- другий, коли встановлюють повний масштаб для фігур. В нашому прикладі 1 картинка може відповідати 5 тис. грн. телевізорів. Ці картинки розташовують у вигляді стрічки. Так, для Львівського заводу стрічка буде складатись з 6 картинок, Київського – 5 картинок, Сімферопольського – 3 картинок.

Фігурні діаграми фіксують на собі увагу, достатньо зрозумілі і дохідливі, а тому вони часто використовуються як агітаційний інструмент.

10.3. Наочне зображення структури і структурних зрушень

Для статистичного дослідження складу сукупності використовують структурні діаграми. Структурні діаграми – це діаграми питомих ваг, які характеризують відношення окремих частин сукупності до її загального обсягу. За видами вони діляться на стовпчикові, стрічкові і секторні.

Стовпчикові і стрічкові діаграми застосовують не тільки для порівняння самих величин між собою, але й для одночасного порівняння частин цих величин. Звернемось до прикладу.

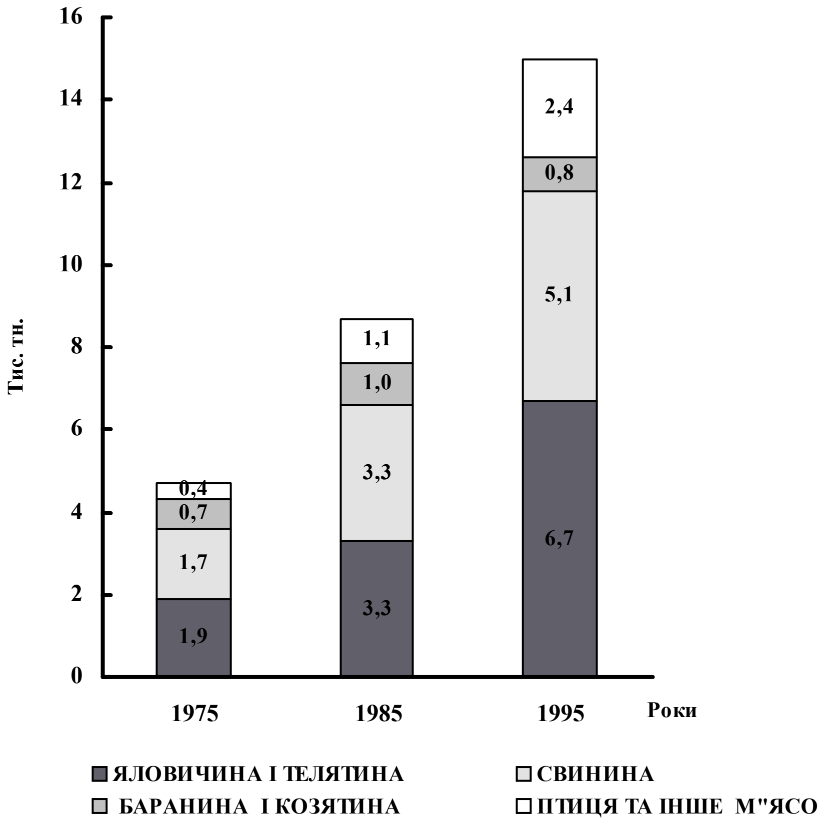
Виробництво м’яса всіма категоріями господарств регіону (дані умовні):

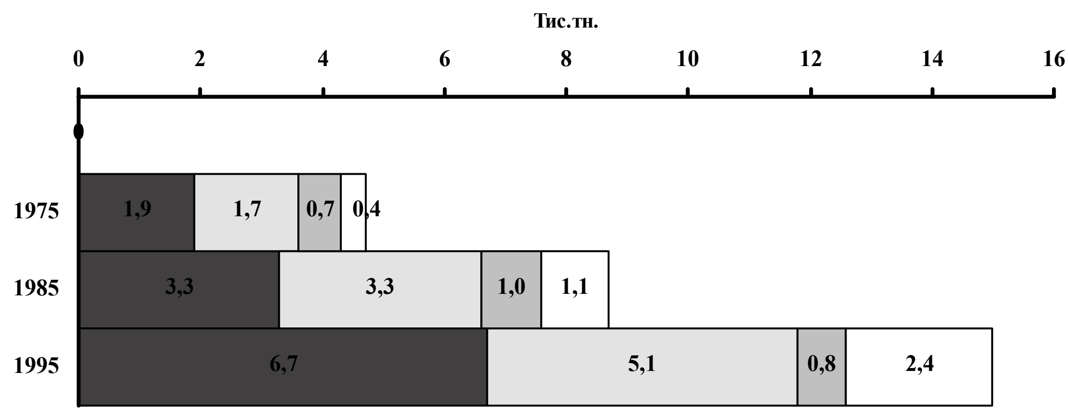
Роки	М'ясо (в забійній вазі), тис. грн.	в тому числі			
		яловичина і телятина	Свинина	баранина і козлятина	м'ясо птиці та інше
1985	4,7	1,9	1,7	0,7	0,4
1995	8,7	3,3	3,3	1,0	1,1
2005	15,0	6,7	5,1	0,8	2,4

З метою характеристики і ілюстрації обсягу і структури виробництва м’яса в регіоні побудуємо стовпчикову діаграму. Виберемо і відкладемо масштаб по осі ординат, в нашому прикладі 1 см відповідає 2 тис. т. м’яса. По осі абсцис, на однаковій віддалі будуємо стовпчики, розбивши їх на частини, величини яких відповідають обсягу виробництва різних категорій м’яса.

Аналогічно будуємо і стрічкову діаграму, тільки в даному випадку масштабна шкала відкладається на осі абсцис, а перпендикулярно до осі ординат малюють полоски (стрічки), які відображають статистичне явище. Для кожної частини стовпчика встановлюємо відповідне штрихування.

а)





б)

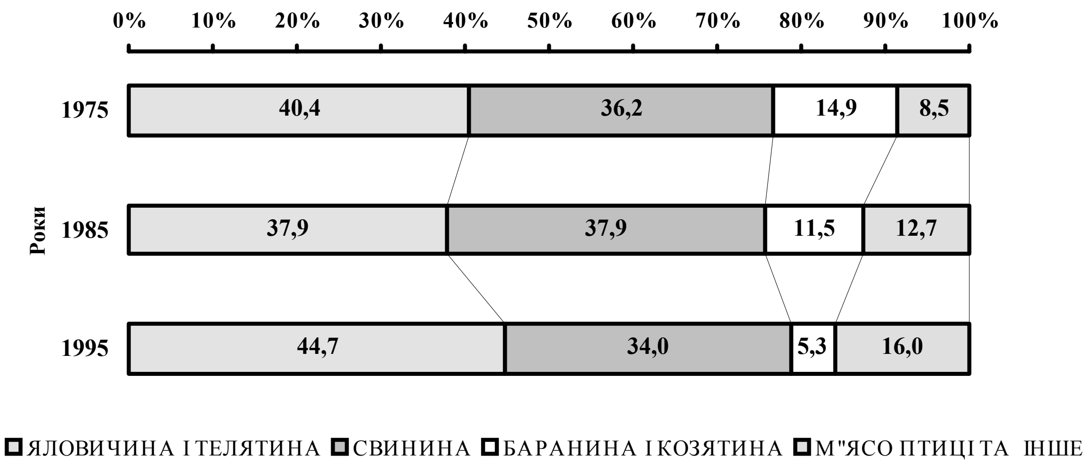
Мал. 10.6. Обсяг і структура виробництва м'яса в регіоні в 1985, 1995 і 2005 рр.

Для більш наочного зображення структури і структурних зрушень на графіку відкладають не самі абсолютні величини, а їх питомі ваги в загальному підсумку. Стовпчики або стрічки в цьому випадку мають однаковий розмір, який відповідає 100 %. В такій діаграмі стовпчики або стрічки розбивають на частини відповідно питомим вагам, котрі інколи для кращого порівняння структурних зрушень з'єднують пунктирними лініями.

Охарактеризуємо структуру виробництва м'яса господарствами регіону в процентах і побудуємо графік:

Таблиця 5.3

Роки	М'ясо (в забійній вазі), тис. грн.	в тому числі			
		яловичина і телятина	свинина	баранина і козлятина	м'ясо птиці та інше
1985	100	40,4	36,2	14,9	8,5
1995	100	37,9	37,9	11,5	12,7
2005	100	44,7	34,0	5,3	16,0



Мал. 10.7. Структура виробництва м'яса в регіоні в 1985, 1995 і 2005 рр.

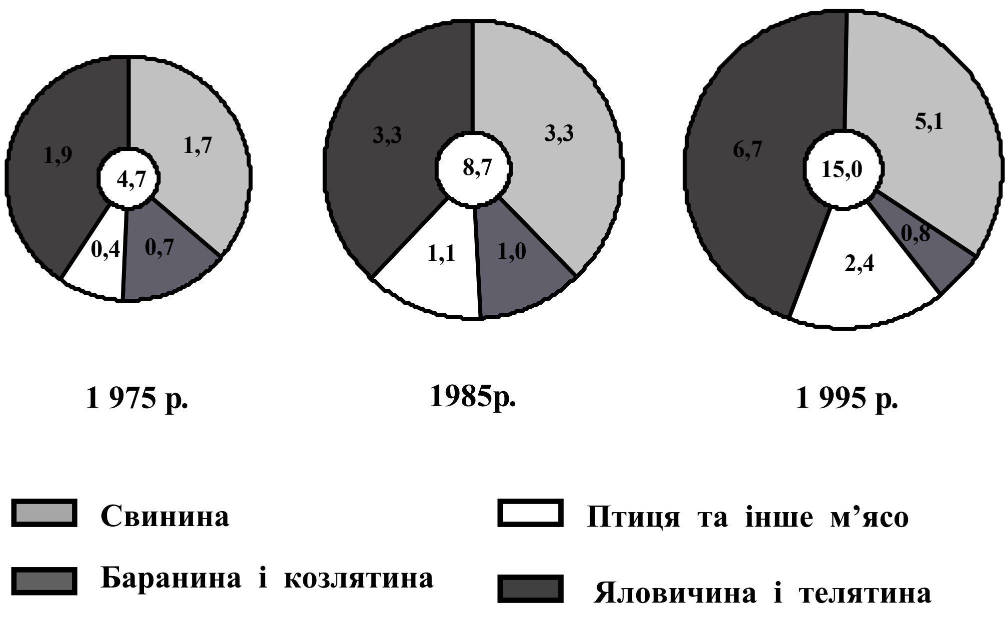
Секторні діаграми являють собою графічні зображення на площі круга, розділеного радіусами на окремі сектори за кількістю різновидів номінальних ознак. Ці діаграми застосовуються для наочної ілюстрації структури явища, для характеристики питомих ваг окремих частин цілого, для виявлення структурних зрушень.

На секторних діаграмах можуть зображуватись частини абсолютних величини явищ, або їх процентний вираз.

Для побудови секторної діаграми, яка характеризує абсолютні величини, спочатку потрібно знайти радіуси кругів, добувши квадратні корені з цих абсолютних величин. Наприклад, для 1985р. $r=\sqrt{4{,}7}=2{,}17$ см, 1995р. – $r=\sqrt{8{,}7}=2{,}95$ см, 2005р. – $r=\sqrt{15{,}0}=3{,}87$ см. При необхідності використовують масштаб.

