А. Т. МАРМОЗА

ТЕОРІЯ СТАТИСТИКИ

Допущено Міністерством аграрної політики України
як навчальний посібник
для підготовки фахівців економічних спеціальностей

Київ
Ельга
Ніка-Центр
2003

ББК 60.6я73
~~М25~~ М28

Допущено Міністерством аграрної політики України як навчальний посібник для підготовки фахівців економічних спеціальностей (лист №18-2-1-13/482 від 15 травня 2003 р.)

Мармоза А. Т.

М25 Теорія статистики.- К.: Ельга, Ніка-Центр, 2003.– 392 с.

ISBN 966-521-225-7

Підручник написаний відповідно до діючої програми курсу «Статистика» для студентів економічних спеціальностей сільськогосподарських вузів. В ньому висвітлено предмет і метод статистики, статистичне спостереження, зведення і групування даних, абсолютні, відносні, середні величини, показники варіації, ряди динаміки, індекси, вибіркове спостереження, дисперсійний і кореляційний аналіз, табличний і графічний методи та ін.

Теоретичний матеріал проілюстрований конкретними розрахунками, схемами, графіками, статистичними таблицями та алгоритмами розв'язання прикладів.

У підручнику вміщено основні математичні таблиці, список рекомендованої літератури.

Для студентів сільськогосподарських вузів, може бути корисним викладачам, аспірантам і науковцям, а також працівникам економічних служб аграрного сектора і статистичних органів

ББК 60.6я73

Рецензенти:

Савчук В.К. – професор Національного аграрного університету, доктор економічних наук.

Липчук В.В. – професор Львівського державного аграрного університету, доктор економічних наук.

Поліщук М.П. – професор Державного агроекологічного університету, доктор економічних наук.

Степаненко М.В. – доцент Національного аграрного університету, кандидат економічних наук.

ISBN 966-521-225-7
9789665212256

Присвячується пам'яті мого вчителя –
видатного вченого-статистика, академіка,
доктора економічних наук, професора
СЕРГІЯ СТЕПАНОВИЧА СЕРГЄЄВА

Передмова

Зростаючий інтерес до статистики в країні викликаний сучасним розвитком економіки, формуванням ринкових відносин, розвитком різноманітних форм господарювання, здійсненням економічних реформ, які зачіпають інтереси кожної людини. У статистичних даних, що відбивають хід цих процесів і які є інформаційною базою для здійснення відповідних управлінських рішень, кожний з нас шукає результати реформ.

Значна роль у вирішенні перелічених завдань належить соціально-економічній статистиці, яка покликана за допомогою системи об'єктивних статистичних показників вірогідно і точно охарактеризувати стан і розвиток суспільного виробництва і соціальної сфери життя, рівень і повноту використання наявних ресурсів (землі, робочої сили, виробничих фондів), економічну ефективність виробництва продукції тощо.

Перед статистикою поставлені важливі завдання щодо подальшого вдосконалення системи статистичних показників, прийомів і методів збирання, *обробки та аналізу масових статистичних даних*, забезпечення усіх рівнів управління народним господарством вичерпною, вірогідною і точною інформацією.

Усе це ставить підвищені вимоги до статистичної підготовки економічних кадрів. Стосовно підготовки економістів та бухгалтерів вищої кваліфікації у вищих аграрних навчальних закладах освіти, це означає підвищення рівня статистичної освіти, зокрема з питань теорії статистики і сільськогосподарської статистики.

Статистична підготовка економістів-аграрників усіх спеціальностей є важливою складовою частиною їх методологічної підготовки у галузі якісно-кількісного аналізу масових процесів і явищ. Вона забезпечується вивченням курсу «Статистика», який включає математичну і загальну теорію статистики і сільськогосподарську статистику з основами соціально-економічної статистики.

Застосування статистичних методів у аналізі сільськогосподарського виробництва дає змогу всебічно вивчити явище, встановити залежність результативного показника від комплексу факторів, виявити тенденції зміни досліджуваних явищ і прогнозувати їх розвиток, виявити резерви підвищення економічної ефективності сільськогосподарського виробництва тощо.

При підготовці навчального посібника враховані основні положення Закону України «Про державну статистику» та інших законодавчих документів щодо статистики, а також останні зміни, що відбулися в обліку і звітності по агропромисловому комплексу.

Навчальний посібник написаний відповідно до діючої програми курсу «Статистика» для студентів економічних спеціальностей вищих аграрних навчальних закладів освіти.

В ньому системно викладаються загальні категорії і методи статистичної науки, теоретичні основи економіко-статистичних методів аналізу масових даних. Тут послідовно розглядаються питання, що виникають на стадії статистичного спостереження, зведення первинного матеріалу та його наступної обробки і аналізу. За цією ж схемою побудований навчальний посібник. В ньому викладаються предмет і метод статистики як суспільної науки, система статистичних показників і способи їх одержання, основні статистичні методи аналізу масових даних: статистичне спостереження, зведення і групування, абсолютні і відносні показники, статистичні таблиці і графіки, середні величини і показники варіації, прийоми аналізу рядів динаміки, індексний, вибірковий, дисперсійний і кореляційний методи тощо.

Наприкінці навчального посібника вміщено основні математично-статистичні таблиці і подано список рекомендованої літератури для самостійного вивчення курсу статистики.

Головне завдання для автора цього навчального посібника полягає у тому, щоб у доступній і логічній формі розкрити змістовну сторону методів статистики, тобто пояснити особливості і специфіку наукового їх застосування, застерегти від можливих помилок. Тому особлива увага приділена економічній інтерпретації одержаних результатів і висновків по них, що сприяє творчому засвоєнню матеріалу, що викладається. Зважаючи на велике прикладне значення окремих статистичних методів аналізу, деякі теми викладені більш поглиблено і ширше. Це насамперед стосується статистичних групувань, вибіркового, дисперсійного і кореляційного аналізу.

У навчальному посібнику врахований характер роботи економістів, бухгалтерів і менеджерів сільськогосподарського виробництва. Тому значна увага приділена питанням конкретного статистичного аналізу масових даних, взаємозв'язку показників, залежності результативних показників від комплексу факторів, виявленню внутрішньогосподарських резервів і т.п.

Наведені в навчальному посібнику статистичні дані є в основному умовними. Проте автор прагнув, щоб зміст завдань мав наближений до реальної дійсності характер, а тому, складаючи їх, значною мірою використовував фактичний матеріал, попередньо обробивши його методично.

Приклади, які дає навчальний посібник, не лише роз'яснюють загальнотеоретичні положення, а й наочно показують можливі області застосування статистичних методів у аналізі масових даних по народному господарству.

У більшості прикладів використовується один і той самий вихідний статистичний матеріал, що дає змогу всебічно охарактеризувати досліджувану сукупність або явище за допомогою комплексу статистичних показників і характеристик.

При викладенні матеріалу автор прагнув показати, що статистика не є нудною і важкою наукою, як інколи її уявляють, вона може надавати задоволення і принести користь.

Навчальний посібник призначається для викладачів і студентів економічних спеціальностей вищих аграрних навчальних закладів освіти II-IV рівнів акредитації. Його можуть використовувати студенти інших факультетів, аспіранти, науковці і практичні працівники економічних служб аграрного сектору, які бажають систематизувати і поглиблювати свої знання щодо прикладного статистичного аналізу сільськогосподарського виробництва.

Автор висловлює щиру подяку рецензентам навчального посібника, а також колективу кафедри статистики Московської сільськогосподарської академії, доцентам Степаненко М.В. (Національний аграрний університет), Двірнику В.М. (Дніпропетровський державний аграрний університет), Кобилкіній С.В. і Кобилкіну О.М. (Сумський державний аграрний університет) за цінні зауваження і пропозиції щодо удосконалення підручника на стадії підготовки його до видання.

Зауваження і пропозиції щодо удосконалення навчального посібника просимо направляти у видавництво або автору за адресою: 10008, м.Житомир-8, вул. Старий бульвар, 7. Державний агроекологічний університет, кафедра аналізу і статистики.

Розділ 1

Предмет і метод статистики

1.1. Предмет статистики, її розділи

Розпочинаючи вивчення статистики, як і будь-якої іншої науки, слід передусім з'ясувати її предмет, основні поняття, метод і завдання.

Термін «статистика» походить від латинського слова «status» (статус), що в перекладі означає положення, стан явищ. Від кореня цього слова утворилось італійське слово «stato» (стато) – держава. Осіб, що володіли знаннями про устрій і стан справ у різних державах, тобто державних діячів, політиків називали «statista» (статиста). Від цього ж кореня утворився іменник «statistika» (статистика – певна сума знань, відомостей про державу). В науковий ужиток слово «статистика» увійшло у XVIII ст. і первісно вживалось у значенні «державознавство». Проте статистична наука стала розвиватися ще раніше, у середині XVII ст.

Нині під терміном «статистика» розуміють три пов'язані між собою значення:

1) цифри, які характеризують рівні, розміри та обсяги тих або інших суспільних явищ;

2) особливу галузь практичної діяльності, спрямовану на збирання, нагромадження, обробку та аналіз даних, які характеризують населення, економіку, культуру, освіту та інші явища суспільного життя;

3) самостійну суспільну науку, яка займається розробкою методів збирання, зведення, обробки, аналізу і теоретичним узагальненням цифрових даних про явища суспільного життя.

У цьому підручнику основну увагу звернено на розгляд принципів і положень статистики як специфічної суспільної науки.

Між статистичною наукою і статистичною практикою існує тісний зв'язок і залежність. Статистична практика використовує розроблені наукою теоретичні положення і методи для вирішення конкретних управлінських завдань. У свою чергу статистична наука використовує дані практики, узагальнює їх і розробляє методи проведення статистичного дослідження.

Статистика має багатовікову історію. Як суспільна наука вона виникла з практичних потреб суспільного життя (підрахунок населення, худоби, визначення розмірів територій держав, земельних угідь, національного багатства тощо). Об'єктом її вивчення є суспільство, явища і процеси суспільного життя.

Специфіка статистики як особливої галузі знань полягає в тому, що вона в змозі виміряти рівень і обсяг явищ, визначити їх структуру, тенденцію та інтенсивність тих або інших процесів. Справді, тільки статистика дає нам можливість визначити вартість валового внутрішнього продукту і валового національного доходу, створених у країні за рік, оцінити ефективність суспільного виробництва, економічних реформ тощо. Статистика за допомогою цифр характеризує фактичний стан (рівень) явища, що вивчається, на певному ступені його розвитку в конкретних умовах місця і часу.

Будь-яка наука являє собою систематизоване знання. Це стосується й статистики. Статистика ніколи не змогла б піднятись до рівня науки, якби вона тільки тим і займалась, що реєструвала явища, нехай навіть і масових процесів, без їх систематизації, наукових узагальнень і висновків, аналізу і синтезу.

Кожна наука володіє рядом специфічних властивостей, що відрізняють її від інших наук і що дають їй право на самостійне існування як особливої галузі знань.

Як вже зазначалось, статистика вивчає явища суспільного життя і тому належить до суспільних наук. Однак суспільство є об'єктом вивчення не тільки статистики, але й багатьох інших суспільних наук. При цьому кожна з них, вивчаючи одну з граней суспільного життя, відрізняється своїм предметом, під яким розуміють ті особливі сторони і властивості об'єкта, що підлягають дослідженню певною наукою. У чому ж відмінність статистики від інших суспільних наук про суспільство? Що вона вивчає, тобто що є предметом її пізнання?

Численні визначення статистики як науки зводяться здебільшого до такого. **Статистика як суспільна наука** вивчає кількісну сторону масових суспільних явищ у нерозривному зв'язку з їх якісною стороною, досліджує кількісний вираз закономірностей суспільного розвитку у конкретних умовах місця і часу. Вона вивчає також вплив природних, технічних, економічних і соціальних факторів на умови і результати виробництва, зворотний вплив розвитку суспільного виробництва на умови життя людей.

З наведеного визначення предмету статистики випливає дві особливості статистики як суспільної науки.

Перша відмінна особливість статистики як суспільної науки полягає в тому, що предметом її вивчення є кількісна сторона масових суспільних явищ. При цьому статистика вивчає кількість не саму по собі, а у зв'язку з її якісним змістом у конкретних умовах місця і часу.

Між кількістю та якістю як філософськими категоріями існує нерозривний зв'язок: у єдності вони складають **міру явища**. У мірі якісна визначеність – кількісна, а кількісна – якісна. Будь-яке суспільне явище являє собою цілісність якісної і кількісної визначеності. Дуже важливо те, що саме міра знаходить вираження в статистичних показниках. Статистичний показник – завжди якась єдність якості і кількості. У зв'язку з цим перед статистикою постає важливе завдання – встановити на певному етапі дослідження за якісними змінами ознак переходи, зародження нових класів, типів і одноякісних сукупностей. Це завдання статистика вирішує за допомогою своїх спеціальних методів дослідження, розрахунку і аналізу відповідних показників.

Другою особливістю статистики як науки є те, що вона вивчає масові суспільні явища. Це означає, що статистичні показники завжди є результатом узагальнення деякої сукупності фактів.

Теоретичною основою статистики є економічна теорія (політична економія, макро- і мікроекономіка), яка формує і досліджує закони розвитку соціально-економічних явищ, з'ясовує їх природу і значення у житті суспільства. Ґрунтуючись на знаннях і принципах економічної теорії, статистика формує статистичні сукупності, встановлює суттєві ознаки для виділення соціально-економічних типів, виявляє кількісні зміни масових суспільних явищ, вивчає конкретні явища і процеси суспільного життя. Наприклад, без наукового розуміння суті таких економічних категорій, як валовий національний продукт, валовий національний доход, продуктивність праці, *собівартість, прибуток та інших, статистика не може правильно визначити* їх обсяг і рівень.

В той же час, керуючись законами і категоріями економічної теорії, статистика збагачує економічні науки фактами, одержаними в статистичному дослідженні, підтверджує або заперечує їх теоретичні положення.

Економічна теорія, спираючись на статистику, формує закони суспільного розвитку. Статистика, характеризуючи кількісну сторону суспільних *явищ у конкретних історичних умовах, створює фундамент з точних і беззаперечних фактів.* Економічні науки використовують статистичну інформацію для перевірки, обґрунтування або ілюстрації своїх теоретичних положень і висновків.

Статистика як самостійна наука пройшла складний шлях свого становлення. Це багатогалузева наука, яка складається з окремих розділів або частин, які, будучи самостійними, тісно пов'язані між собою. Нині закінчене оформлення одержало п'ять розділів (частин) статистики: математична, загальна теорія, соціальна, економічна і галузеві статистики.

Розглянемо коротко зміст і специфіку кожного розділу статистики.

Математична статистика – це розділ математики, присвячений вивченню закономірностей, що мають місце в масових явищах і математичних сукупностях. Зміст математичної статистики складають математичні методи систематизації, обробки і аналізу масових статистичних даних незалежно від їхнього якісного змісту.

Методи математичної статистики можуть бути застосовані для обробки і аналізу будь-яких статистичних даних. Математичні методи безпосередньо пов'язані з імовірнісною оцінкою результатів спостереження і визначенням математичної ймовірності. У зв'язку з цим висновки математичної статистики відносно масових явищ і процесів носять імовірнісний характер і спираються на апарат теорії ймовірностей.

Специфіка математичної статистики як особливої наукової дисципліни полягає в тому, що вона розглядає сукупності незалежно від їх конкретної природи і змісту, в яких варіація визначається випадковими причинами, розглядає їх як абстрактно-математичні сукупності.

Абстрактні сукупності і є предметом вивчення математичної статистики.

Теоретичною основою математичної статистики є теорія ймовірностей (як складова частина вищої математики), яка розглядає закономірності випадкових явищ.

Найважливішими розділами математичної статистики є статистичні ряди розподілу та їх характеристики, оцінка параметрів розподілу, перевірка статистичних гіпотез, дисперсійний і кореляційний аналіз. У цьому підручнику ці розділи математичної статистики розглядаються у прикладному аспекті до аналізу сільськогосподарського виробництва. Останнім часом одержують широке застосування методи багатовимірного статистичного аналізу – факторний і кластерний аналіз, метод головних компонент та ін.

Загальна теорія статистики розглядає категорії статистичної науки, а також принципи, правила і методи, які є загальними для вивчення кількісної сторони будь-яких масових суспільних явищ.

До завдань загальної теорії статистики входять розробка методів збирання, зведення, групування, узагальнення і аналізу статистичних даних, вивчення закономірностей і тенденцій розвитку суспільних явищ і процесів, структури явищ, причинно-наслідкових зв'язків між ними, а також принципів і методів, статистичного моделювання і прогнозування.

У курсі загальної теорії статистики вивчаються такі найважливіші розділи статистичної науки: статистичне спостереження, зведення і групування даних, середні величини і показники варіації, ряди динаміки, індекси, вибірковий, табличний і графічний методи та ін. Розгляду цих розділів статистики присвячений цей підручник.

Соціальна статистика як галузь єдиної статистичної науки вивчає кількісну сторону масових явищ і процесів, що відбуваються у соціальному житті суспільства, у нерозривному зв'язку з їх якісною стороною. Вона розробляє комплексну систему взаємопов'язаних показників, які дозволяють одержати всебічну характеристику стану і розвитку умов життя людей, виявити в ньому тенденції і закономірності, що складаються, розкрити існуючи тенденції і закономірності розвитку соціальних процесів, створити повну картину суспільного устрою і способу життя людини у конкретних історичних умовах розвитку суспільства. Соціальна статистика охоплює політичну, ідеологічну і правову сторони життя людей, вивчає методами статистики державний устрій, політичну систему, соціальну структуру суспільства, соціальні умови і характер праці, питання формування особистості, сім'ї, трудового колективу, підвищення добробуту народу.

Економічна статистика – галузь статистики, яка вивчає кількісну сторону масових суспільних явищ і процесів, що відбуваються в сфері матеріального виробництва, з метою виявлення пропорцій, тенденцій і закономірностей. Вона розглядає систему об'єктивних статистичних показників, що характеризують стан і розвиток народного господарства як єдиного цілого. Економічна статистика вивчає загальні, переважно комплексні і синтетичні показники розміру, структури і динаміки всього народного господарства, розкриває його соціальну, галузеву і регіональну структуру, його міжгалузеві, міжрайонні і соціальні взаємозв'язки, його найважливіші рівні, пропорції і співвідношення.

Об'єктом вивчення економічної статистики є процеси розширеного відтворення, умови його здійснення і кінцеві результати у народному господарстві. Вона розглядає суспільне виробництво у єдності двох його сторін (продуктивні сили і виробничі відносини) і в безперервному потоці його відновлення (виробництво, розподіл, обмін, споживання, знову виробництво і т.д.).

Галузеві статистики розробляють систему статистичних показників, які характеризують стан і розвиток відповідних галузей (промисловості, сільського господарства, будівництва, торгівлі та ін.). Вони розглядають також зміст, специфіку і методику їх обчислення. Так, сільськогосподарська статистика розробляє систему показників, які характеризують стан і розвиток сільськогосподарського виробництва.

Об'єктом вивчення галузевих статистик є відповідні галузі народного господарства.

1.2. Основні поняття в статистиці

З поняттям про предмет статистики тісно пов'язані поняття статистичного показника, статистичної закономірності, статистичної сукупності, ознаки, варіації та ін.

Кількісну сторону масових суспільних явищ статистика виражає у вигляді об'єктивних статистичних показників (чисел). Але числа, що застосовуються у статистиці, це не абстрактні числа математики, які характеризуються тільки величиною, знаком, формою (цілі, дробні і т.п.). Статистика застосовує, власне кажучи, не числа, а показники, точніше – статистичні показники. Показник – одне з основних понять статистики.

Статистичним показником називають узагальнену числову характеристику будь-якого масового явища в поєднанні з його якісною визначеністю в конкретних умовах місця і часу. Прикладом статистичного показника є кількість населення країни на початок року, питома вага міського і сільського населення, урожайність сільськогосподарських культур, площа ріллі, собівартість продукції тощо.

Найважливіше завдання статистичної науки – правильно визначити зміст статистичних показників і розробити методологію їх обчислення.

Величина показника визначається в результаті його вимірювання за допомогою системи одиниць вимірювання і відповідної методології.

Статистичні показники можуть бути виражені у вигляді абсолютних і відносних величин. Якщо статистичний показник стосується окремого явища (наприклад, сільськогосподарського підприємства), його називають **індивідуальним**, якщо ж сукупності явищ (наприклад, сільськогосподарських підприємств району), то **узагальненим**, або **зведеним**. Зведені статистичні показники, які стосуються складного комплексу економічних явищ або об'єктів, називають **синтетичними** (наприклад, виробництво за рік у народному господарстві валового внутрішнього продукту, валового національного доходу тощо).

Індивідуальні та зведені статистичні показники можуть бути виражені у таких одиницях вимірювання: **натуральних** (кг, т, м, км та ін.), **вартісних** (грн.) і **умовно-натуральних** (кормових одиницях, калоріях, еталонних гектарах тощо).

Для відображення властивостей, структури і динаміки складних соціально-економічних явищ недостатньо користуватись яким-небудь одним показником, оскільки він відображає лише одну грань пізнання предмету. Тому для всебічного і цілісного аналізу масових суспільних явищ необхідно, щоб показники знаходились між собою в певному зв'язку, утворюючи систему взаємопов'язаних і таких, що доповнюють один одного, показників. ***Системою статистичних показників*** *називають сукупність взаємопов'язаних і розташованих у логічній послідовності показників.* Наприклад, при проведенні економіко-статистичного аналізу врожайності сільськогосподарських культур потрібно застосовувати систему статистичних показників, які б охарактеризували рівень і динаміку врожайності, вплив на її рівень комплексу природних (якості ґрунту, опадів тощо) і економічних (кількість добрив, фондозабезпеченість тощо) факторів, а також вплив рівня самої урожайності на економічну ефективність виробництва тієї або іншої сільськогосподарської культури (рентабельність, собівартість, трудомісткість тощо).

Система статистичних показників не є незмінною, вона має бути динамічною. У процесі суспільного розвитку одні явища відмирають, інші виникають, цей процес знаходить своє вираження й у системі статистичних показників. Система статистичних показників постійно удосконалюється: одні показники виключаються і додаються нові, що характеризують розвиток сучасного суспільства.

У системі категорій статистичної науки особливе місце належить закону великих чисел і статистичній закономірності.

Маючи в своєму розпорядженні узагальнені статистичні показники та їхню систему, статистика намагається виявити тенденції, взаємозв'язки та закономірності розвитку тих або інших масових суспільних явищ. Вивчаючи тенденції та закономірності розвитку суспільних явищ, статистика спирається на **закон великих чисел**. Його суть полягає в тому, що закономірність масового явища виявляється лише при досить великій кількості спостережень, коли відбувається взаємне погашення впливу випадкових причин, які мають місце в окремих випадках.

Закон великих чисел є одним з проявів діалектичної єдності між випадковістю і необхідністю. Необхідність проявляється через масу випадковостей. Кожен окремий випадок є індивідуальним, бо він зазнає впливу не тільки загальних, а й часткових причин, які виступають як випадковості, за якими приховується необхідність, тобто загальна закономірність масового явища. Наприклад, відомо, що урожайність буде вищою на тих ділянках, де більше вносять добрив. Однак на окремих ділянках така закономірність може й не виявитися внаслідок впливу різних факторів (різна якість ґрунту, насіння, кількість опадів тощо). Якщо для вивчення взяти досить велику сукупність ділянок, то дія випадкових факторів значною мірою зрівноважиться. Зіставляючи середню врожайність по ділянках, можна побачити, що вона буде вищою на тих ділянках, де більше вносилося добрив.

Сукупна дія великої кількості випадкових факторів, спрямована в різні сторони, зумовлює результат, який майже не залежить від випадку, тобто результат, що виражає певну статистичну закономірність. Оскільки згадані закономірності масових суспільних явищ виявляються в статистичних сукупностях, то їх називають **статистичними закономірностями**. Вони є однією з різноманітних форм виявлення загального зв'язку між явищами в природі і суспільстві.

Статистичною сукупністю називають множину одиниць подій або фактів, які об'єднуються однією якісною основою, але відрізняються між собою за рядом ознак. Наприклад, статистичною сукупністю є населення будь-якої області, яке складається з окремих людей, що розрізняються за статтю, віком і багатьма іншими ознаками. В той же час ця сукупність людей єдина в тому відношенні, що вона складається з жителів даної області. Прикладом статистичної сукупності є сільськогосподарські підприємства району, робітники підприємства, студенти академічної групи тощо.

Окремі елементи статистичної сукупності називають **одиницями сукупності,** а загальну їх кількість – **обсягом сукупності**. Елементи сукупності характеризують однією або кількома ознаками. **Ознакою** в статистиці називають властивість, характерну рису або особливість одиниць явищ, які можна спостерігати або виміряти. Так, ознакою сільськогосподарського підприємства можуть бути розмір земельної площі, число робітників, тракторів, худоби, вартість основних фондів, рентабельність виробництва продукції тощо.

Статистичні сукупності, розчленовані за однією ознакою, називають **одновимірними,** за двома – **двовимірними,** за трьома та більше – **багатовимірними.** При цьому ознаки, за якими поділяють сукупності на групи (класи), називають **груповими**, а число одиниць сукупності, які мають однакове значення ознаки, – **частотою** даної варіанти, або її **вагою.**

Залежно від повноти обстеження одиниць сукупності розрізняють генеральну і вибіркову сукупності. Сукупність явищ, з яких відбирають частину одиниць для вибіркового спостереження, називають **генеральною сукупністю.** Та ж частина, яка відібрана з генеральної сукупності для вибіркового спостереження, називається **вибірковою сукупністю.** Наприклад, дослідника цікавить вплив рівня годівлі на продуктивність корів. З цією метою він з 1000 корів відбирає для дослідження 100 корів. Отже, загальна кількість корів (1000 голів) є генеральною сукупністю, а відібрані 100 корів – вибірковою. Основне завдання формування вибіркової сукупності полягає в тому, щоб показники, які характеризують цю сукупність з найбільшим ступенем точності, відтворювали показники генеральної сукупності.

Статистичні сукупності характеризуються рядом специфічних властивостей, відмінних особливостей. Однією з таких особливостей є наявність **варіації,** тобто відмінностей у числових значеннях окремих одиниць сукупності. Статистика вивчає тільки ті ознаки, що варіюють. Якщо значення ознак у окремих одиниць однакові, то вивчення таких ознак у зв'язку з відсутністю варіації безглуздо, достатньо описати одну одиницю.

Ознаки, які набувають різних значень або видозмін в окремих одиниць сукупності, називають **варіюючими,** а окремі їх значення – **варіантами.** Варіанти можуть приймати будь-які значення в межах даних границь варіації.

Варіюючі ознаки поділяють на атрибутивні (якісні) та кількісні. **Атрибутивними** називають ознаки, які не підлягають числовому вираженню. Наприклад, національність, професія, порода худоби, сорт рослин тощо. Якщо якісна ознака набуває лише двох протилежних (взаємовиключних) значень, її називають **альтернативною** (вироби придатні та браковані, стать чоловіча та жіноча, рослина вражена та не вражена хворобою тощо). Ознаку називають **кількісною,** якщо окремі варіанти виражаються числами (урожайність, заробітна плата, продуктивність праці та ін.).

За характером варіювання кількісні ознаки поділяють на дискретні (перервні) і неперервні. **Дискретними** називають такі кількісні ознаки, які можуть набувати тільки перервних (цілочислових) значень (наприклад, кількість членів родини, кількість автомобілів, верстатів тощо).

Кількісні ознаки, які можуть в окремих межах набувати будь-яких значень, називають **неперервними** (наприклад, вік, урожайність, собівартість продукції, стаж роботи тощо).

Розрізняють істотні (основні) та неістотні (другорядні) ознаки. **Істотними** називають такі ознаки, які нерозривно пов'язані з якісним змістом явища, його суттю. Наприклад, істотною ознакою для сільськогосподарського підприємства є обсяг виробленої та реалізованої продукції, кількість працівників, продуктивність праці, рівень рентабельності, урожайність сільськогосподарських культур, продуктивність тварин та ін. **Неістотними** називають такі ознаки, які не пов'язані безпосередньо з внутрішнім змістом явища, його суттю. Так, неістотними ознаками для сільськогосподарського підприємства є його назва, підпорядкування, належність до тієї чи іншої території тощо.

Ознаки поділяють на первинні та вторинні. **Первинні** ознаки одержують в результаті збирання статистичного матеріалу. Вони характеризують безпосередньо одиницю сукупності, її абсолютний розмір (наприклад, обсяг виробленої продукції, площу сільськогосподарських угідь, ріллі та ін.). **Вторинні,** або розрахункові, ознаки отримують в результаті обробки зібраних даних, це ознаки співвідношень. Так, вторинна ознака «виробництво валової продукції на один гектар сільськогосподарських угідь» є результатом співвідношення двох первинних ознак «вартість валової продукції» та «площа сільськогосподарських угідь».

Розрізняють *результативні і факторні ознаки*. **Результативною** називають ознаку, зміна якої залежить від зміни іншої або інших ознак, а **факторною** – ознаку, що є причиною зміни іншої ознаки. Наприклад, з двох ознак – урожайність та якість грунту – перша є результативною, бо її зміна багато в чому залежить від зміни другої ознаки (якість грунту), а друга – факторною ознакою.

Слід мати на увазі, що та сама ознака може бути в одному випадку результативною, а в другому – факторною. Наприклад, при вивченні зв'язку між двома ознаками – урожайністю та собівартістю продукції – урожайність є факторною ознакою (від її величини залежить рівень собівартості продукції), а собівартість, відповідно, виступатиме результативною ознакою.

1.3. Метод статистики

Для вивчення свого предмету – кількісної сторони масових суспільних явищ – статистична наука розробила ряд своїх особливих прийомів, способів, правил і методів дослідження, які в сукупності складають **статистичну методологію**.

Під терміном «метод» (від грец. methodos – шлях дослідження або пізнання, теорія, наука) розуміють сукупність прийомів або теоретичного опанування дійсності, підпорядкованих вирішенню конкретного завдання.

Загальним методом пізнання для всіх наук, у тому числі для статистики, є діалектичний метод. Відповідно до основоположних принципів і законів діалектики всі суспільні явища і процеси, які вивчаються статистикою, знаходяться в постійному русі і розвитку, не ізольовано одне від одного, а у взаємозв'язку і взаємозалежності, що дуже важливо при вивченні причинно-наслідкових взаємозв'язків між явищами.

Спираючись на ці принципи діалектики, статистика виділяє різні типи і форми соціально-економічних явищ і процесів, вивчає їх особливості і оцінює вплив комплексу факторів, які формують варіацію і динаміку явищ, виявляє тенденції і закономірності їх розвитку.

При статистичному вивченні суспільних явищ керуються також діалектичним законом про перехід кількісних змін в якісні. Це має важливе значення при вивченні кількісних змін у масових соціально-економічних явищах для пізнання глибоких якісних змін. Статистика спирається на діалектичні категорії випадкового і необхідного, одиничного і масового, індивідуального і загального.

Загальні принципи і методи наукового пізнання, розроблені в діалектиці, є фундаментом для розуміння і правильного використання статистичної методології.

На основі діалектичного методу статистика розробила свої специфічні методи дослідження. До них належать: статистичне спостереження, зведення і групування даних, абсолютні і відносні показники, середні величини і показники варіації, ряди динаміки, індексний, вибірковий, дисперсійний, кореляційний, табличний і графічний методи та ін.

Докладне викладення змісту перерахованих статистичних методів дається в наступних розділах підручника.

Застосування в статистичному дослідженні конкретних методів визначається поставленими при цьому завданнями, суттю і особливостями досліджуваного явища і залежить від характеру вихідної інформації.

Будь-яке статистичне дослідження складається з трьох послідовно виконуваних етапів: 1) статистичне спостереження; 2) зведення і групування даних статистичного спостереження; 3) аналіз одержаних результатів і формулювання висновків. На кожному з цих етапів застосовуються специфічні статистичні методи.

На першому етапі статистичного дослідження на основі певних правил і відповідно до його програми і плану вирішується завдання по збиранню первинного статистичного матеріалу про кожну одиницю сукупності. Для здійснення цієї початкової стадії дослідження застосовується метод масового статистичного спостереження, який забезпечує загальність, повноту і представництво (репрезентативність) одержаної інформації. Тільки метод масового статистичного спостереження дає змогу виявитися загальним умовам і закономірностям, характерним для всієї сукупності, і уникнути впливу випадкових причин, що діють на окремі одиниці сукупності. Від повноти і якості інформації, одержаної на етапі статистичного спостереження, залежить ефективність наступних етапів і досягнення кінцевої мети дослідження.

На другому етапі статистичного дослідження переходять від характеристики окремих одиниць до їх загальної характеристики, від вивчення індивідуальних значень ознаки до їх узагальнення. З цією метою проводять зведення зібраної в ході масового спостереження статистичної інформації. Суть зведення полягає в систематизації, обробці первинних даних, приведенні їх у певний порядок, підрахунку чисельності одиниць сукупності в цілому і окремих її частинах, а також обсягу ознак, що характеризують їх.

Найважливішим методом другого етапу дослідження є метод статистичних групувань, який дає змогу виділяти якісно однорідні соціально-економічні типи, групи і підгрупи і тим самим дає узагальнену характеристику всієї сукупності.

На третьому, заключному етапі статистичного дослідження проводиться аналіз статистичної інформації і формулювання висновків. У процесі аналізу і виявлення статистичних закономірностей і взаємозв'язків широко застосовується ряд специфічних статистичних методів, які дають змогу одержати узагальнюючі показники за допомогою яких здійснюється вимірювання і кількісна оцінка виявлених при аналізі закономірностей.

До таких узагальнюючих показників відносяться абсолютні і відносні величини, статистичні коефіцієнти, середні величини, показники варіації і рядів динаміки та ін.

Для характеристики причинно-наслідкових взаємозв'язків масових суспільних явищ застосовується індексний, балансовий, кореляційний, дисперсійний та інші методи.

При аналізі статистичної інформації широке застосування мають табличний і графічний методи.

1.4. Зв'язок статистики з іншими науками

Соціально-економічна статистика пов'язана з багатьма науками. При цьому передусім необхідно зазначити тісний і нерозривний зв'язок статистичної науки з політичною економією. У вивченні кількісної сторони економічних явищ статистика спирається на теорію політичної економії, в якій визначається суть економічних категорій і розкриваються в їх загальній формі закони економічного розвитку. Збираючи і узагальнюючи за допомогою числових показників точні факти економічного життя, статистика дає об'єктивне зображення дійсного ходу розвитку економічних явищ в конкретних історичних умовах.

Зв'язок між статистикою і політичною економією не є одностороннім. Не тільки статистика використовує положення, встановлені політичною економією, але й ця остання (рівно як й інші суспільні науки) широко використовує положення, факти і висновки, встановлені статистикою.

Однак було б помилково звести все значення статистичної науки тільки до ілюстрації положень політичної економії, її роль у предметному аналізі ширше і глибше. Вивчаючи закономірності процесів суспільного життя в конкретних умовах місця і часу, соціально-економічна статистика за допомогою певної системи категорій, понять і показників дає цим процесам об'єктивне, точне, основане на масових даних числове вираження. В цьому її відмінна риса як науки, в цьому її особливість порівняно з іншими суспільними науками.

Статистика тісно пов'язана з плануванням. Цей зв'язок виявляється вже на початковому етапі планування. Без статистичних даних, які характеризують досягнутий рівень, ніяке планування неможливе.

Важливо зазначити особливе значення статистики у справі виявлення резервів, які є в народному господарстві. Ретельний аналіз даних за минулий період дає змогу виявити невикористані резерви, які мають бути враховані при плануванні.

На статистику покладається важлива функція контролю за виконанням плану, перевірки ходу його виконання.

Зв'язок і відмінність між статистикою і математикою полягає в тому, що обидві ці науки вивчають кількісну сторону явищ, але математика вивчає кількісну сторону всіх явищ (природи і суспільства) безвідносно до якості, а статистика – кількісну сторону лише суспільних явищ і завжди певної якості.

У статистиці застосовується математика всіх рівнів. Значення математики для розвитку статистики виключно велике, воно особливо зросло в сучасних умовах в зв'язку з широким впровадженням математико-статистичних методів у економічний аналіз, автоматизацією процесів збирання, збереження, передачі і обробки статистичної інформації і нової обчислювальної техніки, за допомогою якої стало можливим ставити і вирішувати найскладніші завдання.

Широке застосування математико-статистичних методів у поєднанні з обробкою статистичної інформації на ЕОМ дає змогу значно поглибити аналіз інформації, зробити його більш оперативним.

Статистика тісно пов'язана з іншими видами обліку суспільних явищ. У системі народногосподарського обліку розрізняють три види обліку: оперативно-технічний, бухгалтерський і статистичний. Кожний з цих видів обліку має свої особливості і завдання, але всі вони тісно пов'язані один з одним єдністю кінцевої мети і створюють єдину централізовану систему обліку і статистики.

Оперативно-технічний облік забезпечує потреби оперативного керівництва роботою підприємства та його підрозділів (цехів, відділень, бригад, ланок, ферм і т.д.). Він являє собою реєстрацію конкретних фактів безпосередньо в момент їхнього здійснення на робочих місцях, виробничих ділянках, у цехах і т.д. Прикладом оперативно-технічного обліку є записи в табелі про вихід робітників на роботу, щоденний облік випуску готової продукції, надоїв молока, витрати пального і кормів, засіяної площі і т.д. Цей облік ведуть переважно в натуральному виразі.

Бухгалтерський облік служить для неперервного і повсякденного контролю за збереженням матеріальних і грошових засобів підприємства, він відображає всі господарські операції, які пов'язані з рухом і використанням матеріальних і грошових засобів. Результати господарської діяльності і фінансового стану підприємств бухгалтерський облік показує у вигляді балансу. Бухгалтерський облік ведуть переважно у грошовій формі.

Статистичний облік – завершальна стадія обліку. Він являє собою закономірне, науково організоване збирання даних про соціально-економічні явища в масштабі народного господарства, галузей, економічних районів і т.д. Джерелом відомостей для статистичних узагальнень передусім є дані оперативно-технічного і бухгалтерського обліку. Важливим джерелом даних є спеціальне збирання відомостей, яке проводять статистичні органи. Узагальнені дані оперативно-технічного і бухгалтерського обліку статистичний облік виражає за допомогою статистичних показників закономірності розвитку відповідних явищ.

У єдиній системі народногосподарського обліку ведучим є статистичний облік, який порівняно з іншими видами обліку є найбільш широким за просторовим охопленням і глибоким за виявленням змісту явищ, що обліковуються.

Статистика пов'язана з мікро- і макроекономікою, економікою окремих галузей, менеджментом, маркетингом, аналізом господарської діяльності, технологічними та іншими дисциплінами.

1.5. Завдання і організація статистики в Україні

Завдання статистичної науки тісно пов'язані з практичними потребами державного управління і керівництва розвитком народного господарства і соціальної сфери. Кожний новий етап розвитку суспільства висуває перед статистикою нові конкретні завдання. На сучасному етапі завдання статистики визначається актуальними проблемами здійснення радикальних економічних реформ, переходу від командно-адміністративних форм управління до економічних, ринкової економіки та ін.

Найважливішими завданнями статистики є:

1. Збирання, розробка, узагальнення і аналіз статистичних даних, які характеризують стан і розвиток суспільного виробництва і соціальної сфери України.

2. Всебічне дослідження на основі науково обґрунтованої системи статистичних показників корінних перетворень економічних і соціальних процесів, що відбуваються у суспільстві.

3. Своєчасне забезпечення законодавчих, виконавчих, управлінських та господарських органів, а також широкої громадськості вичерпною і достовірною інформацією.

4. Розробка статистичних даних, які показують хід виконання народногосподарських планів, а також статистичних даних, необхідних для складання планів і розробки прогнозів розвитку нашого суспільства.

5. Пропаганда передового досвіду, розробка даних, які характеризують ефективність нових технологій, форм організації праці і виробництва, ринкових відносин, міжнародних економічних зв'язків та ін.

6. Всебічне і глибоке вивчення пропорцій і взаємозв'язків у народному господарстві, факторів і резервів підвищення ефективності суспільного виробництва, зростання продуктивності праці і життєвого рівня населення.

7. Удосконалення статистичної методології і системи статистичних показників, які характеризують стан і розвиток народного господарства і культури України, збирання і обробки статистичної інформації, економіко-статистичного аналізу соціально-економічних явищ і процесів, що відбуваються у суспільному житті України.

8. Розробка питань створення загальнодержавної автоматизованої системи збирання, нагромадження, обробки і аналізу статистичної інформації на основі новітньої обчислювальної техніки.

9. Розробка заходів щодо зближення вітчизняної методології статистичних досліджень з методологією і стандартами міжнародної статистики, яку здійснюють Статистична комісія ООН, Міжнародний статистичний інститут та інші міжнародні організації, впровадження у статистичну практику системи національних рахунків ООН.

Це найбільш загальні і важливі завдання нашої статистики; на окремих етапах розвитку суспільства вони насичуються новим конкретним змістом.

Система статистичних органів України відповідає державному устрою і адміністративно-територіальному розподілу країни.

Керівним організаційним і методологічним центром статистики в Україні, який здійснює централізоване керівництво справою обліку і статистики, є Державний комітет статистики України. Він проводить свою роботу через відповідні статистичні органи на місцях. У автономній республіці Крим таким органом є Державний комітет статистики Криму, а в областях – статистичні управління. Низовими органами державної статистики, які знаходяться в безпосередньому підпорядкуванні статистичних управлінь, є районні і міські відділи статистики.

Права, обов'язки і функції органів державної статистики визначені Законом України «Про державну статистику» (1992 р.).

Державний комітет статистики України здійснює державне управління всією, що знаходиться в його веденні, єдиною системою статистичних органів, справою статистики, обліку і звітності в усіх галузях народного господарства, створенням і функціонуванням статистичної інформаційної системи на основі єдиної наукової методології.

Основним завданням усіх органів статистики є збирання, перевірка достовірності, розробка, узагальнення, аналіз і подання в установлені строки законодавчим, виконавчим, управлінським і господарським органам науково обґрунтованих статистичних даних, які характеризують економічний і соціальний розвиток країни, виконання загальнодержавних і регіональних програм, зростання ефективності суспільного виробництва, економічних реформ, використання природних, трудових і матеріальних ресурсів, динаміку життєвого рівня народу тощо.

Органи державної статистики здійснюють соціальні переписи (населення, основних фондів, худоби, багаторічних насаджень тощо) і одночасні обліки, необхідні для глибоко вивчення окремих сторін життя суспільства, виконують роботу щодо подальшого удосконалення організації і методології обліку і статистики, вдосконалення наявної статистичної звітності і доведення її обсягів до потреб системи управління в умовах переходу до ринкових відносин, контролюють стан обліку у народному господарстві, забезпечують прискорене переведення обліку і статистики на базу машинної техніки.

Завдання для самоконтролю до розділу 1

1. Що розуміють під терміном «статистика»?
2. Сформулюйте означення предмета статистики.
3. Дайте характеристику основним рисам статистики як суспільної науки.
4. Що є теоретичною основою статистики?
5. Назвіть розділи статистики.
6. Що вивчає математична статистика?
7. Що вивчає загальна теорія статистики?
8. Що вивчає соціальна статистика?
9. Що вивчає економічна статистика?
10. Що вивчають галузеві статистики?
11. Дайте визначення статистичного показника, вкажіть їх види. Наведіть приклади. Якими статистичними показниками можна охарактеризувати фірми міста, студентів академічної групи, фермерські господарства?
12. Сформулюйте визначення статистичної сукупності, вкажіть їх види. Наведіть приклади. Вкажіть, які сукупності можна виділити серед населення і промислових підприємств для статистичного вивчення?
13. Що називається ознакою в статистиці? Назвіть їх види і наведіть приклади.
14. Розкрийте суть статистичної методології.
15. Назвіть специфічні методи статистики і стадії статистичного дослідження.
16. Назвіть науки, з якими пов'язана статистика.
17. Назвіть завдання державної статистики на сучасному етапі розвитку України.
18. Назвіть принципи організації статистики і систему статистичних органів України.

Розділ 2

Статистичне спостереження

2.1. Поняття про статистичне спостереження. Програмно-методологічні та організаційні питання статистичного спостереження

Для того щоб вивчити кількісну сторону масових суспільних явищ і процесів, насамперед потрібно зібрати про них необхідну статистичну інформацію. З цією метою організовується масове статистичне спостереження, яке є першим етапом статистичного дослідження.

Статистичне спостереження – це планомірне, науково організоване збирання або одержання масових відомостей про явища і процеси суспільного життя (наприклад, переписи населення, основних фондів, багаторічних насаджень, худоби, заповнення анкет, бланків, форм статистичної звітності, вибірка статистичних даних з річних звітів сільськогосподарських підприємств та ін.). Матеріали статистичного спостереження є основою для одержання узагальнюючих показників, які характеризують ті або інші суспільні явища і процеси.

Статистичне спостереження – це складна, трудомістка і відповідальна робота, яка залежно від масштабів дослідження виконується силами багатьох працівників. Воно складається з підготовчих робіт, безпосереднього збирання даних та їх контролю.

Завданням статистичного спостереження є одержання точної і вірогідної інформації, яка об'єктивно відображає фактичний стан речей. Мають бути також забезпечені повнота і порівнянність інформації і одержання її у можливо короткий строк.

У процесі статистичного спостереження формуються дані, від якості яких залежить успіх усього дослідження. Помилки, які допущені на етапі статистичного спостереження, тобто на етапі збирання вихідної інформації, розповсюджуються і на інші етапи статистичного дослідження – зведення, групування, обробку, аналіз і узагальнення одержаних показників.

У зв'язку з цим одним з найважливіших завдань статистичного спостереження є одержання достатньо повних, об'єктивних і порівнянних даних про явища і процеси, що вивчаються.

Щоб забезпечити ці вимоги, потрібно дотримувати певних методичних правил проведення спостереження. Статистичне спостереження здійснюють за заздалегідь розробленим планом, який включає вирішення двох груп питань: програмно-методологічних та організаційних.

До **програмно-методологічних** питань належать: формулювання цілей і завдань спостереження, визначення об'єкта та одиниці спостереження, розроблення програми та інструментарію спостереження, вибір видів і способів отримання статистичних даних, форм контролю одержаного матеріалу тощо; до **організаційних** – встановлення місця, часу і строків спостереження, добір, розміщення та навчання кадрів, розмноження та розсилання інструментарію спостереження (бланків, формулярів, інструкцій тощо) на місця, визначення строків здачі матеріалів, а також інші питання, пов'язані з підготовкою, організацією та безпосереднім проведенням статистичного спостереження.

Готуючись до статистичного спостереження, насамперед формулюють його мету і завдання. **Метою статистичного спостереження** є збирання вірогідної та повної статистичної інформації про досліджувані соціально-економічні явища і процеси. **Завдання статистичного спостереження** безпосередньо випливають із завдань статистичного дослідження і полягають, зокрема, в одержанні масових даних про стан досліджуваного об'єкта, в обліку стану явищ, які впливають на об'єкт, вивченні даних про процес розвитку явищ. Чітке формулювання мети і завдання статистичного спостереження необхідні для правильного встановлення обсягу, змісту і структури вихідної інформації, що збирається. Це дає змогу уникнути збирання зайвих або неповних даних.

Відповідно до поставлених цілей і завдань визначають об'єкт та одиницю спостереження. **Об'єктом** статистичного спостереження називають таку сукупність соціально-економічних явищ і процесів, про яку потрібно зібрати необхідні дані. При визначенні об'єкта статистичного спостереження зазначають його основні відмінні риси, найважливіші ознаки. Об'єктами спостереження залежно від цілей і завдань можуть бути підприємства, організації, установи, населення, продукція, виробниче устаткування, худоба тощо. Наприклад, об'єктом спостережень перепису населення 2001 р. були категорія наявного населення, включаючи тих, хто тимчасово проживають, і категорія постійного населення, включаючи тих, які тимчасово відсутні.

Об'єкт спостереження має бути чітко відмежований від інших явищ, щоб запобігти їхньому змішуванню. В деяких випадках для цього встановлюють **ценз**, під яким у статистиці розуміють ряд обмежувальних ознак, наявність яких є основою для віднесення об'єкта до досліджуваної сукупності. Використання цензу суттєво впливає на формування однорідних сукупностей, забезпечує неможливість змішування різних об'єктів або недообліку деякої частини об'єкта.

Правильне відмежування об'єкта статистичного спостереження здійснюється на основі глибокого знання суті досліджуваної сукупності явищ. Відмежувати об'єкт статистичного спостереження – це значить встановити, по-перше, які саме явища за змістом підлягають спостереженню. Так, наприклад, недостатньо вказати, що спостереженню підлягає сукупність сільськогосподарських підприємств. Необхідно вказати, якщо це необхідно, якої форми власності і організації праці, спеціалізації тощо господарства підлягають спостереженню; по-друге, в яких територіальних межах мають бути узяті ці явища (район, область, зона, регіон, країна і т.п.); по-третє, в яких межах часу брати ці явища (на певний момент, у середньому за певний період і т.п.).

Поряд з визначенням об'єкта статистичного дослідження потрібно встановити одиницю спостереження і одиницю обліку.

Одиниця спостереження – це первинний елемент об'єкта статистичного спостереження, який є носієм ознак, що підлягають реєстрації (наприклад, сільськогосподарське підприємство, сім'я і т.д.).

Від одиниці спостереження слід відрізняти одиницю обліку.

Одиниця обліку - це той первинний осередок, від якого мають бути одержані необхідні відомості (наприклад, людина, машина, тварина і т.д.).

Чисельність одиниць обліку характеризує обсяг і розповсюдженість досліджуваного явища, а зміна цієї чисельності – його розвиток. Інакше кажучи, одиниця спостереження – це джерело одержуваних відомостей, а одиниця обліку – це те, що підлягає спостереженню. Визначення одиниці спостереження важливе при вирішенні питань організації збирання відомостей, а визначення одиниці обліку - при розробці програми спостереження. Так, наприклад, при проведенні перепису худоби одиницею спостереження буде сільськогосподарське підприємство (СТОВ, агрофірма, особисте підсобне господарство тощо), бо від нього одержують відомості про тварин. Одиницею обліку буде окрема тварина, оскільки ознаки, що реєструються в цьому перепису (жива маса, вік, продуктивність та ін.), відносяться не до підприємства в цілому, а до кожної окремої тварини.

Одиниця спостереження і одиниця обліку інколи можуть співпадати, як це буває при перепису населення, коли кожна людина є одночасно і джерелом даних і носієм ознак, що підлягають реєстрації, тобто і одиницею спостереження і одиницею обліку.

Вище зазначалося, що об'єкт статистичного спостереження має бути відмежований не тільки у просторі, але й у часі. Іншими словами, має бути встановлено, за який період або на який момент часу необхідно зібрати відомості.

Час спостереження – це той час, до якого належать дані, що збираються. Час реєстрації даних для всіх одиниць встановлюється єдиний: для попередження неповного обліку або повторного рахунку, а також для порівнянності даних.

При вивченні об'єктів спостереження, чисельність яких постійно змінюється, встановлюється **критична дата,** станом на яку збираються відомості. Наприклад, при переоцінці основних фондів встановлюється критична дата, станом на яку обліковуються основні фонди (будівлі, споруди, устаткування, худоба та ін.).

При переписах звичайно встановлюють час початку і час закінчення реєстрації даних (період проведення спостереження). Так, перепис населення проводився протягом 10 днів – з 5 по 14 грудня 2001 р.

При вивченні такого рухомого об'єкта, як населення, недостатньо встановити час спостереження, тому що склад населення та його характеристики постійно змінюються. Тому дані реєструють станом на певний період часу, який називається **критичним моментом спостереження**. Критичним моментом перепису населення 2001р. було 12 годин ночі з 4 на 5 грудня. Це означає, що всі відомості про кожного жителя країни фіксувались такими, якими вони були станом на критичний момент спостереження. Померлі після 12 годин ночі вносились в переписні листи, а народжені після 12 годин ночі обліку не підлягали і в переписні листи не записувались.

Одним з найскладніших і найвідповідальніших моментів статистичного спостереження є складання **програми спостереження** – переліку ознак, які підлягають реєстрації по кожній одиниці. Іншими словами, програма спостереження – це перелік запитань, на які слід отримати відповідь в процесі проведення спостереження.

Зміст програми спостереження визначається характером і властивостями досліджуваного явища, цілями і завданнями дослідження. Від того, наскільки добре її розроблено, багато в чому залежить цінність зібраного статистичного матеріалу.

Щоб дати повну і всебічну характеристику явищ, що вивчаються, програма спостереження повинна включати все коло суттєвих ознак явищ. Однак через ряд об'єктивних причин це практично неможливо. Тому до програми слід включати тільки ті суттєві ознаки, які мають найбільше практичне і теоретичне значення для досліджуваного об'єкта спостереження, розкривають його зміст та основні взаємозв'язки. Включення до програми несуттєвих питань призводить до її розширення, подорожчання статистичного спостереження і обробки його результатів.

Прагнучи до повноти збирання потрібних даних, необхідно враховувати можливість одержання доброякісних відомостей за широкою програмою. Якщо немає впевненості в одержанні таких даних, то краще обмежитися переліком запитань програми, щоб мати вірогідний матеріал.

Зміст та кількість запитань програми залежить від завдань спостереження. Її потрібно складати так, щоб не було зайвих запитань, аби уникнути додаткових затрат праці і коштів і водночас усе врахувати. Ознаки (запитання), включені до програми, мають становити систему взаємопов'язаних показників, заданих у певній логічній послідовності. Програму спостереження доцільно будувати так, щоб відповідями на одні запитання можна було проконтролювати відповіді на інші.

Для одержання доброякісних матеріалів велике значення має формулювання запитань. Вони мають бути сформульовані чітко і ясно, так щоб їх зміст по можливості розуміли всі однаково.

При складанні програми важливо мати чітке уявлення про систему показників, які сподіваються отримати в процесі збирання даних. Для цього спочатку доцільно проектувати макети кінцевих аналітичних таблиць, де і буде подано необхідні для статистичного аналізу показники. Після розробки макетів підсумкових статистичних таблиць буде ясно, які запитання слід включити до програми статистичного спостереження. Таким чином, складання програми статистичного спостереження, як правило, починається з проектування таблиць, що призначаються для характеристики підсумків спостереження.

Одночасно з програмою статистичного спостереження розробляють його **інструментарій**, який включає формуляри та інструкції щодо їх заповнення.

Статистичний формуляр – це первинний документ, в якому фіксуються відповіді на запитання програми статистичного спостереження по кожній одиниці сукупності. Формуляр, таким чином, є носієм первинної інформації. Формуляри статистичного спостереження, що містять програму спостереження, мають різні назви: бланк, переписний лист, опитувальний лист, картка, фішка, акт, табель, анкета, форма первинного обліку або звітності та ін. Однак для всіх формулярів характерні деякі обов'язкові елементи: змістовна частина, що включає перелік запитань програми, вільна графа або кілька граф для запису відповідей і шифрів (кодів), титульна і адресна частини. У титульній і адресній частинах зазначається найменування документа, ким і коли затверджений, дата подання відомостей, найменування підприємств або прізвище, ім'я та по батькові осіб, що обстежуються, та їхня адреса.

У практиці статистичного спостереження знайшли застосування два види носіїв інформації: карткові і спискові формуляри.

Картковий (індивідуальний) формуляр призначений для занесення до нього відомостей про одну одиницю спостереження.

У **списковому формулярі** містяться дані по кількох одиницях спостереження. Наприклад, при перепису населення члени кожної сім'ї можуть записуватися в один переписний лист.

Обидва види формулярів мають свої переваги і недоліки. Порівняно зі списковою формою картковий формуляр може містити більше запитань, оскільки у ньому характеризується тільки одна одиниця спостереження. Зручні для ручної обробки (особливо при побудові групувань, коли картки можна розгрупувати за окремими ознаками) карткові формуляри в той же час потребують більших трудових і матеріальних витрат, ніж спискові (кожний раз потрібно писати адресну частину, більші витрати паперу і т.д.). Спискові формуляри більш економічні, більш зручні для машинної обробки і контролю даних.

Для правильного заповнення статистичних формулярів, як правило, складається **інструкція** – перелік вказівок і основних положень, якими слід керуватися лічильнику або реєстратору при заповненні статистичних документів. Основна мета інструкції полягає в забезпеченні єдності тлумачення змістовної частини статистичних формулярів, як слід давати на них відповіді і заповнювати формуляри. Інструкція роз'яснює мету статистичного спостереження, характеризує його об'єкт і одиницю, час і тривалість спостереження, порядок оформлення документів і строки їх подання в статистичні органи.

Інструкція може бути видана або у вигляді окремої брошури, або дається в підказах, або на самому формулярі спостереження (звичайно на зворотному боці).

2.2. Форми, види і способи статистичного спостереження

Статистичні дані можна одержати різними шляхами і способами.

Залежно від **організації статистичного спостереження** розрізняють дві основні форми: 1) звітність; 2) спеціально організоване статистичне спостереження.

Звітністю називають такий вид спостереження, при якому відомості надходять у статистичні органи від підприємств, установ та організацій у вигляді обов'язкових звітів про їхню діяльність. Подання звітності за затвердженою програмою у встановлені адреси і строки є обов'язковим для *кожного підприємства, установи та організації. Статистична звітність – основна форма організації статистичного спостереження в Україні.*

Джерелом відомостей для статистичної звітності є первинні облікові записи в документах бухгалтерського та оперативно-технічного обліку.

Звітність здійснюється за строго встановленими формами, затвердженими Державним комітетом статистики України. Звітність за затвердженими статистичними органами формами, як і звітність, що здійснюється частіше встановленої періодичності, вважається незаконною. Органам статистики надано право відміняти незаконну звітність.

Перелік усіх форм із зазначенням їхніх реквізитів називається **табелем звітності.** Обов'язкова сукупність деяких зовнішніх елементів, зафіксованих в кожній формі звітності, називається **реквізитом** даної форми.

Основними реквізитами статистичної звітності є: 1) найменування форми; 2) номер і дата затвердження форми звітності; 3) адреси, в які подається звітність; 4) період, за який подаються відомості або на яку дату; 5) строки подання звітності; 6) назва підприємства або установи, яка подає звіт, і його адреса; 7) *назва міністерства (відомства), якому підпорядковане підприємство; 8) підписи посадових осіб, відповідальних за складання звіту.*

Звітність підрозділяють на загальнодержавну і відомчу. ***Загальнодержавна звітність*** *обов'язкова для всіх підприємств, установ і організацій, вона надходить і узагальнюється в органах загальнодержавної статистики для потреб державного управління.* **Відомча звітність** *збирається для своїх потреб міністерствами, відомствами* (наприклад, звітність по сільському господарству, торгівлі та ін.).

Розрізняють типову і спеціалізовану звітність. **Типова** має однакові показники для однієї або всіх галузей народного господарства. У **спеціалізованій** звітності показники відбивають специфіку тієї або іншої галузі і підгалузі (наприклад, звіти по птахівництву, рибництву, звірівництву і т.д.).

За **періодичністю** подання звітність підрозділяється на річну і поточну: квартальну, місячну і тижневу. Найбільш повною за складом показників, що подаються, є річна звітність.

За **способами подання** розрізняють звітність термінову, коли відомості передаються по телетайпу, телеграфу та іншими швидкими засобами, і **поштову.**

Поряд із звітністю важливим джерелом статистичних даних є спеціально організоване статистичне спостереження.

Спеціально організоване статистичне спостереження *являє собою збирання відомостей за допомогою переписів, одночасних обліків і обстежень. Прикладом спеціально організованого статистичного спостереження є: переписи населення, устаткування, залишків матеріалів, багаторічних насаджень, обстеження сімейних бюджетів населення, облік одержаної* продукції в особистих підсобних господарствах, переоцінка основних фондів у народному господарстві та ін.

У результаті спеціальних статистичних спостережень отримують дані про явища і процеси, які не охоплені статистичною звітністю, а в ряді випадків суттєво уточнюють дані поточного обліку.

Статистичні переписи є другою за значенням організаційною формою статистичного спостереження в Україні. **Перепис** являє спеціально організоване статистичне спостереження, спрямоване на облік чисельності і складу певних об'єктів (явищ), а також на встановлення якісних характеристик їх сукупностей на певний момент часу.

Відмінними особливостями переписів є: одночасність проведення їх на всій передбаченій території, єдність програми спостереження, стислість строків статистичного спостереження, реєстрація всіх одиниць спостереження станом на один і той же момент часу – критичну дату перепису.

Серед статистичних переписів особливе місце належить **переписам населення** – спеціально організованим статистичним спостереженням, метою яких є одержання даних про чисельність, склад і розміщення населення. Перепис населення 2001 р. дав змогу одержати повну і детальну статистичну інформацію про чисельність і склад населення країни за статтю, віком, сімейним складом, національністю, зайнятістю, суспільним групам, а також про розміщення населення по території країни. Програма перепису складалась з 19 запитань суцільного перепису.

Залежно від **ступеня охоплення одиниць** досліджуваної сукупності розрізняють два види статистичного спостереження: суцільне та несуцільне.

Суцільним називають таке спостереження, при якому обстеженню підлягають усі без винятку одиниці сукупності (наприклад, перепис населення, облік виходу продукції та ін.). Матеріали суцільного спостереження дають максимально повне уявлення про все розмаїття форм і можливість одержання точних характеристик про досліджувані соціально-економічні явища і процеси.

Однак через ряд причин (велика трудомісткість, тривалість проведення, висока вартість і т.д.) суцільне спостереження часто буває економічно недоцільним або практично нездійсненним. Тому на практиці переважно застосовують несуцільне спостереження.

Несуцільним називають таке спостереження, при якому обстежується тільки частина сукупності (наприклад, визначення якості насіння, жирності молока, втрат урожаю, вивчення використання робочого часу і устаткування, цін на ринках тощо.).

Несуцільне спостереження базується на обліку деякої частини, як правило, достатньо масової частини одиниць спостереження, яка дає змогу на основі наукового відбору одиниць одержати стійкі узагальнюючі характеристики всієї сукупності.

Якість несуцільного спостереження поступається результатам суцільного, однак досить очевидні і деякі переваги першого: за рахунок зменшення числа одиниць сукупності, що обстежуються, воно потребує менших витрат, сил і засобів, дає змогу застосовувати більш детальну програму спостереження, швидше підводити підсумки обстежень і, отже, підвищує оперативність статистичного матеріалу.

Несуцільне спостереження в свою чергу підрозділяється на вибіркове, монографічне, анкетне і основного масиву.

Вибірковим називають таке спостереження, при якому обстеженню підлягає певна частина сукупності явищ, яку отримали на основі ненавмисного випадкового відбору. Цьому виду спостереження присвячений спеціальний розділ підручника.

Монографічне спостереження – це докладний і всебічний опис окремих одиниць досліджуваної сукупності, які цікавлять дослідника. Монографічне обстеження може бути спрямоване на вивчення процесу розвитку окремого трудового колективу, узагальнення передового досвіду, опис нових *технологій виробництва і форм організації праці та ін. Об'єктом монографічного* опису можуть бути також сім'я, учбовий заклад, місто, регіон та інші об'єкти.

Анкетний спосіб спостереження ґрунтується на розсиланні анкет певному колу осіб або установ. Заповнення і повернення їх до органів, які проводять спостереження, є добровільним. Як правило, заповнених анкет повертається менше, ніж розсилається. Крім того, неможливо проконтролювати правильність відповідей на запитання анкети. Тому такий спосіб спостереження може застосовуватись у тих випадках, коли не вимагається висока точність відомостей, а потрібні приблизні характеристики. До нього удаються при проведенні соціологічних обстежень, у торгівлі для вивчення попиту на окремі товари і т.д.

Обстеження основного масиву являє собою спостереження за частиною найбільш крупних одиниць, питома вага яких переважає в загальному обсязі досліджуваної сукупності. За принципом основного масиву в країні організоване спостереження за міською ринковою торгівлею. Число охоплених нею міст складає менше 5% усіх міст, однак в них мешкає більше половини чисельності всього міського населення країни.

За способом реєстрації або **отримання статистичних даних** розрізняють три основних види спостереження: безпосереднє, документальне та опитування.

Безпосереднім називають спостереження, яке здійснюється шляхом реєстрації досліджуваних одиниць та їхніх ознак на основі безпосереднього огляду, підрахунку, зважування, зняття показників приладів спеціальними особами, які проводять спостереження, інакше кажучи, реєстраторами, в завдання яких входить поряд зі встановленням і оцінкою фактів фіксування їх у документах первинного обліку (наприклад, щорічний облік худоби, перепис багаторічних насаджень, інвентаризація матеріальних цінностей, метеорологічні спостереження – реєстрація температури повітря, ґрунту, снігового покриву, кількості опадів, облік одержаної продукції і т.д.).

Безпосереднє спостереження є досить точним і надійним джерелом статистичних даних, воно потребує великих затрат кваліфікованої праці.

Документальне спостереження ґрунтується на використанні різних документів первинного обліку підприємств, організацій установ (звітності, первинних бухгалтерських документів, річних звітів та ін.). Воно застосовується, наприклад, при переоцінці основних фондів підприємств, аналізі використання тракторів і автомобілів, урожайності і продуктивності тварин та ін. Цей вид спостереження так само, як і безпосереднє спостереження, забезпечує найбільшу вірогідність статистичних даних.

Опитування – це спосіб спостереження, при якому відомості отримують зі слів опитуваних осіб. Опитування може бути усним та письмовим.

Розрізняють три способи опитування: експедиційний, самореєстрації і кореспондентський.

Експедиційний спосіб опитування полягає в тому, що спеціально підготовлені реєстратори на основі опитування обстежуваних осіб заповнюють переписні формуляри, одночасно контролюючи правильність отриманих відповідей (наприклад, перепис населення). Цей спосіб опитування потребує найбільших витрат, проте гарантує високу якість матеріалу і дає змогу включати до програми такі запитання, які ризиковано ставити при інших способах опитування. Найважливіші статистичні обстеження населення проводяться експедиційним способом.

При способі самореєстрації (самообчисленні) спеціально виділені особи безпосередньо вступають в контакт з тими, від кого потрібно одержати відомості, роздають бланки, інструктують про порядок їх заповнення і призначають час, до якого вони мають бути заповненими; потім в призначений час ці особи отримують заповнені бланки, перевіряють повноту і правильність їх складання. Прикладом такого спостереження є обстеження сімейних бюджетів, при якому сім'ї самі ведуть записи про свої доходи і витрати, а статистики-бюджетники регулярно (два рази на місяць) при відвідуванні сімей перевіряють повноту і правильність цих записів.

При способі самореєстрації забезпечується суттєва економія робочого часу, оскільки лічильники звільнені від необхідності заповнювати формуляри. Якість отриманих даних поступається усе ж даним, отриманим експедиційним способом.

Суть **кореспондентського способу спостереження** полягає в тому, що необхідні відомості надають особи, які добровільно виявили бажання відповісти на поставлені в анкетах запитання. Цей спосіб не потребує великих витрат, але він не забезпечує високої якості матеріалів, так як перевірити точність даних, що повідомляються, безпосередньо на місцях не завжди можливо.

За **часом проведення** статистичне спостереження поділяють на *безперервне* і *перервне*.

Під **безперервним** (поточним) розуміють спостереження, при якому встановлення і реєстрація фактів проводиться мірою їх здійснення (наприклад, облік випуску продукції, виконання робіт, реєстрація народжень і смертей, шлюбів і розлучень, доходів і витрат у сім'ях при бюджетному обстеженні та ін.). Поточне спостереження проводиться на основі первинних документів, які містять інформацію, необхідну для достатньо повної характеристики досліджуваного явища.

Перервним *називають таке спостереження, яке проводиться або регулярно через певні проміжки часу, або у разі потреби.* Його поділяють на періодичне і одночасне.

Періодичним називають таке спостереження, яке здійснюють регулярно через певні проміжки часу (наприклад, щорічні обліки худоби, заключний облік посівних площ, який проводять один раз на рік після закінчення сівби ярих культур та ін.).