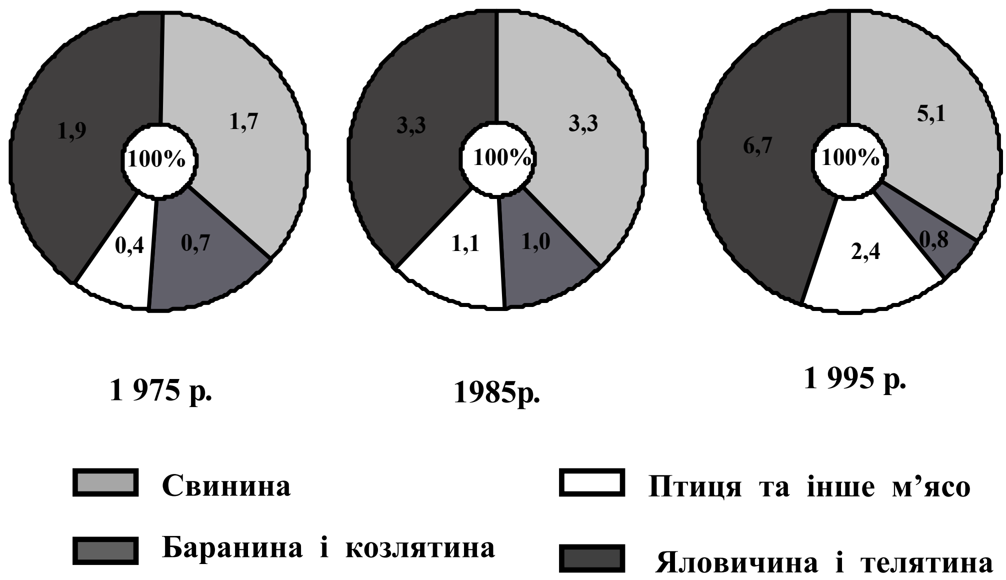
Щоб розбити круг на сектори, які відповідають величинам частин цілого, потрібно 360^0 розділити на обсяг цілого (цим самим ми знайдемо скільки градусів припадає на одиницю явища) і отриманий результат перемножити на величину частин. Наприклад, для 2005р. 360^0:15=24^0, тобто 1 тонна м'яса дорівнює 24^0, в такому випадку сектор яловичини і телятини становитиме $160{,}9^0$ $(6{,}7\cdot 24)$; свинини – $122{,}4^0$ $(5{,}1\cdot 24)$; баранини і козлятини – $19{,}1^0$ $(0{,}8\cdot 24)$; м'яса птиці та іншого – $57{,}6^0$ $(2{,}4\cdot 24)$. Аналогічно визначають сектори виробництва різних категорій м'яса і за інші роки.

Кути секторів відкладаємо за допомогою транспортира, прийнявши який-небудь радіус за початок відліку.



Мал. 10.8. Обсяг і структура виробництва м'яса в регіоні в 1985, 1995 і 2005 рр.

Якщо секторна діаграма враховує лише питомі ваги частин явища, абстрагуючись від розмірів явища, креслять круги однакових діаметрів. Вся величина явища приймається за 100 %, розраховуються долі окремих його частин в процентах. Круг розробляється на сектори пропорціонально частинам зображуваного цілого. Таким чином на 1 % припадає $3{,}6^0$. Для отримання кутів секторів, які зображають долі частин цілого, потрібно їх процентний вираз перемножити на $3{,}6^0$. Наприклад, для 2005р.: кут сектора виробництва м'яса яловичини і телятини становитиме $160{,}9^0$ $(44{,}7 \cdot 3{,}6)$; свинини – $122{,}4^0$ $(34 \cdot 3{,}6)$; баранини і козлятини – $19{,}1^0$ $(3{,}3 \cdot 3{,}6)$; м'яса птиці та іншого – $57{,}6^0$ $(16 \cdot 3{,}6)$. Аналогічно визначаємо кути секторів виробництва різних категорій м'яса в 1985 і 1995 рр. Покажемо цю діаграму на мал. 10.9.

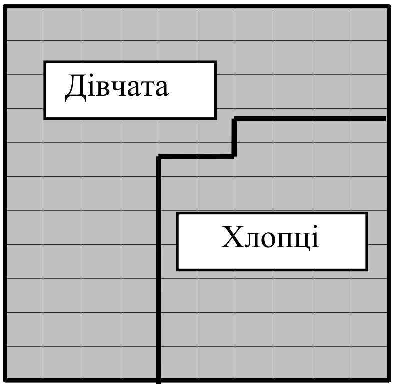


Мал. 10.9. Структура виробництва м'яса в регіоні в 1985, 1995 і 2005 рр.

Секторні діаграми виразні в тих випадках, коли досліджувана сукупність ділиться не більше ніж на 4-5 частин і спостерігаються помітні структурні зрушення. Якщо ж структурні зрушення незначні, або сукупність ділиться на більше число секторів, тоді для графічного зображення структури доцільно використовувати стовпчикові або стрічкові діаграми.

В деяких випадках для характеристики структури сукупності використовують також квадратні і кругові діаграми.

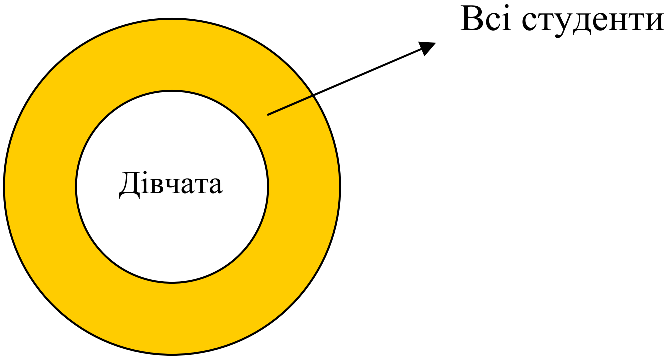
Для зображення структури сукупності, яка складається (в більшості випадків) з двох частин, беруть квадрат. Площу квадрата ділять на 100 рівних частин. Кожний маленький квадратик дорівнює одній сотій всієї площі великого квадрата. Потім ці квадратики заштриховують у відповідності із процентною структурою досліджуваної сукупності. Звернемось до прикладу. Нехай статева структура студентів економічного вузу в 2006р. відповідно становила: дівчата – 60 %, хлопці – 40 %. Покажемо побудову такої діаграми на графіку:



Мал. 10.10. Питома вага дівчат і хлопців в економічному вузі в 2006р.

В тому випадку, коли частину і ціла зображають при допомозі кругової діаграми, тоді круги креслять не окремо один від одного, а накладається один на другий.

Якщо для нашого прикладу візьмемо радіус R=2 см., то площа круга становитиме $S = ПR^2 = 3{,}1416 \cdot 4 = 12{,}5664$ см^2, яка відповідає 100% всіх студентів. Для 60% дівчат площа круга – $S_q = ПR^2 = 3{,}1416 \cdot 4 \cdot 0{,}6 = 7{,}5398$ см^2. Радіус такої площі дорівнює R=1,55 см. ($R = \sqrt{S:П} = \sqrt{7{,}5398:3{,}1416}$). Після розрахунків будуємо кругову діаграму статевої структури студентів економічного вузу.



Мал. 10.11. Статева структура студентів економічного вузу в 2006 р.

Інколи доцільно показати три круги, один в другому, але навколо різних центрів.

Зустрічається також комбінування кругових діаграм із секторними, коли круги різної величини подають з розбивкою на сектори.

10.4. Графічне зображення динаміки статистичних показників

Графіки, які ілюструють зміну статистичних явищ в часі називаються **динамічними**. Для зображення динаміки явищ часто використовують стовпчикові, стрічкові, квадратні, кругові і картинні діаграми, в яких кожний стовпчик, стрічка, квадрат і т.д. зображають величину статистичного явища на певну дату, або за відповідний проміжок часу.

Крім названих вище графіків нерідко застосовуються і лінійні графіки.

Лінійні графіки використовуються для характеристики зміни явищ в часі, виявлення залежності між двома показниками і деяких інших завдань. Вони будуються при допомозі прямокутної системи координат, на осі абсцис якої розміщують шкалу характеристик часу, а на осі ординат – рівні ряду динаміки.

У лінійній діаграмі динаміки шкала на осі ординат повинна починатися з нуля, інакше діаграма буде не правильно відображати характер розвитку явища.

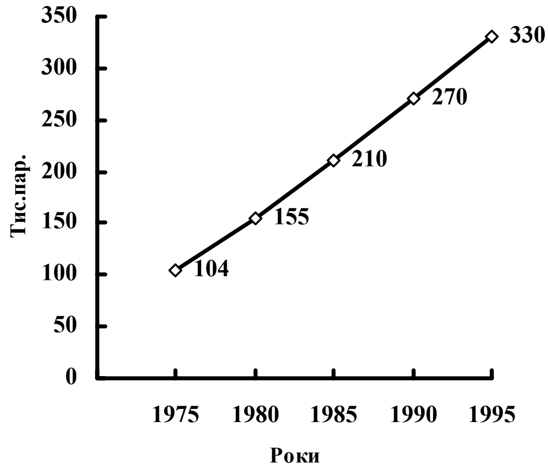
Оскільки при великих значеннях рівні динамічного ряду діаграма з початковим нульовим рівнем ординати буде невиразною і некомпактною, тоді на осі ординат слід зробити розрив шкали. Для базисних характеристик швидкості зміни досліджуваного явища початковий рівень ординати може починатись із 100.

Лінійні діаграми дають можливість наочно визначити періоди часу, коли явища зростали (зменшувались) більш чи менш інтенсивно, або залишались без змін.

Особливістю лінійного графіка наочного зображення даних, які характеризують підсумки розвитку явища за певний період часу є те, що динаміка показується у вигляді неперервної лінії, котра характеризує неперервність процесу.

Покажемо побудову лінійного графіка на основі наступних даних: виробництво шкіряного взуття швейним об'єднанням в 1985 р. становило 104 тис.

пар, в 1990 р. – 155 тис. пар, в 1995 р. – 210 тис. пар, в 2000 р. – 270 тис. пар і в 2005 р. — 330 тис. пар. Зобразимо ці дані графічно:



Мал. 10.12. Динаміка виробництва шкіряного взуття швейним об'єднанням в 1985-2005рр.

Часто на одному лінійному графіку приводиться декілька кривих, котрі дають порівняльну характеристику динаміки різних показників або одного і того ж показника, але різних об’єктів. В таких випадках спочатку потрібно показники рядів динаміки , які будемо наносити на графік, привести до однієї основи, тобто абсолютні показники кожного ряду замінити базисними темпами росту, прийнявши для всіх рядів один і той же період в якості бази порівняння. В цих графіках лінії всіх рядів розходяться із однієї точки, прийнятої за 1 або 100 %. Розглянемо побудову такого графіка.

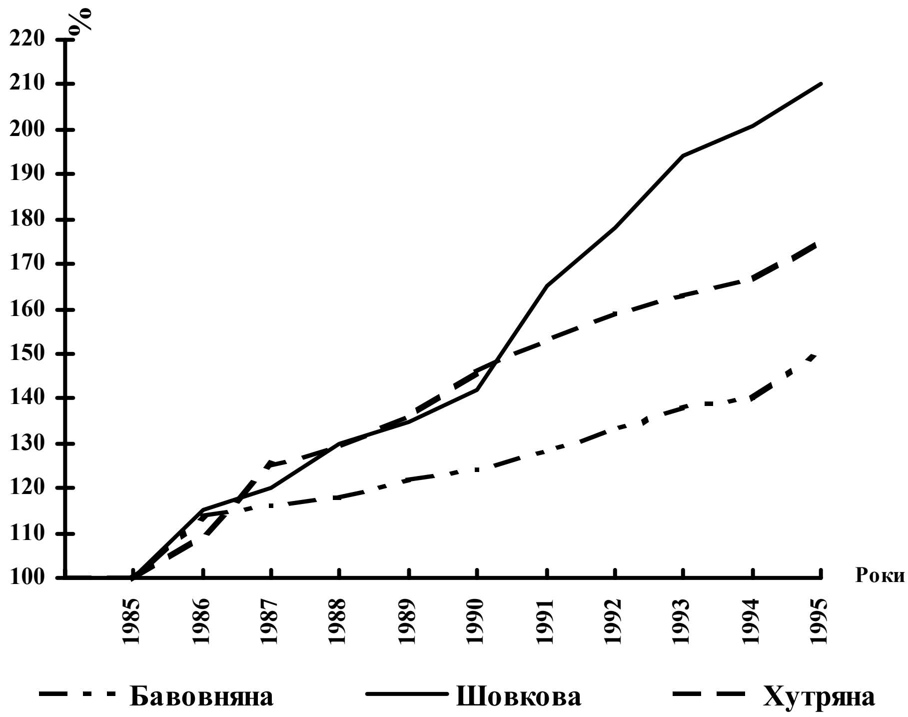
Темпи росту загального обсягу продукції деяких галузей легкої промисловості за період 1995-2005рр. (в процентах до 1995р.):

Роки	Галузі		
	бавовняна	Шовкова	хутряна
1995	100	100	100
1996	114	115	110
1997	116	120	125
1998	118	130	129
1999	122	135	136
2000	124	142	146
2001	128	165	153
2002	133	178	159
2003	138	194	163
2004	140	201	167

2005	151	210	175

Лінійні графіки використовують в статистиці не тільки для ілюстрації динаміки якого-небудь явища, але і для наочного зображення рядів розподілу. В цьому випадку на осі абсцис відкладаються варіанти, а на осі ординат – частоти ряду розподілу.

Лінійними графіками користуються також для наглядного зображення залежності однієї варіаційної ознаки від іншої.



Мал. 10.13. Динаміка виробництва продукції бавовняною, шовковою і хутряною галузями легкої промисловості за 1995-2005рр.

В статистичній практиці побудови графіків для аналізу темпів динаміки явища використовують лінійні графіки на напівлогарифмічній сітці. **Напівлогарифмічною** називається сітка, в котрій на осі абсцис нанесений звичайний масштаб, а на осі ординат – логарифмічний.

Перевага напівлогарифмічної сітки в аналізі динаміки явища заключається в тому, що вона дає більш коректну уяву про темпи динаміки. Діаграму на напівлогарифмічній сітці називають ще **діаграмою темпів**.

Для побудови лінійного графіка з напівлогарифмічною шкалою по осі ординат замість звичайної шкали відкладають логарифмічну з рівними інтервалами. Далі по таблиці логарифмів, знаходять логарифми для цілих чисел, які проставляють з правої сторони осі ординат для кращої наочності. За масштабом логарифмічної шкали знаходять відповідні точки, які проставляють на графіку і з'єднують їх лініями.

Різновидністю лінійної діаграми є **радіальні діаграми**, побудовані в полярних координатах і призначені для відображення процесів і явищ, які періодично повторюються в часі (переважно сезонних коливань).

За вісь ординат, в полярних координатах, приймаються радіуси, а за вісь абсцис – коло. Пунктом відліку служить центр кола, або його окружність.

Радіальні діаграми бувають двох видів – замкнуті і спіральні.

Замкнуті діаграми відображують весь внутрішньорічний цикл зміни явища за один рік.

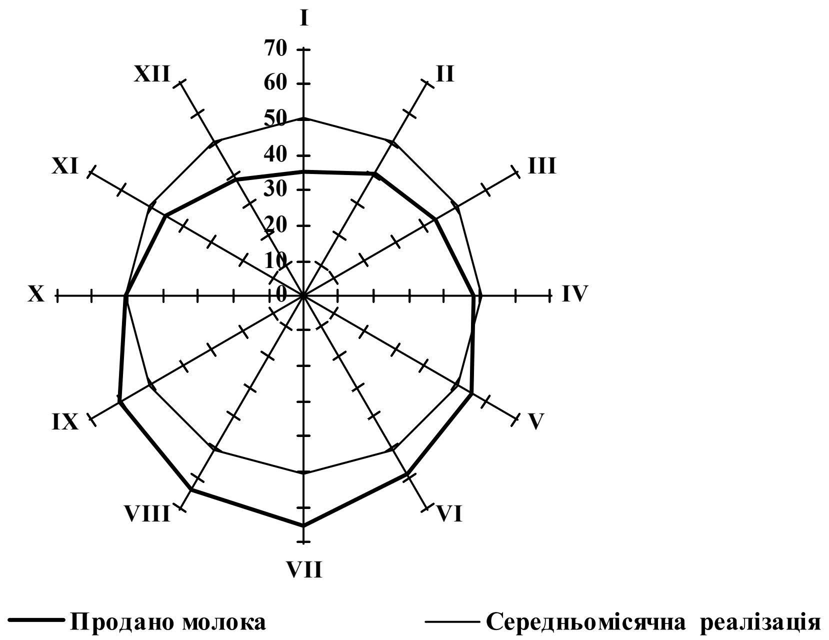
Для того щоб побудувати радіальну діаграму замкнутого виду, у якої пунктом відліку служитиме центр кола, креслять коло радіусом, рівним середньомісячному показнику. Усе коло ділять на стільки частин, скільки внутрішньорічних періодів і відповідно їм проводять радіуси. Періоди часу розміщують за годинниковою стрілкою, причому розміщення місяців (якщо коло розбите на 12) аналогічне циферблату годинника. На кожному радіусі відповідно до прийнятого масштабу відкладають від центра кола відрізки пропорційно рівням показників конкретного місяця. Дані, які перевищили середньомісячний рівень, відкладаються за межами кола на продовжені радіуса. Потім кінці відрізків на радіусах з'єднуються лініями, причому точка грудня з'єднується із точкою січня одного і того ж року. Побудову радіальної замкнутої діаграми розглянемо на прикладі.

Маємо умовні дані про реалізацію молока на ринку одного міста по місяцях 2005 року:

Місяць	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Разом
Продано молока, т.	35;	40;	43;	48;	55;	58;	65;	63;	60;	50;	45;	38	600

Середньомісячна реалізація молока за рік складає 50 т. (600:12).

Креслимо коло з радіусом, який дорівнює середньомісячному показнику (R=50 т). На горизонтальному діаметрі побудуємо шкалу, взявши довжину радіуса 2,5 см. Відповідно, 1 см = 20 т. (50:2,5). Все коло ділимо на 12 радіусів, за кількістю місяців в році і відкладаємо наведені в умові прикладу за масштабом дані. Мітки на радіусах різних місяців з'єднуємо між собою.



Мал. 10.14. Сезонність продажу молока на ринку в 2005 р. по місяцях

Спіральна радіальна діаграма будується в тому випадку, коли є дані по місяцях за ряд років. Принцип їх побудови той же, що і замкнутих, однак різниця лише в тому, що в спіральних діаграмах грудень одного року з'єднується не з січнем даного ж року, а з січнем наступного року, в результаті чого виходить крива у вигляді спіралі.

10.5. Контрольно-планові графіки

Графічний метод широко використовується для поточного контролю за ходом виконання плану. Форми графічного зображення для порівняння планових і фактичних показників досить різноманітні. Розглянемо два основних види цих графіків:

а) лінійні графіки виконання плану;

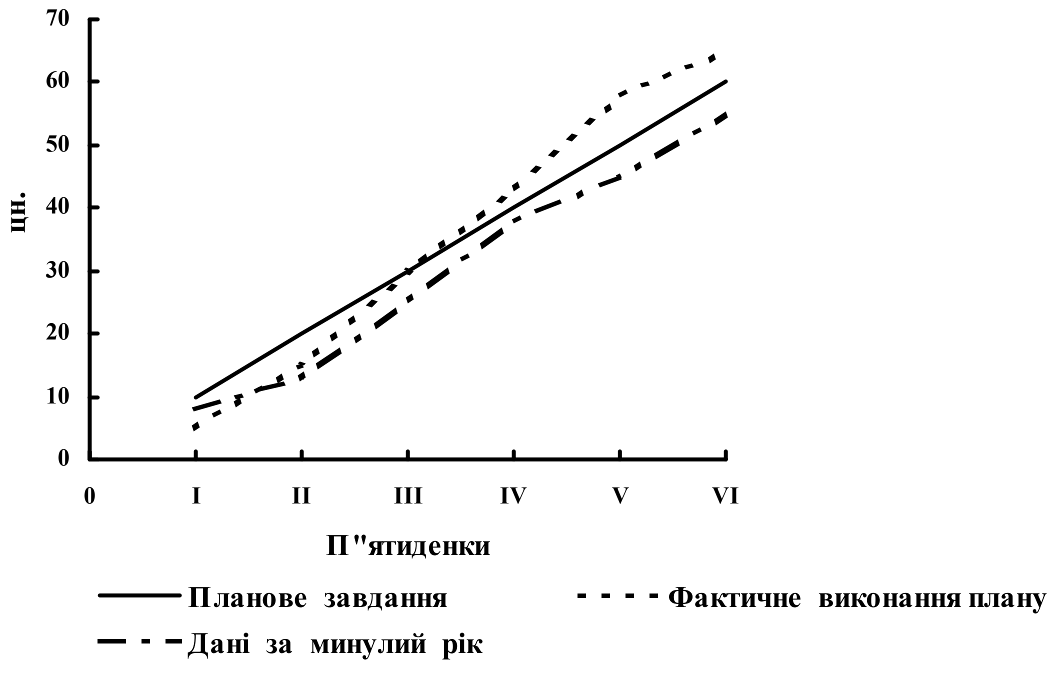
б) обліково-планові графіки.

Лінійні графіки є зручним засобом контролю виконання плану по одному якому-небудь об’єкту або показнику. При цьому для аналізу на графіку доцільно показати наростаючим підсумком не тільки планові і фактичні показники у звітному періоді, але й фактичні – за минулий рік. Покажемо це на прикладі.

Виконання плану заготівлі лікарських рослин райспоживспілкою
в серпні 2005 р.

П’ятиденка	План	Фактично заготовлено лікарських рослин (тонн – наростаючим підсумком)	
		в поточному році	в минулому році
I	10	5	8
II	20	15	13
III	30	30	25
IV	40	43	38
V	50	58	45
VI	60	65	55

Побудуємо графік виконання плану:



Мал. 10.15. Виконання плану заготівлі лікарських рослин райспоживспілкою в серпні 2005 р.

На графіку наочно видно хід виконання плану заготівлі лікарських рослин по п’ятиденках: в першій і другій п’ятиденках фактичні заготівлі відстають від плану, третя п’ятиденка співпала з планом і в подальшому фактична лінія йде вище планової. Видно також, що обсяг заготівель поточного року (крім першої п'ятиденки) весь час вищий минулого.

В тих випадках, коли потрібно організувати наочний контроль виконання плану одночасно на декількох об'єктах, будують **обліково-планові графіки**.

Їх будують на спеціально розграфленій сітці, яка має форму таблиці, і на якій по горизонталі відкладають одиниці часу (день, п'ятиденку, декаду, місяць, квартал), а по вертикалі розміщують об'єкти дослідження.

Кожний відрізок по горизонталі відповідає 100 % виконання планового завдання, який, в свою чергу ділиться на п'ять рівних частин (кліток) по 20 % на кожну.

Ступінь виконання плану по кожному об'єкту зображається двома лініями: тонкою переривчастою, яка показує ступінь виконання плану за одиницю часу і жирною суцільною, яка характеризує виконання плану за звітний період в цілому.

Розглянемо порядок побудови обліково-планового графіка за такими даними:

Виконання плану по цехах заводу

Число місяця	Цех 1				Цех 2			
	план на день, шт.	555роб-лено за день, шт.	вико-нання плану, %	фактично вироблено з початку місяця, шт.	план на день, шт.	вироб-лено за день, шт.	вико-нання плану, %	фактично вироблено з початку місяця, шт.
1	500	450	90	450	500	500	100	500
2	500	500	100	950	510	561	110	1061
3	500	525	105	1470	520	546	105	1607
4	500	530	106	2000	530	535	101	2142
5	500	500	100	2500	540	590	109	2732
і т.д.								

Цех 1 щоденно за планом повинен був виготовляти однакову кількість деталей – 200 шт., а цех 2 не однакову їх кількість: за перший день – 500 шт., другий – 510 шт., третій – 520 шт., і т.д.

Побудова графіка виконання плану за цими двома варіантами дещо відрізняється.

Нанесення тонких ліній на графіку не складене. Розглянемо спосіб їх нанесення в таких випадках: 1) планове завдання не виконано; 2) планове завдання виконано на 100 %; 3) планове завдання перевиконано.

Цех 1 планове завдання за перший день виконав на 90 % і тонка лінія займає 4 клітки (80 %) і половину п'ятої (10 %). За другий день планове завдання цех виконав на100 %, а тому тонка лінія займає всі 5 клітинок. Планове завдання третього числа місяця виконано на 105 %, а тому тонка лінія займає всі п'ять кліток третього дня (100 %) і під лінією в першій клітці проведена маленька лінія, яка доходить до чверті (5 %). Ця додаткова лінія показує перевиконання денного планового завдання.