Одночасне – таке спостереження, яке проводять у разі потреби для вирішення якогось завдання (наприклад, перепис багаторічних насаджень, переоцінка основних фондів у галузях народного господарства, перепис житлового фонду, перепис і бонітування тварин, паспортизація полів і т.п.).

2.3. Помилки статистичного спостереження і способи контролю зібраних даних

У процесі збирання статистичних даних можуть виникнути похибки і неточності, які називають **помилками спостереження**. Кількісно вони визначаються різницею між зафіксованою величиною ознаки і дійсною її величиною.

Помилки статистичного спостереження мають різну природу і характер і по різному відбиваються на результатах спостереження.

Розрізняють передусім дві групи помилок статистичного спостереження – помилки реєстрації і помилки репрезентативності (представництва). *Кожна з цих груп помилок поділяється на випадкові* та систематичні.

Помилки реєстрації виникають внаслідок неправильного встановлення фактів у процесі спостереження або помилкового запису їх. Ці помилки притаманні як суцільному, так і несуцільному спостереженню.

Випадкові помилки реєстрації виникають у результаті дії різних випадкових причин: помилки рахунку, закруглення чисел, описки, обмовки і т.п.

Систематичні помилки реєстрації виникають внаслідок дії певних постійних причин (свідоме перекручення фактів у сторону зменшення або збільшення їх величини, неточність вимірювальних приладів). Прикладом систематичної помилки є широко відоме у статистиці явище вікової акумуляції, яке полягає у закругленні віку особами середньої і старшої вікових груп до чисел кратних 5 і особливо 10. Багато опитуваних, наприклад, замість віку 48-49 і 51-52 років кажуть, що їм 50 років.

Систематичні помилки реєстрації можуть бути навмисними і ненавмисними.

Навмисні помилки (свідомі, тенденційні перекручення) виникають внаслідок того, що опитуваний, знаючи дійсний стан речей, у цілях отримання користі свідомо повідомляє неправильні дані. Прикладом таких помилок є перекручення даних у статистичній звітності. **Ненавмисні** помилки викликаються різними випадковими причинами (наприклад, недбалістю або неуважністю реєстратора, несправністю вимірювальних приладів тощо).

Якщо випадкові помилки мають різну спрямованість і в силу дії закону великих чисел взаємно зрівноважуються (погашаються), то систематичні помилки спрямовані в один бік і тому зменшують або збільшують значення ознак, які реєструються, що впливає на кінцевий результат спостереження.

Помилки **репрезентативності** (представництва) притаманні тільки несуцільному спостереженню. Вони виникають внаслідок того, що вибіркова сукупність, як би ретельно і правильно не була сформована, недостатньо точно відображує характеристики генеральної сукупності.

Помилки репрезентативності так само, як і помилки реєстрації можуть бути випадковими і систематичними. **Випадкові помилки репрезентативності** – це відхилення, які виникають при несуцільному спостереженні внаслідок того, що сукупність відібраних одиниць неповно відтворює генеральну сукупність. Величину випадкової помилки репрезентативності можна виміряти за допомогою відповідних методів математичної статистики. **Систематичні помилки репрезентативності** – це помилки, які виникають внаслідок порушення принципів відбору одиниць у вибіркову сукупність. Докладніше помилки репрезентативності розглядаються у розділі 6.

Для виявлення і усунення допущених при реєстрації помилок можна застосувати лічильний і логічний контроль зібраного матеріалу.

Лічильний контроль полягає в лічильній перевірці підсумкових і розрахункових показників, а також арифметичній ув'язці пов'язаних між собою або які виводяться один з одного показників.

Логічний контроль ґрунтується на логічному взаємозв'язку між ознаками. На відміну від лічильного контролю логічний контроль безпосередньо не встановлює помилку спостереження, а тільки сигналізує про її можливість і потребує додаткового зв'язку з одиницею спостереження. Прийоми логічного контролю залежать від особливостей досліджуваного явища, організаційної форми спостереження, способу реєстрації даних та ін.

Найбільш розповсюдженими прийомами логічного контролю є: 1) зіставлення відповідей на різні взаємопов'язані запитання одного і того ж формуляра. Наприклад, зіставляючи відповіді на запитання про вік, освіту, сімейний стан та інші, можна логічним шляхом встановити помилку. Якщо, наприклад, виявиться, що громадянин десяти років одружений або шестирічна дитина має вищу освіту, то ясно, що при заповненні переписного листа припущені помилки або при запису віку, або другої характеристики; 2) порівняння записів у документі, що перевіряється, з аналогічними записами в інших документах; 3) порівняння даних спостереження з плановими або фактичними даними за попередні періоди, нормативними показниками, даними раніше проведених спостережень.

Крім того, логічний контроль спирається на уявлення про межі можливих значень ознаки: мінімуму і максимуму.

Завдання для самоконтролю до розділу 2

1. Що таке статистичне спостереження, в чому полягає його суть і основні завдання?

2. Назвіть основні програмно-методологічні та організаційні питання статистичного спостереження.

3. Що таке об'єкт і одиниця спостереження?

4. Що таке критичний момент і час спостереження?

5. Дайте визначення програми спостереження.

6. Яких вимог дотримуються під час формування програми статистичного спостереження?

7. Сформулюйте об'єкт, одиницю і мету статистичного спостереження і розробіть програму: а) обстеження фермерських господарств; б) аналізу продуктивності праці на підприємстві; в) аналізу використання робочої сили і робочого часу на підприємстві.

8. Складіть анкету опитування студентів з метою вивчення:

а) їхнього вікового і статевого складу, успішності, сімейного положення; б) їхньої оцінки якості викладання окремих дисциплін і майстерності викладачів; в) їхніх побутових умов і матеріального положення.

9. Перерахуйте основні форми статистичного спостереження.

10. Розкажіть про види статистичного спостереження: а) за організацією проведення; б) за ступенем охоплення одиниць статистичної сукупності; в) за способом реєстрації даних; г) за часом реєстрації даних.

11. Що таке статистична звітність, її види?

12. Що таке спеціально організоване статистичне спостереження?

13. Назвіть і охарактеризуйте види несуцільного спостереження.

14. Назвіть види носіїв статистичної інформації.

15. Що таке помилки статистичного спостереження та які їх види?

16. Як здійснюється контроль результатів спостереження? Назвіть способи усунення помилок статистичного спостереження.

Розділ 3

Зведення і групування статистичних даних. Статистичні таблиці

3.1. Поняття про статистичне зведення

Після того як одержаний в результаті статистичного спостереження первинний матеріал проконтрольований, можна перейти до зведення і групування даних – другого етапу статистичного дослідження.

Одержані в результаті статистичного спостереження дані про кожну одиницю спостереження необхідно систематизувати, привести в необхідний порядок, обробити, узагальнити і за допомогою системи узагальнюючих показників дати характеристику досліджуваного явища. Це завдання вирішується на етапі зведення статистичних даних.

Під **зведенням** розуміють сукупність прийомів наукового узагальнення і обробки даних статистичного спостереження з метою отримання статистичних показників і подальшого їх аналізу. На основі цих показників має бути дана характеристика чисельності сукупності, розміру притаманних їй ознак, структури сукупності та її якісний склад, встановлені специфічні особливості і закономірності досліджуваного явища, взаємозв'язку між ознаками та ін.

У результаті зведення здійснюється перехід від даних, які зібрані по кожній окремій одиниці об'єкта спостереження, до підсумкових даних по сукупності в цілому або групах, що виділені в її межах. Цими даними заповнюються складені макети таблиць або бланки статистичних формулярів.

Розрізняють зведення у вузькому і широкому розумінні слова.

Під статистичним **зведенням у вузькому розумінні слова** розуміють підрахунок підсумків у групах і підгрупах і оформлення одержаного матеріалу в таблицях.

Статистичне зведення в широкому розумінні слова включає такий комплекс операцій: 1) групування даних статистичного спостереження, яке включає відбір групувальних ознак, визначення числа груп і величини інтервалу, формування груп і підгруп; 2) підсумовування (зведення у вузькому розумінні слова) показників по окремих групах і по всій сукупності, тобто одержання абсолютних статистичних показників; 3) розрахунок на основі абсолютних показників середніх і відносних величин; 4) табличне і графічне оформлення результатів зведення та їх аналіз.

Статистичне зведення – відповідальний етап статистичного дослідження. Від якості зведення залежить і зміст отриманих на його основі висновків. Як і будь-яка велика робота, статистичне зведення має здійснюватись за заздалегідь розробленим планом. У план зведення включаються питання, пов'язані з послідовним здійсненням його окремих етапів, з черговістю обробки матеріалів спостереження. При складанні плану зведення розробляються макети зведених статистичних таблиць, на основі яких дається всебічна характеристика досліджуваних явищ і процесів. У плані зведення вказується також, хто і в які строки здійснює зведення, яким способом, куди надходять зведені дані, хто проводить їх подальшу обробку, аналіз та оформлення його результатів у таблицях, публікаціях, статистичних збірниках та ін.

Статистичні зведення відрізняються рядом ознак: за складністю побудови, способом розробки матеріалів статистичного спостереження і місцем проведення.

За складністю побудови зведення можуть бути простими і груповими. **Просте зведення** полягає в одержанні зведеного підсумку по всьому масиву вихідної інформації. При цьому будь-яке попереднє групування і систематизація вихідної інформації не виконуються. Тому просте зведення має в основному допоміжні цілі. **Групове зведення** на відміну від простого розробляється на основі вихідної інформації, яка попередньо піддається систематизації і групуванню. Отже, групове зведення відрізняється від простого своєю інформативністю, вмістом більшого числа групових підсумків.

За способом розробки матеріалів статистичного спостереження зведення підрозділяють на ручні і машинні. При **ручному зведенні** всі основні операції (шифровка, сортування, підрахунок підсумків і т.д.) виконуються вручну за допомогою карток або списків. Нині розробка статистичних матеріалів вручну застосовується дуже рідко, як виключення. В основному здійснюється **машинне зведення** даних за допомогою електронно-обчислювальних машин. При машинному зведенні первинні дані переносять зі статистичних формулярів на технічні носії інформації (магнітні диски, магнітні стрічки, перфострічки і т.д.), які потім вводять у машину разом з програмою обробки інформації.

За місцем проведення зведення може бути централізованим і децентралізованим. Централізованим називається зведення, при якому всі первинні статистичні матеріали зосереджуються в одному місці (наприклад, у Державному комітеті статистики України), де вони розробляються за єдиною програмою в потрібних розрізах і групах. Децентралізованим називається таке зведення, при якому підсумкові дані одержують на основі їх обробки послідовними етапами. Наприклад, спочатку виконується зведення даних по району, потім порайонні дані об'єднуються в областях, потім обласні зведення об'єднуються у Державному комітеті статистики України. Цим способом розробляються дані державної статистичної звітності.

Кожний вид зведення має свої переваги і недоліки.

При централізованому зведенні головною перевагою є наявність більшої можливості проведення його за єдиною методологією розробки даних з включенням додаткових групувань і розрахунків похідних показників, а також можливості використання найбільш ефективних технічних засобів обробки даних. Цей вид зведення відрізняється відносно невеликими затратами праці і високою точністю розрахунків. Однак при централізованому зведенні трудніше проконтролювати вірогідність первинних даних, має місце суттєвий розрив у часі між збиранням даних і результатом їх обробки, що знижує їх оперативну значущість.

При децентралізованому зведенні місцеві органи бистріше отримують потрібні матеріали для оперативного управління і прийняття рішень, а також бистріше здійснюється виправлення виявлених помилок у первинних документах. Однак при цьому матеріал розпиляється, не створюється достатньо великого масиву документів для впровадження більш ефективних засобів для обробки статистичної інформації.

У статистичній практиці частіше застосовують децентралізоване зведення, бо його результати можна бистріше використовувати на місцях. Інколи централізоване і децентралізоване зведення поєднують. Так, при розробці матеріалів перепису населення короткі попередні підсумки щодо загальної чисельності і складу населення в окремих населених пунктах і областях отримують за допомогою децентралізованого зведення, а кінцеві детальні підсумки перепису – в результаті централізованого зведення.

3.2. Статистичні групування, їх зміст, завдання і види

Зведення статистичних даних, як правило, не обмежується простим підрахунком загальних підсумків по досліджуваній сукупності. Найчастіше вихідна інформація на цій стадії статистичного дослідження впорядковується, систематизується, ділиться на групи за суттєвими ознаками. Це досягається за допомогою статистичних групувань – основного і вирішального моменту зведення.

Статистичне групування являє собою розчленування сукупності масових суспільних явищ на однорідні типові групи за суттєвими для них ознаками з метою всебічної характеристики їх стану, розвитку і взаємозв'язків.

Значення і необхідність групувань випливають з самого предмету статистики. В кожному складному масовому суспільному явищі є якісно відмінні групи, є старе, що відмирає, і народжуване нове, яке треба виділити і вивчити. Наприклад, у сільському господарстві сформувались різні соціальні типи господарств: приватні і державні сільськогосподарські підприємства, кооперативи, підсобні господарства підприємств і населення та ін. Кожному з цих типів притаманні характерні риси і закономірності. Тому при групуванні не можна змішувати різні типи господарств, потрібно виділяти їх якісно однорідні групи. Чим досконаліше статистика виділить такі групи, тим повніше вона їх охарактеризує, тим глибше розкриє суть процесів, що відбуваються, направлення і темпи їх розвитку і т.д.

Найбільш відповідальним моментом групування є відбір ознак, які дозволять відокремити один від одного дійсно суттєво відмінні групи одиниць. Кожна одиниця спостереження володіє багатьма ознаками. Одні з цих ознак виражають суть, найбільш характерне у даному явищі, інші – другорядне, поверхневе, нетипове.

Якщо в основу групувань покладені найбільш суттєві ознаки, то будуть виділені дійсно типові для даного явища групи. Якщо за основу групування узяти малозначущі ознаки, то ми отримаємо вкрай поверхневе або взагалі перекручене уявлення про явище.

Метод статистичних групувань, який є одним з найбільш ефективних методів обробки масових даних, відкриває широкі можливості для вивчення взаємозв'язків між явищами, виявлення об'єктивних закономірностей досліджуваних явищ і процесів, встановлення на певному етапі переходу кількісних змін у якісні. Побудувавши групування досліджуваних об'єктів за будь-якою суттєвою ознакою і охарактеризувавши виділені групи різними показниками, можна прослідкувати залежність між ознаками, що покладені в основу групування, і вибраними показниками.

Метод статистичних групувань дає змогу так розробити первинний статистичний матеріал, щоб всі суттєві риси і особливості досліджуваних суспільних явищ отримали чітке вираження. Цим визначається роль групувань як наукової основи зведення.

У сільському господарстві важлива роль належить групуванням сільськогосподарських підприємств за виходом валової продукції на одиницю ресурсів *сільськогосподарського виробництва*, продуктивністю праці, собівартістю продукції, рентабельністю, урожайністю, продуктивністю тварин та ін.

Використання методу групувань створює умови для застосування багатьох інших статистичних методів наукового пізнання, передусім відносних і середніх величин, індексного, кореляційного, дисперсійного методу та ін. Перераховані методи ефективні тільки на основі групувань і в поєднанні з ними.

За допомогою групувань вирішують різні завдання. Найважливішими з них є: 1) виділення і всебічна характеристика різних соціально-економічних явищ; 2) характеристика структури досліджуваних явищ; 3) вивчення взаємозв'язків між окремими ознаками сукупності. Відповідно до цього розрізняють три види групувань: типологічні, структурні і аналітичні.

Групування, що приводять до виділення соціально-економічних типів, класів, одноякісних груп або сукупностей, називають **типологічними**. До них відносяться групування підприємств за формами власності, виробничим напрямком (зернові, молочні, відгодівельні та ін.); населення – за класовою належністю або соціальними групами; робітників – на зайнятих переважно фізичною і переважно розумовою працею.

Типологічні групування широко застосовуються в економічних, соціологічних і демографічних дослідженнях. Вони не тільки служать меті *виділення типів явищ*, але й забезпечують можливість аналізу специфіки і особливостей розвитку окремих типів, зміни їх співвідношень в рамках загального економічного процесу. Проілюструємо це на прикладі (табл. 3.1).

Таблиця 3.1

Основні фонди народного господарства України за формами власності, %

Форма власності	1994 р.	1997 р.	1999 р.
Всі основні фонди			
в тому числі за формами власності:	100,0	100,0	100,0
державна	59,1	45,8	39,0
комунальна	6,2	10,9	14,3
приватна	7,7	1,9	2,4
колективна	26,7	41,3	44,1
власність міжнародних організацій			
та юридичних осіб інших держав	0,3	0,1	0,2

Дані табл. 3.1 свідчать про те, що в динаміці збільшується частка основних фондів комунальної і *колективної власності і зменшується* частка основних фондів державної і приватної форм власності.

Структурні групування *характеризують розподіл одиниць досліджуваної сукупності за будь-якою ознакою.* За їх допомогою проводиться аналіз структури сукупності і структурних зрушень в розвитку соціально-економічних явищ і процесів. Порівняння структурних групувань у часі дає уявлення про структурні зрушення. До них відносяться групування населення за статтю, віком, сільськогосподарських підприємств за площею сільськогосподарських угідь, поголів'ям худоби та ін. Прикладом структурного групування можуть бути дані табл. 3.2.

Таблиця 3.2

Групування населення України за статтю в 2001 р. (на 5.12.2001 р.)

Населення	Чисельність населення, млн чол.	В % до підсумку
Чоловіки	22,5	46,5
Жінки	25,9	53,5
Разом	48,4	100,0

Відомо, що явища суспільного життя знаходяться в певному взаємозв'язку і взаємозалежності. Тому важливим завданням статистичних групувань *є вивчення взаємозв'язків і взаємозалежностей між ознаками.* Таке завдання вирішується за допомогою аналітичних групувань.

Аналітичними називають групування, за допомогою яких вивчаються взаємозв'язки між окремими ознаками статистичної сукупності. Прикладом таких групувань можуть бути групування, в яких вивчається взаємозв'язок між собівартістю та її факторами, продуктивністю тварин та її факторами тощо. Аналітичні групування можуть бути побудовані за **результативною** (урожайністю, собівартістю, виходом продукції з одиниці земельної площі та ін.) і **факторною** (якість ґрунту, кількість добрив, опадів та ін.) ознакою. Відповідно одержимо результативні і факторні групування.

Наведемо приклад аналітичного групування за результативною ознакою (табл. 3.3).

Проведемо аналіз групової таблиці. Для цього проведемо деякі зіставлення факторів *і умов виробництва по окремих групах.* Спочатку порівняємо показники крайніх груп – I і IV (нижчої і вищої).

З даних табл. 3.3 видно, що різниця в результативній ознаці (урожайності) між цими групами досягає 16,3 ц/га. Урожайність зернових культур в господарствах IV групи вище, ніж в господарствах I групи в 1,7 рази головним чином за рахунок більш високого рівня *виробництва: у господарствах* IV групи порівняно з господарствами I групи вносять мінеральних добрив на 1 га зернових більше на 1,34 ц діючої речовини, або в 1,95 рази, працівників на 100 га ріллі на 3,7 чол., або в 1,24 рази більше. Відрізняється *ця група і більш високою якістю ґрунту – на 29 балів*, або в 1,56 рази.

Таблиця 3.3

Залежність урожайності зернових культур від її факторів в господарствах лісостепової зони області

Групи господарств за урожайністю, ц/га	Кількість господарств	Середня урожайність, ц/га	Якість ґрунту, балів	Внесено мінеральних добрив на 1 га зернових культур, ц діючої речовини	Припадає середньо-річних працівників на 100 га ріллі, чол.
I – до 25,0	16	22,1	52,0	1,41	15,6
II – 25,1-30,0	21	27,7	58,2	1,64	17,4
III – 30,1-35,0	38	32,6	69,4	2,12	17,5
IV – понад 35,0	15	38,4	81,0	2,75	19,3
У середньому	90	31,5	67,7	2,07	17,3

Спостерігаються суттєві відмінності в рівнях середньої урожайності по групах господарств. Зіставлення між собою показників окремих груп дозволяє зробити висновок, що більш високій рівень урожайності від групи до групи (від I до IV) зумовлюється за рахунок кращої забезпеченості добривами, робітниками, кращою якістю ґрунту. Так, наприклад, зіставлення урожайності і факторів та умов виробництва II та I групи показує, що урожайність в II групі господарств порівняно з I групою господарств вище в 1,25 рази. Ця група господарств має вище забезпеченість працівниками (в 1,12 рази), більше вносять добрив (в 1,16 рази), має більш родючі ґрунти (в 1,12 рази).

В цілому можна зробити висновок про те, що більш високій рівень урожайності зернових культур забезпечується високим рівнем інтенсифікації виробництва, кращою якістю земель і більш раціональним і ефективним використанням самих факторів і умов виробництва.

Розглянуте вище результативне групування дає змогу виявити фактори урожайності і визначити їх спільний комплексний вплив.

З метою оцінки ступеня впливу окремих факторів на урожайність необхідно провести факторне групування господарств за якістю ґрунту, як одного з факторів, що найбільш коливається, по типових групах (див. табл. 3.3).

Для побудови факторного групування використаємо дані тих самих 90 господарств лісостепової зони області (табл. 3.4).

Таблиця 3.4

Економічна ефективність виробництва зерна при різній якості ґрунту

Показник	Групи господарств за якістю ґрунту, балів				У середньому
	I до 56	II 57-66	III 67-76	IV понад 76	
Кількість господарств	19	25	26	20	90
Середня якість ґрунту, балів	52,3	63,1	73,8	81,2	67,7
Урожайність, ц/га	21,5	28,0	34,7	39,4	31,5
Собівартість 1 ц зерна, грн	8,70	7,55	6,66	6,01	7,15
Затрати праці на 1 ц зерна, людино-годин	1,71	1,58	1,38	1,22	1,47

З даних табл. 3.4 видно, що з поліпшенням якості ґрунту підвищується економічна ефективність виробництва зерна. Ця закономірність спостерігається *при порівнянні всіх груп між собою*. Так, *порівняння вищої* (IV) і нижчої (I) групи, де середня якість ґрунту відповідно складає 52,3 і 81,2 бали показує, що у вищій групі урожайність на 17,9 ц/га, або в 1,83 рази вище, а собівартість і трудомісткість 1 ц зерна (затрати праці на 1 ц) нижче відповідно на 2,69 грн. і 0,49 людино-години, або на 30,9 і 28,7%.

Отже, можна зробити висновок, що в даних умовах більш висока якість ґрунту забезпечує підвищення економічної ефективності виробництва зерна.

За кількістю ознак, покладених в основу групування, розрізняють прості і комбінаційні групування. Групування, проведені за однією ознакою, називають **простими або одновимірними** (див. табл. 3.3 і 3.4), а за двома і більшим числом ознак – **комбінаційними, або багатовимірними**.

При побудові комбінаційного групування сукупність спочатку підрозділяється на групи за однією ознакою, а потім отримані групи поділяються в свою чергу на підгрупи за другою, третьою і т.д. ознаками. В цьому випадку можна простіджувати вплив на результат сукупності факторів, покладених в основу групування, і оцінити вплив кожного окремо узятого фактора при вирівняних значеннях інших факторів.

Комбінаційне групування дає змогу також оцінити спільний вплив на результат групувальних ознак.

Порядок побудови комбінаційного групування той самий, що і простого, за однією ознакою. Спочатку вирішується питання, про першу групувальну ознаку, потім про другу і т.д.

При комбінаційному групуванні важливо забезпечити достатнє число одиниць в кожній групі і підгрупі. При комбінації ознак число підгруп швидко зростає зі збільшенням числа групувальних ознак: воно дорівнює *добутку кількості груп, утворених за кожною ознакою із ознак окремо.* Наприклад, якщо за першою ознакою виділити три групи, за другою – три і за третьою – дві, то в підметі таблиці одержимо 18 рядків (3Ч3Ч2), не рахуючи підсумкових рядків по групах і загального підсумку. Звідси зрозуміло, що комбінаційне групування при невеликій чисельності доцільно будувати за 2-3 ознаками. При виділенні більшого числа групувальних ознак підгрупи *можуть виявитися нечисленними, що утруднить виявлення існуючої між* ознаками залежності.

Порівняно з простими комбінаційні групування володіють рядом додаткових аналітичних можливостей. Розглянемо це на прикладі комбінаційного групування 90 господарств лісостепової зони за якістю грунту і кількістю внесених мінеральних добрив (табл. 3.5). При цьому вивчимо вплив на ефективність виробництва зерна різних доз мінеральних добрив при вирівняній якості грунту.

Таблиця 3.5

Вплив якості ґрунту і різних доз добрив на економічну ефективність виробництва зерна

Групи господарств за якістю ґрунту, балів	Підгрупи господарств за внесенням добрив на 1 га зернових культур, ц діючої речовини	Кількість господарств	Урожайність, ц/га	Собівартість 1 ц зерна, грн.	Затрати праці на 1 ц зерна, людино-годин
I - до 56	а) до 2,1 б) понад 2,1	11 8	20,4 22,9	9,15 8,50	1,85 1,65
	у середньому	19	21,5	8,70	1,71
II – 57-66	а) до 2,1 б) понад 2,1	10 15	24,6 31,1	8,27 7,02	1,61 1,57
	у середньому	25	28,0	7,55	1,58
III – 67-76	а) до 2,1 б) понад 2,1	14 12	33,8 35,2	6,95 6,29	1,50 1,21
	у середньому	26	34,7	6,66	1,38
IV – понад 76	а) до 2,1 б) понад 2,1	7 13	36,0 43,7	6,28 5,64	1,39 1,16
	у середньому	20	39,4	6,01	1,22
	Разом	90	31,5	7,15	1,47
Підгрупи за добривами	а) до 2,1 б) понад 2,1	42 48	28,5 33,1	7,60 6,82	1,57 1,40

Групи господарств за якістю ґрунту візьмемо ті самі, що і в табл. 3.4.

З даних табл. 3.5 чітко видна закономірність: з поліпшенням якості ґрунту і збільшенням доз добрив економічна ефективність виробництва зерна підвищується від групи до групи і від підгрупи до підгрупи (урожайність зростає, а собівартість і затрати праці на 1 ц знижуються).

Проведемо аналіз взаємозв'язку групувальних ознак з результативними показниками. Розглянемо взаємозв'язок урожайності з якістю ґрунту і дозами добрив на 1 га зернових культур. Для цього на основі табл. 3.5 складемо більш зручну для аналізу і читання табл. 3.6.

Таблиця 3.6

Вплив якості ґрунту і доз добрив на урожайність зернових культур

Групи господарств за якістю ґрунту, балів	Підгрупи господарств за внесенням добрив на 1 га зернових культур, ц діючої речовини		У середньому
	а) до 2,1	б) понад 2,1	
I – до 56	20,4	22,9	21,5
II – 57-66	24,6	31,1	28,0
III – 67-76	33,8	35,2	34,7
IV – понад 76	36,0	43,7	39,4
У середньому	28,5	33,1	31,5

Дані табл. 3.6 свідчать про те, що на урожайність зернових культур суттєво впливає як якість ґрунту, так і добрива. Вплив якості ґрунту можна оцінити при порівнянні урожайності по групах в межах кожної підгрупи при однакових дозах добрив. Прибавки урожайності за рахунок якості ґрунту при порівнянні I і II груп склали: по I підгрупі 4,2 ц/га (24,6 – 20,4), по II підгрупі 8,2 ц/ га (31,1 – 22,9).

Ці прибавки при порівнянні II і III груп складають: по I підгрупі 9,2 ц/га (33,8 – 24,6), по II підгрупі 4,1 ц/га (35,2 – 31,1), а між крайніми групами (I і IV) відповідно по I підгрупі 15,6 ц/га (36,0 – 20,4), по II підгрупі 20,8 ц/га (43,7 – 22,9).

Вплив на урожайність доз добрив прослідковується при порівнянні її рівня по підгрупах у межах тієї чи іншої групи за якістю ґрунту. При переході від першої підгрупи до другої приріст урожайності становить:

а) в I групі (до 56 балів) 2,5 ц/га (22,9 – 20,4);

б) в II групі (57-66 балів) 6,5 ц/га (31,1 – 24,6);

в) в III групі (67-76 балів) 1,4 ц/га (35,2 – 33,8);

г) в IV групі (понад 76 балів) 7,7 ц/га (43,7 – 36,0).

Порівняння одержаних приростів урожайності за рахунок кожного з факторів показує, що найвищі прибавки урожайності за рахунок добрив одержані в IV групі господарств, де найвища якість ґрунту.

За даними табл. 3.6 оцінимо спільний вплив на урожайність двох групувальних ознак, так, в підгрупі з самою високою якістю ґрунту і найвищими дозами добрив (4б) урожайність порівняно з підгрупою з самим низьким рівнем факторів (1а) вище на 23,3 ц/га (43,7 – 20,4). Отже, за рахунок спільного впливу двох факторів урожайність підвищилась на 23,3 ц/га.

Аналогічно можна провести аналіз взаємозв'язку інших результативних показників (собівартості і затрат праці на 1 ц зерна) з групувальними ознаками (якістю ґрунту і дозами добрив).

Завдання всебічного комплексного аналізу соціально-економічних явищ не може бути вирішене шляхом побудови одного групування: необхідна їх **система** – ряд взаємопов'язаних групувань, які доповнюють і конкретизують один одне. У системі групувань першорядна роль належить результативним, факторним і комбінаційним групуванням.

Розглянемо деякі особливості застосування зазначених групувань, їх взаємозв'язки, переваги і недоліки.

Результативні групування дають змогу вивчити відмінності в результатах, виділити виробничі типи і охарактеризувати їх особливості.

При групуванні за результативною ознакою групи мають бути охарактеризовані комплексом найважливіших факторів, які формують досліджуваний результативний показник.

Результативне групування дає змогу вивчити спільний вплив на результат комплексу найбільш суттєвих факторів, ступінь відмінностей груп за окремими факторами, відібрати головні з них, визначити напрямок і оцінити силу їх дії. Такі фактори відбирають на основі теоретичного аналізу досліджуваного явища, причому це можуть бути найістотніші, такі, що формують даний результат.

При побудові аналітичного групування для характеристики взаємозв'язків і подальшого аналізу між ознаками по групах можна обмежитися відбором 2-3 найважливіших факторів.

Результативні групування, що відіграють значну роль у виділенні соціально-економічних типів і основних факторів виробництва, водночас мають істотний недолік: в групах з різними результатами за ними можна визначити лише рівнодійну всіх факторів, отже в одній групі об'єднуються одиниці з однаковим результатом, незалежно від того, за рахунок яких факторів цей результат досягнуто. Наприклад, високої врожайності в одному випадку досягають за рахунок високого рівня інтенсивності виробництва при низькій якості ґрунту, а в іншому – на ґрунтах високої якості при низькому рівні інтенсивності виробництва. У групі ж у цілому ці фактори можуть вирівнюватись і на перше місце виходять інші фактори. Тому, щоб точніше оцінити вплив комплексу та окремих факторів на результат, результативні групування треба застосовувати в комплексі з факторними, які дають змогу в узагальненому вигляді розкрити ступінь впливу на результат окремих факторів, що діють у різноманітних умовах.

При групуванні за факторною ознакою групи характеризують комплексом найважливіших результативних показників, а також важливими факторами. Це дає змогу обґрунтовано віднести відмінності в результатах за рахунок факторної ознаки, покладеної в основу групування. Факторні групування значною мірою дають змогу усереднити вплив інших факторів на результат і вивчити середню зміну результативної ознаки під впливом ознаки, покладеної в основу групування. Проте відмінності в результатах можна віднести за рахунок досліджуваного фактора в тому випадку, якщо інші умови (фактори) у групах у середньому буде вирівняно. На практиці при вивченні економічних явищ внаслідок тісного взаємозв'язку факторів такого вирівнювання немає і досягти цього дуже важко.

Оскільки вплив сукупних умов (факторів) у більшості випадків не розглядають, то висновки про дію факторів (у разі групування за однією ознакою) будуть неточними, а відмінності в результатах цілком відносять за рахунок досліджуваного фактора. Ця особливість і зумовлює основний недолік факторних групувань, застосованих на практиці.

Точно і обґрунтовано можна визначити вплив факторів на результат на основі комбінаційного групування, здійснивши спеціальне статистичне вирівнювання інших умов.

3.3. Методологія статистичних групувань

Статистичні групування здійснюють у кілька послідовних етапів: 1) теоретичний аналіз досліджуваного явища або процесу; 2) вибір групувальної ознаки (ознак); 3) визначення кількості груп і величини інтервалу; побудова інтервального ряду розподілу одиниць сукупності за досліджуваною групувальною ознакою (ознаками); 4) визначення та обґрунтування системи статистичних показників для виділення і характеристики типових груп; складання макетів таблиць; 5) обчислення абсолютних, відносних і середніх показників; 6) табличне і графічне оформлення результатів групування; 7) аналіз одержаних результатів; формулювання висновків та пропозицій.

Безпосередній побудові групування має передувати глибокий теоретичний аналіз досліджуваного явища або процесу, в якому провідна роль належить з'ясуванню тенденцій і закономірностей розвитку явища, характеру його рушійних сил, специфіці виникнення в ході цього розвитку нових типів та форм явищ. Важлива роль в теоретичному аналізі відводиться також вивченню взаємозв'язку досліджуваного явища з іншими явищами, встановленню впливу окремих факторів на результативні показники.

Принципове значення при побудові групувань має вибір групувальної ознаки, визначення кількості груп і величини інтервалу. Вибір групувальної ознаки, тобто ознаки, на основі якої виділяють різні типи, групи, підгрупи, є одним з найважливіших моментів побудови групувань.

Вибір групувальної ознаки має бути оснований на аналізі якісної природи досліджуваного явища. Всебічний теоретично-економічний аналіз суті явища має бути спрямований на те, щоб у відповідності з метою і завданнями дослідження покласти в основу групування суттєві ознаки.

Групувальними ознаками можуть бути кількісні, атрибутивні (якісні), результативні та факторні ознаки.