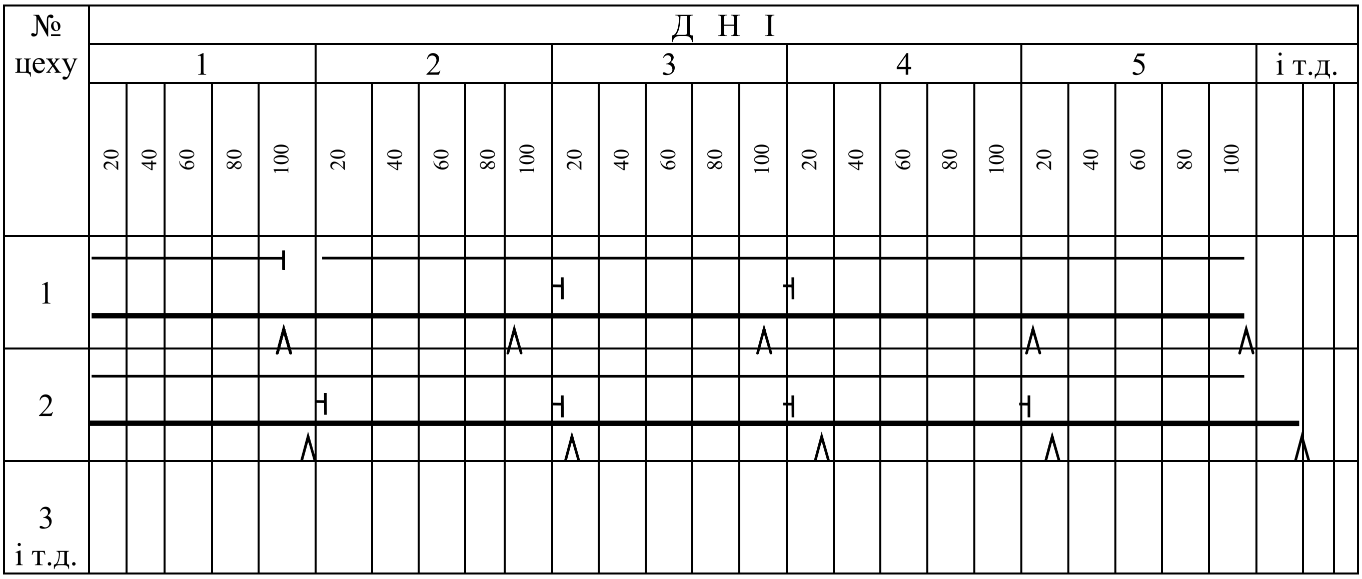
Техніка нанесення суцільної жирної лінії проста в тих випадках, коли планове завдання за кожний день не змінюється на протязі тривалого періоду, на котрий складається графік. Цех 1 щоденно планував виготовляти по 500 деталей, і в перший день він виконав план на 90 %, другий – 100 %, третій – 105 %, четвертий – 106 % і в п'ятий – 100 %.

За перший день жирна лінія буде такої довжини, що й тонка. В другий день план був виконаний на 100 %, за два дні – на 190 % (90+100), тобто суцільна жирна лінія доводиться за перший день до 100 %, а за другий день – до поділки, котра відповідає 90 %. Це показує, що планове завдання за перші два дні не виконано.

За третій день планове завдання виконано на 105 %, а за три дні – на 295 % (90+100+105). Жирна лінія займе по 5 кліток першого і другого дня і 95 % третього. Всі дальші розрахунки проводять аналогічно до попередніх.

По цеху 2 принцип проведення тонкої лінії такий же, як і по цеху 1. Складніше нанести на графік жирну лінію, тому що планові завдання за кожний день змінювались і не можна додавати проценти щоденного виконання завдання. В даному випадку ступінь виконання планового завдання за окремі періоди часу знаходять відношенням наростаючих підсумків абсолютних даних фактичного виконання до плану.

Суцільна жирна лінія повинна показати виконання планового завдання за всі попередні дні з початку звітного періоду. За перший день жирна лінія співпаде з тонкою. За другий день виготовлено продукції 561 шт., із яких 510 зараховують у виконання плану другого дня і 51 шт., або 9,8 % (51:520), – в рахунок виконання плану третього дня. За третій день виготовлено 546 деталей, до яких добавляють ще 51 шт. деталей виготовлений за другий день в рахунок третього, разом – 597 деталей, що складає 114,8 % (597:520) денного планового завдання третього дня. Жирна лінія займе п'ять кліток третього дня і ще майже ¾ першої клітки, відведеної для четвертого дня (відповідає приблизно 15 %). Це означає, що за три дні другий цех повністю виконав триденне планове завдання і продовжував працювати в рахунок четвертого дня. Таким чином, жирна лінія графіка наочно демонструє хід виконання плану за кожний день з початку звітного періоду:



Мал. 10.16. Виконання плану виробництва деталей цехами заводу за першу п'ятиденку місяця

Ступінь виконання плану за кожний день на жирній лінії позначається крапкою або штрихом.

Зображення на одному графіку декількох об'єктів дає можливість порівнювати підсумки їх роботи в цілому і оцінювати рівномірність виконання плану.

10.6. Графіки просторового розміщення і просторового розповсюдження

Для вивчення розміщення, рівня і ступеня розповсюдження якого-небудь явища в просторі використовується три види графіків: а) картограма; б) картодіаграма; в) центрограма.

Картограма – це схематична географічна карта, на якій розподіл зображуваних явищ по території дається за допомогою розмальовування, штриховки або крапок.

В залежності від використовуваних символів розрізняють фонові і крапкові картограми. Для побудови **фонових картограм** використовується вся поверхня карти в кордонах досліджуваної території. На цій карті повинні бути чітко позначені контури меж адміністративного поділу країни, області, району.

Географічний ряд, призначений для картограмування, потрібно оптимально розбити на групи (райони), що дозволить простежити закономірності розміщення зображуваного явища. Кожній групі (району) надається певний тип штрихування або колір, а потім їх наносять на карту. Так, наприклад, якщо ми хочемо дати картограму розміщення садів і виноградників в Україні, то ми повинні всі дані про це по окремих областях розбити, припустимо, на чотири групи з відповідних штрихуванням. В першу групу увійдуть всі категорії господарств з площею садів і виноградників питома вага яких в загальній площі сільськогосподарських угідь до 1 %; другу – 1,1-2,0 %; третю – 2,1-3,0 % і четверту – понад 3,0 %. Після на кожну область у відповідності з тією групою, до котрої вона попала, наноситься вказане штрихування. Інколи в якості умовного знаку замість штрихування користуються кольором, тільки при цьому вибирають кольори таким чином, щоб була витримана зростаюча інтенсивність по мірі переходу від нижчих груп до вищих.

При зображенні деяких явищ, які вивчаються статистикою, розподіл за адміністративними районами не має великого значення, а тому в подібних випадках виділяються райони з однаковим показником досліджуваного явища за допомогою **ізолінійних картограм**. Такі картограми використовуються в метеорології і геодезії.

В економіці ізолінійні картограми застосовуються для встановлення часу виконання основних сільськогосподарських робіт (ізотопи), для зображення регіонів з однаковими цінами (ізопрайси) і т.д. На ізолінійних картограмах замкнутими плавними лініями зображаються контури приблизно рівних величин статистичного показника.

Недоліком штрихових картограм є те, що повний географічний регіон штрихується однаково, без переходу по густоті штрихів, хоч в дійсності розподіл любої ознаки на місцевості не завжди рівномірний. А тому замість фарби і штрихування в якості графічних символів в картограмах використовують крапки.

В **крапкових картограмах** графічним знаком статистичних даних є крапки строго визначеного розміру, розміщені в заданих межах. Кожна крапка відповідає певній числовій величині і є носієм елементу обліку. Крапки на картограмі надають обліку наочність і природність. Вони добре ілюструють ступінь концентрації об’єктів промисловості і сільського господарства в різних районах і можуть використовуватись в багатьох галузях статистики.

Крапки на розрахунковій картограмі розміщуються на контурах території з врахуванням їх фактичного розподілу на окремих ділянках цієї території. Це дозволяє порівнювати щільність розміщення досліджуваних об’єктів на різних ділянках території за гущеністю крапок. Виразність крапкової діаграми залежить від розміру крапки. Якщо зменшити розмір крапки, тоді на цій же площі можна розмістити більше крапок, і не буде нашарування крапок однієї на другу. Крапкові діаграми за своєю суттю близькі до фонових. Однак фонові картограми, як правило, використовуються для зображення середніх і відносних показників. Крапкові ж картограми використовуються для об’ємних (кількісних) показників. Їх застосовують в тих випадках, коли сума ваг статистичного розподілу по районах має економічний зміст. У фонових діаграмах сума ваг економічного змісту немає.

Якщо після за штриховки, фарбування або нанесення крапок на відповідні ділянки карти виявляється певна закономірність в географічному розміщенні території з однаковою величиною зображуваного показника, тоді можна судити про залежність даного показника від географічного фактора. Якщо ж райони з однаковим зображенням розміщені на карті в хаотичному порядку, це свідчить про відсутність певної закономірності в просторовому розміщенні даного показника, тобто розповсюдження або рівень не зв'язані з географічним положенням району.

Картодіаграма – це поєднання схематичної географічної карти із діаграмою. Основне завдання картодіаграм заключається в тому, щоб показати географічний розподіл зображуваного статистикою явища. Головна їх особливість заключається в розміщенні на контурній географічній карті спеціальних знаків-символів у вигляді стовпчиків, квадратів, кругів та інших. Величина геометричного знаку залежить від розміру даного явища в зображуваному районі. Знаки і символи на картодіаграмі розміщуються не в простій лінійній послідовності, а орієнтуються в географічному просторі.

Основна перевага картодіаграм перед звичайними діаграмами заключається в точній географічній орієнтації статистичних величин, у встановленні їх взаємної відповідності і просторовому розподілі.

Певну перевагу картодіаграми мають також і перед картограмами. На картодіаграмі зображуються самі досліджувані величини, що сприяє більш точному їх відображенню. На картограмі зображуються головним чином середні, крайні значення, або значення інтервалів.

Основним видом знаків-символів при побудові картодіаграм є кругові і секторні діаграми. За допомогою цих символів на картодіаграмі зображують одночасно як обсяг так і структуру (склад) статистичного явища розміщеного в просторі.

Якщо на картодіаграмі зображують лише структуру досліджуваного явища без врахування його обсягу, тоді будують круги однакового радіуса.

Для зображення розподілу по території абсолютних величин на карту наносять прямокутники у вигляді стовпчиків або стрічок. Ці прямокутники або стрічки можна використати також для графічного зображення структури явища.

В економічних дослідженнях доводиться інколи поєднувати картодіаграми з картограмами. Діаграми якби накладаються на картограми. Картограми в поєднанні з картодіаграмами при вмілій їх побудові є важливим засобом наочного зображення і аналізу суспільно-економічних явищ і процесів. Наприклад, такий графік дає можливість проаналізувати територіальне розміщення міського населення за чисельністю в поєднанні із щільністю розселення сільського населення. Міста за чисельністю жителів на карті зображують за допомогою кругових діаграм, а щільність сільського населення – за вибраною штриховкою відповідних територій.

Соціально-економічне районування на географічній карті наочно зображують і аналізують також за допомогою поєднання цих двох графіків. Об'єм і структуру промисловості на карті показують за допомогою секторних діаграм, а напрямки сільськогосподарського виробництва – через заштриховку відповідних регіонів.

Центрограма – це контурна карта на якій розміщуються короткі цифрові таблиці з інформацією про історико-географічний розвиток і розміщення досліджуваного явища чи процесу. Центрограми ще називають історико-географічними картами. Вони дозволяють скласти цілі статистико-географічні описи нанісши цифрові ряди на карті для різних територій, що дає можливість наочно уявити окремі сторони протікання досліджуваного процесу в динаміці.

Центрограми дозволяють визначити питому вагу окремих регіонів, а також тенденцію переміщення центру розвитку в розташуванні окремих явищ.

ЛЕКЦІЯ 11. СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗ

ПЛАН:

11.1. Загальні поняття, вибір типу критичної ділянки.

11.2. Перевірка гіпотези про належність спостережень, які виділяються від досліджуваної генеральної сукупності.

11.3. Перевірка гіпотези про відповідність емпіричного розподілу типу кривих нормального розподілу.

11.4. Перевірка гіпотези про величину центру розподілу.

11.5. Елементи дисперсійного аналізу.