Відібравши групувальну ознаку (ознаки) і побудувавши ранжированний ряд за цією ознакою, встановлюють кількість груп, на які буде поділено сукупність, що вивчається, і величину інтервалу.

Кількість груп залежить від загальної чисельності одиниць сукупності, характеру групувальної ознаки і виду групувань. Разом з тим при вирішенні цього питання слід дотримуватися двох важливих умов побудови групувань: 1) виділені групи мають відрізнятися якісною однорідністю; 2) кількість одиниць у кожній групі має бути досить великою. Ця вимога випливає із закону великих чисел.

Визначення числа груп і інтервалів у групуванні передусім залежить від того, якою є групувальна ознака – атрибутивною чи кількісною. Якщо групування здійснюють за атрибутивною (якісною) ознакою (стать, сорт, порода, професія тощо), то виділяють стільки груп, скільки є градацій ознаки. Аналогічно виділяють групи і при групуванні за дискретною кількісною ознакою, яка змінюється в невеликих межах (кількість членів родини, оцінки студентів, кількість приплоду від однієї матки, кількість бригад у ТОВ тощо).

Якщо ж групують за кількісною ознакою (урожайність, собівартість тощо), що змінюється безперервно і набуває в певних межах будь-яких дрібних значень, то групи виділяють шляхом встановлення для кожної з них інтервалів, зазначених верхньою і нижньою межами величини ознаки для даної групи.

При встановленні числа груп і меж інтервалів важливо встановити за кількісними змінами якісні переходи, щоб виділити типи, не змішати суттєво відмінні одиниці спостереження в одній групі. Це завдання вирішується на основі теоретичного аналізу досліджуваного явища (процесу), порівняння групувальної ознаки з раніше оціненими величинами, для яких якісні переходи відомі.

Якщо ж заздалегідь характер зміни кількісної ознаки і якісні переходи в ньому оцінити важко, то групування доцільно проводити в такій послідовності.

1. Побудувати ранжирований ряд розподілу за обгрунтовано виділеною ознакою, в якому всі одиниці спостереження розташовуються у порядку зростання або зменшення групувальної ознаки. Ранжирований ряд показує інтенсивність зміни величини групувальної ознаки. Різка зміна його величини при переході від однієї одиниці спостереження до другої є свідоцтвом якісних відмін між ними. Аналізуючи ранжирований ряд, особливо його графічне зображення, можна виділити якісно відмінні групи.

2. Побудувати інтервальний варіаційний ряд розподілу з виділенням достатньо великого числа груп, який дасть змогу одержати уявлення про склад досліджуваної сукупності і характер розподілу. При встановлені числа груп і величини інтервалу слід уникати як надмірного подрібнювання сукупності на групи (не виявиться властива масовим даним закономірність в розподілі), так і занадто малого числа груп (об'єднуються в групу якісно відмінні одиниці). Інтервальний ряд при поступовій (плавній) зміні ознаки будується звичайно з рівними інтервалами. Якщо групувальна ознака змінюється нерівномірно, то інтервали можуть бути нерівними.

3. Побудувати проміжне аналітичне групування і шляхом об'єднання дрібних однорідних груп перейти до типологічного групування. Аналітичне групування дає змогу на основі аналізу його показників дати якісну оцінку виділеним в інтервальному ряду групам. Якщо цей аналіз показує однорідність кількох (двох, трьох і т.д.) послідовно розташованих в інтервальному ряду груп, то є підстави для їх об'єднання в одну типову групу.

Для побудови інтервального варіаційного ряду необхідно встановити число груп і величину інтервалу.

Питання щодо числа груп і величини інтервалу слід вирішувати з урахуванням множини обставин, передусім виходячи з цілей дослідження, особливостей досліджуваного явища та ін. При цьому число груп і величину інтервалу слід встановити такими, які б дозволили більш рівномірно розподілити одиниці сукупності по групах і досягти при цьому їх представництва, якісної однорідності.

При визначенні числа груп потрібно брати до уваги розмах варіації ознаки, тобто різницю між його максимальним і мінімальним значенням. Чим більший цей розмах, тим, як правило, більше утворюється груп. Необхідно також враховувати чисельність досліджуваної сукупності. Доцільно, щоб число груп не було занадто великим чи малим і щоб в кожну групу попало достатньо велике число одиниць спостереження.

Число груп наближено можна визначити за формулою Стерджесса

$$n = 1 + 3{,}322 \lg N,$$

де N – чисельність сукупності.

На основі цієї формули можна встановити число груп для сукупностей, що мають різну чисельність:

чисельність сукупності (N) 25-40; 40-60; 60-100; 100-200; 200-500 понад 500;

число рекомендованих груп (n) 5-6; 6-7; 7-8; 8-9; 9-10; 10-15.

Виведена на основі теоретичних доведень формула Стерджесса не враховує особливостей і характеру варіації і розподілу досліджуваної ознаки. Тому механічне її застосування може призвести до неправильних висновків.

Визначення числа груп за формулою Стерджесса обґрунтоване в тих випадках, коли розподіл одиниць сукупності за даною ознакою наближається до нормального, застосовуються рівні інтервали в групах і при незначній варіації ознаки. В решті випадків кількість груп має бути визначено у вищевикладеній послідовності (побудова ранжированого ряду, інтервального ряду, проміжного аналітичного групування).

Після встановлення числа груп визначають величину інтервалу. **Інтервалом групування** називають різницю між максимальним і мінімальним значенням ознаки в кожній групі. Однак цю величину можна визначити як різницю між верхніми і нижніми межами значень ознаки в суміжних групах. У практиці статистичних групувань правильне встановлення величини інтервалу має першорядне значення для утворення якісно однорідних груп.

Коли значення групувальної ознаки в ранжированому ряду має плавний наростаючий характер, величина інтервалу визначається за формулою:

$$h = \frac{x_{\max} - x_{\min}}{n},$$

де $x_{\max}$ і $x_{\min}$ – максимальне і мінімальне значення ознаки; n – число груп.

З формули видно, що величина інтервалу знаходиться у прямій залежності від розмаху варіації і в оберненій від числа груп: чим більше розмах варіації, тим більше величина інтервалу, чим більше число груп, тим менше величина інтервалу.

Застосування цієї формули правомірно для випадків, коли всі значення ознаки мають плавний, поступовий характер наростання, а чисельність сукупності достатньо велика. Якщо ж невелика частина сукупності значно віддалена за розміром групувальної ознаки від сукупності основного масиву, то замість $x_{\max}$ всієї сукупності необхідно взяти $x_{\min}$ основного масиву сукупності, а різко відмінні одиниці виділити в особливу групу. Невиконання цієї вимоги може призвести до того, що переважна частина одиниць сукупності сконцентрується в одній-двох групах, в той час, як в решту груп увійде дуже невелике число одиниць або взагалі не увійде жодної одиниці.

За способом побудови інтервали можуть бути рівними і нерівними, відкритими і закритими, спеціалізованими. Вибір того або іншого виду інтервалу залежить від характеру розподілу одиниць досліджуваної сукупності.

Рівними називають інтервали, у яких різниці між верхньою і нижньою межами однакові. Групування з рівними інтервалами застосовуються тоді, коли варіація ознаки проявляється у порівняно вузьких межах і розподіл носить більш або менш рівномірний характер. Визначення величини інтервалу у випадку групування із застосуванням рівних інтервалів здійснюється за вищенаведеною формулою. *Припустимо, що площа хмелю* в групі господарств коливається від 10 до 110 га. Необхідно побудувати групування за розміром площі, утворивши 5 груп з рівними інтервалами. Величина інтервалу становитиме:

$$h = \frac{x_{max} - x_{min}}{5} = \frac{110 - 10}{5} = 20 \text{ га.}$$

Додаючи до мінімального значення ознаки (в даному випадку 10 га), знайдемо значення інтервалу, одержимо верхню межу першої групи: 10 + 20 = = 30 га. Додаючи далі величину інтервалу до верхньої межі першої групи, одержимо верхню межу другої групи: 30 + 20 = 50 га і т.д. У результаті одержимо такі групи господарств за розміром площі хмелю (га): 10-30; 30-50; 50-70; 70-90; 90-110.

Нерівними називають інтервали, у яких різниця між верхньою і нижньою межами неоднакові. Нерівні інтервали обчислюються в тих випадках, коли досліджувана ознака змінюється в широких межах. Прикладом нерівних інтервалів може бути групування господарств за чисельністю великої рогатої худоби (голів): 1-99; 100-299; 300-599; 600-999; 1000-3000; понад 3000.

Відкритими називають інтервали, у яких наперед невідомі максимальне і мінімальне значення. Тому при групуванні перший і останній інтервал залишаються відкритими. Наприклад, групування господарств за урожайністю вівса (ц/га): до 15; 15-20; 20-25; 25-30; понад 30.

Закритими називають інтервали, у яких максимальне і мінімальне значення відомі. Наприклад, групування працівників за стажем роботи (років): 0-5; 5-10; 10-15; 15-20; 20-25; 25-30; 30-35; 35-40.

У групуваннях, що мають за мету відобразити якісну своєрідність груп, застосовуються **спеціалізовані** інтервали. В цьому випадку в кожній групі є свій особливий зміст і межа інтервалу встановлюється там, де відбувається перехід від однієї якості до другої. Кількість груп встановлюється відповідно до теорії *питання. Наприклад, при характеристиці відгодівель-*

них господарств за чисельністю поголів'я виділяють мілкі, середні і крупні господарства; за рівнем рентабельності – збиткові, низькорентабельні, середньорентабельні, високорентабельні.

При групуванні даних за кількісною ознакою важливе значення має правильне позначення нижньої і верхньої меж кожної групи. Це пов'язано з наступним віднесенням окремих одиниць спостереження до відповідної групи. Так, якщо до наведених вище груп господарств за площею хмелю (10-30; 30-50; 50-70; 70-90; 90-110 га) не дати спеціальних вказівок, то виникають труднощі при побудові та аналізі даного групування. Куди, наприклад, віднести господарства з площею, рівною 30 га, в першу чи в другу групу? Для усунення подібної невизначеності потрібні додаткові вказівки про те, рахувати верхні межі груп «включно» (при побудові групувань звичайно користуються цим правилом), то тоді господарства з площею хмелю до 30 га, 50 га і т.д. мають бути віднесені відповідно до першої, другої і т.д. групи. Якщо ж рахувати верхні межі «виключно», то господарства із зазначеними площами мають бути віднесеними до другої, третьої і т.д. групи.

Той самий результат досягається і за допомогою відкритих інтервалів у першій і останній групах. Щоб показати, що господарства з площею хмелю, рівній верхній межі інтервалу, включаються в дану групу, останню групу слід позначити «понад 90 га». Навпаки, щоб показати, що верхні межі інтервалів не входять в дану групу, останню групу потрібно позначити «90 га і більше».

Часто для надання групуванням більшої визначеності верхню межу попередньої і нижню межу наступної групи позначають по-різному. Якщо групувальна ознака може приймати тільки цілі значення, то нижня межа наступної групи відрізняється від верхньої межі попередньої групи на одну цілу одиницю. Наприклад, групування тих же господарств за площею хмелю може бути виконане за такими інтервалами (га): 10-30; 31-50; 51-70; 71-90; 91-110. Аналогічно будуються інтервали з десятими і сотими частинами одиниць. Наприклад, інтервали господарств за собівартістю 1 ц зерна можуть мати вигляд (грн.): до 14,00; 14,01-15,00; 15,01-16,00 і т.д.

3.4. Вторинне групування

Поряд з первинним групуванням у статистиці знаходить широке застосування вторинне групування. **Вторинним групуванням** називають утворення нових груп на основі раніше проведеного групування.

Вторинне групування використовують для вирішення різних завдань, найважливішими з яких є: 1) утворення на основі групувань за кількісними ознаками якісно однорідних груп (типів); 2) приведення двох (або більше) групувань з різними інтервалами до єдиного виду з метою порівнянності та аналізу; 3) утворення більш укрупнених груп, в яких ясніше проявляється характер розподілу.

Суть цього прийому полягає в одержанні порівнянних даних по різних групуваннях, для чого: чисельний склад групи (за %) фіксується на одному рівні у всіх групуваннях; по всіх групуваннях встановлюється також рівне число груп і однаковий зміст групових таблиць. Порівнянню і зіставленню підлягають не абсолютні показники по групах, а відносні величини, процентні відношення.

Розрізняють два способи вторинного групування: 1) шляхом перетворення інтервалів первинного групування (частіше простим укрупненням інтервалів) і 2) шляхом закріплення за кожною групою певної частини одиниць сукупності (часткове перегрупування). При використанні цих способів вторинного групування звичайно припускають, що розподіл ознаки всередині інтервалів буде рівномірним.

Застосування вторинного групування для приведення двох групувань з різними інтервалами до єдиного виду з метою порівняння проілюструємо на такому прикладі. Для цього використаємо дані первинного групування двох районів за чисельністю працівників тваринництва (табл. 3.7).

Таблиця 3.7

Групування господарств двох районів за чисельністю працівників тваринництва

Район I		Район II	
Групи господарств за чисельністю працівників, чол.	**в % до підсумку**	**Групи господарств за чисельністю працівників, чол.**	**в % до підсумку**
до 140	4	до 160	8
140-160	12	160-190	16
160-180	18	190-220	30
180-200	27	220-250	21
200-220	24	250-280	15
220-240	6	280-310	6
240-260	4	понад 310	4
260-280	2	–	–
понад 280	3	–	–
Разом	**100**	**Разом**	**100**

Безпосередньо дані групувань двох районів непорівнянні, оскільки господарства розподілені по групах з різними інтервалами: 20 чол. у районі I і 30 чол. у районі II. Число виділених груп також неоднакове.

Для приведення двох групувань в порівнянний вид проведемо вторинне групування. З цією метою перегрупуємо матеріали в групи, єдині для обох районів: візьмемо інтервал 40 чол. (табл. 3.8).

Таблиця 3.8

Вторинне групування господарств двох районів за чисельністю працівників тваринництва

Групи господарств за чисельністю працівників, чол.	В % до підсумку	
	район I	**район II**
до 160	16 (4 + 12)	8
160-200	45 (18 + 27)	26 (16 + $\frac{1}{3}\cdot 30$)
200-240	30 (24 + 6)	34 ($\frac{2}{3}\cdot 30 + \frac{2}{3}\cdot 21$)
240-280	6 (2 + 4)	22 ($\frac{1}{3}\cdot 21$ + 15)
понад 280	3	10 (6 + 4)
Разом	**100**	**100**

Оскільки є можливість вторинне групування господарств району I здійснити способом простого укрупнення інтервалів (має місце збіг нижніх і верхніх інтервалів у двох групуваннях), використаємо цей спосіб для вирішення поставленого завдання.

Пояснимо послідовність розрахунків. У першу групу господарств з чисельністю працівників до 160 чол. увійдуть господарства I і II груп.

Питома вага господарств цих груп у загальному підсумку становитиме 16% (4 + 12). У другу групу господарств з чисельністю працівників від 160 до 200 чол. увійдуть господарства III і IV груп. Їхня питома вага в загальному підсумку становитиме 45% (18 + 27). Аналогічно виконуються розрахунки при утворенні решти груп.

Перегрупуємо господарства району II. Оскільки укрупнення інтервалів для господарств району II не підходить і завдання не вирішує використаємо спосіб часткового перегрупування даних первинного групування.

У першу, заново створену групу господарств району II з чисельністю працівників тваринництва до 160 чол., повністю увійдуть господарства первинного групування з таким же інтервалом. Питома вага господарств цієї групи становить 8%.

У другу групу господарств вторинного групування з чисельністю працівників від 160 до 200 чол. повністю увійдуть господарства II групи (16%) і частина господарств III групи. Для визначення частини господарств, яку потрібно взяти з III групи, необхідно її розчленити на підгрупи з чисельністю працівників 190-200, 200-210, 210-220 чол. Показники питомої ваги господарств в цих підгрупах визначаються пропорційно діленню величини інтервалу. Величина інтервалу, яку ми розглядаємо, становить 30 чол. і ділиться на три рівні частини. Для одержання потрібного інтервалу 160-200 чол. до величини інтервалу II групи (160-190 чол.) слід додати одну третину величини інтервалу III групи (190-220 чол.) і таку ж частину господарств цієї групи.

Отже, в другу, заново створену групу господарств, увійдуть 16% господарств другої групи і одна третина III групи – 10% (1/3 · 30), що становитиме 26% загальної чисельності господарств району II.

У III групу господарств вторинного групування (200-240 чол.) увійде частина господарств III групи (190-220 чол.), що залишилась, – 20% ($^2/_3$ · 30) і дві третини господарств IV групи (220-250 чол.) – 14% ($^2/_3$ · 21), тобто 34% всієї чисельності господарств району II.

Аналогічні розрахунки виконуються і при утворенні решти заново створених груп господарств: 240-280 і понад 280 чол. Якби в табл. 3.7 поряд з даними про питому вагу господарств по групах були наведені дані про їх чисельність, то розрахунки по заново створених групах виконувались би в тих же співвідношеннях, що і за питомою вагою господарств.

Після вторинного групування первинний матеріал стає порівнянним, оскільки для двох районів узяті однакові групи за чисельністю працівників. З даних табл. 3.8 видно, що розподіл господарств за чисельністю працівників тваринництва в двох районах суттєво відрізняється: в районі І переважають господарства з чисельністю працівників тваринництва до 200 чол. (61% загальної чисельності господарств), в районі II – господарства з чисельністю працівників тваринництва – понад 200 чол. (64% загальної чисельності господарств).

3.5. Ряди розподілу

Особливим видом групувань в статистиці є ряди розподілу, які є найпростішим способом упорядкування і узагальнення статистичних даних. Групування, в якому виділені групи характеризуються тільки їхньою чисельністю або питомою вагою в загальному обсязі сукупності, називають **статистичним рядом розподілу** (наприклад, розподіл господарств району за урожайністю, продуктивністю тварин, робітників за тарифним розрядом та ін.).

Статистичні ряди розподілу являють собою упорядкований розподіл одиниць досліджуваної сукупності на групи за групувальною ознакою. Вони характеризують структуру (склад) досліджуваного явища, дають змогу судити про однорідність сукупності, про варіювання досліджуваної ознаки.

Ряди розподілу складаються з двох елементів: найменування групи з відповідними значеннями досліджуваної ознаки і чисельності одиниць, що увійшли до кожної групи. В цьому їх відміна від статистичних групувань, при побудові яких кожна група характеризується системою пов'язаних і взаємозалежних між собою показників.

Найпростішим видом статистичних рядів розподілу є **ранжирований ряд**, в якому значення досліджуваної ознаки розташовані у порядку зростання або зменшення. Однак ранжирований ряд ще не дає загальної картини розподілу, оскільки не видно, яка закономірність закладена в розподілі, навколо якої величини концентруються варіанти. Тому виникає потреба подальшого узагальнення статистичних даних, об'єднання їх в окремі групи і підрахунку частот для кожної групи. В результаті здійснення цієї операції одержимо **варіаційний ряд розподілу**.

Ряди розподілу, утворені за якісною ознакою називають **атрибутивними**. Прикладом таких рядів можуть бути розподіли населення за статтю (див табл. 3.2), освітою, тварин за породою, рослин за сортом і т.д.

Різновидом атрибутивних рядів розподілу є альтернативні ряди. **Альтернативними** називають ряди якісних ознак, які приймають тільки два значення, що виключають одне одного: так або ні. Прикладом таких рядів може бути розподіл господарств району на прибуткові і збиткові, або на такі, що виконали і не виконали план виробництва продукції тощо.

Ряди розподілу, побудовані за кількісними ознаками, називають **варіаційними**. Варіаційний ряд розподілу являє собою упорядковану статичну сукупність, в якій значення варіант розташовані в ранжирований ряд із зазначенням для кожного інтервалу (групи) відповідних частот (частостей).

Варіаційні ряди розподілу складаються з двох елементів: варіант і частот. **Варіанта** – це окреме значення ознаки, яке вона приймає в ряду розподілу. **Частотами** називають чисельності окремих варіант або кожної групи варіаційного ряду. Частоти можуть бути виражені як в абсолютних величинах, тобто числом будь-яких одиниць, так і у відносних величинах у вигляді часток і процентів до підсумку. Частоти, що виражені в частках одиниці або в процентах до підсумку, називають **частостями**. Суму частот варіаційного ряду називають його **обсягом**. Сума частот дорівнює одиниці, якщо вони виражені в частках одиниці, і 100%, якщо виражені в процентах. У математичній статистиці для визначення деяких характеристик (наприклад, медіани) розраховують **нагромаджені частоти** – сума частот (частостей) варіантів від мінімального значення до даного значення. Нагромаджені частоти визначаються шляхом послідовного додавання до частот (частостей) першої групи частот наступних груп ряду розподілу (див. табл. 3.10).

Варіаційні ряди розподілу підрозділяються на дискретні (перервні) та інтервальні (безперервні).

Дискретні – це такі варіаційні ряди розподілу, в яких варіанти приймають значення тільки цілих чисел. Прикладом дискретного ряду розподілу може бути розподіл господарств району за кількістю бурякозбиральних комбайнів (табл. 3.9).

Таблиця 3.9

Розподіл господарств району за кількістю бурякозбиральних комбайнів

Кількість комбайнів у господарстві, шт.	Кількість господарств	В % до підсумку
2	3	11,5
3	4	15,4
4	5	19,2
5	6	23,1
6	5	19,2
7	3	11,6
Разом	**26**	**100,0**

Інтервальними називають ряди розподілу, в яких варіанти дані у вигляді інтервалів. Приклад інтервального ряду розподілу наведено в табл. 3.10, де наводиться розподіл 100 господарств за надоєм молока на корову.

Таблиця 3.10

Інтервальний варіаційний ряд розподілу 100 господарств за надоєм молока на корову

Номер групи	Межі інтервалів за надоєм молока на корову, ц	Кількість господарств (частота)	Питома вага		Нагромаджені частоти
			в % до підсумку	в частках (частість)	
I	26-28	8	8,0	0,080	8
II	28-30	16	16,0	0,160	24 (8 + 16)
III	30-32	17	17,0	0,170	41 (24 + 17)
IV	32-34	25	25,0	0,250	66 (41 + 25)
V	34-36	18	18,0	0,180	84 (66 + 18)
VI	36-38	11	11,0	0,110	95 (84 + 11)
VII	38-40	5	5,0	0,050	100 (95 + 5)
Разом	**х**	**100**	**100,0**	**1,000**	**–**

У наведеному прикладі варіантами є значення надою на корову, а частотами – чисельність господарств.

При побудові рядів розподілу виникають запитання щодо числа груп, величини інтервалу, його межі. Методологія побудови рядів розподілу грунтується на викладеній вище методології побудови статистичних групувань.

Якщо варіаційний ряд розподілу має групи з нерівними інтервалами, то частоти в окремих інтервалах безпосередньо неспівставні, тому що залежать від ширини інтервалу. Для того щоб частоти можна було порівнювати, обчислюють **щільність розподілу та відносну щільність розподілу.** Перша характеристика визначається відношенням частоти до величини інтервалу, друга – відношенням частості до величини інтервалу.

Для наочності та полегшення аналізу рядів розподілу їх зображують графічно у вигляді: огіви, полігону, гістограми і кумуляти.

Графічне зображення варіаційного ряду розподілу називають **кривою розподілу.**

Статистична сукупність, представлена у вигляді ранжированого ряду, графічно зображується у вигляді огіви. Огіва будується так: на ось абсцис наносять номери елементів сукупності за ранжиром, а на осі ординат відкладаються значення ознаки (варіант). Огіва наочно показує зміну досліджуваної ознаки.

Послідовність побудови огіви покажемо на такому прикладі ранжированого ряду урожайності хмелю у 18 господарствах району (ц/га):

8,0; 8,1; 8,3; 8,5; 8,6; 9,2; 9,4; 9,5; 10,1; 10,6; 11,0; 12,1; 12,5; 13,0; 13,1; 13,5; 14,0.

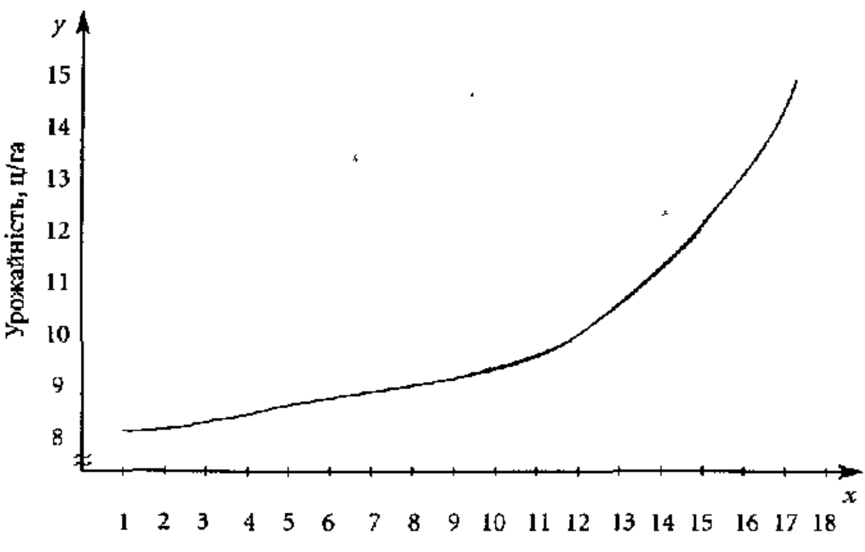


Рис. 3.1. Огіва розподілу господарств за урожайністю хмелю

У вигляді полігону (многокутника розподілу) звичайно зображують дискретні варіаційні ряди розподілу. При цьому на осі абсцис відкладаються значення варіант, а на осі ординат – частота або частість. За даними табл. 3.9 побудуємо полігон розподілу господарств району за чисельністю бурякозбиральних комбайнів (рис. 3.2).

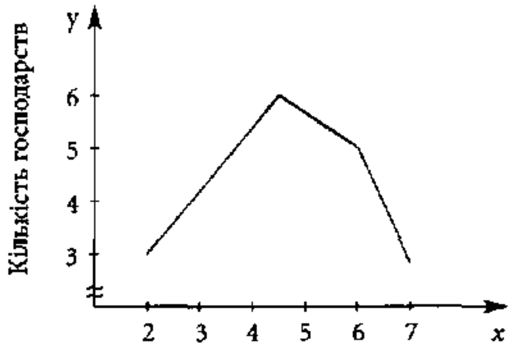


Рис. 3.2. Полігон розподілу господарств району за чисельністю бурякозбиральних комбайнів

Для зображення інтервальних варіаційних рядів розподілу застосовується **гістограма**, яка являє собою фігуру у вигляді прямокутників, що прилягають один до одного. Порядок побудови гістограми такий: на осі абсцис відкладають інтервали варіантів, а на осі ординат – частоти (частості).

Над віссю абсцис будуються прямокутники, площа яких відповідає величинам добутків інтервалів на їх частоти. Ширина стовпчиків при рівних інтервалах однакова, при нерівних – неоднакова. Якщо середини верхніх сторін прямокутників (середини інтервалів) з'єднати, то одержимо полігон розподілу.

За даними табл. 3.10 побудуємо гістограму розподілу 100 господарств за надоєм молока на корову (рис. 3.3).

При зображені інтервальних рядів розподілу з нерівними інтервалами гістограму будують не за частотами (частостями) інтервалів, а за показниками щільності розподілу. При побудові гістограми за абсолютною щільністю розподілу загальна її площа дорівнює чисельності сукупності. При побудові графіка відносної щільності площа гістограми дорівнює одиниці.

При розв'язуванні деяких задач зручніше користуватись нагромадженими частотами. При цьому значення чисельностей окремих варіант замінюється нагромадженими частотами, які одержують підсумовуванням частоти даної варіанти з попередніми частотами.

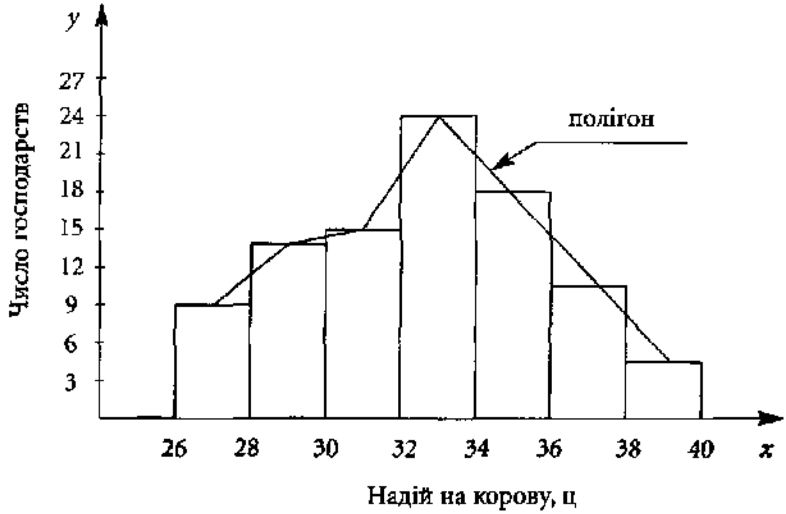


Рис. 3.3. Гістограма розподілу господарств за надоєм молока на корову

Варіаційний ряд з нагромадженими частотами на графіку зображується у вигляді кривої, яка одержала назву **кумуляти розподілу**.

Для побудови кумуляти спочатку підраховують нагромаджені частоти, послідовно підсумовуючи їх (див. графу 6 табл. 3.10). Якщо розподіл має дискретний (перервний) характер, то на графіку на осі абсцис відкладають значення варіант, а на осі ординат – нагромаджені частоти (частості). Якщо розподіл має безперервний характер і поданий у вигляді інтервального ряду розподілу, то будують точки, абсциси яких є праві (верхні) межі інтервалів, а ординати – відповідні їм нагромаджені частоти (частості).

За даними табл. 3.10 побудуємо кумулятивну криву розподілу 100 господарств за надоями молока на корову (рис. 3.4).

Кумулята зв'язана з огівою таким чином: якщо лист паперу, на якому зображена кумулята, повернути на 90° і подивитися на нього з протилежного боку на світло, то можна побачити огіву.

Побудувавши полігон або гістограму, можна одержати перше уявлення про **форму розподілу**, під якою розуміють форму його графіка в границі (в математичному розумінні), тобто форму **кривої розподілу**.

Рис. 3.4. Кумулята розподілу 100 господарств за надоєм молока на корову

Розрізняють передусім одновершинні (одномодальні) і багатовершинні (багатомодальні) розподіли. До **одновершинних** відносять криві, які *мають один максимум* частот. *Одновершинний* розподіл може бути *гостровершинним* і плосковершинним. Для багатовершинних розподілів характерна наявність кількох максимумів і мінімумів частот, що перемежаються між собою. Багатовершинність розподілу, як правило, є ознакою неоднорідності досліджуваної сукупності, вказує на наявність диференціації (розшарування) сукупності, а найчастіше є наслідком змішування якісно відмінних сукупностей. В цьому випадку досліджувану сукупність необхідно розчленувати на окремі однорідні сукупності і *вивчати їх окремо*.

Серед різноманіття форм одновершинних розподілів, що найчастіше зустрічаються на практиці, можна виділити такі характерні розподіли: *симетричні, помірноасиметричні, крайньоасиметричні* (*І-подібні*), *утнуті* (*U*-подібні) та ін.

Симетричним називають такий розподіл, в якому частоти варіант з мірою віддалення від якогось центра розсіяння зменшуються, залишаючись рівними між собою по обидві сторони до кінців розподілу. Криві таких розподілів *симетричні відносно ординати*, встановленої у точці, яка відповідає математичному сподіванню.

Симетричний розподіл може бути гостровершинним і плосковершинним. Для гостровершинних розподілів одиниці сукупності зосереджуються біля центральної варіанти, для плосковершинних – навпаки, розосереджуються.

Криві розподілу, побудовані на основі фактичних даних, звичайно рідко бувають ідеально симетричними, хоча ця форма розподілу притаманна багатьом явищам. Емпіричні розподіли, як правило, є асиметричними (скошеними). Такі помірноасиметричні розподіли на практиці зустрічаються частіше. **Помірноасиметричними (скошеними)** називають такі розподіли, в яких частоти по один бік від центра розсіювання зменшуються помітно бистріше, ніж по другий, внаслідок чого ординати рівновіддалених від центра значень ознаки неоднакові.

При цьому, якщо більш довша гілка кривої припадає на більші значення ознаки, що лежать на правому боці графіка, то таку асиметрію називають **правосторонньою**, або **додатною**. У протилежному випадку асиметрія вважається **лівосторонньою**, або **від'ємною**.

Асиметричний розподіл в границі стає крайньоасиметричним.

Крайньоасиметричними (І-подібними) називають такі розподіли, в яких найбільша частота розташована на одному з кінців розподілу. Такі розподіли за формулою нагадують І й тому називаються **І-подібними**.

Інколи зустрічаються розподіли, які мають криву угнутої форми, що нагадує латинську букву U, такі розподіли називають **U-подібними**. U-подібні розподіли характерні тим, що мінімальна частота знаходиться звичайно поблизу центра розсіювання, а по мірі віддалення від неї до кінців розподілу частоти зростають. Такі розподіли на практиці зустрічаються рідко.

Зустрічаються Z-подібні розподіли.

Усі викладені вище форми статистичних розподілів подано у вигляді такої схеми (рис. 3.5).

Зображення варіаційних рядів розподілу в табличній і графічній формах дає змогу одержати лише перше уявлення про найбільш загальні характерні властивості досліджуваного розподілу.

Всебічна характеристика рядів розподілу передбачає з'ясування умов, у яких сформувався досліджуваний розподіл, вираження його основних особливостей числовими характеристиками.

Комплексний опис статистичних розподілів полягає в знаходженні передусім найважливіших узагальнюючих характеристик: середньої величини, ступеня варіації ознаки, скошеності, гостровершинності розподілу. Для їх визначення використовуються відповідні кількісні характеристики. Про них мова буде йти у наступних розділах підручника (див. розділи 4 і 5).