11.1. Загальні поняття, вибір типу критичної ділянки

Під **статистичною гіпотезою** розуміють різні передбачення щодо виду або параметрів розподілу випадкової перемінної, які можна перевірити, спираючись на результати спостережень у власне випадковій вибірці. Статистична перевірка гіпотез носить ймовірний характер, так як їх висновки ґрунтуються на вивченні властивостей розподілу випадкової перемінної за даними вибірки, а тому завжди є ризик допустити помилку. Однак, з допомогою статистичної перевірки гіпотез можна визначити ймовірність прийняття помилкового висновку. Якщо його ймовірність незначна, то вважається, що застосований критерій забезпечує малий ризик помилки.

При перевірці гіпотез існує можливість допустити помилку двоякого роду:

а) гіпотеза, яка перевіряється (нульова гіпотеза «H_0»), в дійсності є вірною, але результати перевірки приводять до відказу від неї;

б) гіпотеза, яка перевіряється в дійсності є помилковою, але результати перевірки проводять до її прийняття.

Для побудови статистичного критерію, який дозволяє перевірити певну гіпотезу, потрібно виконати наступне:

1. Сформулювати гіпотезу «H_0» яка перевіряється. Поряд з нею висувається альтернативна гіпотеза;

2. Вибрати рівень значущості «α», який контролюватиме допустиму ймовірність помилки першого роду;

3. Визначити критичну ділянку допустимих значень;

4. Прийняти відповідне рішення.

Рівнем значущості називається таке мале значення ймовірності попадання критерію в критичну ділянку при умові справедливості гіпотези, що появу цієї події можна розцінювати як наслідок суттєвого розходження висунутої гіпотези з результатами вибірки. Як правило, рівень значущості приймають рівним 0,01 або 0,05. Виходячи з величини рівня значущості визначають критичну ділянку, під якою розуміють таку ділянку значень вибіркової характеристики, попадаючи в котру вони будуть свідчити про те, що гіпотеза яка перевірялась повинна бути відкинута.

До критичної ділянки відносяться такі значення, поява котрих при умові правильності гіпотези були б малоймовірними. Припустимо, що обчислене за емпіричними даними значення критерію попало в китичну ділянку, тоді за умови ймовірності перевірочної гіпотези «H_0» ймовірність цієї події буде не більша значущості «α». Оскільки «α» вибирається достатньо малим, то така подія є малоймовірною, і, отже, перевірочна гіпотеза «H_0» може бути відкинута. Якщо ж досліджуване значення критерію не належить до критичної ділянки, а знаходиться в ділянці допустимих значень, то перевірочна гіпотеза «H_0» не відкидається. Ймовірність попадання критерію в ділянку допустимих значень при справедливості перевірочної гіпотези «H_0» дорівнює $1-\alpha$. Потрібно враховувати, що попадання критерію в ділянку допустимих значень не означає строгого доведення гіпотези «H_0». Воно лише вказує, що між висунутою гіпотезою і результатами вибірки не існує суттєвих розбіжностей. Із зменшенням рівня значущості, зменшується ймовірність забракувати перевірочну гіпотезу, коли вона вірна, тобто менша ймовірність допустити помилку першого роду. Але при цьому розширюється ділянка допустимих значень і, значить, збільшується ймовірність зробити помилки другого роду.

За даним рівнем значущості можливо по-різному встановлювати критичну область, яка гарантує цей рівень. Нехай ми задаємося рівнем значущості $\alpha = 0{,}05$. В якості критичної ділянки, яка відповідає даному рівню значущості, можемо взяти (для нормального розподіленого критерію):

1. Ділянку великих позитивних відхилень так, щоб

$$P_I = P(x > \bar{x} + t_I \mu) = 0{,}05.$$

Тоді $P_I = \dfrac{1}{\sqrt{2\pi}} \cdot \int\limits_t^\infty e^{-\frac{t^2}{2}} dt = \dfrac{1}{2} - \dfrac{1}{\sqrt{2\pi}} \cdot \int\limits_0^t e^{-\frac{t^2}{2}} dt,$

звідки $\dfrac{1}{\sqrt{2\pi}} \cdot \int\limits_0^t e^{-\frac{t^2}{2}} dt = \dfrac{1}{2} - 0{,}05 = 0{,}45$

і за таблицею нормованої функції Лапласа знаходимо, що $t_I = 1{,}64$.

2. Ділянка великих від’ємних відхилень

$$P_{II} = P(x < \bar{x} - t_{II} \mu) = 0{,}05, \text{ тобто } P_{II} = P(x < \bar{x} - 1{,}64\mu).$$

3. Ділянка великих за абсолютною величиною відхилень

$$P_{III} = P(\left| x - \bar{x} \right| > t_{III}\mu) = 0{,}05, \text{ тобто } 0{,}05 = 1 - 2\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt.$$

Тоді $\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt = \frac{1-0{,}05}{2} = 0{,}475.$

За таблицею нормованої функції Лапласа знаходимо, що $t_{III} = 1{,}96.$

4. Ділянка малих за абсолютною величиною відхилень

$$P_{IV} = P(\left| x - \bar{x} \right| < t_{VI}\mu) = 2\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt = 0{,}05.$$

Тоді $\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt = 0{,}025.$

Величина t, відповідна ймовірність 0,025, дорівнює 0,065.

11.2. Перевірка гіпотези про належність спостережень, ї виділяються від досліджуваної генеральної сукупності

В складі зібраних даних можуть зустрічатись поодинокі спостереження, в яких зареєстровані значення ознаки які помітно відрізняються від загального рівня. Вони виникають в результаті:

а) помилок спостереження;

б) випадкового збігу різного роду окремих несуттєвих обставин;

в) порушення однорідності досліджуваної сукупності.

Суб’єктивне відкидання варіантів значень ознаки, які різко відрізняються від середньої арифметичної, не мають ніякого принципіального виправдання. Для виключення спостережень які виділяються із подальшої обробки потрібне застосування обґрунтованих критеріїв. Припустимо, що розподіл результатів спостережень, які проводяться в звичайних умовах, відповідає нормальному закону з параметрами «$\bar{x}$» і «σ». В результаті проведення однієї із серій спостережень отримані «*n*» значень $x_1 \le x_2 \le x_3 \le ... \le x_n$, серед яких максимальне значення «x_n» (або мінімальне «x_1») різко відрізняється за своєю величиною від решти «$n-1$» спостережень. Виникає запитання, чи відносяться ці значення до даної, яка спостерігається за певних умов, генеральної сукупності, чи вони виникли в результаті яких-небудь екстраординарних обставин? Нульовою гіпотезою в даному випадку є передбачення про те, що «x_n»,«x_1» належать до тієї самої сукупності, як і всі інші «$n-1$» спостережень, тобто «x_n» або «x_1» не є результатом помилки спостереження або зміни загальних умов формування рівня ознаки в сукупності.

Перевірка цієї гіпотези заключається в тому, що «x_n»,«x_1» порівнюється за величиною з деякою критичною межею «x». Якщо спостереження яке виділяється є найбільше, то «x_n» порівнюється з верхньою допустимою межею, вибраною таким чином, щоб ймовірність перевершивши її дорівнювала рівню значущості, тобто в даному випадку маємо справу з критичною ділянкою виду $P(x_n > \bar{x} + t\sigma) = \alpha$. Гіпотеза «$H_0$» забраковується, якщо «$x_n$» перевершує за величиною вказану межу. Якщо ж спостереження яке виділяється є найменшим «x_1», то його порівнюють з нижньою допустимою межею, яка рівна «$\bar{x} - t\sigma$», тобто $P(x_1 < \bar{x} - t\sigma) = \alpha$. Якщо ж випробуванню одночасно підлягає і максимальне, і мінімальне значення ознаки, то критична ділянка буде мати вигляд $P(|x - \bar{x}| > t\sigma) = \alpha$.

Розглянемо приклад. Нехай маємо наступні дані (табл.11.1.).

Таблиця 11.1

Кількість спостережень	Мінімальне значення		Максимальне значення		Різниця суміжних значень		Середнє значення	Середнє квадратичне відхилення у вибірці
	x_1	x_2	$x_n - 1$	x_n	$x_2 - x_1$	$x_n - x_{n-1}$	$\bar{x}$	S
100	10	16	140	188	6	48	67,3	45,3

Використовуємо критичну ділянку виду $P(x_n > \bar{x} + t\sigma)$, так як мінімальне значення «x_1» мало відрізняється від наступного за ним значення в рангованому ряду, тобто перевіримо, чи належить спостереження яке виділяється $x_n = 188$ до сукупності результатів що розглядаються. При рівні значущості $\alpha = 0{,}01$ значення нормованої функції Лапласа для критичної ділянки яка розглядається дорівнює

$$\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt = \frac{1}{2} - 0{,}01 = 0{,}49.$$

Цьому значенню в таблиці нормованої функції Лапласа відповідає $t = 2{,}33$. Тоді верхня допустима межа значень ознаки, яка не може бути перевищена з ймовірністю 0,99, буде рівною $67{,}3 + 2{,}33 \cdot 45{,}3 = 172{,}8$. Значення $x_n = 188$ виходить за розраховану межу, а тому з ймовірністю 0,99 можна вважати $x_n = 188$ не належить до досліджуваної сукупності і повинно бути включене з подальших розрахунків.

Часто бувають такі випадки, коли параметри генеральної сукупності «$\bar{x}$» і «σ» невідомі, і для перевірки гіпотези про спостереження, які виділяються, використовуються параметри отримані при вибірковому спостереженні. Проте належить враховувати, особливо при малих вибірках, що ці оцінки є не цілком надійними. У зв'язку з цим для відкидання спостережень що виділяються за даними малих вибірок використовують критерій Ф. Груббса.

Критерій Ф. Груббса базується на відношенні двох сум квадратів відхилень.

1. Для випробування найбільшого спостереження, яке виділяється у вибірці обсягом «n» із нормально розподіленої сукупності, розраховується відношення:

$$\frac{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_n\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2},$$

де $x_1 \le x_2 \le x_3 \le \ldots \le x_n$; $\quad \bar{x}_n = \frac{\sum_{i=1}^{n-1} x_i}{n-1}$; $\quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

2. Для випробування найменшого спостереження ,яке виділяється у вибірці обсягом «n», розраховується відношення: $\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_1\right)^2}{\sum_{i=2}^{n}\left(x_i - \bar{x}\right)^2}$,

де $\bar{x}_1 = \frac{\sum_{i=2}^{n} x_i}{n-1}$.

Обчислена величина відношення ($К_р$) порівнюється з табличною величиною ($К_т$) при певному числі спостережень і завданому рівні значущості. $К_т$ характеризує ту граничну величину, яка з ймовірністю $(1-\alpha)$ пояснюється випадковими причинами. Якщо $К_р$ дорівнює або менше табличного, то найменше або найбільше спостереження відкидаються. Якщо $К_р < К_т$, то ймовірність того, що розходження в сумах квадратів відхилень пояснюється випадковими причинами, рівна рівню значущості α і в силу малої ймовірності вважається подією гранично не можливою. В такому випадку спостереження які виділяються потрібно відкинути і для подальшого обчислення використати спостереження $n-1$, що залишились.

Розглянемо застосування критерію Ф. Груббса за наступними даними: відхилення деталей від нормальної ваги (г): 0,06; 0,09; 0,10; 0,11; 0,13 0,14; 0,15; 0,16; 0,24. Потрібно визначити чи не містять результати помилки спостереження, передбачаючи, що розподіл ваги деталей в генеральній сукупності відповідає закону нормального розподілу.

Середні відхилення по всіх деталях:

$$\bar{x} = \frac{\sum x}{n} = \frac{1{,}18}{9} = 0{,}1311\,г.$$

Виключаємо максимальне відхилення 0,24:

$$\bar{x}_n = \frac{\sum x - x_n}{n-1} = \frac{1{,}18 - 0{,}24}{9-1} = \frac{0{,}94}{8}\, 0{,}1175\,г.$$

Сума квадратів відхилень від «$\bar{x}$»:

$$\sum\left(x - \bar{x}\right)^2 = 0{,}021289.$$

Сума квадратів відхилень $(n-1)$ значень від $\bar{x}_n$:

$$\sum\left(x - \bar{x}_n\right)^2 = 0{,}007952.$$

Відношення двох сум квадратів відхилень:

$$К_p = \frac{\sum(x - x_n)^2}{\sum(x - \bar{x})} = \frac{0{,}007952}{0{,}021289} = 0{,}3735.$$

Витяг із таблиці Ф. Груббса

Таблиця 11.2

Число спостережень (n)	При рівні значущості (α)		Число спостережень (n)	При рівні значущості (α)	
	0,01	0,05		0,01	0,05
3	0,0001	0,0027	…	...	...
4	0,0100	0,0494	15	0,4401	0,5559
5	0,0442	0,1270	...	...	...
6	0,0928	0,2032	20	0,5393	0,6379
7	0,1447	0,2696	...	...	...
8	0,1948	0,3261	25	0,6071	0,6923
9	0,2411	0,3742	...	...	...
10	0,2831	0,4154	...	...	...