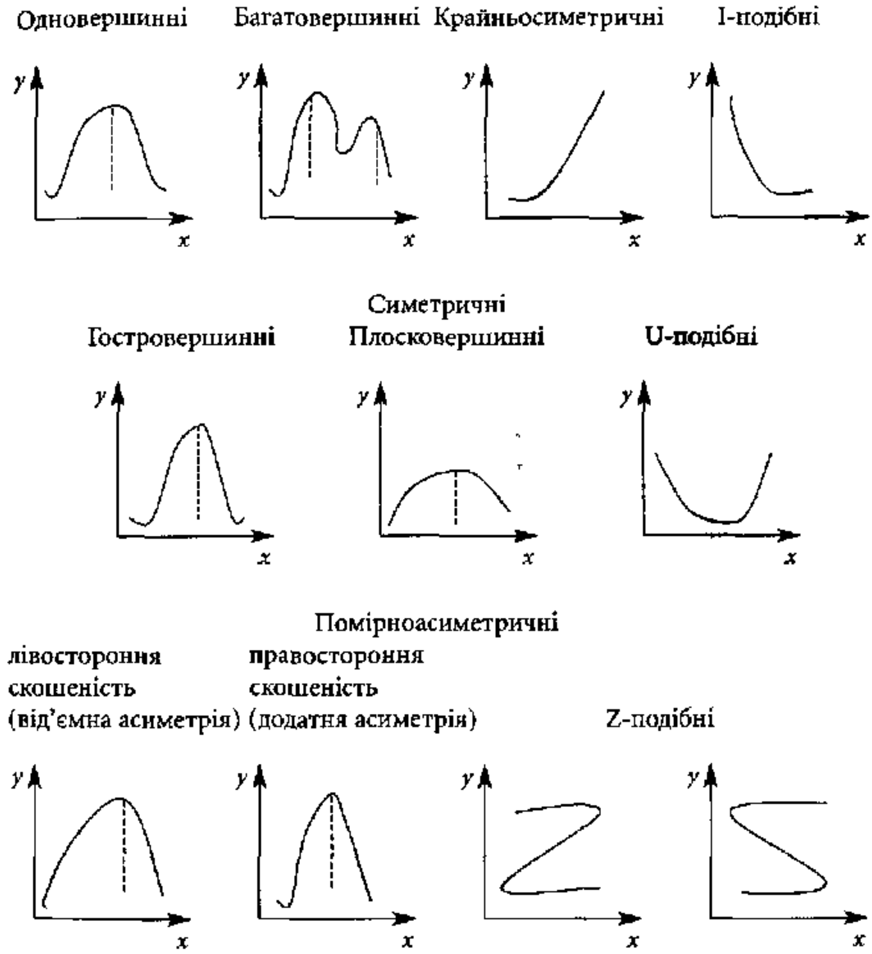


Рис. 3.5. Графіки форм статистичних розподілів

3.6. Статистичні таблиці

Результати статистичного зведення і групування, як правило, оформляються у вигляді статистичних таблиць.

Статистичні таблиці – це форма систематизованого, раціонального і наочного викладення статистичних даних про явища і процеси суспільного життя.

Значення статистичних таблиць полягає в тому, що вони дають змогу охопити матеріали статистичного зведення в цілому. Статистична таблиця по суті є системою думок про досліджуваний об'єкт, що викладається за допомогою цифр – об'єктивних статистичних показників.

Достоїнством статистичних таблиць є виразність, наочність і компактність. Змістом статистичної таблиці є та сукупність відомостей, яка викладена в системі показників. Являючись підсумком статистичного спостереження, зведення і групування і частково аналізу, таблиці мають велике пізнавальне, наукове і практичне значення.

За зовнішнім виглядом статистична таблиця являє собою ряд горизонтальних і вертикальних ліній, які, перетинаючись по горизонталі, утворюють рядки, а по вертикалі – графи (стовпчики, колонки), які в сукупності утворюють мовби макет таблиці.

В утворені всередині таблиці графо-клітини записується відповідна інформація. Кожна клітина має свій певний якісний зміст і кількісну міру, властиву статистичному показнику. Складену таблицю, але не заповнену цифрами, прийнято називати **макетом таблиці**.

Макет таблиці *містить загальний заголовок,* який коротко і точно характеризує зміст таблиці, боковий і верхній заголовки як сукупність рядків і граф. В цих найменуваннях виражається зміст і форма статистичних показників. Розробка макета статистичної таблиці є найважливішим етапом побудови статистичної таблиці, що багато в чому визначає її якість.

По суті статистична таблиця являє собою статистичне речення, яке має підмет і присудок. **Підметом таблиці** є одиниці статистичної сукупності або їх групи, які підлягають характеристиці і вивченню. **Присудком таблиці** – цифрові дані, що характеризують підмет. Наприклад, в табл. 3.3, 3.4 і 3.5 підметом таблиці є групи господарств, а присудком – показники, які характеризують ці господарства (урожайність, якість ґрунту, собівартість та ін.).

Звичайно складові частини досліджуваного об'єкта, що утворюють підмет, розташовують в лівій частині таблиці, а показники, що складають присудок, розміщують справа. Але буває і обернене розташування підмета і присудка таблиці, що зумовлюється метою дослідження, характером вихідної інформації.

В процесі економічних досліджень застосовуються різні види статистичних таблиць. Вони відрізняються числом одиниць і об'єктів, що характеризуються в них, формою підмета і присудка і т.д.

Залежно від побудови (розробки) підмета розрізняють три види статистичних таблиць: прості, групові та комбінаційні.

Простими називають такі статистичні таблиці, в підметі яких міститься простий перелік будь-яких об'єктів, територіальних підрозділів або хронологічних дат. Відповідно таблиці можуть бути названі простими переліковуваними, територіальними або хронологічними (див. табл. 3.1, 3.9 і 10.1).

Груповими називають статистичні таблиці, в яких статистичний підмет складається з груп, виділених за будь-якою однією суттєвою ознакою, а присудок містить ряд ознак, які характеризують зазначені групи (див. табл. 3.3, 3.4, 3.7, 3.8).

Комбінаційними називають статистичні таблиці, в яких підмет являє собою комбінацію, сполучення двох або кількох ознак, а в присудку наводяться ознаки, що характеризують виділені групи і підгрупи. Комбінаційні *таблиці отримують внаслідок комбінаційних групувань (див. табл. 3.5, 3.6).* Комбінаційні таблиці мають дуже велике аналітичне значення. Вони дають змогу за допомогою комбінування різних групувальних ознак найбільш правильно охарактеризувати вплив окремих факторів на результативні показники. Вирівнюючи сукупність в певних межах за однією ознакою і диференціюючи за другою і т. д., комбінаційні таблиці дають змогу не тільки встановити наявність зв'язку, але й виміряти ступінь цього зв'язку.

Залежно від завдання дослідження і характеру інформації присудок статистичних таблиць буває простим і складним. Показники присудка при простій розробці застосовуються послідовно один за другим. Розподіляючи показники на групи за однією або кількома ознаками в певному сполученні, одержують складний присудок.

При розробці і заповненні макетів таблиць необхідно строго дотримуватися правил їх побудови.

1. Статистичні таблиці не повинні бути надмірно громіздкими і ускладненими, вони повинні полегшувати, а не утруднювати їх аналіз. У зв'язку з цим по можливості таблицю слід складати невеликою за розміром, легкодоступною для огляду. Іноді доцільно замість однієї великої таблиці побудувати декілька зв'язаних між собою, послідовно розташованих таблиць.

2. Всі таблиці можуть бути пронумеровані арабськими цифрами. Номер таблиці *вказують перед її заголовком. При цьому знак «№» не пишуть.*

3. Кожна таблиця повинна мати загальний заголовок, в якому коротко і ясно відображається основний зміст таблиці, вказано до якої територі і до якого періоду або моменту часу відносяться дані, що наведені в ній. Вимога точності, чіткості та ясності відноситься і до заголовків рядків і граф.

4. Показники таблиці обов'язково повинні супроводжуватись одиницями вимірювання. Якщо для всіх показників використовується одна одиниця вимірювання, то її пишуть в кінці заголовка таблиці, а якщо їх кілька – в кінці рядків або граф. Одиниця вимірювання відокремлюється від назви показника комою.

5. Слова в таблиці пишуться повністю. Можна використовувати тільки загальноприйняті скорочення.

6. Таблиці, як правило, мають бути замкненими, тобто мати підсумкові результати (в цілому, по групах і підгрупах).

7. При заповненні таблиць потрібно використовувати такі умовні позначення: при відсутності явища пишеться прочерк (–), якщо ж немає інформації про явище, ставиться три крапки (…) або пишеться «немає відомостей», в тих випадках, коли клітинка не підлягає заповненню в зв'язку з відсутністю осмисленого змісту – ставиться знак множення (Ч). При наявності інформації по досліджуваному явищу, числові значення якого складають величину менше критичної в таблиці точності, прийнято записувати 0,0.

8. Однаковий ступінь точності, обов'язковий для всіх чисел, забезпечується дотриманням правил їх заокруглення. Всі значення однойменних показників мають бути записані з однаковим ступенем точності (до цілих; до 0,1; до 0,01 і т.д.).

9. Коли одна величина перевищує другу багатократно, то отримані показники динаміки краще виражати не в процентах, а в разах (коефіцієнтах). Наприклад, замість 288% слід написати «в 2,9 рази більше». В аналітичних таблицях значність абсолютних цифр має бути найменшою. Тому великі числа необхідно заокруглювати до тисяч, мільйонів і т.д. Наприклад, замість числа 1 200 000 грн. краще написати 1,2 млн грн.

10. Якщо в таблиці поряд із звітними даними наводяться відомості розрахункового порядку, то про це слід зробити відповідне застереження. По можливості ці пояснення краще зробити в самій таблиці або в заголовку до неї.

11. Якщо є потреба, до таблиці можуть бути застосовані і виноски. Примітки даються у вигляді необхідності додаткових пояснень змісту окремих показників таблиці. У виносках звичайно вказують джерела одержаних у таблиці відомостей.

Аналіз і читання даних статистичних таблиць має велике пізнавальне і практичне значення. Перед тим як приступити до аналізу даних таблиць, слід ознайомитися з її назвою, заголовками граф і рядків і з'ясувати їх суть; цифрові дані необхідно починати читати з підсумків і тільки після цього переходити до аналізу даних окремих рядків і граф, тобто до оцінки частин досліджуваного об'єкта, вивчаючи при цьому спочатку важливі, а потім вже і решту елементів таблиці. Аналіз таблиць полягає в аналітичному осмисленні і тлумаченні табличних даних і спрямований на виявлення взаємозв'язків і взаємозалежностей між ознаками.

3.7. Абсолютні показники

У процесі статистичного спостереження отримують дані про значення тих чи інших ознак, що характеризують кожну одиницю досліджуваної сукупності. Для характеристики сукупності в цілому або окремих її частин дані по окремих одиницях сукупності піддають зведенню. Шляхом безпосереднього підсумовування первинних даних отримують узагальнюючі абсолютні показники, які характеризують чисельність сукупності і обсяг (розмір) досліджуваного явища в конкретних межах часу і місця.

Абсолютні показники мають велике пізнавальне і практичне значення. Знання рівнів, розмірів і обсягів абсолютних статистичних показників необхідне для планування, управління і аналізу господарської діяльності народного господарства, його галузей і підприємств. В абсолютних показниках встановлюється більшість планових завдань по розвитку народного господарства, задоволення потреб суспільства в різноманітних продуктах і послугах, здійснюється контроль за їх виконанням.

За допомогою абсолютних показників характеризують обсяг виробленого у країні валового внутрішнього продукту, валового національного доходу, вартість основних фондів, чисельність працівників, фонд заробітної плати підприємства, виробництво продукції в господарстві та інші соціально-економічні явища.

Абсолютні показники – це величини, які виражають розміри суспільних явищ як таких без відношення їх до інших явищ. Наприклад, на 5 грудня 2001 р. чисельність населення України становила 48,4 млн чол., а фермерських господарств – на 1.01.2002 р. – 42350.

Абсолютні показники виражають розміри суспільних явищ в певних межах часу і території. Вони є іменованими числами, завжди мають певну розмірність і одиниці вимірювання.

Залежно від характеру явища і завдань дослідження абсолютні показники *виражаються в натуральних, вартісних, трудових і умовно-натуральних* одиницях вимірювання.

Абсолютні показники можуть виражати розміри, обсяги і рівні суспільних явищ на певний момент (на 1.01.2003 р. поголів'я корів в господарстві становило 770 гол.) і за певний період часу (виробництво молока в господарстві за 2002 р. становило 21 600 ц).

За способом вираження розмірів досліджуваних явищ абсолютні показники підрозділяються на індивідуальні, групові та загальні.

Індивідуальними *називають такі абсолютні показники, які виражають* розміри кількісних ознак у окремих одиниць сукупності. Наприклад, чисельність працівників підприємства, виробництво валової продукції в агрофірмі, прибуток підприємства та ін.

Групові абсолютні показники виражають розміри ознаки або чисельність одиниць у окремих частин (груп) сукупності. Їх отримують при обробці матеріалів статистичного спостереження шляхом підсумовування абсолютних розмірів ознаки у окремих одиниць сукупності або підрахунку числа одиниць сукупності, що входять в окремі групи.

Загальними називають абсолютні показники, які виражають розміри ознаки у всіх одиниць сукупності. Вони є результатом зведення даних статистичного спостереження. Наприклад, фонд заробітної плати господарств району, вартість основних виробничих фондів в СТОВ області, валовий збір картоплі в країні та ін.

3.8. Поняття про відносні величини, їх види

Абсолютні показники відіграють важливу роль в системі статистичних *показників. Разом з тим при вивченні соціально-економічних явищ статистика* не може обмежуватись обчисленням тільки абсолютних показників, тому що вони часто не дають достатньо повного уявлення про досліджуване явище. Так, наприклад, при зіставленні абсолютного показника виробництва яловичини в господарстві, припустимо 3600 ц з плановим рівнем, рівнем минулого року або встановленим проектом, стають добре видними успіхи і недоліки в роботі господарства. Якщо приріст виробництва порівняно з минулим роком становив + 12%, а план виконано на 97% і по відношенню до проектної потужності становить 92%, то стає ясним, що господарство має достатні резерви для збільшення виробництва яловичини. Тому в статистичному аналізі поряд з абсолютними величинами виникає потреба розрахунку похідних узагальнюючих показників – середніх і відносних показників. Середні величини докладно розглядаються в розділі 4. Тут же зупинимося на характеристиці відносних показників.

Відносними називають показники, які виражають кількісні співвідношення між соціально-економічними явищами. Їх одержують в результаті ділення двох абсолютних або середніх величин. Так, урожайність зернових культур (відносний показник) одержують в результаті зіставлення двох абсолютних показників валового збору і посівної площі: 48 000 ц : 1200 га = = 40,0 ц/га.

При цьому ту величину, з якою порівнюють, називають **основою**, або **базою порівняння**, а порівнювану величину – **поточною**, або **звітною**.

При обчисленні відносних величин слід мати на увазі, що в чисельнику завжди знаходиться показник, що відображає те явище, яке вивчається, тобто порівнюваний показник, а в знаменнику – показник, з яким порівнюють, що приймається за основу, або базу порівняння.

Відносні показники мають велике аналітичне значення. Вони обчислюються для одержання характеристики різноманітних сторін суспільного життя. За їх допомогою виражають ступінь виконання народногосподарських планів, ефективність та інтенсивність суспільного виробництва, продуктивність праці, ступінь задоволення матеріальних і культурних потреб людей, структуру і динаміку виробництва та ін.

За допомогою відносних показників можуть порівнюватися однойменні та різнойменні величини.

Одним з найважливіших достоїнств відносних показників є те, що вони дають змогу порівнювати такі явища, абсолютні розміри яких безпосередньо неспівставні. Наприклад, виробництво валової продукції сільського господарства на 100 га земельних угідь, щільність населення, виробництво окремих видів продуктів харчування на душу населення та ін.

Залежно від бази порівняння відносні показники можуть бути виражені різними формами: коефіцієнтами (частками), процентами (%), проміле (‰), продецеміле (‱).

Якщо база порівняння приймається за одиницю (прирівнюється до одиниці), то відносна величина (результат порівняння) називається **коефіцієнтом** (часткою) і показує в скільки разів досліджувана величина більше основи. Якщо значення основи, або базу порівняння, прийняти за 100%, результат обчислення відносної величини буде виражений в процентах.

Щоб уникнути важкосприйнятних дрібних відносних величин, базисна величина приймається іноді за 1000 або 10 000 одиниць. В тих випадках, коли базу порівняння приймають за 1000 (наприклад, при обчисленні демографічних коефіцієнтів), результат порівняння виражається в проміле, а коли за 10 000 – продецеміле. Використовуються при порівняннях явищ, які рідко зустрічаються, щоб придати відносним величинам зручний для сприйняття вид. Наприклад, замість числа тракторів на 100 га ріллі 1,87 застосовують проміле 18,7 шт. на 1000 га.

В тих випадках, коли величина, що порівнюється, більше основи, відносний показник може бути виражений або коефіцієнтом, або в процентах. Коли порівнюваний показник менше основи, відносний показник краще виражати в процентах, якщо порівняно малі за числовим значенням величини зіставляються з великими, відносні показники виражаються в проміле або продецеміле. Так, в цих формах вираження розраховуються коефіцієнти народжуваності, смертності, природного і механічного приросту населення, шлюбності, розлучень, число осіб з вищою освітою і число лікарняних ліжок на 10 000 чол. населення та ін.

Залежно від змісту і пізнавального значення розрізняють такі основні види відносних показників: структури, планового завдання, виконання плану, динаміки, інтенсивності, координації, диференціації, порівняння та ін.

Відносні показники структури являють собою відношення частини до цілого або питому вагу частини одиниць в загальному обсязі сукупності. Вони характеризують структуру і склад досліджуваної сукупності, що дає змогу виділити в складному явищі головні ланки, елементи і зосередити на них увагу при подальшому аналізі. Їх одержують в результаті ділення значення кожної частини сукупності на їх загальний підсумок. Ці показники виражаються в частках одиниці (коефіцієнтах) або процентах. Показники структури за будь-якою ознакою, що в сумі дають 100%, **складають структурний ряд**. Прикладом відносних показників структури можуть бути склад населення України за статтю (див. табл. 3.2), питома вага корів в загальній чисельності великої рогатої худоби, структура посівних площ, собівартості, затрат праці, продукції та ін.

Відносний показник планового завдання являє собою відношення величини показника, який встановлюється на плановий період, до його величини, досягнутої за попередній період або будь-який інший період, що приймається за базу порівняння.

Відносний показник виконання плану являє собою відношення фактично досягнутого рівня до планового завдання.

Відносні показники динаміки характеризують зміну суспільних явищ у часі. Вони визначаються як відношення досліджуваного рівня до рівня, прийнятого за базу порівняння (до попереднього року або до постійної бази порівняння). Відносні показники динаміки виражаються у вигляді коефіцієнтів (темпів) зростання, абсолютних і відносних приростів. Більш докладно цей вид відносних показників розглядається в розділі 10, який присвячений статистичній обробці і аналізу динамічних рядів.

Відносні показники динаміки, планового завдання і виконання плану пов'язані між собою такою рівністю: відносний показник динаміки дорівнює добутку відносних показників планового завдання і виконання плану. Розглянемо цей взаємозв'язок на такому прикладі. По господарству є дані щодо середньодобового приросту молодняку великої рогатої худоби (г): за базисний рік (y_0) – 420, за планом ($y_{пл}$) – 450 і фактично (y_1) – 465.

Відносні показники інтенсивності характеризують відношення різнойменних, але пов'язаних між собою певною залежністю величин.

Відносний показник динаміки = *Відносний показник планового завдання* · *Відносний показник виконаного плану*

$$\frac{y_1}{y_0} = \frac{y_{пл}}{y_0} \cdot \frac{y_1}{y_{пл}}$$

$$\frac{465}{420} = \frac{450}{420} \cdot \frac{465}{450}$$

$$1{,}1071 = 1{,}0714 \cdot 1{,}0333.$$

Розраховуються відносні показники інтенсивності діленням абсолютної величини досліджуваного явища на абсолютну величину, яка характеризує обсяг середовища, в якому здійснюється розвиток і розповсюдження явища. Відносна величина показує, скільки одиниць однієї сукупності припадає на одиницю другої сукупності. Прикладом відносних величин інтенсивності можуть бути щільність населення на 1 км^2, вихід валової продукції на 100 га ріллі, одну гривню основних виробничих фондів, середньорічного працівника, поголів'я корів на 100 га сільськогосподарських угідь та ін. Показники такого роду часто називають якісними, оскільки вони відображають найважливіші якісні сторони виробництва: рівень інтенсифікації, озброєність праці, продуктивність землі і тварин, окупність витрат і т.п.

Відносні показники координації характеризують співвідношення різних структурних одиниць однієї і тієї ж сукупності (наприклад, співвідношення між чисельністю міського і сільського населення, чоловіків і жінок, робітників і службовців, основними і оборотними фондами, силовими і робочими машинами і т.д.). Відносні показники координації найчастіше виражаються числом одиниць однієї частини на 100 або 1000 одиниць другої частини.

Відносні показники диференціації одержують в результаті зіставлення двох структурних рядів, один з яких характеризує співвідношення частин сукупності за чисельністю одиниць, а другий – за величиною будь-якої ознаки. Наприклад, порівняння питомої ваги господарства за чисельністю і питомої ваги в цих господарствах валової продукції, землі, працівників і т.д.

Відносні величини порівняння отримують внаслідок порівняння однойменних показників, що стосуються різних об'єктів, взятих за той самий період чи момент часу (наприклад, порівняння урожайності хмелю в двох господарствах за звітний період, продуктивності свиней в господарствах двох областей за п'ять років тощо).

Однією з важливих умов правильного обчислення, порівняння і аналізу відносних показників є забезпечення порівнянності даних. Це означає, що взяті для розрахунків, порівнянь та аналізу абсолютні і відносні показники повинні: 1) відноситись до одного й того самого кола об'єктів і одиниць спостереження або тієї самої сукупності; 2) визначатися за єдиною методикою, що забезпечує порівняння їх за змістом; 3) відноситись до однієї території; 4) характеризувати дані за той самий період або момент часу; 5) мати однакові одиниці вимірювання.

Відносні показники можуть бути простими і складеними. При статистичному аналізі складені відносні показники, які являють собою рівнодійну кількох простих показників, доцільно розкласти на ряд простих відносних показників, що мають самостійне значення.

Таке розкладання дає можливість вивчити залежність складеного відносного показника від його факторів. Сам взаємозв'язок при цьому має вигляд певного рівняння. Найчастіше прийом розкладання складених показників застосовують при вивченні виходу продукції на одиницю ресурсів виробництва (землі, основних фондів, робочої сили), виробітку на машину і працівника, витрат на одиницю площі або голову тварин. Схеми розкладання показників можуть змінюватися залежно від характеру інформації та завдань аналізу. Так, наприклад, виробництво зерна на 1 га сільськогосподарських угідь – складений відносний показник – можна подати як добуток таких простих показників:

Виробництво зерна на 1 га сільськогосподарських угідь = *Урожайність, ц/га* · *Частка посіву зернових в загальній посівній площі* · *Коефіцієнт використання ріллі під посів* · *Коефіцієнт розораності угідь*

$$\frac{\text{Валовий збір зерна}}{\text{Площа сільськогосподарських угідь}} = \frac{\text{Валовий збір зерна}}{\text{Площа посіву зернових}} \cdot \frac{\text{Площа посіву зернових}}{\text{Площа посіву всіх культур}} \cdot \frac{\text{Площа посіву всіх культур}}{\text{Площа ріллі}} \cdot \frac{\text{Площа ріллі}}{\text{Площа сільськогосподарських угідь}}$$

Співвідношення цих показників розглянемо на такому прикладі (табл. 3.11).

Таблиця 3.11

Дані для аналізу виробництва зерна на 1 га сільськогосподарських угідь в господарстві

Показники	Базисний рік	Звітний рік
Вихідні дані		
Валовий збір зерна, ц	74375	88926
Площа посіву зернових, га	1750	1826
Посівна площа всіх культур, га	3500	3518
Площа ріллі, га	3646	3625
Площа сільськогосподарських угідь, га	4558	4540
Розрахункові дані		
Виробництво зерна на 1 га сільськогосподарських угідь, ц	16,3	19,6
Урожайність, ц/га	42,5	48,7
Частка посіву зернових в загальній посівній площі	0,50	0,52
Коефіцієнт використання ріллі під посів	0,96	0,97
Коефіцієнт розораності сільськогосподарських угідь	0,80	0,80

Перевіримо взаємозв'язок між обчисленими показниками і зробимо висновки:

базисний рік $16{,}3 = 42{,}5 \cdot 0{,}50 \cdot 0{,}96 \cdot 0{,}80$;

звітний рік $19{,}6 = 48{,}7 \cdot 0{,}52 \cdot 0{,}97 \cdot 0{,}80$;

звітний рік до базисного $1{,}2024 = 1{,}1459 \cdot 1{,}0400 \cdot 1{,}0104 \cdot 1{,}0000$.

У звітному році на 1 га сільськогосподарських утідь вироблено зерна на 3,3 ц більше, ніж у базисному, в основному внаслідок зростання врожайності зернових культур з 42,5 до 48,7 ц/га, тобто на 14,59%. Частково виробництво зерна збільшилось внаслідок підвищення питомої ваги посівів зернових культур на загальній посівній площі (на 4%) і більш інтенсивного використання ріллі під посів (питома вага посівів в ріллі підвищилась на 1,04%).

Завдання для самоконтролю до розділу 3

1 Що таке статистичне зведення, його завдання, суть і види? Коротко їх охарактеризуйте.

2. Що таке централізоване і децентралізоване зведення?

3. Яка роль і завдання статистичних групувань?

4. Що називають статистичним групуванням? Які є види групувань? Коротко їх охарактеризуйте і наведіть приклади.

5. В чому полягає перевага комбінаційних групувань порівняно з простими групуваннями?

6. Які можливості результативного групування?

7. Які можливості факторного групування?

8. Що таке групувальна ознака? Назвіть їх види та наведіть приклади.

9. До яких групувальних ознак – атрибутивних чи кількісних – відносяться: а) вік людини; б) професія; в) форма власності; г) заробітна плата?

10. Яке з нижче наведених групувань є типологічним:

а) групування населення за національністю;

б) групування підприємств за рівнем рентабельності;

в) групування підприємств за формою власності.

11. Викладіть основні положення методології статистичних групувань.

12. Як визначають кількість груп і межі інтервалів між ними? Які види інтервалів Ви знаєте?

13. *Що називають вторинним групуванням? Які є способи побудови* вторинного групування?

14. Що таке ряди розподілу, їх види і за якими ознаками вони можуть утворюватися?

15. *Які є форми розподілу?* В яких випадках ряд розподілу зображують за допомогою огіви, полігону, гістограми і кумуляти?

16. Що називають статистичною таблицею? Які завдання ставляться перед статистичними таблицями?

17. Що таке макет статистичної таблиці? Назвіть його складові елементи.

18. Які види статистичних таблиць Ви знаєте? Наведіть приклади.

19. За якими правилами будують статистичні таблиці?

20. Що таке абсолютні статистичні величини, яка їх роль і значення в статистиці? Назвіть їх види.

21. Що називають відносними величинами, їх роль і значення в статистиці? Назвіть форми, якими їх можна виражати.

22. Назвіть види відносних величин, охарактеризуйте їх та наведіть приклади.

23. Дайте характеристику складним відносним показникам, їх ролі і призначенню в статистичному аналізі.

Розділ 4

Середні величини

4.1. Поняття про середні величини

Статистична сукупність складається з множини одиниць, об'єктів або явищ однорідних в деякому відношенні і одночасно відмінних за величиною ознак. Величина ознаки кожного об'єкта визначається як загальними для всіх одиниць сукупності, так і індивідуальними її особливостями.

Аналізуючи впорядковані ряди розподілу (ранжировані, інтервальні та ін.), можна помітити, що елементи статистичної сукупності явно концентруються навколо деяких центральних значень. Така концентрація окремих значень ознаки навколо деяких центральних значень, як плавило, має місце у всіх статистичних розподілах. Тенденцію окремих значень досліджуваної ознаки групуватися навколо центра розподілу частот називають **центральною тенденцією**. Для характеристики центральної тендеції розподілу застосовуються узагальнюючі показники, які отримали назву середніх величин.

Середньою величиною у статистиці називають узагальнюючий показник, який характеризує типовий розмір ознаки в якісно однорідній сукупності. Обчислюється середня величина у більшості випадків шляхом ділення загального обсягу ознаки на число одиниць, що володіють цією ознакою. Якщо, наприклад, відомий фонд місячної заробітної плати і кількість робітників за місяць, то середню місячну заробітну плату можна визначити шляхом ділення фонду заробітної плати на кількість робітників.

Середні величини, які обчислюються як з абсолютних, так і з відносних величин, є показниками іменованими і виражаються в тих самих одиницях вимірювання, що і усереднювана ознака. Вони характеризують одним числом значення досліджуваної сукупності. В середніх величинах знаходить відображення об'єктивний і типовий рівень соціально-економічних явищ і процесів.

У статистичній науці і практиці середні величини мають виключно велике значення. Метод середніх величин є одним з найважливіших статистичних методів, а середня величина – однією з основних категорій статистичної науки. Теорія середніх величин займає одне з центральних місць в теорії статистики.Середні величини є основою для розрахунку показників варіації (розділ 5), помилок вибірки (розділ 6), дисперсійного (розділ 8) і кореляційного аналізу (розділ 9).

В усіх випадках, коли виникає потреба охарактеризувати одним числом сукупність значень ознаки, що змінюється, користуються його середнім значенням.

В статистичній сукупності значення ознаки змінюється від об'єкта до об'єкта, тобто варіює. Усереднюючи ці значення і надаючи урівняне значення ознаки кожному члену сукупності, ми абстрагуємось від індивідуальних значень ознаки, тим самим мовби замінюємо ряд розподілу значень ознаки одним і тим самим значенням, рівним середній величині. Однак така абстракція правомірна лише в тому випадку, якщо усереднення не змінює основної властивості по відношенню до даної ознаки в цілому. Ця основна властивість статистичної сукупності, пов'язана з окремими значеннями ознаки, і яка при усередненні має бути збережена незміннною, називається **визначальною властивістю середньої** по відношенню до досліджуваної ознаки. Інакше кажучи, середня, замінюючи індивідуальні значення ознаки, *не повинна змінювати загального обсягу явища*, тобто обов'язкова така рівність: обсяг явища дорівнює добутку середньої величини на чисельність сукупності. Наприклад, якщо з трьох значень урожайності ячміню (x_i = 20,0; 24,6; 23,6 ц/га) обчислена середня (20,0 + 24,6 + 22,3) : 3 = 22,3 ц/га, то за визначальною властивістю середньої має бути дотримана така рівність:

$$n\bar{x} = \Sigma x_i, \text{ тобто } 3 \cdot 22{,}3 = 20{,}0 + 24{,}6 + 22{,}3;\ 66{,}9 = 66{,}9.$$

Головне значення середніх величин полягає в їх узагальнюючій функції, тобто заміні множини різних індивідуальних значень ознаки середньою величиною, яка характеризує всю сукупність явищ. Властивість середньої характеризувати не окремі одиниці, а виразити рівень ознаки з розрахунку на кожну одиницю сукупності є її відмінною спроможністю. Ця особливість робить середню узагальнюючим показником рівня варіюючої ознаки, тобто показником, який абстрагується від індивідуальних значень розміру ознаки у окремих одиниць сукупності. Але те, що середня є абстрактною, не позбавляє її наукового дослідження. Абстракція є необхідним ступенем усякого наукового дослідження. В середній величині, як у будь-якій абстракції, здійснюється діалектична єдність індивідуального і загального. Взаємозв'язок середніх і окремих значень усереднюваної ознаки служить вираженням діалектичного зв'язку індивідуального і загального.

Застосування середніх має базуватися на розумінні та взаємозв'язку діалектичних категорій загального і індивідуального, масового і одиничного.

Середня величина відображає те загальне, що складається в кожному окремому, одиничному об'єкті. Завдяки цьому середня отримує велике значення для виявлення закономірностей, притаманних масовим суспільним явищам і не помітних в одиничних явищах.

У розвитку явищ необхідність поєднується з випадковістю. Тому середні величини пов'язані із законом великих чисел. Суть цього зв'язку полягає в тому, що при розрахунку середньої величини випадкові коливання, що мають різну спрямованість, через дію закону великих чисел взаємно урівноважуються, погашаються і у величині середньої чітко відображається основна закономірність, необхідність, вплив загальних умов, характерних для даної сукупності. В середній знаходить відображення типовий, реальний рівень досліджуваних явищ. Оцінка цих рівнів і зміна їх в часі і просторі – одне з головних завдань середніх величин. Так, через середні виявляється, наприклад, закономірність підвищення продуктивності праці, урожайності сільськогосподарських культур, продуктивності тварин. Отже, середні величини являють собою узагальнюючі показники, в яких знаходить своє відображення дія загальних умов, закономірність досліджуваного явища.

За допомогою середніх величин вивчають зміну явищ у часі і просторі, тенденції в їх розвитку, зв'язки і залежності між ознаками, ефективність різних форм організації виробництва, праці і технологій, впровадження науково-технічного прогресу, виявлення нового, прогресивного в розвитку тих чи інших соціально-економічних явищ і процесів.

Середні величини широко застосовуються в статистичному аналізі сільськогосподарського виробництва, оскільки саме в них знаходять своє виявлення закономірності і тенденції розвитку масових суспільних явищ, варіюючих як у часі, так і у просторі. Так, наприклад, закономірність підвищення продуктивності праці в сільському господарстві знаходить своє відображення в зростанні середнього виробництва продукції з розрахунку на одного працівника, зайнятого в сільськогосподарському виробництві, збільшення валових зборів – у зростанні середньої урожайності сільськогосподарських культур тощо.

Середня величина дає узагальнену характеристику досліджуваного явища тільки за однією ознакою, яка відображає одну з найважливіших його сторін. У зв'язку з цим для всебічного аналізу досліджуваного явища необхідно будувати систему середніх величин за рядом взаємопов'язаних і доповнюючих один одного суттєвих ознак.

Для того щоб середня відображала дійсно типове і закономірне в досліджуваних суспільних явищах, при її розрахунку необхідно дотримуватись таких умов.

1. Ознака, за якою обчислюється середня, має бути істотною. В противному випадку буде отримана несуттєва або спотворена середня.

2. Середню потрібно обчислювати тільки за якісно однорідною сукупністю. Тому безпосередньому обчисленню середніх має передувати статистичне групування, яке дає змогу розчленувати досліджувану сукупність на якісно однорідні групи. У зв'язку з цим науковою основою методу середніх величин є метод статистичних групувань.

3. Розрахунок середньої величини має базуватися на охопленні всіх одиниць даного типу або досить великої сукупності об'єктів, щоб випадкові коливання взаємно зрівноважували один одного і проявлялася закономірність, типові і характерні розміри досліджуваної ознаки.

4. Загальною вимогою при розрахунку будь-якого виду середніх величин є обов'язкове збереження незмінним загального обсягу ознаки в сукупності при заміні індивідуальних її значень середнім значенням (так звана визначальна властивість середньої).

4.2. Види середніх величин і способи їх обчислення

Залежно від характеру усереднюваної ознаки і наявної вихідної інформації в статистиці застосовуються різні види середніх величин, серед яких найбільше використовуються такі: середня арифметична, середня гармонічна, середня геометрична і середня квадратична.

Поряд з переліченими видами середніх величин в статистичній практиці знаходять застосування також середня хронологічна, середня ковзна, середня прогресивна, середня багатовимірна і так звані структурні середні: мода, медіана та ін.

Кожну середню можна визначити як просту, коли значення варіант спостерігаються тільки один раз або однакову кількість разів, і як зважену, коли значення варіант повторюється різну кількість разів.

Уведемо такі позначення і поняття середніх:

$\bar{x}$ – середнє значення досліджуваної ознаки;

x – окремі значення усереднюваної ознаки (варіанти);

n – число одиниць досліджуваної сукупності;

f – частота повторень (вага) варіант;

$W = xf$ – обсяг явищ.

Ознаку, за якою знаходять середню, називають **усередненою ознакою.** Величину ознаки кожної одиниці сукупності називають **варіантою,** або **значенням досліджуваної ознаки**. Частоту повторень варіантів у сукупності називають **статистичною вагою.**

Середні величини, що застосовуються в статистиці, належать до загального типу степеневих середніх. Відрізняються вони тільки показником степеня. Математична статистика виводить різні середні з формули степеневої *середньої, яка являє собою корінь k-го степеня з частки від ділення* суми індивідуальних значень ознаки k-го степеня на число індивідуальних значень:

проста зважена

$$\bar{x}=\sqrt[k]{\frac{\sum_{i=1}^{n}x_i^k}{n}} \qquad \bar{x}=\sqrt[k]{\frac{\sum_{i=1}^{n}x_i^k\cdot f_i}{\sum f_i}},$$

де k – показник степеня, який визначає тип середньої. Підставляючи у наведену формулу замість k відповідні значення показника степеня, одержимо такі середні:

	проста	зважена
при $k = 1$ – арифметичну	$\bar{x}=\frac{\sum x}{n}$[*];	$\bar{x}=\frac{\sum xf}{\sum f}$;
при $k = -1$ – гармонічну	$\bar{x}=\frac{n}{\sum\frac{1}{x}}$;	$\bar{x}=\frac{\sum w}{\sum\frac{w}{x}}$;
при $k = 0$ – геометричну	$\bar{x}=\sqrt[k]{x_1\cdot x_2\cdot\ x_n}$;	$\bar{x}=\sqrt[\sum f]{x_1^{f_1}\cdot x_2^{f_2}\cdot\ .\ .\cdot x_n^{f_n}}$;
при $k = 2$ – квадратичну	$\bar{x}=\sqrt{\frac{\sum x^2}{n}}$;	$\bar{x}=\sqrt{\frac{\sum x^2 f}{\sum f}}$.

[*] В подальшому для зручності і спрощення запису суми $\sum_{i=1}^{n}$ замінимо знаком $\sum$.

Вибір того чи іншого виду середньої визначається цілями і завданнями дослідження і наявною інформацією.

Загальною умовою правильного обчислення усіх видів середніх є збереження незмінним загального обсягу варіюючої ознаки при заміні індивідуальних значень ознаками їхньої середньої. Так, середня арифметична застосовується тоді, коли обсяг варіюючої ознаки утворюється як сума окремих варіант; середня гармонічна – коли обсяг варіюючої ознаки утворюється як сума обернених значень окремих варіант; середня геометрична – коли обсяг варіюючої ознаки утворюється як добуток окремих варіант; середня квадратична – коли обсяг варіюючої ознаки утворюється як сума квадратів окремих варіант.

Розглянемо перелічені вище види середніх більш докладно.

Середня арифметична. Середня арифметична – найпоширеніший вид середньої. Середня арифметична проста являє собою частку від ділення суми індивідуальних значень ознаки на їх загальне число. Її обчислюють за формулою

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

Середня арифметична проста застосовується в тих випадках, коли відомі дані про окремі значення ознаки та їх число в сукупності. В статистичній практиці вона застосовується, як правило, для розрахунку середніх рівнів ознак, представлених у вигляді абсолютних показників. Наприклад, якщо є дані про посівну площу овочів у трьох бригадах господарства (га): 47; 65; 38 і необхідно визначити середній розмір посівної площі, то розрахунок середньої величини необхідно здійснювати за формулою середньої арифметичної простої, оскільки значення усереднюваної ознаки зустрічаються однакове число раз (по одному разу):

$$\bar{x} = \frac{\sum x}{n} = \frac{47 + 65 + 38}{3} = \frac{150}{3} = 50 \text{ га}.$$

Отже, середній розмір посівної площі з розрахунку на одну бригаду становить 50 га.

Середня арифметична зважена обчислюється із значень варіюючої ознаки з урахуванням ваг. Її застосовують у тих випадках, коли значення ознаки представлені у вигляді варіаційного ряду розподілу, в якому чисельність одиниць по варіантах не однакова, а також при розрахунку середньої із середніх при різному обсязі сукупності. Зважування в даному випадку здійснюється за частотами, які показують, скільки разів повторюється та або інша варіанта.

Формула середньої арифметичної зваженої має вигляд:

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

Отже, при обчисленні середньої арифметичної зваженої необхідно всі значення варіант помножити на їхню частоту, одержані добутки підсумувати і цю суму розділити на суму частот, тобто загальний обсяг сукупності.

За аналогічною формулою визначається загальна середня ($\bar{x}_{заг}$) з групових середніх ($\bar{x}_{гр}$), якщо чисельність одиниць по групах ($f_{гр}$) неоднакова:

$$\bar{x}_{заг} = \frac{\sum \bar{x}_{гр} f_{гр}}{\sum f_{гр}}.$$

Розглядаючи формулу середньої арифметичної зваженої, можна помітити, що вона не має принципової відміни від простої середньої арифметичної. Тут підсумовування f разів одного і того самого варіанта (x) замінюють множенням його на число повторень (частоту – f).

Порядок розрахунку середньої арифметичної у варіаційному ряду розподілу покажемо на прикладі середнього настригу вовни по групі господарств (табл. 4.1).

Таблиця 4.1

Дані для розрахунку середньої арифметичної зваженої

Господарство	Вихідні дані		Розрахункові дані				
	настриг вовни на вівцю, кг	число овець, гол.	добуток (валовий настриг), кг	показники структури: питома вага, %	показники структури: частість	добуток	
	x	f	xf	d	w	xd	xw
1	2	80	160	8	0,08	16	0,16
2	3	230	690	23	0,23	69	0,69
3	4	400	1600	40	0,40	160	1,60
4	5	220	1100	22	0,22	110	1,10
5	6	70	420	7	0,07	42	0,42
Разом	x	1000	3970	100	1,00	397	3,97

Оскільки значення усереднюваної ознаки (настриг вовни) повторюється неоднакове число разів, то середній настриг вовни визначимо за формулою середньої арифметичної зваженої:

$$x = \frac{\sum xf}{\sum f} = \frac{3970}{1000} = 3{,}97 \text{ кг.}$$

При розрахунку середньої арифметичної зваженої частотами (вагами) можуть бути використані відносні показники структури, виражені в процентах або коефіцієнтах (частках). Методика розрахунку середньої і кінцевий результат при цьому не зміняться.

Якщо частоти виражені в процентах, то формула середньої арифметичної зваженої може бути записана в такому виді:

$$\overline{x} = \frac{\sum x_i d_i}{\sum d_i},$$

де $d_i = \frac{f_i}{\sum f_i} \cdot 100$ – питома вага кожної частини в загальному обсязі всіх частот (в процентах).

Оскільки для всієї сукупності $\sum d_i = 100\%$, то формулу можна записати так:

$$\overline{x} = \frac{1}{100} \cdot \sum x_i d_i = 0{,}01 \sum x_i d_i .$$

Якщо частоти виражені в коефіцієнтах (частках), $\sum w_i = 1$, тоді формула середньої спрощується:

$$\overline{x} = \sum x_i w_i .$$

Порядок і послідовність розрахунку середньої арифметичної для випадків, коли вагами використовуються відносні показники структури, розглянемо на даних того ж самого прикладу (*табл. 4.1*).

Якщо вагами взяті частоти, виражені в процентах, то середній настриг вовни на вівцю становитиме:

$$\overline{x} = \frac{\sum xd}{\sum d} = \frac{397}{100} = 3{,}97 \text{ кг},$$

а якщо частості:

$$\overline{x} = \sum x_i w_i = 0{,}16 + 0{,}69 + 1{,}60 + 1{,}10 + 0{,}42 = 3{,}97 \text{ кг.}$$

Отже, одержані ті ж самі результати, як при розрахунку середньої арифметичної зваженої звичайним способом.

Для інтервальних варіаційних рядів розподілу, в яких значення ознаки дано в межах «від – до», середню арифметичну зважену знаходять в такій послідовності. Спочатку необхідно інтервальний ряд розподілу перетворити в дискретний. Для цього по кожному інтервалу знаходять його середину (центр). Серединне значення інтервалу звичайно визначають як півсуму його нижньої і верхньої меж. Наприклад, для інтервального ряду розподілу господарств за надоєм молока на корову (ц): 26-28, 28-30, 30-32 і т.д. серединами інтервалів будуть (ц): 27 = (26 + 28) : 2; 29 = (28 + 30) : 2; 31 = (30 + 32) : 2 і т.д.

Якщо є інтервали з нечітко вираженими межами, з так званими «відкритими межами» (перший інтервал «до» і останній – «понад»), то для визначення серединного значення потрібно встановити умовні межі цих інтервалів. Звичайно в цих випадках вирішують так: для першого інтервалу беруть величину другого інтервалу, а для останнього – величину передостаннього інтервалу.

Покажемо перехід від інтервалів з відкритими межами до інтервалів із закритими межами на такому прикладі розподілу господарств за середньодобовим прирістом відгодівельного поголів'я свиней (г):

відкриті інтервали	закриті інтервали
до 350	300-350
350-400	350-400
400-450	400-450
450-500	450-500
понад 500	500-550.

Після того як знайдені середини інтервалів, середню арифметичну зважену обчислюють так, як і в дискретному ряду розподілу: значення варіант множать на частоти і одержану суму добутків ділять на суму частот.

Порядок розрахунку середньої арифметичної в інтервальному ряду розподілу розглянемо на прикладі розподілу 100 господарств за надоєм молока на корову (табл. 3.10). Всі розрахунки зведемо в табл. 4.2.

Таблиця 4.2

Дані для розрахунку середньої арифметичної в інтервальному ряду розподілу

Група	Групи господарств за надоєм молока на корову, ц	Число господарств	Середина інтервалу, ц	Добуток
		f	x	xf
I	26-38	8	27	216
II	28-30	16	29	464
III	30-32	17	31	527
IV	32-34	25	33	825
V	34-36	18	35	630
VI	36-38	11	37	407
VII	38-40	5	39	195
Разом	×	100	×	3264

Середній надій на корову знайдемо за середньою арифметичною *зваженою:*

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{3624}{100} = 32{,}64 \text{ ц, або } 3624 \text{ кг.}$$

Середня гармонічна. Середня гармонічна є оберненою до середньої арифметичної, обчисленої з обернених значень усереднюваної ознаки. Залежно від характеру наявного матеріалу її застосовують тоді, коли ваги доводиться не множити, а ділити на варіанти, або, що те ж саме, множити на обернене їх значення. Таким чином, середня гармонічна розраховується, коли відомі дані про обсяг ознаки ($W = xf$) та індивідуальні значення ознаки (x) і невідомі ваги (f). Так як обсяги ознак являють собою добуток значень ознаки (x) на частоту (f), то частоту (f) визначають як $f = W : x$.

Формули середньої гармонічної простої і зваженої мають вигляд:

проста зважена

$$\bar{x} = \frac{n}{\sum \frac{1}{x}}; \qquad \bar{x} = \frac{\sum w}{\sum \frac{w}{x}}.$$

Як видно, середня гармонічна є перетвореною формою середньої арифметичної. Замість гармонічної завжди можна розрахувати середню арифметичну, попередньо визначивши ваги окремих значень ознаки. При обчисленні середньої гармонічної вагами є обсяги ознак.

Середня гармонічна проста застосовується у випадках, коли обсяги явищ по кожній ознаці рівні.

Наприклад, три комбайнера працюють на збиранні зернових культур. Перший комбайнер на збирання 1 га протягом 7-годинної зміни затрачав 35 хв, другий – 31 хв, третій – 33 хв. Потрібно визначити середні затрати праці на збирання 1 га зернових культур.

Розрахунок середніх затрат часу на збирання 1 га зернових культур за формулою середньої арифметичної простої

$$(\bar{x} = \frac{\sum x}{n} = \frac{35 + 31 + 33}{3} = \frac{99}{3} = 33 \text{ хв})$$

був би правильним тоді, коли б усі комбайнери протягом зміни зібрали по 1 га або однакову кількість гектарів зернових культур. Проте протягом зміни окремими комбайнерами була зібрана різна площа зернових культур.

Неправомірність застосування формули середньої арифметичної пояснюється ще й тим, що показник затрат праці на одиницю робіт (збирання 1 га зернових культур) є оберненим до показника продуктивності праці (збирання зернових культур за одиницю часу).

Середній час, потрібний для збирання 1 га зернових культур, визначимо як відношення затрат часу усіма комбайнерами до загальної кількості зібраних гектарів. У нашому прикладі немає відомостей про кількість фактично зібраних гектарів кожним комбайнером. Однак ці величини можна обчислити за таким співвідношенням:

$$\frac{\text{Кількість зібраних гектарів}}{\text{одним комбайнером}} = \frac{\text{весь затрачений час}}{\text{затрати часу на збирання 1 га}},$$

де весь затрачений час для кожного комбайнера становитеме 420 хв (7 год · · 60 хв).

Тоді середні затрати часу на збирання 1 га зернових культур можна визначити за формулою:

$$\bar{x} = \frac{7\cdot 60 + 7\cdot 60 + 7\cdot 60}{\frac{7\cdot 60}{35} + \frac{7\cdot 60}{31} + \frac{7\cdot 60}{33}} = \frac{1260\ \text{хв}}{38{,}28\ \text{га}} = 32{,}9\ \text{хв/га},$$

$$\text{або}\quad \bar{x} = \frac{3\cdot 7\cdot 60}{7\cdot 60(\frac{1}{35} + \frac{1}{31} + \frac{1}{33})} = \frac{3}{0{,}911} = 32{,}9\ \text{хв/га}.$$

Розрахунки можна значно спростити, якщо використати формулу середньої гармонічної простої:

$$\bar{x} = \frac{n}{\sum \frac{1}{x}} = \frac{3}{\frac{1}{35} + \frac{1}{31} + \frac{1}{33}} = \frac{3}{0{,}911} = 32{,}9\ \text{хв/га}.$$

Отже, по цій сукупності комбайнерів на збирання 1 га зернових культур у середньому витрачається 32,9 хв.

Порядок розрахунку середньої гармонічної зваженої розглянемо на такому прикладі (табл. 4.3).

Таблиця 4.3

Дані для розрахунку середньої гармонічної зваженої

Господарство	Вихідні дані		Розрахункові дані
	урожайність картоплі, ц/га	валовий збір, ц	посівна площа, га
	x	w	$\frac{w}{x}$
1	130	11700	90
2	165	13200	80
3	142	7100	50
Разом	×	**32000**	**220**

Оскільки середня урожайність являє собою відношення валового збору до площі посіву, то спочатку визначимо площу посіву картоплі по кожному господарству, а потім середню урожайність:

$$\bar{x} = \frac{\sum W}{\sum \frac{W}{x}} = \frac{32000}{220} = 145{,}4 \text{ ц/га}.$$

Відповідно до однієї з властивостей середня гармонічна не зміниться, якщо обсяги явищ, які є вагами окремих варіант, помножити або розділити на яке-небудь довільне число. Це дає змогу при її обчисленні користуватися не абсолютними показниками, а їх питомими вагами. Припустимо, потрібно визначити середню ціну реалізації картоплі за такими даними (табл. 4.4).

Таблиця 4.4

Дані для розрахунку середньої ціни реалізації картоплі

Сорт картоплі	Ціна реалізації 1 ц, грн.	Питома вага в загальній сумі виручки, %	Відношення питомої ваги до цін реалізації
	x	d	$\frac{d}{x}$
Ранній	80	16	0,200
Середньостиглий	54	14	0,259
Пізній	32	70	2,188
Разом	**x**	**100**	**2,647**

У наведеному прикладі відсутні дані про виручку від реалізації окремих сортів картоплі, яка являє собою добуток ціни реалізації 1 ц на кількість реалізованої картоплі. Тому замість обсягів явищ можна використати їх співвідношення, тобто питому вагу окремих сортів картоплі в загальній виручці. Використовуючи дані таблиці, визначимо середню ціну реалізації картоплі:

$$\bar{x} = \frac{\sum d}{\sum \frac{d}{x}} = \frac{100}{2{,}647} = 37{,}8 \text{ грн.}$$

Середню гармонічну застосовують також для визначення середньої урожайності по групі однорідних культур, якщо відомі валовий збір і урожайність окремих культур, для обчислення середнього процента виконання плану виробництва і реалізації продукції по однорідній сукупності, якщо відомі дані про фактично вироблену або реалізовану продукцію і процент виконання плану по окремих об'єктах і т.д.

Середня геометрична. Середню геометричну застосовують, коли загальний обсяг явища є не сума, а добуток значень ознаки. Ця середня використовується здебільшого для розрахунку середніх коефіцієнтів (темпів) зростання і приросту при вивченні динаміки явищ (див. розділ 10) і має такий вигляд:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\Pi(x)}, \text{ або } \bar{x} = \sqrt[n-1]{\frac{y_n}{y_1}},$$

де n – число коефіцієнтів зростання; y_1 і y_n – початковий і кінцевий рівні динамічного ряду.

Величина середньої геометричної залежить тільки від співвідношення кінцевого і початкового рівнів. Якщо не змінюються в цих межах інші рівні, величина середньої не зміниться.

Розглянемо такий приклад. За даними про посівну площу цукрових буряків у господарстві за 5 років знайти середній коефіцієнт зростання площі посіву цукрових буряків за 1997–2001 рр. Всі розрахунки зведемо в табл. 4.5.

Таблиця 4.5

Дані для розрахунку середньої геометричної

Рік	Площа посіву цукрових буряків, га	Коефіцієнт зростання	Логарифм коефіцієнта зростання
	y_i	x_i	$\lg x_i$
1997	250	–	–
1998	275	1,1000	0,0414
1999	290	1,0545	0,0230
2000	310	1,0690	0,0290
2001	320	1,0323	0,0138
Сума	–	–	0,1072

Середнє значення логарифма коефіцієнта зростання становитиме: 0,1072 : 4 = 0,0268. За таблицями антилогарифмів знайдемо середній коефіцієнт зростання посівної площі цукрових буряків: antilg x = 1,0636, або 106,36%.

Такий саме результат одержимо і за другою формулою:

$$\bar{x} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5-1]{\frac{320}{250}} = \sqrt[4]{1,280};$$

$$\lg\overline{x} = \frac{\lg 320 - \lg 250}{5-1} = \frac{2{,}5051 - 2{,}3979}{4} = \frac{0{,}1072}{4} = 0{,}0268,$$

antilg $\overline{x}$ = 1,0636, або 106,36%.

Отже, середній коефіцієнт зростання посівної площі цукрових буряків у господарстві за 1997–2001 рр. становив 1,0636. Інакше кажучи, посівна площа цукрових буряків у господарстві щорічно збільшувалась в середньому на 6,36%.

Середня квадратична. Середня квадратична використовується переважно для розрахунку показників варіації (коливання) ознаки – дисперсії і середнього квадратичного відхилення, які обчислюються на основі квадратів відхилень індивідуальних значень ознаки від їхньої середньої арифметичної. Крім того, вона застосовується для узагальнення ознак, виражених лінійними мірами яких-небудь площ (при обчисленні середніх діаметрів стовбурів дерев, кошиків, листків, клубнів тощо).

Формули її такі:

проста зважена

$$\overline{x} = \sqrt{\frac{\sum x^2}{n}}; \qquad \overline{x} = \sqrt{\frac{\sum x^2 f}{\sum f}}.$$

Наприклад, є дані про розмір діаметрів стовбурів трьох яблунь (x_i): 17; 22; 19 см. Потрібно обчислити середній розмір діаметра стовбура яблуні. Оскільки вихідні дані представлені у вигляді квадратних функцій, середній розмір діаметра стовбура яблуні визначимо за формулою середньої квадратичної простої:

$$\overline{x} = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{17^2 + 22^2 + 19^2}{3}} = \sqrt{\frac{1134}{3}} = \sqrt{378} = 19{,}44 \text{ см}.$$

Якби у наведеному прикладі окремі значення діаметра стовбура повторювались неоднакове число разів, то середній розмір діаметра стовбура слід було б розраховувати за формулою середньої квадратичної зваженої.

Досліджуючи статистичну сукупність, можна виявити, що поряд з ознаками, які притаманні усім одиницям досліджуваного явища, є й такі ознаки, якими одні одиниці володіють, а інші ні. Такими ознаками, наприклад, будуть наявність в партії продукції бракованої продукції, рослини уражені хворобами та ін. Такі виключаючі одна одну ознаки називають **альтернативними.** При альтернативній варіації, коли є лише два виключаючих один одного випадки, наявність ознаки у одиниці сукупності прийнято позначати 1, а її відсутність – 0. Частку одиниць, що володіють досліджуваною ознакою, позначають p, а долю одиниць, не володіючих цією ознакою, – q. Очевидно, що $p + q = 1$, а $q = 1 - p$.

Середнє значення альтернативної ознаки, обчислене за формулою середньої арифметичної, буде дорівнювати:

$$\bar{x}=\frac{1\cdot p+0\cdot q}{p+q}=\frac{p}{p+q}=\frac{p}{1}=p.$$

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

Якщо обчислити різні типи середніх величин, одержаних із степеневої середньої для одного і того самого варіаційного ряду, то їх чисельні значення будуть відрізнятися один від одного, а самі середні розташуються таким чином:

$$\bar{x}_{\text{гарм}}<\bar{x}_{\text{геом}}<\bar{x}_{\text{ар}}<\bar{x}_{\text{квадр}},$$

тобто найбільшою буде середня квадратична, а найменшою – середня гармонічна. Порядок зростання середніх визначається значенням степеня k в степеневій середній.

Ця властивість степеневих середніх одержала назву **властивості мажорантності середніх.**

Приклад. Нехай маємо такі значення x_i : 2; 3; 36. Обчислимо вказані середні величини:

$$\bar{x}_{\text{гарм}}=\frac{n}{\sum\frac{1}{x}}=\frac{3}{\frac{1}{2}+\frac{1}{3}+\frac{1}{36}}=\frac{3}{0{,}8587}=3{,}50;$$

$$\bar{x}_{\text{геом}}=\sqrt[n]{\chi_1\cdot\chi_2\cdot\ldots\cdot\chi_n}=\sqrt[3]{2\cdot3\cdot36}=\sqrt[3]{216}=6{,}00;$$

$$\bar{x}_{\text{ар}}=\frac{\sum x}{n}=\frac{2+3+36}{3}=\frac{41}{3}=13{,}67;$$

$$\bar{x}_{\text{квадр}}=\sqrt{\frac{\sum x^2}{n}}=\sqrt{\frac{2^2+3^2+36^2}{3}}=\sqrt{463{,}3}=20{,}89.$$

Одержані середні розташуються у такому порядку: 3,50 < 6,00 < 13,67 < < 29,89, що відповідає вимозі властивості мажорантності середніх:

$$\bar{x}_{\text{гарм}}<\bar{x}_{\text{геом}}<\bar{x}_{\text{ар}}<\bar{x}.$$

Інші види середніх величин. Крім розглянутих вище видів середніх величин, статистикою розроблено й інші види.

Середня хронологічна являє собою середню величину з показників, що змінюються у часі. Вона розраховується із рівнів моментного або інтервального рядів динаміки за принципом середньої арифметичної простої і зваженої.

Для інтервального ряду динаміки середня хронологічна проста обчислюється за формулою:

$$\bar{y} = \frac{\sum y}{n},$$

де y – рівень ряду динаміки, n – число рівнів у ряду динаміки.

Для моментного ряду динаміки (при рівній відстані періодів, наприклад, місяць, квартал і т.д.) середня хронологічна проста обчислюється за формулою:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \ldots + \frac{1}{2}y_n}{n-1}.$$

Середня хронологічна зважена має вигляд $\bar{y} = \frac{\sum yt}{\sum t}$, якщо відомий час, протягом якого зберігалось кожне значення y.

Тут t – період часу, який відокремлює один рівень від іншого.

Для виявлення тенденції зміни досліджуваного явища у часі розраховують середню ковзну. Суть способу її розрахунку полягає в тому, що склад періоду безперервно і постійно змінюється – відбувається зсув на одну дату при збереженні постійного інтервалу (трирічча, п'ятиріччя і т.д.).

Приклади і методика розрахунку середньої хронологічної і середньої ковзної наведені при розгляді рядів динаміки (див. розділ 10).

В аналізі і плануванні сільськогосподарського виробництва застосовується також середня прогресивна. Цей вид середньої на відміну від загальної середньої дає узагальнену характеристику не всієї сукупності, а тільки тієї її частини, яка представлена показниками, вищими за загальну середню.

Середню прогресивну обчислюють у такій послідовності: 1) з усіх варіант обчислюють загальну середню; 2) відбираються варіанти, що за величиною перевищують загальну середню; 3) за відібраними варіантами обчислюють середню. Вона й буде середньою прогресивною. Наприклад, якщо сукупність представлена рядом чисел $x_1, x_2, \ldots, x_8$ та їх середнім значенням $\bar{x}$, серед яких x_1, x_2 і x_8 виявляться більшими за розміром, ніж загальна середня, то середня прогресивна становитиме:

$$\bar{x}_{\text{прогр}} = \frac{x_1 + x_2 + x_8}{3}.$$

Особливим видом середніх величин є **середня багатовимірна**, яка являє собою середню величину кількох ознак для однієї одиниці сукупності. Оскільки неможливо розрахувати середню величину за абсолютними значеннями різних ознак (різноякісних, виражених у різних одиницях виміру), то багатовимірна середня визначається з відносних величин (часток, процентів і т.п.), як правило, з відношень абсолютних значень для одиниці сукупності до середніх значень цих ознак.

Середня багатовимірна – похідна величина, що розраховується для статистичної сукупності чисельністю N одиниць з порядковими номерами i (i = 1, 2, 3,..., N), які володіють k ознаками (x) з порядковими номерами j (j =1, 2, 3,..., k) у такий спосіб. Спочатку обчислюють відношення P_{ij} значень кожної ознаки (x) у кожної величини сукупності до її середнього значення за формулою

$$P_{ij} = \frac{x_{ij}}{\bar{x}_j},$$

де x_{ij} – значення j-ої ознаки у i-ої одиниці сукупності; – її **середнє значення.**

Після цього визначають середню з цих відношень для **кожної одиниці** сукупності (), яку і називають багатовимірною середньою:

$$\bar{p}_i = \frac{\sum p_{ij}}{k}.$$

Багатовимірні середні дають узагальнену характеристику кожної одиниці сукупності за кількома ознаками одночасно. При цьому значущість ознаки для багатовимірної оцінки одиниці сукупності вважається однаковою, що економічно, звісно, неточно.

Розрахунок середньої багатовимірної розглянемо на прикладі порівняльної оцінки якості ґрунтів по групі господарств (табл. 4.6).

Розраховані багатовимірні середні дають змогу здійснити порівняльну оцінку якості грунтів по групі господарств за трьома найважливішими їхніми ознаками. З даних таблиці видно, що кращі грунти мають дев'яте і друге господарства, а гірші – третє і п'яте господарства.

Середню багатовимірну використовують для аналізу господарської діяльності підприємств, зокрема при визначенні ефективності використання виробничого потенціалу (землі, трудових ресурсів, виробничих фондів) та ін.

Таблиця 4.6

Дані для розрахунку середньої багатовимірної

Господарство	Вміст гумусу		Вміст доступного фосфору		Кислотність		$P_{ij} = P_{i1} + P_{i2} + P_{i3}$	$P_i = \frac{\sum P_{ij}}{3}$
	% до сухої маси x_1	$P_{i1} = \frac{x_{i1}}{\bar{x}_1}$	мг на 100 г грунту x_2	$P_{i2} = \frac{x_{i2}}{\bar{x}_2}$	pH x_3	$P_{i3} = \frac{x_{i3}}{\bar{x}_3}$		
1	3,1	0,92	19,8	0,91	6,5	1,03	2,86	0,95
2	4,0	1,18	25,5	1,16	6,5	1,03	3,37	1,12
3	2,4	0,71	18,2	0,83	5,8	0,91	2,45	0,82
4	3,7	1,10	22,8	1,04	6,3	0,99	3,13	1,04
5	2,3	0,68	19,8	0,91	6,1	0,96	2,55	0,85
6	3,9	1,15	20,6	0,94	6,2	0,98	3,07	1,02
7	2,9	0,86	21,9	1,00	6,2	0,98	2,84	0,95
8	4,0	1,18	23,4	1,07	6,5	1,03	3,28	1,09
9	3,9	1,15	25,3	1,16	7,0	1,10	3,41	1,14
10	3,6	1,07	21,4	0,98	6,3	0,99	3,04	1,02
Разом	**33,8**	**10,00**	**218,7**	**10,00**	**63,4**	**10,00**	**30,00**	**10,00**
У середньому	**3,38**	–	**21,87**	–	**6,34**	–	–	–

4.3. Властивості середньої арифметичної. Розрахунок середньої арифметичної способом моментів

Середня арифметична має ряд математичних властивостей, які можна використати, щоб спростити її розрахунки. Основні властивості середньої арифметичної такі.

1. Середня арифметична постійної величини дорівнює цій постійній:

$$\overline{A} = A \text{ при } A = \text{const}.$$

2. Сума квадратів відхилень від середньої арифметичної завжди менша, **ніж** сума квадратів відхилень від будь-якої іншої величини:

$$\sum (x - \overline{x})^2 f < (x - A)^2 f.$$

3. Величина середньої не зміниться, якщо частоти ряду розподілу замінити частостями.

4. Сума відхилень окремих значень ознаки від середньої, перемножених на ваги (частоти), дорівнює нулю:

$$\sum(x-\bar{x})=\sum x-n\bar{x}=0 \text{ – для простої середньої;}$$

$$\sum(x-\bar{x})f=\sum xf-\bar{x}\sum f=0 \text{ – для зваженої середньої.}$$

5. Якщо усі значення ознак збільшити або зменшити у ту саму **кількість** разів (h), то середня ($\bar{x}$) збільшиться або зменшиться у стільки ж разів:

$$\frac{\sum\frac{x}{h}f}{\sum f}=\frac{1}{h}\cdot\frac{\sum xf}{\sum f}=\frac{\bar{x}}{h},$$

тобто середня зменшилася в (h) разів.

6. Якщо з усіх значень варіант (x) відняти або додати до них ту саму постійну величину (x_0), то середня ($\bar{x}$) зменшиться або збільшиться на таку саму величину (x_0):

$$\frac{\sum(x-x_0)f}{\sum f}=\frac{\sum xf}{\sum f}-\frac{\sum x_0 f}{\sum f}=\bar{x}-\frac{x_0\sum f}{\sum f}=\bar{x}-x_0,$$

тобто середня зменшилася на постійне число x_0.

7. Якщо частоти (ваги) поділити або помножити на будь-яке постійне число (k), то середня не зміниться:

$$\bar{x}=\frac{\sum xkf}{\sum kf}=\frac{k\sum xf}{k\sum f}=\frac{\sum xf}{\sum f}=\bar{x},$$

тобто значення *середньої не змінилося.*

8. Добуток середньої на суму частот дорівнює сумі добутків варіант на частоти:

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

Ця рівність випливає з визначальної властивості середньої арифметичної, згідно з якою, зрівнюючи варіанти, надаючи їм однакові значення шляхом заміни їх середнім значенням, незмінним залишається загальний обсяг ознаки.

9. Загальна *середня дорівнює середній* із часткових середніх, зважених за чисельністю відповідних частин (груп) сукупності:

$$\bar{x}=\frac{\bar{x}_1 f_1+\bar{x}_2 f_2+\ldots+\bar{x}_n f_n}{f_1+f_2+\ldots+f_n}=\frac{\sum\bar{x}_i f_i}{\sum f_i}.$$

Викладені вище властивості середньої арифметичної дають змогу спростити її розрахунки: можна з усіх значень ознаки відняти довільну постійну величину, одержану різницю поділити на величину інтервалу, а потім обчислену середню помножити на величину інтервалу і додати довільну постійну величину, що прийнята за початок відліку.

Формула обчислення середньої арифметичної спрощеним способом має такий вигляд:

$$\bar{x} = \bar{x}'h + x_0 = \frac{\sum\left(\frac{x - x_0}{h}\right)f}{\sum f}h + x_0,$$

де $\bar{x}' = \frac{\sum x'f}{\sum f}$ – зменшена середня арифметична;

$x' = \frac{x - x_0}{h}$ – відхилення в інтервалах;

x_0 – початок відліку;
h – величина інтервалу.