При числі спостережень $n = 9$ і рівні значущості 1 % $К_т$ = 0,2411, тобто $K_p > K_T$ $(0{,}3735 > 0{,}2411)$, тому відхилення від номінальної ваги 0,24 г не можна віднести до помилок спостереження.

Одним із етапів статистичної обробки результатів спостережень є перевірка гіпотези про можливість віднесення емпіричного розподілу, отриманого за даними вибірки, до типу кривих нормального розподілу.

11.3. Перевірка гіпотези про відповідність емпіричного розподілу типу кривих нормального розподілу

Між частотами емпіричного і теоретичного розподілу завжди є певне розходження. В деяких випадках ці розходження не є суттєвими і пояснюються випадковостями вибірки. В інших випадках ці розходження пояснюються тим, що розподіл ознаки в генеральній сукупності не відповідає нормальному розподілу. Для того щоб дати обґрунтовану відповідь про причини розходжень емпіричних і теоретичних частот, використовуються критерії згоди.

Одним з найпоширеніших критеріїв згоди є **критерій x_i-квадрат (χ^2) К. Пірсона**:

$$\chi^2 = \sum \frac{(f_i - f_i')^2}{f_i'},$$

де f_i, f_i' – відповідно частоти емпіричного і теоретичного розподілу в певному інтервалі.

Із збільшенням різниці між фактичними і теоретичними частотами зростає величина критерію К. Пірсона. Щоб відрізнити суттєві значення «χ^2_ϕ» від значень, які виникають в результаті випадковостей вибірки, вирахуване значення критерію порівнюється з табличним значенням «χ^2_T» при відповідному числі ступенів вільності і завданим рівнем значущості. Рівень значущості вибирається таким чином, що $P\left(\chi^2_\phi > \chi^2_T\right) = \alpha$ (величина α приймається 0,01 або 0,05).

Визначаючи значення критерію К. Пірсона за даними конкретної вибірки, можна зустрітись з такими варіантами:

1. $\chi^2_\phi > \chi^2_T$, тобто χ^2_ϕ попадає в критичну ділянку. Це означає, що розходження між емпіричними і теоретичними частотами суттєве і його не можна пояснити випадковими коливаннями вибіркових даних. В такому випадку нульова гіпотеза відхиляється:

2. $\chi^2_\phi \le \chi^2_T$, тобто вирахуваний критерій не перевищує максимально можливу величину розходжень емпіричних і теоретичних частот, яка може виникнути в силу випадкових коливань вибіркових даних.

Табличне значення критерію К. Пірсона визначається при фіксованому рівні значущості і відповідному числі ступенів вільності. Число ступенів вільності дорівнює $k = m - r$,
де m – число груп; r – число обернених зв'язків.

Розглянемо розрахунок критерію χ^2 (x$_i$-квадрат) для господарів району за врожайністю озимої пшениці

Таблиця 11.3

Номер груп	Емпіричні частоти (f_i)	Теоретичні частоти (f_i')	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
1	4	2	4	2,00
2	7	11	16	1,45
3	28	25	9	0,36
4	35	31	16	0,52
5	16	21	25	1,19
6	6	8	4	0,50
7	4	2	4	2,00

Разом	100	100	х	8,02

За результатами наших розрахунків:

$$\chi^2_{\phi} = \sum \frac{\left(f_i - f_i'\right)^2}{f_i'} = 8{,}02; \qquad k = m - r = 7 - 3 = 4; \qquad \chi^2_{T(0{,}99)} = 13{,}28.$$

Оскільки фактичний критерій «χ^2_{ϕ}» значно менший «χ^2_T» ($8{,}02 < 13{,}28$), то з ймовірністю 0,99 можна вважати доведеним, що тип розподілу вибраний правильно, тобто, що розподіл господарств за врожайністю озимої пшениці є нормальним.

Критерій згоди А.Н. Колмогорова (λ) оцінює близькість фактичного розподілу до теоретичного шляхом знаходження величини (D), тобто максимальної різниці нагромаджених (кумулятивних) часток або частот фактичного і теоретичного розподілів.

Критерій А.Н. Колмогорова визначається за формулою:

$$\lambda = D \cdot \sqrt{n},$$

де D – абсолютна максимальна різниця кумулятивних часток;

$D = \max\left|S_d - S_{d'}\right|$, або частот $D = \max\left|S_f - S_{f''}\right|$ емпіричного і теоретичного розподілів;

n – число спостережень (чисельність одиниць сукупності).

Якщо розподіл задано в частотах, тоді формула матиме вигляд:

$$\lambda = \frac{D}{\sqrt{n}}.$$

Методику розрахунку цього показника розглянемо на прикладі даних табл.11.3.

Таблиця 11.3.

Номер групи	Нагромаджені частоти		Відхилення $\left\|S_f - S_{f'}\right\|$
	Емпіричні S_f	Теоретичні $S_{f'}$	
1	4	2	2
2	11	13	2
3	39	38	1
4	74	69	5
5	90	90	0
6	96	98	2

7	100	100	0

$$\lambda = \frac{D}{\sqrt{100}} = \frac{5}{10} = 0{,}5.$$

По спеціальній таблиці ймовірностей для критерію згоди «λ» знаходимо, що значенню $\lambda = 0{,}5$ відповідає ймовірність 0,9639, це означає, що з ймовірністю 0,9639 можна стверджувати про нормальний розподіл господарств за врожайністю озимої пшениці.

Критерій згоди В.І. Романовського (R) також використовується для оцінки наближення фактичного розподілу до теоретичного. Він визначається за формулою:

$$R = \frac{\chi^2 - k}{\sqrt{2k}}.$$

Скориставшись розрахунками прикладу за даними табл. 11.3, визначаємо критерій згоди В.І. Романовського.

$$R = \frac{\chi^2 - k}{\sqrt{2k}} = \frac{8{,}02 - 4}{\sqrt{2 \cdot 4}} = \frac{4{,}02}{2{,}8284} = 1{,}42 < 3.$$

Якщо при дослідженні наближення фактичного розподілу до теоретичного величина цього виразу менша трьох (1,42<3), це дає підставу для стверджування про можливість розподілу за законом даного розподілу. Тобто розподіл господарств за врожайністю озимої пшениці є нормальним.

Критерій згоди П.С. Ястремського (L**)** використовується для прямої відповіді на питання про міру розбіжності між фактичним і теоретичним розподілом. Для визначення критерію П.С. Ястремського використовується критерій К. Пірсона. В загальному вигляді критерій П.С. Ястремського має вигляд:

$$L = \frac{\chi^2 - n}{\sqrt{2n + 4Q}},$$

де $Q = \frac{(f_i - f_i')^2}{f_i(1 - P_i)}$ – при кількості груп менша 20 ($n < 20$), дорівнює 0,6.

За даними попередніх прикладів покажемо розрахунок цього показника:

$$L = \frac{\chi^2 - n}{\sqrt{2n + 4Q}} = \frac{8{,}02 - 7}{\sqrt{2 \cdot 7 + 4 \cdot 0{,}6}} = \frac{1{,}02}{4{,}02} = 0{,}25.$$

Так як величина $L < 3(0{,}25 < 3)$, то з ймовірністю 0,997 можна стверджувати, що розподіл господарств за врожайністю озимої пшениці є нормальним.

Всі розглянуті нами критерії згоди дають загальну оцінку ступеня відхилення емпіричного розподілу від нормального, але не визначають його характеру, а тому при суттєвих їх відхиленнях аналіз розподілу доцільно доповнювати характеристиками асиметрії і ексцесу.

Таким чином, для перевірки висунутої гіпотези про відповідність чи невідповідність теоретичного закону розподілу емпіричному можна використати любий із наведених критеріїв, які забезпечують дослідження законів розподілу з різною точністю, надійністю і трудоємкістю.

11.4. Перевірка гіпотези про величину центру розподілу

Припустимо, що ми маємо дані про витрачання повного виду сировини на одиницю продукції до і після модернізації устаткування. До модернізації устаткування таке витрачання становило 30 кг, після модернізації – 28 кг. Середня помилка вибірки 1 кг. Потрібно встановити, чи привела модернізація устаткування до зниження витрат сировини на одиницю продукції.

Нульова гіпотеза заключається в тому, що виробництво продукції до і після модернізації устаткування з точки зору впливу на матеріалоємкість суттєво не змінилось, тобто що між генеральними середніми до і після модернізації устаткування немає суттєвої різниці. Тобто нульова гіпотеза означає, що $\bar{z}_0 = \bar{z}_1$, де $\bar{z}_0$ і $\bar{z}_1$ – середнє витрачання сировини на одиницю продукції відповідно до і після модернізації устаткування.

Альтернативна гіпотеза в даному випадку може бути сформульована двояко.

1. Модернізація устаткування веде до зміни витрачання сировини на одинцю продукції, тобто нульова гіпотеза заключається в тому, що $\bar{z}_1 \neq \bar{z}_0$. Приймемо рішення значущості $\alpha = 0{,}05$, тоді $P\left(\left|\bar{z}_1 - \bar{z}_0\right|\right) \geq t_{\mu \bar{z}} = 0{,}05$, і критична ділянка задається нерівністю $\left|\bar{z}_1 - \bar{z}_0\right| > t_{\mu \bar{z}}$. За таблицями інтегральної функції Лапласа використовуємо коефіцієнт довір'я $t = 1{,}96$. Таким чином, розмір граничного

розходження двох середніх з ймовірністю 0,95, не повинен перевищувати $t_{\mu\bar{z}} = 1{,}96(1{,}96 \cdot 1{,}00)$. Отже, з ймовірністю 0,95 можна стверджувати, що довірчі межі для генеральної середньої дорівнюватимуть: $28{,}04 \le \bar{z}_1 \le 31{,}96$. Середнє витрачання матеріалів після модернізації устаткування складає 28 кг і попадає в критичну ділянку. Дані спостереження не співпадають з висунутою нульовою гіпотезою про те, що між виробництвом продукції до і після модернізації устаткування відсутні суттєві розбіжності з точки зору їх впливу на матеріалоємкість.

2. Модернізація устаткування веде до зниження витрат сировини на виробництво одиниці продукції, тобто нульова гіпотеза заключається в тому, що $\bar{z}_1 < \bar{z}_0$. В цьому випадку розглядається ділянка великих від'ємних відхилень, тобто при $\alpha = 0{,}05$ ($\bar{z}_1 < \bar{z}_0 - t_{\mu\bar{z}}$)=0,05. В цьому варіанті критична ділянка визначається нерівністю $\bar{z}_1 < \bar{z}_0 - t_{\mu\bar{z}}$. Нульова гіпотеза не буде відхилятись, якщо середні витрати матеріалу на одиницю продукції будуть більші за 28,36 кг ($30 - 1 \cdot 1{,}64$). Після модернізації устаткування витрачання сировини на одиницю продукції склало 28 кг, тобто з ймовірністю 0,995 можна стверджувати, що нульова гіпотеза повинна бути відхилена і що модернізація устаткування привела до зниження витрачання сировини на виготовлення продукції.

Якщо гіпотеза про величину центру розподілу перевіряється за результатами малої вибірки, то потрібно врахувати, що відношення різниці середніх до стандартної помилки вибірки $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\mu_{\bar{x}}}$ має розподіл Стьюдента з *n*-1 ступенями волі. Припустимо, що дані про витрачання сировини на одиницю продукції були отримані за результатами перевірки 17 виробів. За результатами спостережень значення емпіричного середнього кореляційного відношення склало 5,0. Обчислимо величину t-критерію за формулою:

$$t_{\phi} = \frac{|\bar{z}_1 - \bar{z}_0|}{\dfrac{S}{\sqrt{n-1}}} = \frac{|28-30| \cdot 4}{5} = 1{,}6.$$

За таблицею розподілу Стьюдента значення t-критерію для числа для числа ступенів вільності (n-1)=16 і рівня значущості 0,05 дорівнює $t_T = 2{,}12$. Так як фактичне значення не перевищує табличне [$t_ф < t_T \left(1{,}6 < 2{,}12\right)$], то нульова гіпотеза не відхиляється, тобто у нас відсутні достатні підстави вважати, що модернізація устаткування веде до зниження матеріалоємності.