Середня $\bar{x}'$ із значенням $\frac{x - x_0}{h}$ називається моментом першого порядку, а спосіб обчислення середньої – **способом моментів**, або **способом відліку від умовного початку.**

За умовний початок відліку (x_0) звичайно приймають одне із значень варіюючої ознаки, яке, як правило, знаходиться в центрі ряду розподілу або таке, що має найбільшу частоту.

Розглянемо приклад визначення середньої арифметичної в інтервальному ряду розподілу способом моментів, використовуючи дані про розподіл 100 господарств за надоєм молока на корову (табл. 4.7).

За умовний початок відліку (x_0) візьмемо одне із значень інтервалу, розташованого в центрі ряду розподілу і яке має найбільшу частоту. В нашій задачі таким значенням буде x_0= 33 ц. Величина інтервалу h = 2 ц.

За даними таблиці визначимо умовну (зменшену) середню арифметичну:

$$\bar{x}' = \frac{\sum x'f}{\sum f} = \frac{-18}{100} = -0{,}18 \text{ ц.}$$

Щоб одержати дійсну середню продуктивність корів, необхідно внести відповідні поправки:

$$\bar{x} = \bar{x}'h + x_0 = -0{,}18 \cdot 2 + 33 = -0{,}36 + 33 = 32{,}64 \text{ ц.}$$

Таким чином, одержано такий самий результат, як і за даними табл. 4.2. Результати розрахунків середньої арифметичної двома способами повністю співпали.

Таблиця 4.7

Дані для розрахунку середньої арифметичної в інтервальному ряду розподілу способом моментів

Група	Групи господарств за надоєм молока на корову, ц	Число господарств	Середина інтервалу	Відхилення від умовного початку		Добуток
				в ц	в інтервалах	
		f	x	$x - x_0$	$x' = \frac{x - x_0}{h}$	$x'f$
I	26-28	8	27	– 6	– 3	– 24
II	8-30	16	29	– 4	– 2	– 32
III	30-32	17	31	– 2	– 1	– 17
IV	32-34	25	33	0	0	0
V	34-36	18	35	2	1	18
VI	36-38	11	37	4	2	22
VII	38-40	5	39	6	3	15
Разом	–	**100**	–	–	–	**– 18**

4.4. Мода, медіана, квартилі і децилі

Крім перелічених вище середніх, у статистичному аналізі, як узагальнюючі характеристики сукупності, використовують такі значення ознаки, які відрізняються особливим розташуванням у варіаційному ряду розподілу. Це так звані **структурні (позиційні) середні**. Із них найчастіше застосовують моду і медіану.

Величина моди і медіани залежить лише від характеру частот, тобто від структури розподілу. Якщо величина середньої арифметичної залежить від усіх значень ознаки, то величина моди і медіани не залежить від крайніх значень ознаки. Це особливо важливо для рядів розподілу, в яких крайні значення ознаки мають нечітко виражені межі (до і понад).

Модою називають значення ознаки, що має найбільшу частоту в статистичному ряду розподілу. Спосіб обчислення моди залежить від того, в якому вигляді дано значення ознаки: дискретного чи інтервального ряду розподілу. В дискретних варіаційних рядах моду обчислюють без додаткових розрахунків за значенням варіанти з найбільшою частотою. Наприклад, відомий змінний виробіток деталей робітниками цеху:

виробіток деталей, шт. 30; 33; 35; 38;

число робітників, чол. 7; 10; 15; 12.

В даному прикладі модальною величиною є 35 деталей, оскільки ця величина у досліджуваній сукупності має найбільшу частоту – 15 випадків.

Модальною ціною на той або інший продукт на ринку є та ціна, яка спостерігається найчастіше.

В інтервальному варіаційному ряду розподілу модою наближенно вважають центральний варіант так званого *модального інтервалу*, тобто того інтервалу, який має найбільшу частоту. В межах інтервалу необхідно знайти те значення ознаки, яке є модою.

В інтервальних варіаційних рядах розподілу моду визначають за формулою:

$$\mathrm{Mo} = x_0 + h\frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)},$$

де x_0 – нижня (мінімальна) межа модального інтервалу; h – величина інтервалу; f_1 – частота передмодального інтервалу; f_2 – частота модального інтервалу; f_3 – частота післямодального інтервалу.

Формула ґрунтується на припущенні, що відстані від нижньої межі модального інтервалу до моди і від моди до верхньої межі модального інтервалу прямо пропорційні різницям між чисельностями (частотами) модального інтервалу і інтервалів, що прилягають до нього.

Розрахунок моди в інтервальному варіаційному ряду розподілу покажемо на прикладі розподілу 100 господарств за надоєм молока на корову (табл. 4.8).

Таблиця 4.8

Дані для розрахунку моди і медіани в інтервальному ряду розподілу

Група	Групи господарств за надоєм молока на корову, ц	*Число господарств*	*Нагромаджені частоти*
I	26-28	8	8
II	28-30	16	24 (8 + 16)
III	30-32	17	41 (17 + 24)
IV	32-34	25	66 (25 + 41)
V	34-36	18	84 (18 + 66)
VI	36-38	11	95 (11 + 84)
VII	38-40	5	100 (5 + 95)
Разом	–	100	–

Інтервал, в якому міститься мода, буде 32-34 ц, тому що цей інтервал має найбільшу частоту.

Підставивши відповідні числові значення у формулу моди, одержимо:

$$\mathrm{Mo} = x_0 + h\frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)} = 32 + 2\frac{25 - 17}{(25 - 17) + (25 - 18)} = 33{,}07 \text{ ц}.$$

Отже, у досліджуваній сукупності найбільше число господарств має продуктивність корів 33,07 ц.

Медіаною *називають таке значення ознаки, яке поділяє ранжирований* ряд розподілу на дві рівні частини, тобто значення, яке перебуває всередині ряду розподілу. Якщо в дискретному варіаційному ряду 2 m + 1 випадків, то *значення ознаки у випадку m + 1 є медіанним. Якщо в ряду парне число 2 m* випадків, медіану визначають як середню арифметичну з двох серединних значень. Наприклад, якщо 15 комбайнерів держгоспу розташувати у порядку зростання, тобто в ранжирований ряд за кількістю намолоченого ними зерна, то намолот зерна у восьмого комбайнера буде медіанним. Якщо ж число комбайнерів буде 16 чол., то медіаною буде середнє значення намолоту зерна восьмого і дев'ятого комбайнерів.

Медіану з парним і непарним числом варіант у дискретному ряду розподілу обчислюють за формулами:

$$Me = \frac{x_m + x_{m+1}}{2};\; Me = x_{m+1}.$$

В інтервальному варіаційному ряду розподілу медіану визначають за *формулою*:

$$Me = x_0 + h\frac{0{,}5\sum f - S_{me-1}}{f_{me}},$$

де x_0 – нижня (мінімальна) межа медіанного інтервалу; h – величина інтервалу; $0{,}5\sum f$ – половина суми нагромаджених частот інтервального ряду розподілу; S_{me-1} – сума нагромаджених частот інтервалу, що передує медіанному; f_{me} – частота медіанного інтервалу.

Для визначення медіани в інтервальному варіаційному ряду розподілу треба обчислити нагромаджені частоти і відшукати медіанний інтервал. Під **нагромадженими частотами** розуміють наростаючий підсумок частот, починаючи з першого інтервалу. Медіанним є той інтервал, на який припадає перша нагромаджена частота, що перевищує половину всього обсягу сукупності.

Обчислимо медіану за даними цього ж інтервального ряду розподілу (табл. 4.8). За даними таблиці побудуємо ряд нагромаджених частот і знайдемо медіанний інтервал. Медіанним інтервалом є інтервал 32-34 ц, тому що на цей інтервал припадає перша нагромаджена частота, що перевищує половину всього обсягу сукупності (66 перевищує $\sum f : 2 = 100 : 2 = 50$).

Медіанне значення продуктивності корів становитиме:

$$Me = x_0 + h\frac{0{,}5\sum f - S_{me-1}}{f_{me}} = 32 + 2\frac{100:2 - 41}{25} = 32{,}72 \text{ ц}.$$

Отже, продуктивність корів, рівна 32,72 ц, і є варіантою, що поділяє варіаційний ряд розподілу 100 господарств на дві рівні частини (50 господарств має надій на корову менше 32,72 ц і 50 господарств – більше 32,72 ц).

В одномодальних симетричних рядах розподілу середня арифметична, мода і медіана збігаються: $\bar{x} = Mo = Me$.

Для помірно асиметричних розподілів К.Пірсон встановив таке наближенне співвідношення між цими характеристиками:

$$Me = \frac{1}{3}Mo + \frac{2}{3}\bar{x}_{ар};\ Mo = \bar{x}_{ар} - 3(\bar{x}_{ар} - Me),\ \text{або}\ Mo = \bar{x}_{ар} + 3(Me - \bar{x}_{ар}).$$

Моду і медіану застосовують звичайно в тих випадках, коли визначати середню арифметичну недоцільно. Так, немає сенсу обчислювати середній розмір одягу і взуття, що їх виробляють фабрики. Для цього досить знати модальні розміри одягу і взуття, тобто ті, які користуються найбільшим попитом у населення з тим, щоб фабрики, плануючи своє виробництво, могли якомога краще задовольнити попит покупців саме на ці розміри одягу і взуття.

Медіана широко використовується при проектуванні місць будівництва об'єктів масового обслуговування населення (шкільних та дошкільних закладів, кінотеатрів, підприємств служби побуту і торгівлі тощо). Наприклад, продовольчий магазин у сільському селищі доцільно розташувати в такій точці, щоб він обслуговував половину кількості мешканців селища, а не розташовувався точно всередині його.

Додатково до медіани для характеристики структури варіаційного ряду розподілу обчислюють квартилі, які поділяють ранжирований ряд на 4 рівні частини, і децилі, які поділяють ранжирований ряд на 10 рівних частин. Другий квартиль Q_2 дорівнює медіані, а перший – Q_1 і третій – Q_3 обчислюють аналогічно розрахунку медіані, тільки замість медіанного інтервалу беруть для першого квартиля інтервал, в якому знаходиться варіанта, що відокремлює 1/4 кількості частот, а для третього квартиля – інтервал, в якому знаходиться варіанта, що відокремлює 3/4 кількості частот.

В інтервальному ряду розподілу перший і третій квартиль розраховують за такими формулами:

$$\text{перший квартиль}\ Q_1 = x_0 + h\frac{0{,}25\sum f - S_{Q_1-1}}{f_{Q_1}};$$

$$\text{третій квартиль}\ Q_3 = x_0 + h\frac{0{,}75\sum f - S_{Q_3-1}}{f_{Q_3}},$$

де x_0 – нижні (мінімальні) межі квартильних інтервалів; h – величина інтервалу; $\sum f$ – сума нагромаджених частот ряду розподілу; S_{Q_1-1} і S_{Q_3-1} –

нагромаджені частоти інтервалу, що передує інтервальному відповідно для першого і третього квартилів; f_{Q_1} і f_{Q_3} – частоти квартильних інтервалів.

Розрахунок першого і третього квартилів розглянемо на прикладі табл. 4.8.

Обчислимо перший квартиль. Для знаходження інтервалу, в якому знаходиться перший квартиль, використаємо нагромаджені частоти. Перший квартиль знаходиться в інтервалі, в який входить перша нагромаджена частота, що перевищує чверть загального обсягу сукупності ($0{,}25 \cdot 100 = 25$). Отже, перший квартиль Q_1 знаходиться в третьому інтервалі (з надоєм молока від 30 до 32 ц), який має суму нагромаджених частот 41.

Значення першого квартиля:

$$Q_1 = x_0 + h\frac{0{,}25\sum f - S_{Q_1-1}}{f_{Q_1}} = 30 + 2\frac{0{,}25 \cdot 100 - 24}{17} = 30{,}12 \text{ ц}.$$

Це означає, що одна чверть господарств має надій 30,12 ц, а три чверті – більше ніж 30,12 ц.

Щоб визначити третій квартиль, знайдемо інтервал, в якому він знаходиться. На цей інтервал припадає перша нагромаджена частота, що перевищує три чверті загального обсягу сукупності ($0{,}75 \cdot 100 = 75$). Отже, третій квартиль знаходиться в інтервалі 34-36 ц, який має суму нагромаджених частот, рівну 84.

Значення третього квартиля:

$$Q_3 = x_0 + h\frac{0{,}75\sum f - S_{Q_3-1}}{f_{Q_3}} = 34 + 2\frac{0{,}75 \cdot 100 - 66}{18} = 35{,}00 \text{ ц}.$$

Отже, три чверті господарств мають надій молока на корову до 35,00 ц, а одна чверть – більше ніж 35,00 ц.

В інтервальному ряду розподілу **децилі** визначають за такою формулою:

$$D = x_0 + h\frac{0{,}1\sum f - S_{D-1}}{f_D},$$

де x_0 – нижня (мінімальна) межа відповідного децильного інтервалу; h – величина інтервалу; S_{D-1} – сума нагромаджених частот інтервалів, що передують децильним; f_D – частоти відповідних децильних інтервалів.

Підставивши дані табл. 4.8 у формулу, визначимо перший дециль:

$$D_1 = x_0 + h\frac{0{,}1\sum f - S_{D_1-1}}{f_{D_1}} = 28 + 2\frac{0{,}1 \cdot 100 - 8}{16} = 28{,}25 \text{ ц}.$$

Отже, десята частина всіх господарств має надій на корову 28,25 ц і менше, а решта (90%) – більше ніж 28,25 ц.

Аналогічно розраховуються і решта децилів (другий, третій і т.д.).

4*

Завдання для самоконтролю до розділу 4

1. Дайте визначення середньої величини, її суті і значення в статистичному аналізі масових даних.

2. Які види середніх Ви знаєте? Напишіть їх формули.

3. Коли застосовуються прості і зважені формули середніх?

4. Які умови застосування середніх?

5. Назвіть основні властивості середньої арифметичної. В чому полягає суть спрощеного розрахунку середньої арифметичної?

6. В яких випадках застосовується середня арифметична і середня гармонічна?

7. Для розрахунку яких статистичних характеристик використовується середня квадратична і середня геометрична?

8. Що таке структурні середні?

9. Дайте визначення моди і медіани. Наведіть їх формули.

10. Як визначають моду і медіану в дискретних та інтервальних рядах розподілу?

11. За наведеними даними визначте середню арифметичну, моду і медіану.

Підприємство	Місячна заробітна плата, грн.	Чисельність працюючих, чол.
1	до 200	50
2	200-250	62
3	250-300	84
4	300-350	60
5	понад 350	54

Розділ 5

Показники варіації

5.1. Поняття варіації ознак. Показники варіації

Середні величини ($\overline{x}$, Mo, Me) як показники центральної тенденції, характеризуючи варіаційний рід одним числом, не враховують варіацію (коливання) ознаки, хоча вона має місце. В середній відображаються загальні умови, притаманні всій даній сукупності, але не відображаються індивідуальні часткові умови, що породжують варіацію у окремих одиниць сукупності. Середні величини також не показують як окремі значення досліджуваної ознаки групуються навколо середньої, чи зосереджені вони поблизу, чи значно відхиляються від неї. Середня, даючи узагальнену характеристику всієї сукупності, не показує характер варіації ознаки та ступінь її коливання. Два ряди розподілу, що мають однакову середню величину, можуть значно відрізнятися один від одного за ступенем варіації величини досліджуваної ознаки. Це можна показати на такому прикладі.

Припустимо, що на оранці зябу працюють дві бригади, кожна з трьох трактористів. Нехай кількість гектарів, зораних за зміну окремими трактористами, становила в першій бригаді 6, 7, 8 га, а в другій – 5, 7, 9 га. Середній виробіток на одного тракториста в обох бригадах однаковий і становить 7 га, однак коливання виробітку окремих трактористів у першій бригаді значно менше (2 га), ніж у другій (4 га). Очевидно, що ці відмінності зумовлені рядом факторів: виробничим стажем, рівнем кваліфікації, віком, умовами роботи і т.д.

В зв'язку з цим середня величина як показник центральної тенденції не дає вичерпної характеристики положення статистичного розподілу. Виникає необхідність вивчення варіації ознак, використовуючи для цієї мети специфічні показники міри розсіювання.

Аналізуючи одержанні в процесі статистичного спостереження дані про ту чи іншу ознаку, можна виявити чисельні відмінності між окремими одиницями сукупності. Відомо, що в певних межах коливаються показники урожайності, продуктивності тварин, продуктивності праці, заробітної плати та ін.

Варіацією ознаки називають наявність відмінностей в чисельних значеннях ознак у одиниць сукупності. Термін «варіація» походить від латинського слова *variatio* – зміна, коливання, відмінність. Варіація є властивістю статистичної сукупності. Вона зумовлена множиною взаємопов'язаних між собою необхідних та випадкових, внутрішніх і зовнішніх факторів, серед яких є основні та другорядні. Основні фактори формують центр розподілу, другорядні – варіацію ознак, спільна їх дія – форму розподілу. Наприклад, продуктивність тварин залежить від рівня годівлі, породності, рівня механізації та інших об'єктивних і суб'єктивних факторів. Спільна їх дія зумовлює той чи інший рівень продуктивності тварин в окремих господарствах, а також закономірність розподілу господарств за цією ознакою.

За ступенем варіації можна робити висновки про багато сторін процесу розвитку досліджуваних явищ, зокрема про однорідність сукупності, стійкість індивідуальних значень ознаки, типовість середньої, взаємозв'язок між ознаками одного і того самого явища і ознаками різних явищ.

Вимірювання варіації дає змогу оцінити ступінь впливу на досліджувану ознаку інших варіюючих ознак. Показники варіації слугують характеристикою типовості, надійності середньої величини. Чим менша варіація, тим середня більш типова, і навпаки – чим більш індивідуальні значення ознаки варіюють, коливаються навколо середньої, тим вона менш типова.

Основними завданнями вивчення варіації ознак є: 1) визначення міри варіації, тобто кількісного її вимірювання. Це завдання вирішується за допомогою розрахунку спеціальних показників варіації; 2) вивчення причин, які викликають варіацію, причинно-наслідкова оцінка характеру розсіювання, що передбачає дослідження закономірностей випадкової варіації в статистичних сукупностях; 3) розкладання загальної варіації ознаки на варіацію, що породжується систематичними та випадковими причинами.

Для вимірювання варіації ознаки застосовуються різні показники. Відповідно з визначенням варіація вимірюється ступенем коливання варіантів ознаки від рівня їх середньої величини. Саме на цьому і грунтується більшість показників, які застосовуються в статистиці для вимірювання варіації ознаки в сукупності.

Всі показники варіації поділяються на дві групи: абсолютні та відносні. До абсолютних відносяться розмах варіації, середнє лінійне відхилення, дисперсія, середнє квадратичне відхилення, середнє квартильне відхилення. Друга група показників розраховується як відношення абсолютних показників варіації до середньої арифметичної (або медіани). Відносними показниками варіації є коефіцієнт осциляції, варіації, відносне лінійне відхилення та ін. Кожний з названих показників має певні аналітичні переваги при вирішенні тих чи інших статистичних задач. Вибір того чи іншого показника залежить від конкретних завдань статистичного аналізу і наявної інформації.

Розглянемо перелічені показники варіації більш докладно.

Розмах варіації. Найпростішим показником варіації є розмах варіації, який представляє собою різницю між максимальним і мінімальним значеннями ознаки.

$$R = x_{max} - x_{min}.$$

В інтервальних рядах розподілу розмах варіації визначають як різницю між верхньою межею останнього та нижньою межею першого або як різницю між серединами інтервалів.

Безумовною перевагою показника розмаху варіації є простота його розрахунку. Однак він не може повною мірою охарактеризувати варіацію ознаки, оскільки не враховує всіх значень ознаки, проміжних між максимальним та мінімальним значеннями. Не враховує він і частот. Особливість показника розмаху варіації полягає у тому, що він залежить лише від двох крайніх значень ознаки, які можуть виявитися не достатньо типовими. В зв'язку з цим розмах варіації відображає інколи випадкове, а не типове для даного ряду коливання. Зазначені недоліки розмаху варіації звужують область його практичного застосування. В основному він використовується для попередньої оцінки варіації. Тому необхідні інші показники варіації, які ґрунтуються на всіх значеннях ознаки в даній сукупності.

Більш досконалим показником вимірювання варіації є середнє лінійне та середнє квадратичне відхилення, які усувають зазначені вище недоліки розмаху варіації.

Середнє лінійне відхилення являє собою середню арифметичну з абсолютних значень відхилень окремих варіант від середньої арифметичної.

просте зважене

$$\bar{l} = \frac{\sum|x-\bar{x}|}{n}; \qquad \bar{l} = \frac{\sum|x-x|f}{\sum f}.$$

Прямі дужки означають, що абсолютні значення відхилень беруться по модулю, тобто підсумовування виконується без врахування знаків (плюс або мінус). Така умовність пояснюється тим, що оскільки сума відхилень індивідуальних значень ознаки від середньої в першому ступені дорівнює нулю (*нульова властивість середньої арифметичної*), то для одержання суми всіх відхилень, відмінної від нуля, кожне відхилення слід брати як додатну величину.

Показник середнього лінійного відхилення більш обґрунтований порівняно з розмахом варіації. Він не залежить від випадкових коливань крайніх значень, оскільки спирається на всі значення ознаки, враховує всю суму відхилень індивідуальних варіантів від середньої арифметичної та частоти.

Однак і цей показник варіації має суттєві недоліки. Основним є те, що в ньому не враховуються знаки (спрямованість) відхилень. Довільне відкидання алгебраїчних знаків відхилень призводить до того, що математичні властивості цього показника є далеко не елементарними, а це значно ускладнює використання середнього лінійного відхилення при розв'язанні задач, пов'язаних з імовірнісними розрахунками. Тому середнє лінійне відхилення використовується рідко.

Намагання скласти показник варіації, який би усував недоліки розмаху варіації та середнього лінійного відхилення призводить до дисперсії та середнього квадратичного відхилення.

Дисперсією називають середній квадрат відхилень індивідуальних значень ознаки від середньої арифметичної. Її визначають за формулами:

проста зважена

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n}; \qquad \sigma^2 = \frac{\sum(x-\bar{x})^2 f}{\sum f}.$$

Середнє квадратичне відхилення одержують шляхом добування кореня квадратного з дисперсії:

просте зважене

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}; \qquad \sigma = \sqrt{\frac{\sum(x-\bar{x})^2 f}{\sum f}}.$$

Змістовне значення середнього квадратичного відхилення таке ж саме, як і середнього лінійного відхилення. Воно показує, на скільки в середньому відхиляються індивідуальні значення варіант від їх середнього значення.

Середнє квадратичне відхилення є критерієм надійності середньої. Чим воно менше, тим краще середня арифметична відображає всю досліджувану сукупність. Перевага середнього квадратичного відхилення порівняно з середнім лінійним відхиленням полягає у тому, що при розрахунку ніякого умовного припущення про підсумовування відхилень без врахування знаків не допускається, оскільки всі відхилення підносяться до квадрату.

Середнє квадратичне відхилення ще називають **стандартним відхиленням.** Воно як розмах варіації й середнє лінійне відхилення є величиною іменованою та виражається в тих самих одиницях вимірювання, що і варіанти досліджуваної ознаки і середня величина (ц, кг, грн., м, ц/га і т.д.).

Дисперсія і середнє квадратичне відхилення широко застосовуються на практиці. Пояснюється це тим, що вони входять в більшість теорем, які є фундаментом математичної статистики. Крім того, дисперсія може бути розкладена на складові елементи, які дають змогу оцінити вплив різних факторів, що зумовлюють варіацію досліджуваної ознаки. В наступних розділах буде показано, як дисперсія використовується для оцінки результатів вибіркового спостереження, побудови показників тісноти кореляційного зв'язку, в дисперсійному аналізі і т.д.

Середнє квадратичне відхилення відіграє важливу роль в аналізі рядів розподілу. В умовах нормального розподілу існує така залежність між величиною середнього квадратичного відхилення і кількістю спостережень: в межах $x \pm 1\sigma$ розташовується 0,683, або 68,3% кількості спостережень; в межах $x \pm 2\sigma$ – 0,954, або 95,4%; в межах $x \pm 3\sigma$ – 0,997, або 99,7% кількості спостережень. В дійсності на практиці майже не зустрічаються відхилення, які перевищують $\pm 3\sigma$. Відхилення 3σ може вважатися максимально *можливим. Це положення називають* ***правилом трьох сигм***.

Якщо показником центру розподілу використовується медіана, то для характеристики варіації можна застосувати так зване **квартильне відхилення**:

$$Q = \frac{Q_3 - Q_1}{2},$$

де Q_1 і Q_3 – відповідно перший і третій квартилі розподілу.

Цей показник також можна застосувати замість розмаху варіації, щоб запобігти недоліків, пов'язаних з використанням крайніх значень ознаки.

Між середнім квадратичним відхиленням, середнім лінійним відхиленням, квартильним відхиленням і розмахом варіації в нормально розподіленій сукупності існує таке співвідношення:

$$6\sigma \approx 7{,}5\bar{l} \approx 9{,}0Q \approx R, \text{ або } \sigma \approx 1{,}25\bar{l} \approx 1{,}5Q \approx 1/6R.$$

Поряд з варіацією кількісних ознак в соціально-економічних явищах має місце і варіація якісних ознак. При цьому якщо є тільки два взаємовиключаючих варіанти, то таку варіацію називають **альтернативною.** При альтернативній мінливості одні одиниці сукупності володіють даною ознакою, а інші не володіють. Наприклад, розгляд сільськогосподарських тварин з точки зору їх статевого та породного складу (бички та телички, породна та безпородна худоба), придатності продукції (придатна і бракована) і т.д. дає альтернативну ознаку. Наявність ознаки у одиниці сукупності позначають 1, а відсутність – 0; частку одиниць, що володіють даною ознакою, позначають p, а не володіючих – q. Очевидно, що $p + q = 1$. Звідки $p = 1 - q$, а $q = 1 - p$.

Дисперсія альтернативної ознаки визначається за формулою:

$$\sigma^2 = \frac{\sum(x-\bar{x})^2 f}{\sum f} = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = \frac{q^2 p + p^2 q}{p+q} = pq.$$

Таким чином, дисперсія альтернативної ознаки дорівнює добутку частки одиниць, що володіють даною ознакою, на частку одиниць, що не володіють нею.

Середнє квадратичне відхилення альтернативної ознаки дорівнює:

$$\sigma = \sqrt{\sigma^2} = \sqrt{pq}\text{, або } \sigma = \sqrt{p(1-p)}.$$

Оскільки $p + q$ не може бути більше одиниці (0,5 + 0,5), то дисперсія не може перевищувати 0,25.

Наприклад, при огляді партії сільськогосподарської продукції 2% виявилося бракованою. Позначимо наявність браку – 1, а відсутність – 0, частку бракованої продукції – p, а доброякісної – q.

Тоді $\sigma^2 = pq = 0{,}02 \cdot 0{,}98 = 0{,}0196$; $\sigma = \sqrt{pq} = \sqrt{0{,}0196} = 0{,}14$.

Отже, середнє квадратичне відхилення, яке показує як в середньому відхиляються індивідуальні значення ознаки від середньої арифметичної, дорівнює 0,14, або 14%.

При порівнянні коливання сукупностей, що мають різні одиниці вимірювання та значення середніх величин, робити висновки про ступінь варіації за середнім лінійним і середнім квадратичним відхиленнями важко. Тому з метою одержання порівнянних даних необхідно від абсолютних показників варіації перейти до відносних. Ці показники розраховуються як відношення абсолютних показників варіації до середньої арифметичної (медіани). Використовуючи за абсолютні показники варіації розмах варіації, середнє лінійне відхилення, середнє квадратичне відхилення та квартильне відхилення, одержимо відносні показники коливання (найчастіше вони виражаються у відсотках):

коефіцієнт осциляції:

$$V_R = \frac{R}{\bar{x}} \cdot 100\%;$$

відносне лінійне відхилення:

$$V_{\bar{l}} = \frac{\bar{l}}{\bar{x}} \cdot 100\%;$$

коефіцієнт варіації:

$$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%;$$

відносне квартильне відхилення:

$$V_Q = \frac{Q}{\bar{x}} \cdot 100\%\text{, або } V_Q = \frac{Q}{\text{Me}} \cdot 100\%\text{, або } V_Q = \frac{Q_3 - Q_1}{2Q_2},$$

де Q – квартильне відхилення; Q_1 – перший квартиль; Q_2 – медіана; Q_3 – третій квартиль.

Враховуючи, що середнє квадратичне відхилення дає узагальнену характеристику коливання всіх варіантів сукупності, коефіцієнт варіації є показником відносної варіації, що найчастіше застосовується. Його застосовують не тільки для порівняльної оцінки варіації, але й для характеристики однорідності сукупності. При цьому виходять з того, що якщо коефіцієнт варіації менше 33%, то сукупність вважається однорідною (для розподілів близьких до нормального).

Розрахунок перелічених показників варіації здійснимо за даними розподілу 100 господарств за надоєм молока на корову (табл. 5.1).

Таблиця 5.1

Дані для розрахунку середнього лінійного і середнього квадратичного відхилень

Група	Інтервали за надоєм молока на корову, ц	Середина інтервалу, ц	Число господарств	Середнє лінійне відхилення		Середнє квадратичне відхилення	
		x	f	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert f$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
I	26-28	27	8	5,64	45,12	31,81	254,48
II	28-30	29	16	3,64	58,24	13,25	212,00
III	30-32	31	17	1,64	27,88	2,96	45,73
IV	32-34	33	25	0,36	9,00	0,13	3,25
V	34-36	35	18	2,36	42,48	5,57	100,26
VI	36-38	37	11	4,36	47,96	19,01	209,11
VII	38-40	39	5	6,36	31,80	40,45	202,25
Разом	×	×	100	×	262,48	×	1027,08

Нагадаємо, що раніше, в розділі 4, за даними досліджуваного розподілу були обчислені такі характеристики: середня арифметична $\bar{x} = 32{,}64$ ц, перший квартиль – $Q_1 = 30{,}17$ ц, медіана – $Q_2 = 32{,}72$ ц, третій квартиль – $Q_3 =$ $= 35{,}00$ ц.

Абсолютні показники варіації

Розмах варіації

$$R = x_{max} - x_{min} = 40 - 26 = 14 \text{ ц}.$$

Середнє лінійне відхилення

$$\bar{l} = \frac{\sum|x-x|f}{\sum f} = \frac{262{,}48}{100} = 2{,}6 \text{ ц}.$$

Дисперсія

$$\sigma^2 = \frac{\sum(x-\bar{x})^2 f}{\sum f} = \frac{1027{,}08}{100} = 10{,}27 \text{ ц}.$$

Середнє квадратичне відхилення

$$\sigma = \sqrt{\sigma^2} = \sqrt{10,27} = 3,2 \text{ ц}.$$

Квартильне відхилення

$$Q = \frac{Q_3 - Q_1}{2} = \frac{35,00 - 30,17}{2} = 2,4 \text{ ц}.$$

Відносні показники варіації

Коефіцієнт осциляції

$$V_R = \frac{R}{\bar{x}} \cdot 100\% = \frac{14}{32,64} \cdot 100\% = 42,9\%.$$

Відносне лінійне відхилення

$$V_{\bar{l}} = \frac{\bar{l}}{\bar{x}} \cdot 100\% = \frac{2,6}{32,64} \cdot 100\% = 8,0\%.$$

Коефіцієнт варіації

$$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{3,2}{32,64} \cdot 100\% = 9,8\%.$$

Відносне квартильне відхилення

$$V_Q = \frac{Q}{\text{Me}} \cdot 100\% = \frac{2,4}{32,72} \cdot 100\% = 7,3\%.$$

Отже, надої молока по даній сукупності господарств коливаються в межах ±3,2 ц (по σ), або на 9,8% по відношенню до середнього надою.

5.2. Математичні властивості дисперсії та спрощені способи її розрахунку

Дисперсія володіє рядом математичних властивостей, які дають змогу спростити розрахунки. Розглянемо їх.

1. Дисперсія постійної величини дорівнює нулю:

$$\sigma^2_{A\ \text{const}} = 0.$$

Ця властивість випливає з того, що дисперсія є показником розсіювання варіант навколо середньої арифметичної, а середня арифметична постійної величини дорівнює нулю.

2. Якщо з усіх значень варіант відняти постійну величину (x_0), то дисперсія не зміниться:

$$\sigma^2_{(x_i - x_0)} = \sigma^2.$$

Це означає, що дисперсію можна розрахувати не за даними значення ознаки, а за відхиленнями від будь-якого постійного числа.

3. Якщо всі значення варіант зменшити (збільшити) в одне й те ж саме число разів (h), то дисперсія зменшиться (збільшиться) в h^2 разів, а середнє квадратичне відхилення в h разів:

$$\sigma^2_{\frac{x}{h}} = \sigma^2 : h^2.$$

Це означає, що всі значення ознаки можна поділити на постійне число (наприклад, на величину інтервалу), обчислити середнє квадратичне відхилення, а потім помножити його на це постійне число:

$$\sigma_x = \sigma_{\frac{x}{h}} \cdot h.$$

4. Якщо обчислити середній квадрат відхилень від будь-якої величини A в тому чи іншому ступені відмінності від середньої арифметичної ($\bar{x}$), то він завжди буде більше середнього квадрата відхилень, обчисленого від середньої арифметичної:

$$\sigma^2_A > \sigma^2.$$

При цьому більше на цілком певну величину – квадрат різниці між середньою та цією умовно взятою величиною, тобто на

$$(\bar{x} - A)^2; \sigma^2_A = \sigma^2 + (\bar{x} - A)^2,$$

або

$$\sigma^2_A = \sigma^2 - (\bar{x} - A)^2 = \frac{\sum (x - A)^2 f}{\sum f} - (\bar{x} - A)^2,$$

де σ^2_A – середній квадрат відхилень від довільної величини A; σ^2 – середній квадрат відхилень від середньої арифметичної.