На наступному прикладі розглянемо перевірку гіпотези про суттєвість різниці двох вибіркових середніх (для випадку малої вибірки).

Маємо дані про живу вагу телят дослідної і контрольної групи в 3-місячному віці (табл. 11.5.).

Таблиця 11.5

Спостереження	Жива вага телят, кг	
	дослідна група, x_1	контрольна група, x_2
1	94	66
2	99	74
3	108	78
4	115	76
5	99	69
6	100	75
7	128	73
8	96	77
9	110	70
10	111	72
Разом	1060	730

Потрібно встановити різницю між двома середніми живої ваги телят в дослідній і контрольній групах і в якій мірі ця різниця викликана згодовуванням в дослідній групі крім незбираного молока ще й канцкормів.

Визначимо середню вагу телят по групах:

$$\tilde{x}_1 = \frac{\sum x_1}{n} = \frac{1060}{10} = 106\,кг;\ \ \tilde{x}_2 = \frac{\sum x_2}{n} = \frac{730}{10} = 73\,кг.$$

Знайдемо різницю між середніми двох вибірок:

$$\Delta_ф = \tilde{x}_1 - \tilde{x}_2 = 106 - 73 = 33\,кг\,.$$

Обчислимо середнє квадратичне відхилення ваги для кожної групи телят:

$$S_1 = \sqrt{\frac{\Sigma\left(x_1 - \tilde{x}_1\right)^2}{n_1 - 1}} = \sqrt{\frac{988}{9}} = 10{,}45\,кг;$$

$$S_2 = \sqrt{\frac{\sum(x_2 - \tilde{x}_2)^2}{n_2 - 1}} = \sqrt{\frac{130}{9}} = 3{,}80\,кг.$$

Розрахуємо середні помилки вибіркових середніх по групах:

$$\mu_1 = \frac{S_1}{\sqrt{n_1}} = \frac{10{,}45}{3{,}162} = 3{,}3\,кг; \qquad \mu_2 = \frac{S_2}{\sqrt{n_2}} = \frac{3{,}8}{3{,}162} = 1{,}2\,кг.$$

Середню помилку різниці двох вибіркових середніх визначаємо за формулою:

$$\overline{\mu}_{1-2} = \sqrt{\mu_1^2 + \mu_2^2} = \sqrt{3{,}3^2 + 1{,}2^2} = \sqrt{12{,}33} = 3{,}51.$$

Число ступенів вільності двох вибірок буде становити:

$$(n_1 - 1) + (n_2 - 1) = 9 + 9 = 18.$$

При рівні значущості 0,05 і 18 ступенях вільності:

$$t_T = 2{,}1009 \approx 2{,}1.$$

Гранична помилка для двох вибіркових середніх становитиме:

$$\Delta_{0,05} = t \cdot \overline{\mu}_{1-2} = 2{,}1 \cdot 3{,}51 = 7{,}37\,кг.$$

Порівнявши фактичну різницю між обома середніми $\Delta_ф = 33\,кг,$ з граничною похибкою $\Delta_{0,05} = 7{,}37\,кг,$ бачимо, що перша значно перевищує другу. Це свідчить про те, що різниця в середній вазі телят в дослідній і контрольній групах зумовлена дією досліджуваного чинника.

До такого самого висновку можна прийти зіставивши фактичне нормоване відхилення з табличним. Для нашого прикладу фактичне нормоване відхилення становить:

$$t_ф = \frac{\Delta_ф}{\overline{\mu}_{1-2}} = \frac{33{,}0}{3{,}51} = 9{,}4.$$

Табличне нормоване відхилення $t_T = 2{,}1$ показує максимальну величину відношень граничних помилок вибіркових середніх до їх середньої помилки. В нашому прикладі фактичне нормоване відхилення значно перевищує табличне, а тому можна зробити висновок, що різниця ваги двох середніх є не випадковою, а цілком достовірна.

11.5. Елементи дисперсійного аналізу

Основною метою дисперсійного аналізу є виявлення на основі загальної дисперсії впливу окремих чинників чи умов, які визначають варіацію ознаки. Для оцінки частки варіації, зумовленої тією чи іншою ознакою, сукупність розподіляють на групи за ознакою, вплив якої досліджується. Це дозволяє розкласти загальну варіацію ознаки на дві дисперсії, з яких одна частина варіації визначається впливом чинника, закладеного в основу групування, а друга – варіацією, зумовленою впливом усіх інших чинників, крім того, що вивчається. Отже, згідно з правилом складання дисперсій для розрахунку використовують загальну, міжгрупову і внутрішньогрупову (залишкову) дисперсії.

Загальна дисперсія характеризує варіацію ознаки у статистичній сукупності в результаті впливу усіх чинників, міжгрупова – вплив чинника, покладеного в основу групування, залишкова – впливом усіх інших чинників.

Таким чином, загальна дисперсія результативної ознаки складається з двох частин: міжгрупової та середньої з групових (залишкової).

Для кількісної ознаки зв'язку між явищами на базі матеріалів аналітичного групування вираховують коефіцієнт детермінації і емпіричне кореляційне відношення.

Коефіцієнт детермінації показує ступінь варіації ознаки під впливом чинника покладеного в основу групування. Він визначається як відношення міжгрупової дисперсії до загальної.

$$\eta^2 = \frac{\delta^2}{\sigma_з^2},$$

де η^2 – коефіцієнт детермінації;

δ^2 – міжгрупова дисперсія;

$\sigma_з^2$ – загальна дисперсія.

Критерієм суттєвості і сили зв'язку між факторною і результативною ознаками виступає **емпіричне кореляційне відношення**, яке визначається за формулою:

$$\eta = \frac{\delta^2}{\sigma_з^2}.$$

Розглянемо приклад. Нехай маємо згруповані дані (5 груп) за стажем робітників-відрядників і їх місячною виробіткою (табл. 11.6).

Таблиця 11.6

Розрахунок міжгрупової дисперсії

Групи робітників-відрядників за стажем роботи, років		Число робітників, чол. f_i	Середня місячна виробітка одного робітника, тис. грн. $\bar{y}_i$	$\bar{y}_i - \bar{y}_з$	$(\bar{y}_i - \bar{y}_з)^2$	$(\bar{y}_i - \bar{y}_з)^2 f_i$
I	1-4	3	154,67	-27,93	780,08	2340,25
II	4-7	6	171,83	-10,77	115,99	695,96
III	7-10	5	191,00	8,40	70,56	352,80
IV	10-13	4	193,50	10,90	118,81	475,24
V	13-16	2	214,00	31,40	985,96	1971,52
Разом:		20	$\bar{y}_з = 182,60$	x	x	5936,17

Міжгрупову дисперсію визначаємо за формулою:

$$\delta^2 = \frac{\Sigma(\bar{y}_i - \bar{y}_з)^2 f_i}{\sum f_i} = \frac{5936,17}{20} = 291,81.$$

Обчислимо загальну дисперсію за формулою:

$$\sigma_з^2 = \overline{y_2} - (\bar{y})^2 = 33666,40 - 33342,76 = 323,64.$$

Визначимо коефіцієнт детермінації і емпіричне кореляційне відношення:

$$\eta^2 = \frac{\delta^2}{\sigma_з^2} = \frac{291,81}{323,64} = 0,9, \text{ або } 90\ \%;$$

$$\eta = \sqrt{\frac{\delta^2}{\sigma_з^2}} = \sqrt{\frac{291,81}{323,64}} = \sqrt{0,9} = 0,95.$$

Коефіцієнт детермінації показує, що середня місячна виробітка робітників-відрядників на 90 % залежить від стажу їх роботи і на 10 % від інших чинників.

Емпіричне кореляційне відношення свідчить про те, що зв'язок між стажем роботи і середньою місячною виробіткою робітників-відрядників дуже сильний.

Емпіричне кореляційне відношення повинно мати високий рівень надійності. Для оцінки надійності кореляційних характеристик використовують критерій Р. Фішера (F-критерій), або Стьюдента (t-критерій).

Критерій Р. Фішера (F-критерій) визначається за формулою:

$$F_{\phi} = \frac{\delta^2}{\overline{\sigma_i^2}} \cdot \frac{k_2}{k_1}, \text{ або } F_{\phi} = \frac{\delta^2 \cdot k_2}{\overline{\sigma_i^2} \cdot k_1},$$

де δ^2 – міжгрупова дисперсія;

$\overline{\sigma_i^2}$ – середня групова (залишкова) дисперсія;

k_1, k_2 – ступені вільності для великої і малої дисперсії.

Р. Фішер встановив розподіл відношень дисперсій і розробив відповідні математичні таблиці. В них наводиться F-критерій теоретичний (F_T) при двох ймовірностях 0,95 і 0,99. Якщо $F_{\phi} > F_T$, то з прийнятим ступенем ймовірності можна стверджувати про наявність впливу чинника, який вивчається. Коли ж $F_{\phi} \le F_T$, можна стверджувати, що різниця між дисперсіями обумовлена впливом випадкових чинників.

Розподіл в таблицях Фішера для знаходження F_T залежить від ступенів вільності міжгрупової (k_1) і середньої з групових (k_2) дисперсій. В аналітичному групуванні вони обчислюються за формулами:

$$k_1 = m - 1; \qquad k_2 = n - m,$$

де n – кількість елементів досліджуваної сукупності;

m – число груп.

За даними нашого прикладу:

$$\overline{\sigma_i^2} = \sigma_з^2 - \delta^2 = 323{,}64 - 291{,}81 = 31{,}83.$$

Тоді $F_{\phi} = \frac{\delta^2}{\overline{\sigma_i^2}} \cdot \frac{k_2}{k_1} = \frac{291{,}81 \cdot 15}{31{,}83 \cdot 4} = 34{,}38.$

Для оцінки отриманого відношення, його порівнюють з табличним.

Ступінь вільності для великої (міжгрупової) дисперсії $k_1 = m - 1 = 5 - 1 = 4,$ для малої (внутрішньогрупової) – $k_2 = n - m = 20 - 5 = 15.$

Знаходимо F_T при ймовірності 0,95 і даних ступенях вільності по математичній таблиці Р. Фішера. Воно становить $F_{T(0,95)} = 3{,}06$.

Таким чином $F_{ф} > F_{T}$ $(34,38 > 3,06)$, що свідчить про суттєвість впливу стажу роботи робітників-відрядників на середньомісячну виробітку.

До аналогічного висновку можна прийти за оцінкою надійності кореляційного відношення за критерієм Стьюдента (t-критерієм), який визначається за формулою:

$$t_{\eta} = \frac{\eta}{\mu_{\eta}},$$

де μ_{η} – середня похибка кореляційного відношення.

Вона визначається за формулою:

$$\mu_{\eta} = \frac{1-\eta^2}{\sqrt{\eta}}.$$

Якщо критерій Стьюдента дорівнює або більший $3 (t_{\eta} \geq 3)$ показник кореляційного відношення вважають вірогідним (тобто зв'язок між досліджуваними явищами є доведеним). Якщо ж цей критерій менший $3 (t_{\eta} < 3)$, то не можна зробити висновок про вірогідність зв’язку між досліджуваними явищами.

Для нашого прикладу:

$$\mu_{\eta} = \frac{1-\eta^2}{\sqrt{\eta}} = \frac{1-0,9}{\sqrt{0,95}} = 0,1026; \qquad t_{\eta} = \frac{\eta}{\mu_{\eta}} = \frac{0,9487}{0,1026} = 9,25.$$

Так як критерій Стьюдента значно більший за 3, то кореляційне відношення вважається вірогідним, а зв'язок між стажем роботи робітників-відрядників і середньомісячною виробіткою доведеним.

Тематичний план з курсу «Статистика»

Номер теми	Назва теми	Кількість годин	
		лекційних	практичних
Тема 1.	Статистика як наука. Статистичне спостереження.	4	2
Тема 2.	Зведення і групування статистичних даних. Статистичні таблиці.	6	4
Тема 3.	Абсолютні, відносні та середні величини.	4	6
Тема 4.	Показники варіації.	4	4
Тема 5.	Ряди розподілу.	6	6
Тема 6.	Статистичні методи вимірювання взаємозв’язку.	6	6
Тема 7.	Ряди динаміки.	6	6
Тема8.	Індекси.	4	6
Тема 9.	Вибіркове спостереження.	4	6
Тема 10.	Статистичні графіки.	4	4
Тема 11.	Статистична перевірка гіпотез.	4	4
Разом:	–	48	48

Глосарій

1. **Абсолютний показник** – це показник у формі абсолютної величини, яка відображає фізичні властивості, часові та вартісні характеристики соціально-економічних процесів та явищ.
2. **Агрегатний індекс** – складний відносний показник, що характеризує середню зміну соціально-економічних явищ, але досліджується з декількох видів одиниць (однорідних або неоднорідних).
3. **Відносний показник** – показник у формі відносної величини; результат порівняння одного абсолютного показника з іншими; характеризує співвідношення між кількісними характеристиками процесів і явищ, що вивчаються, чи міру кількісного співвідношення різнойменних чи однойменних показників.
4. **Групування в статистиці** – це розподіл одиниць об'єкта спостереження на однорідні групи за суттєвими для них ознаками.
5. **Зведений (загальний) індекс** – показник, що характеризує динаміку складного явища, елементи якого не підлягають безпосередньому підсумовуванню в часі, просторі чи порівняно з планом.
6. **Звітність** – це форма спостереження, при якій кожний об'єкт діяльності подає свої дані в державні органи статистик та відомств у вигляді документів, звітів соціально затвердженої теми.
7. **Зведення** – це науково організована обробка матеріалів спостереження (згідно з наперед розробленою програмою), яка містить систематизацію, класифікацію або групування матеріалу, складання таблиць, отримання підсумкових результатів і похідні показники (середні, відносні величини і т.д.).
8. **Зведення** – це комплекс послідовних операцій, спрямованих на упорядкування первинного статистичного матеріалу з метою виявлення характерних рис та певних типових знак тих чи інших явищ, а також закономірностей процесі, що досліджуються.
9. **Індекс** – це відносний показник, що характеризує зміну рівня будь-якого явища чи процесу в часі, просторі або порівняно з планом, нормою, стандартом.
10. **Основна тенденція (тренд)** – це достатньо стійка зміна рівня явища в часі, більш-менш вільна від випадкових коливань. Основну тенденцію можна подати аналітично – у вигляді рівняння (моделі, тренду або графічно.
11. **Показник варіації** – це ті показники, що вивчають міру варіації (коливання) окремих значень ознаки від середньої величини.
12. **Ряд динаміки** – це розміщення в часі значення явища, тобто послідовність чисел, які характеризують зміни розмірів суспільних явищ в часі.
13. **Ряд динаміки** – це ряд статистичних показників, що розташовані в хронологічній послідовності і характеризують зміну явища в часі.
14. **Ряд розподілу** – це упорядкований розподіл одиниць сукупності на групи за певною варіаційною ознакою.
15. **Середня величина** – це узагальнюючий показник, що характеризує рівень варіаційної ознаки в якісно однорідній сукупності.

16. **Середня величина** – це узагальнююча характеристика сукупності явищ за ознакою, що варіює, тобто це узагальнюючий показник, який характеризує типовий рівень ознаки, що варіює, в розрахунку на одиницю сукупності.

17. **Спеціально організоване спостереження** – це форма спостереження, яка охоплює сфери життя та діяльності, що не відображаються звітністю. до числа таких спостережень належать: переписи, обліки, соціальні обстежування, опитування.

18. **Статистика** – це суспільна наука, що вивчає кількісний бік якісно визначених масових соціально-економічних явищ та процесів, їх структуру та розподіл, розміщення в просторі, руху в часі, досліджує діючі кількісні залежності, тенденції та закономірності в конкретних умовах місця та часу.

19. **Статистика** – це наука, яка вивчає кількісну сторону масових суспільних, соціально-економічних та інших явищ у нерозривному зв'язку з їх якісними показниками в певних умовах місця і часу.

20. **Статистична закономірність** – це форма виявлення причинного зв'язку, який знаходить відображення в послідовності, регулярності, повторювальності подій з достатньо високим ступенем імовірності, якщо причини, що породжують подію, не змінюються, або змінюються неістотно.

21. **Статистична інформація** – це сукупність статистичних даних, що відображають соціально-економічні процеси і які використовуються в процесі управління економікою.

22. **Статистична інформація** – це первинний статистичний матеріал, який формується в процесі статистичного спостереження, групується, аналізується і на основі якого роблять висновки.

23. **Статистична методологія** – це комплекс спеціальних, притаманних лише статистиці, методів, засобів дослідження. Вона ґрунтується на загальних філософських (діалектична логіка) і загальнонаукових (порівняння, аналіз, синтез) принципах.

24. **Статистична сукупність** – це маса певним чином однорідних елементів, які мають єдину якісну основу, але різняться між собою певними ознаками і підлягають певному закону розподілу.

25. **Статистична сукупність** – це певна множина елементів, поєднана умовами існування і розвитку.

26. **Статистична таблиця** – це форма найбільш раціонального, наочного та систематизованого викладання результатів зведення і групування матеріалів статистичного спостереження.

27. **Статистичне спостереження** – це спланована, науково організаційна реєстрація масових даних про будь-які соціально-економічні явища та процеси.

28. **Статистичні показники** – це число в сукупності з набором ознак, що характеризують обставини, яких вони стосуються, що, до, коли, і яким чином підлягає вимірюванню.

29. **Статистичний показник** – це кількісна характеристика соціально-економічних явищ і процесів в умовах якісної визначеності.

ЗМІСТ

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Адамов В.Е. Факторный индексный анализ. – М.: Статистика, 1987. – 199 с.
2. Аллен Р. Экономические индексы. – М.: Статистика, 1980. – 256 с.
3. Бек В.Л. Теорія статистики: Курс лекцій. Навчальний посібник. – Київ: ЦІЛ, 2003. – 288 с.
4. Вайну Я.Я. – Ф. Корреляция рядов динамики. – М.: Статистика, 1977. – 119 с.
5. Головач А.В., Черноскулова З.А. Экономико-статистический анализ потребления и спроса. – К.: «Техника», 1978. – 184 с.
6. Герчук Я.П. Графические методы в статистике. – М.: Статистика, 1972. – 78 с.
7. Єрина А.М., Пальян З.О. Теорія статистики: Практикум. – К.: Товариство «Знання», КОО, 1997. – 432 с.: іл.
8. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. – М.: ИНФРА. – М. 1998. – 416 с.
9. Збірник задач з теорії статистики: Навчальний посібник / Є.І.Ткач, І.М.Шост, З.О.Насінник і ін. / За ред. Є.І.Ткача. – Тернопіль: Друк “Лідер”, 2003. – 72 с.
10. Кендэл М. Временные ряды. – М.: Финансы и статистика, 1981. – 199 с.
11. Кендэл М. Ранговые корреляции. – М.: Статистика, 1975. – 241 с.
12. Кильдишев Г.С., Аболонцев Ю.И. Многомерные группировки. – М.: Статистика, 1978. – 160 с.
13. Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. – М.: Статистика, 1973. – 103 с.
14. Ковтун Н.В., Столяров Г.С. Загальна теорія статистики: Курс лекцій. – К.: Четверта хвиля, 1996. – 144 с.: іл.
15. Копрен Г. Методы выборочного исследования. – М.: Статистика, 1976. – 440 с.
16. Кулинич О.І. Теорія статистики: Підручник. 2-е доп. и дооп. видання. – К. – д.: Державне Центрально-Українське видавництво, 1996. – 228 с.

17. Липкин М.И. Кривые распределения в экономических исследованиях. – М.: Статистика, 1972. – 114 с.
18. Лугінін О.Є., Білоусова С.В. Статистика: Підручник.- К.: Центр навчальної літератури, 2005.- 580с.
19. Лігунун О.Є., Фомішин С.В. Статистика навчальної економіки та світового господарства: Навчальний посібник.- К.: Центр навчальної літератури, 2006.- 502с.
20. Мармоза А.Т. теорія статистики: Навчальний посібник.- К.: Ельга, Ніка-Центр, 2003.- 392с.
21. Михель В.М. Динамические ряды: Учеб. пособие. – М.: МИНХ, 1969. – 95 с.
22. Михок Г., Урсяну В. Выборочный метод и статистическое оценивание. – М.: Статистика, 1982. – 245 с.
23. Монстеллер Ф., Тьюки Д.К. Анализ данных и регрессия. – М.: Финансы и статистика, 1982. – 239 с.
24. Общая теория статистики: Статистическая методология в изучении комплексной деятельности: Учебник / А.И.Харламов, О.Э.Башина, В.Т.Бабурин и др.: под ред. А.А.Спирина, О.Э.Башиной. – М.: Финансы и статистика, 1996. – 296 с.: ил.
25. Общая теория статистики: Учебник / Г.С.Кильдишев, В.Е.Овсиенко, П.М.Рабинович, Т.В.Рябушкин. – М.: Статистика, 1980. – 423 с.
26. Опря А.Т. Статистика (з програмованою формою контролю знань). Математична статистика. Теорія статистики. Навчальний посібник.- К.: Центр навчальної літератури, 2005.- 472с.
27. Пасхавер И.С. Средние величины в статистике. – М.: Статистика, 1979. – 279 с.
28. Ряузов Н.Н. Общая теория статистики: Учебник. – М.: Финансы и статистика, 1984. – 343 с.
29. Сборник задач по общей теории статистики: Учеб. пособие / В.Е.Голованова, Ю.Г.Королев и др. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 1986. – 191 с.

30. Сигел, Ендрю. Практическая бизнес-статистика.: Пер. с англ.–М.: Издательский дом «Вильямс», 2004.- 1056с.: ил.- Парал. Тит. англ.
31. Статистика підприємництва: Підручник / П.Г.Вашків, П.І.Пастер, В.П.Сторожук, Є.І.Ткач: за ред. П.Г.Вашкова, В.П.Сторожука. – К.: "Слобожанщина", 1999. – 600 с.
32. Статистика: Навч.-метод. посібник для самост. вивч. дисц. / А.М.Єріна, Р.М.Моторин, А.В.Головач та ін.: За заг. ред. А.М.Єріної, Р.М.Моторина. – К.: КНЕІ, 2001. – 448 с.
33. Статистика: Підручник / А.В.Головач, А.М.Єріна, О.В.Козирев та ін. / За ред. А.В.Головача. – К.: Вища школа, 1993. – 623 с.
34. Статистика: Підручник / С.С.Герасименко, А.В.Головач, А.М.Єріна та ін. / За наук. ред. д-ра екон. наук С.С.Герасименка. – 2-е вид., перероб. и доп. – К.: КНЕОІ, 2000. – 467 с.
35. Статистика: теоретичні засади і прикладні аспекти. Навчальний посібник. / Р.В.Фещур, А.Ф.Барвінський, В.П.Кічор та ін.; За наук. ред. Р.В.Фещура. – 2-е вид. оновлене і доповнене. – Львів: «Інтелект-Захід», 2003. – 576 с.
36. Теорія статистики: Навчальний посібник / Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. – К.: Либідь, 2001. – 320 с.
37. Ткач Є.І. Загальна теорія статистики: Підручник. – Тернопіль: Лідер, 2004. – 388 с.
38. Уманець Т.В. Загальна теорія статистики: Навч. посіб. -: Знання, 2006.- 239с.
39. Уманець Т.В., Пігарєв Ю.Б. Статистика: Навч. посіб. – К.: Вікар, 2003. – 623 с.
40. Урланис Б.Ц. Общая теория статистики: Учебник. – М.: Статистика, 1973. – 439 с.
41. Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1977. – 200 с.
42. Штангрет А.М., Копилюк О.І. Статистика: Навчальний посібник. – К.: Центр навчальної літератури, 2005. – 232 с.

43. Юл. Дж., Кендэл М. Дж. Теория статистики. – М.: Госполитиздат, 1960. – 779с.