Це означає, що дисперсія від середньої завжди менша дисперсій, обрахованих від будь-яких інших довільних величин, тобто вона має властивість мінімальності.

Ряд властивостей дисперсії ґрунтується на рівності: $\sigma^2 = \overline{x^2} - (\bar{x})^2$, тобто дисперсія дорівнює різниці між середньою з квадратів варіант та квадратом середньої. Ця рівність випливає з того, що якщо довільну величину A прирівняти до нуля, то попередня формула дисперсії приймає вигляд:

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

Ця формула широко використовується в статистиці для спрощеного розрахунку дисперсії (табл. 5.2).

$$\sigma^2 = \overline{x}^2 - (\overline{x})^2 = \frac{\sum x^2 f}{\sum f} - (\overline{x})^2 = \frac{107564}{100} - 32{,}64^2 = 10{,}27;$$

$$\sigma = \sqrt{10{,}27} = 3{,}2 \text{ ц}.$$

Отже, одержано такий самий результат, що і при розрахунку дисперсії звичайним способом.

Використання зазначених вище властивостей дисперсії дає змогу спростити її обчислення. Так, використовуючи другу і третю властивості в ряду розподілу з рівними інтервалами, дисперсію можна обчислити способом відліку від умовного нуля (способом моментів) за формулою:

$$\sigma^2 = \frac{\sum \left(\frac{x - x_0}{h}\right)^2 f}{\sum f} h^2 - (\overline{x} - x_0)^2,$$

де x_0 - початок відліку; h - величина інтервалу.

Перетворюючи наведену формулу, дисперсію і середнє квадратичне відхилення можна визначити через моменти першого та другого порядків:

$$\sigma^2 = h^2 \left[(\overline{x}')^2 - (\overline{x}')^2\right]$$

або в розгорнутому вигляді

$$\sigma^2 = h^2 \left[\frac{\sum (x')^2 f}{\sum f} - \left(\frac{\sum x' f}{\sum f}\right)^2\right],$$

де $x' = \frac{x - x_0}{h}$ – відхилення в інтервалах;

$$\frac{\sum (x')^2 f}{\sum f} = \frac{\sum \left(\frac{x - x_0}{h}\right)^2 f}{\sum f} = m_2$$ – умовний момент другого порядку;

$$\frac{\sum (x') f}{\sum f} = \frac{\sum \left(\frac{x - x_0}{h}\right) f}{\sum f} = m_1$$ – умовний момент першого порядку.

Тоді формули для обчислення дисперсії та середнього квадратичного відхилення можна записати в такому вигляді:

$$\sigma^2 = h^2(m_2 - m_1^2) \text{ і } \sigma = h\sqrt{m^2 - m_1^2}.$$

Отже, дисперсія, обчислена з використанням умовних моментів, дорівнює добутку квадрата величини інтервалу на різницю умовних моментів першого і другого порядків. Такий спосіб розрахунку дисперсії дістав назву **способу моментів,** або **способу відліку від умовного нуля.**

Розрахуємо дисперсію цим способом для нашого прикладу (табл. 5.2).

Таблиця 5.2

Дані для розрахунку дисперсії спрощеним способом і способом відліку від умовного нуля

Номер групи	Серединні значення інтервалів по надою, ц	Число господарств	Спрощеним способом		Способом відліку від умовного нуля				
	x	f	x^2	x^2f	$x-x_0$	$\frac{x-x_0}{h}$	$\left(\frac{x-x_0}{h}\right)f$	$\left(\frac{x-x_0}{h}\right)^2$	$\left(\frac{x-x_0}{h}\right)^2 f$
I	27	8	729	5832	– 6	– 3	– 24	9	72
II	29	16	841	13456	– 4	– 2	– 32	4	64
III	31	17	961	16337	– 2	– 1	– 17	1	17
IV	33	25	1089	27225	0	0	0	0	0
V	35	18	1225	22050	2	1	18	1	18
VI	37	11	1369	15059	4	2	22	4	44
VII	39	5	1521	7605	6	3	15	9	45
Разом	×	100	×	107564	×	×	– 18	×	260

При цьому візьмемо $h = 2$ ц, $x_0 = 33$ ц.

$$\sigma^2 = \frac{\sum\left(\frac{x-x_0}{h}\right)^2 f}{\sum f}h^2 - (\bar{x} - x_0)^2 = \frac{260}{100}2^2 - (32{,}64 - 33{,}00) = 10{,}27.$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{10{,}27} = 3{,}2 \text{ ц}.$$

Такий самий результат одержимо і через умовні моменти першого і другого порядків.

$$m_1 = \frac{\sum\left(\frac{x-x_0}{h}\right)f}{\sum f} = \frac{-18}{100} = -0{,}18;$$

$$m_2 = \frac{\sum\left(\frac{x - x_0}{h}\right)^2 f}{\sum f} = \frac{260}{100} = 2{,}60;$$

$$\sigma^2 = h^2(m_2 - m_1^2) = 2^2\left[2{,}60 - (-0{,}18)^2\right] = 4 \cdot 2{,}5676 = 10{,}27;$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{10{,}27} = 3{,}2 \text{ ц.}$$

Отже, розрахунки дисперсії і середнього квадратичного **відхилення трьома** способами збіглися і дають одні й ті самі результати.

5.3. Види дисперсій і правило їх додавання

Вивчаючи коливання ознаки в цілому по всій сукупності і спираючись на загальну дисперсію, ми не можемо визначити вплив окремих факторів на варіацію ознаки, що нас цікавить. Це завдання можна вирішити за допомогою побудови статистичних групувань. Якщо досліджувану сукупність поділити на окремі сукупності (групи) за ознакою, що нас цікавить, то це дасть змогу розкласти загальну дисперсію ознаки на дві дисперсії, одна з яких буде характеризувати частину варіації, зумовлену впливом фактора, покладеного в основу групування, а друга – варіацію, що виникає під впливом інших факторів (крім фактора, покладеного в основу групування). Таким чином, для сукупності, поділеної на групи за якою-небудь ознакою, можна визначити такі види дисперсій: загальну, міжгрупову і внутрішньогрупову.

Загальна дисперсія ($\sigma^2_{\text{заг}}$) характеризує коливання (варіацію) ознаки під впливом всіх умов (факторів), що викликали цю варіацію. Вона обчислюється як відношення суми квадратів відхилень індивідуальних значень ознаки (x_i) від загальної середньої ($\bar{x}_0$) до числа одиниць сукупності:

проста	зважена
$\sigma^2_{\text{заг}} = \frac{\sum(x_i - \bar{x}_0)^2}{n};$	$\sigma^2_{\text{заг}} = \frac{\sum(x_i - \bar{x}_0)^2 f}{\sum f}.$

Міжгрупова (факторна) дисперсія ($\sigma^2_{\text{м.гр}}$) характеризує варіацію ознаки під впливом досліджуваного фактора (умови), покладеного в основу групування. Вона обчислюється як відношення суми квадратів відхилень групових середніх ($\bar{x}_i$) від загальної середньої до числа одиниць сукупності:

$$\text{проста} \qquad \sigma^2_{\text{м.гр}} = \frac{\sum(\bar{x}_i - \bar{x}_0)^2}{n}; \qquad \text{зважена} \qquad \sigma^2_{\text{м.гр}} = \frac{\sum(\bar{x}_i - \bar{x}_0)^2 f_i}{\sum f_i},$$

де $\bar{x}_i$ і f_i – групові середні і чисельності по окремих групах.

Внутрішньогрупова дисперсія ($\sigma^2_{\text{в.гр}}$) характеризує варіацію ознаки, зумовлену не врахованими при групуванні факторами. Вона залежить від умови (фактора), покладеного в основу групування і характеризує варіацію ознаки тільки за рахунок умов і факторів, що діють всередині групи. Для окремих груп внутрішньогрупова варіація розраховується як відношення суми квадратів відхилень індивідуальних значень ознаки (x_i) від групових середніх ($\bar{x}_i$) до числа одиниць сукупності:

$$\text{проста} \qquad \sigma^2_{\text{в.гр}} = \frac{\sum(x_i - \bar{x}_i)^2}{n}; \qquad \text{зважена} \qquad \sigma^2_{\text{в.гр}} = \frac{\sum(x_i - \bar{x}_i)^2 f_i}{\sum f_i}.$$

Вона може бути також визначена як середня арифметична зважена з групових дисперсій (σ^2_i):

$$\bar{\sigma}^2_{\text{в.гр}} = \frac{\sum \sigma_i^2 f_i}{\sum f_i}.$$

Усі три згадані дисперсії пов'язані між собою такою рівністю: величина загальної дисперсії дорівнює сумі величин міжгрупової та внутрішньогрупової дисперсій:

$$\sigma^2_{\text{заг}} = \sigma^2_{\text{м.гр}} + \sigma^2_{\text{в.гр}}.$$

Ця рівність дістала назву **правила додавання дисперсій.**

Знаючи будь-які два види дисперсій, завжди можна знайти або перевірити правильність розрахунку третього виду:

$$\sigma^2_{\text{м.гр}} = \sigma^2_{\text{заг}} - \sigma^2_{\text{в.гр}}; \sigma^2_{\text{в.гр}} = \sigma^2_{\text{заг}} - \sigma^2_{\text{м.гр}}.$$

Зіставленням міжгрупової та загальної дисперсій (відповідно до обсягів варіації) можна визначити ступінь впливу факторної ознаки, покладеної в основу групування, на коливання результативної ознаки. При цьому визначають так зване **кореляційне відношення:**

$$\eta^2 = \frac{\sigma^2_{\text{м.гр}}}{\sigma^2_{\text{заг}}} = \frac{W_{\text{м.гр}}}{W_{\text{заг}}},$$ яке характеризує частку варіації, зумовлену факторною ознакою.

Решта варіації $\frac{\sigma^2_{\text{в.гр}}}{\sigma^2_{\text{заг}}} = \frac{W_{\text{в.гр}}}{W_{\text{заг}}} = 1 - \eta^2$ визначається неврахованими при групуванні випадковими причинами.

Правило додавання дисперсій знаходить широке практичне застосування в статистичному аналізі оцінки істотності і ступеня впливу окремих факторів на загальне коливання результативних ознак (див. дисперсійний та кореляційний аналіз).

Розглянемо порядок визначення загального обсягу варіації та дисперсій, їх розкладання на міжгрупову та внутрішньогрупову на прикладі даних польового досліду (табл. 5.3).

Таблиця 5.3

Вплив різних доз добрив на урожайність льону-довгунця (соломка, ц/га)

Варіант досліду	Добрива	Урожайність по повтореннях, x_i				Сума $\sum x_i$	Середні $\bar{x}_i$
		I	II	III	IV		
1	Контроль	26,0	27,3	29,0	29,7	112,0	28,0
2	$N_{30}P_{60}K_{60}$	34,8	33,8	36,6	36,4	141,6	35,4
3	$N_{45}P_{90}K_{120}$	44,1	43,9	46,2	46,6	180,8	45,2
Суми	×	**104,9**	**105,0**	**111,8**	**112,7**	**434,4**	**36,2**

Аналіз даних таблиці показує, що урожайність льону-довгунця коливається як під впливом доз добрив (по варіантах досліду), так і в межах того самого варіанта досліду (по повтореннях). Отже, на урожайність крім добрив впливають й інші фактори.

Потрібно визначити загальний обсяг варіації урожайності льону-довгунця, розбивши його на варіацію, зв'язану з дією добрив (міжгрупову варіацію) і варіацію, зумовлену неврахованими в досліді факторами (внутрішньогрупову варіацію).

Введемо умовні позначення: m – число варіантів досліду ($m = 3$); n – число повторень ($n = 4$); N – загальне число спостережень ($N = m \cdot n = 3 \cdot 4 = 12$).

Для визначення відповідних сум квадратів відхилень та дисперсій піднесемо до квадрату урожайність (табл. 5.4).

Таблиця 5.4

Квадрати урожайності по повтореннях

Варіант досліду	Добрива	Урожайність по повтореннях, x_i				Сума $\sum x_i$	Середні $\bar{x}_i$
		I	II	III	IV		
1	Контроль	676,00	745,29	841,00	882,09	3144,38	784,00
2	$N_{30}P_{60}K_{60}$	1211,04	1142,44	1339,56	1324,96	5018,00	1253,16
3	$N_{45}P_{90}K_{120}$	1944,81	1927,21	2134,44	2171,56	8178,02	2043,04
Суми	×	**3831,85**	**3814,94**	**4315,00**	**4378,61**	**16340,40**	**4080,20**

Обчислимо суми квадратів відхилень, що характеризують загальну, міжгрупову і внутрішньогрупову варіації:

а) загальна $W_{заг} = \sum x_i^2 - N \cdot \bar{x}_0^2 = 16340,40 - 12 \cdot 36.2^2 = 615,12;$

б) міжгрупова $W_{м.гр} = n\left(\sum \bar{x}_i^2 - m\bar{x}_0^2\right) = 4(4080,20 - 3 \cdot 36,2)^2 = 595,52;$

в) внутрішньогрупова

для першого варіанту досліду

$W'_{в.гр} = \sum x_1^2 - n\bar{x}_1^2 = 3144,38 - 4 \cdot 28,0^2 = 8,38;$

для другого варіанту досліду

$W''_{в.гр} = \sum x_2^2 - n\bar{x}_2^2 = 5018,00 - 4 \cdot 35,4^2 = 5,36;$

для третього варіанту досліду

$W'''_{в.гр} = \sum x_3^2 - n\bar{x}_3^2 = 8178,02 - 4 \cdot 45,2^2 = 5,86.$

Загальна сума внутрішньогрупової варіації становитиме:

$W_{в.гр} = W'_{в.гр} + W''_{в.гр} + W'''_{в.гр} = 8,38 + 5,36 + 5,86 = 19,60.$

Цю ж суму можна знайти й іншим способом, виходячи з правила додавання (розкладання) варіації:

$$W_{в.гр} = W_{заг} - W_{м.гр} = 615,12 - 595,52 = 19,60.$$

Таким чином, можна записати, що

$$W_{заг} = W_{м.гр} + W_{в.гр};$$

$$615,12 = 595,52 + 19,60;$$

$$100,00\% = 96,8\% + 3,2\%.$$

Отже, загальна варіація урожайності льону-довгунця (615,12) розчленована на систематичну, зумовлену впливом різних доз добрив (595,52) і випадкову, викликану дією неврахованих у досліді факторів (19,60).

За зазначеними сумами квадратів відхилень визначимо загальну, міжгрупову та внутрішньогрупову дисперсії:

$$\sigma^2_{заг} = \frac{W_{заг}}{N} = \frac{615,12}{12} = 51,26;$$

$$\sigma^2_{м.гр} = \frac{W^2_{м.гр}}{N} = \frac{595,52}{12} = 49,63;$$

$$\sigma^2_{в.гр} = \frac{W^2_{в.гр}}{N} = \frac{19,60}{12} = 1,63.$$

За правилом додавання дисперсій можна записати:

$$\sigma^2_{заг} = \sigma^2_{м.гр} + \sigma^2_{в.гр};$$

$$51{,}26 = 49{,}63 + 1{,}63.$$

Отже, доведено, що загальна дисперсія дорівнює сумі міжгрупової і внутрішньогрупової дисперсій.

Співставляючи між собою міжгрупову та загальну дисперсії, визначимо кореляційне відношення, яке характеризує силу впливу досліджуваного фактора на результативну ознаку:

$$\eta^2 = \frac{\sigma^2_{м.гр}}{\sigma^2_{заг}} = \frac{49{,}63}{51{,}26} = 0{,}968, \text{ або } 96{,}8\%.$$

Отже, 96,8% загального коливання урожайності льону-довгунця припадає на частку добрив, а 3,2% зумовлено іншими неврахованими в досліді факторами.

5.4. Моменти статистичних розподілів

Розглянуті вище середні величини і показники варіації є частковими випадками єдиної системи узагальнюючих статистичних характеристик розподілу, що одержала назву моменту статистичного розподілу.

Моментом розподілу називають середню арифметичну величину з піднесених до заданого ступеня відхилень окремих варіант від деякої постійної величини ($0, \bar{x}, x_0$):

$$M_k = \frac{\sum (x_i - A)^k f_i}{\sum f_i} = \overline{(x_i - A)^k},$$

де A – постійна величина, від якої визначаються відхилення (за постійну величину можуть бути взяті нуль, середня арифметична $\bar{x}$, або умовний початок відліку x_0); k – показник степеня, що визначає порядок моменту.

Для вивчення характеристик статистичних розподілів найчастіше використовуються моменти перших п'яти порядків (k дорівнює 0, 1, 2, 3, 4).

Залежно від того, що приймається за постійну величину, від якої визначаються відхилення, розрізняють три види моментів: початкові, центральні та умовні.

Моменти розподілу, при обчисленні яких за вихідну величину приймається нуль ($A = 0$), називають **початковими моментами** (M):

$$M_k = \overline{(x_i - 0)^k} = \frac{\sum x_i^k f_i}{\sum f_i}.$$

Моменти розподілу, при обчисленні яких за вихідну величину приймаються відхилення від середньої арифметичної ($A = \bar{x}$), називають **центральними моментами** (μ):

$$\mu_k = \overline{(x_i - \bar{x})^k} = \frac{\sum (x_i - \bar{x})^k f_i}{\sum f_i}.$$

Моменти розподілу, при обчисленні яких за вихідну величину приймаються відхилення від довільно взятої величини (x_0), тобто від так званого умовного початку відліку, називають **умовними моментами** (m):

$$m_k = \overline{(x_i - x_0)^k} = \frac{\sum (x_i - x_0)^k f_i}{\sum f_i}.$$

Початкові моменти другого, третього і четвертого порядків так само, як і умовні моменти, самостійного значення не мають, а використовуються для спрощеного обчислення центральних моментів.

Аналізуючи формули моментів, можна помітити, що початковий момент першого порядку $M_1 = \frac{\sum x_i f_i}{\sum f_i}$ являє собою середню арифметичну ($\bar{x}$) і використовується як показник центру розподілу. Центральний момент першого порядку завжди дорівнює нулю (нульова властивість середньої арифметичної $\sum (x_i - \bar{x}) = 0$). Центральний момент другого порядку $\mu_2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$ дорівнює дисперсії. Центральний момент третього порядку μ_3 дорівнює нулю в симетричному розподілі і використовується для визначення показника асиметрії (скошеності). Центральний момент четвертого порядку застосовується при обчисленні показника ексцесу (гостровершинності).

У зв'язку з тим, що теоретична форма розподілу найчастіше невідома, викликає інтерес вивчення деяких властивостей кривої, побудованої за даними емпіричного розподілу. Зокрема, велике значення має вимірювання ступеня відхилення даного розподілу від симетричного та характеристика особливості побудови вершини кривої розподілу (ступеня гостровершинності). З цією метою обчислюються показники асиметрії (скошеності) і гостровершинності (ексцесу).

Оскільки моменти залежать від прийнятої системи одиниць, в статистичній практиці виявляється більш доцільним брати не абсолютні значення моментів, а їх відношення до стандартного відхилення (середнього квадратичного відхилення σ) у відповідному степені.

За міру асиметрії (скошеності) прийнято брати стандартизоване відхилення $k_{ск} = \frac{\mu_3}{\sigma_3}$, **тобто коефіцієнт скошеності (асиметрії)**, який являє собою відношення центрального моменту третього порядку до середнього квадратичного відхилення в третьому степені.

Розрізняють також **нормовані моменти**, під якими розуміють відношення k-го порядку до середнього квадратичного відхилення в k-му степені. Відповідно до цього коефіцієнт скошеності можна розглядати як третій нормований центральний момент розподілу.

Про наявність асиметрії в досліджуваному розподілі можна судити і за неспівпаданням показників центру розподілу ($\overline{x}$ і Мо): чим більше між ними різниця, тим більше асиметрія ряду розподілу. Для симетричних розподілів частоти будь-яких двох варіант, рівновіддалених по обидві сторони від центру розподілу, рівні між собою. Розраховані для таких розподілів середня, мода і медіана також рівні.

Одним з найбільш простих показників асиметрії (скошеності), що ґрунтуються на співвідношеннях середньої арифметичної і моди, є показник

$$k_{ск} = \frac{\overline{x} - \text{Mo}}{\sigma}.$$

Його величина може бути додатною чи від'ємною. В першому випадку мова йде про правосторонню асиметрію, в другому – про лівосторонню.

При наявності додатної (правосторонньої) скошеності (права гілка кривої довша) між показниками центру розподілу існує таке співвідношення: $\text{Mo} < \text{Me} < \overline{x}$, *відповідно при наявності від'ємної (лівосторонньої) скошеності* (ліва гілка кривої довша) спостерігається обернене співвідношення: $\text{Mo} > \text{Me} > \overline{x}$.

В практичних розрахунках по визначенню асиметрії перевага надається третьому нормованому центральному моменту.

В симетричному ряді розподілу $k = 0$, при правосторонній скошеності $k > 0$, при лівосторонній $k < 0$. Прийнято вважати, що асиметрія вища 0,5 (незалежно від знаку) рахується значною; якщо вона менша 0,25, то незначною.

Для характеристики ступеня гостровершинності (ексцесу) використовується четвертий нормований центральний момент, тобто відношення μ_4 / σ^4.

В нормальному розподілі існує таке співвідношення між центральним моментом четвертого порядку та центральним моментом другого порядку (дисперсією): $\mu^4 = 3\sigma^4$, тобто для нормального розподілу четвертий нормований момент дорівнює 3 ($\mu_4/\sigma^4 = 3$).

Тому дане співвідношення можна використати як міру гостровершинності. Якщо показник гостровершинності (ексцесу) представити у вигляді $Ex = \frac{\mu_4}{\sigma^4} - 3$, то в нормальному розподілі $Ex = 0$, при гостровершинному або додатному ексцесі $Ex > 0$ і при плосковершинному або від'ємному ексцесі $Ex < 0$.

Обчислимо для досліджуваного ряду розподілу 100 господарств за надоєм молока на корову коефіцієнти скошеності і ексцесу, попередньо визначивши центральні моменти через умовні (табл. 5.5).

Таблиця 5.5

Дані для розрахунку умовних моментів

Серединні значення інтервалу, т	Перші чотири степеня відхилень варіант від умовного початку				Число господарств f_i	Добуток			
	$x' = \frac{x - x_0}{h}$	$(x')^2$	$(x')^3$	$(x')^4$		$x'f$	$(x')^2 f$	$(x')^3 f$	$(x')^4 f$
2,7	– 3	9	– 27	81	8	– 24	72	– 216	648
2,9	– 2	4	– 8	16	16	– 32	64	– 128	256
3,1	– 1	1	– 1	1	17	– 17	17	– 17	17
3,3	0	0	0	0	25	0	0	0	0
3,5	1	1	1	1	18	18	18	18	18
3,7	2	4	8	16	11	22	44	88	176
3,9	3	9	27	81	5	15	45	135	405
Разом	–	–	–	–	100	– 18	260	– 120	1520

За умовний початок відліку приймемо серединне значення інтервалу з надоєм рівним 3,3 т () і який має найбільшу частоту. Величина інтервалу $h = 0{,}2$ т.

Використовуючи дані табл. 5.5, визначимо значення моментів відносно початку відліку, виражені в частках інтервалу:

$$m_1' = \frac{\sum(x - x_0)f}{\sum f} = \frac{-18}{100} = -0{,}18;$$

$$m_2' = \frac{\sum(x - x_0)^2 f}{\sum f} = \frac{260}{100} = 2{,}60;$$

$$m_3' = \frac{\sum(x - x_0)^3 f}{\sum f} = \frac{-120}{100} = -1{,}20;$$

$$m_4' = \frac{\sum (x - x_0)^4 f}{\sum f} = \frac{1520}{100} = 15{,}20.$$

Визначимо значення умовних моментів, виражених у вихідній системі одиниць вимірювання, вносячи при цьому поправку на величину інтервалу у відповідному степені, виходячи зі співвідношення $m_k = m'_k \cdot h^k$, де k – порядок моменту (показник степені); h – величина інтервалу;

$$m_1 = m_1' \cdot h = -0{,}18 \cdot 0{,}2 = -0{,}036;$$

$$m_2 = m_2' \cdot h^2 = 2{,}60 \cdot 0{,}2^2 = 0{,}104;$$

$$m_3 = m_3' \cdot h^3 = -1{,}20 \cdot 0{,}2^3 = -0{,}0096;$$

$$m_4 = m_4' \cdot h^4 = 15{,}20 \cdot 0{,}2^4 = 0{,}02432.$$

Розрахуємо центральні моменти через умовні, використовуючи формули взаємозв'язку між моментами:

$\mu^2 = m^2 - m_1^2 = 0{,}104 - (-0{,}036)^2 = 0{,}1027$, тобто центральний момент другого порядку дорівнює дисперсії ($\sigma^2 = 0{,}1027; \sigma = 0{,}32$ т $= 3{,}2$ ц);

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3 = -0{,}096 - 3 \cdot 0{,}104(-0{,}036) + 2(-0{,}036)^3 = 0{,}0015;$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m^2m_1^2 - 3m_1^4 = 0{,}02432 - 4(-0{,}0096)(-0{,}036) + 6 \cdot 0{,}104 \times$$
$$\times (-0{,}036)^2 - 3(-0{,}036)^4 = 0{,}024.$$

Визначимо коефіцієнт скошеності (асиметрії):

$$k_{\text{ск}} = \frac{\mu_3}{\sigma^3} = \frac{0{,}0015}{0{,}32^3} = \frac{0{,}0015}{0{,}0328} = 0{,}046.$$

Звідси випливає, що даний ряд розподілу господарств за надоєм молока на корову близький до симетричного, але має невелику додатну скошеність.

Розрахуємо коефіцієнт гостровершинності (ексцесу):

$$Ex = \frac{\mu_4}{\sigma^4} - 3 = \frac{0{,}024}{0{,}32^4} - 3 = -0{,}714,$$

Тобто досліджуваний ряд розподілу характеризується істотною плосковершинністю побудови вершини кривої розподілу.

Визначивши комплекс середніх величин і показників варіації, ми отримали систему статистичних характеристик, які дають можливість всебічно описати досліджуваний ряд розподілу і зробити загальні висновки.

Випишемо основні статистичні характеристики ряду розподілу 100 господарств за надоєм молока на корову, ц:

середня арифметична мода	$\bar{x} = 32,64;$
мода	$Mo = 33,07;$
медіана	$Me = 32,72;$
розмах варіації	$R = 14,0;$
середнє лінійне відхилення	$\bar{l} = 2,6;$
дисперсія	$\sigma^2 = 10,27;$
середнє квадратичне відхилення	$\sigma = 3,2;$
коефіцієнт скошеності	$k_{ск} = 0,046;$
коефіцієнт гостровершинності	$Ex = -0,714.$

Аналіз наведених статистичних характеристик дає змогу зробити загальний висновок щодо форми розподілу 100 господарств за надоєм молока на корову: досліджуваний ряд є майже симетричним, незначно додатньо скошеним і плосковершинним; розподіл за формою близький до нормального. Однак доведення цього положення потребує спеціальної статистичної оцінки близькості досліджуваного ряду розподілу нормальному на основі відповідних критеріїв. Одержані характеристики ряду розподілу є лише попередніми оцінками відповідних характеристик генеральної сукупності і тому потрібна оцінка їх надійності. Ці питання розглядаються в наступних розділах підручника.

Завдання для самоконтролю до розділу 5

1. Що таке варіація ознак? Наведіть приклади.

2. Які показники використовують для вимірювання варіації? Назвіть їх і наведіть формули.

3. *Які недоліки притаманні розмаху варіації і середньому лінійному відхиленню?*

4. Розкажіть про дисперсію і середнє квадратичне відхилення та їх місце в системі показників варіації.

5. Назвіть основні математичні властивості дисперсії.

6. Наведіть формули спрощених розрахунків дисперсії і поясніть їх суть.

7. Як вимірюють варіацію альтернативної ознаки?

8. Які види дисперсій Ви знаєте? Розкрийте їх суть.

9. Розкажіть про правило додавання (розкладання) варіації. Де воно застосовується?

10. За наведеними даними розрахуйте такі показники варіації: розмах варіації, середнє лінійне відхилення, дисперсію, середнє квадратичне відхилення, коефіцієнт варіації. Зробіть висновки.

Група	Вартість основних фондів, млн грн.	Кількість підприємств
I	0,5-0,8	8
II	0,8-1,1	15
III	1,1-1,4	30
IV	1,4-1,7	12
V	1,7-2,0	7

Розділ 6

Вибіркове спостереження

6.1. Поняття вибіркового спостереження та його теоретичні основи

Як зазначалось у розділі 2, за ступенем охоплення одиниць досліджуваної сукупності статистичне спостереження може бути суцільним і несуцільним.

Суцільне спостереження передбачає обстеження усіх без винятку одиниць генеральної сукупності. Наприклад, для визначення загальної чисельності населення під час перепису збирають дані про кожну окрему людину, яка проживає в країні, для встановлення обсягу виробленої продукції (зерна, молока, м'яса тощо) ведуть щоденний облік її виходу і т.д.

Проте з деяких причин (великої трудомісткості, тривалості проведення, високої вартості тощо) суцільне спостереження часто буває економічно недоцільним або практично нездійсненним. Тому на практиці переважно застосовують несуцільне спостереження, різновидом якого є вибіркове.

Вибірковим спостереженням в статистиці називають такий вид спостереження, який дає можливість зробити висновок про всю сукупність одиниць при обстеженні тільки її частини. Прикладом вибіркового спостереження може бути вибіркове обстеження діяльності домогосподарств, визначення втрат урожаю, якості продукції, польові досліди тощо.

Сукупність методів математичної статистики, що застосовуються для обґрунтувань та висновків при проведенні вибіркового спостереження, називають **вибірковим методом**.

Вибіркове спостереження є одним з видів несуцільного спостереження, що найбільш широко застосовується. Вибірковими даними користуються досить широко в різних сферах суспільного життя. Наприклад, для оцінки якості сільськогосподарської продукції (зерна, молока тощо) немає необхідності в дослідженні всього обсягу продукції, для цього досить лише взяти певну кількість проб.

Вибіркове дослідження одержало широке поширення в державній і відомчій статистиці. В статистичній практиці вибіркове спостереження застосовується для обстеження домогосподарств населення, житлових умов сімей робітників і службовців, їх заробітної плати, цін на ринках, в науковій роботі при статистичній обробці результатів дослідження, для вивчення і контролю якості продукції, громадської думки тощо.

В сільському господарстві вибіркове спостереження застосовується для визначення якості продукції, втрат урожаю, засміченості посівів, схожості, вологості, засміченості зерна, використання робочого часу, контрольних перевірок перепису худоби тощо.

При дотриманні правил наукової організації обстеження вибіркове спостереження дає досить точні результати, тому його часто застосовують для уточнення даних суцільного обліку. Так, при проведенні перепису населення організовують вибіркові контрольні обходи для перевірки правильності записів суцільного спостереження, під час перепису (обліку) худоби на основі вибіркових даних визначають процент недообліку худоби у населення та ін.

В деяких випадках вибіркове спостереження використовують для дослідження процесів і явищ, суцільне спостереження яких недоцільне внаслідок занадто великого обсягу робіт, наприклад обстеження сімейних бюджетів, реєстрація цін на ринках, вивчення споживчого попиту населення тощо.

У проведенні ряду досліджень вибіркове спостереження є єдино можливим, наприклад при контролі якості продукції, якщо перевірка супроводжується знищенням або псуванням зразків, що обстежуються (визначення жирності молока, схожості зерна, міцності тканин на розрив, тривалості горіння електричних ламп тощо).

Широкому використанню вибіркового спостереження в статистичній практиці сприяє сучасна обчислювальна техніка, яка дає змогу значно ускладнити процедури обробки вихідної інформації, а тому підвищити надійність статистичних показників, що одержують за даними вибірки.

З усіх видів несуцільного спостереження головним є вибіркове спостереження, оскільки тільки дані цього спостереження дають змогу науково обґрунтовано і з необхідною вірогідністю поширити дані, одержані по частині сукупності на всю сукупність.

Науково організоване вибіркове спостереження має ряд суттєвих переваг перед суцільним.

1. Економія трудових, матеріальних ресурсів і коштів за рахунок скорочення обсягів робіт по збиранню, обробці і узагальненні даних. Наприклад, якщо вибірці підлягає 10% загальної чисельності одиниць, то обсяг робіт скорочується порівняно з суцільним обстеженням в 10 разів.

2. Проведення спостереження у стислі строки і за більш широкою програмою, одержання кінцевих результатів дослідження в короткі строки.

3. Зведення до мінімуму псування або навіть знищення досліджуваних зразків при перевірці їх якості.

4. Досягнення більшої точності результатів спостереження завдяки скороченню помилок, що виникають при реєстрації. В процесі проведення вибіркового спостереження необхідно дати правильне уявлення про зведені показники всієї сукупності на основі обстеження її частини за умови дотримання всіх вимог і принципів проведення статистичного спостереження і науково організованої роботи по відбору одиниць.

Сукупність відібраних для обстеження одиниць прийнято називати **вибірковою**, а сукупність одиниць, з якої проводиться відбір, – **генеральною**.

Властивість вибіркової сукупності відтворювати генеральну сукупність дістала назву **репрезентативності**, що означає представництво з певною точністю і вірогідністю. В зв'язку з тим, що при вибірковому спостереженні обстежується тільки частина одиниць генеральної сукупності, то характеристики вибіркової сукупності, як правило, відрізняються від характеристик вибіркової сукупності, тобто мають місце так звані **помилки репрезентативності** (відповідності, відображення).

Тому одним із основних завдань вибіркового методу є одержання таких вибіркових характеристик, які б якомога точніше відтворювали характеристики генеральної сукупності, тобто давали найменші помилки репрезентативності.

Об'єктивну гарантію репрезентативності одержаної вибіркової сукупності дає застосування відповідних науково обґрунтованих способів відбору одиниць, що підлягають обстеженню. В процесі формування вибіркової сукупності має бути забезпечений строго об'єктивний підхід до відбору одиниць, що гарантує рівну можливість кожній одиниці генеральної сукупності потрапити у вибірку. Кількість відібраних при цьому одиниць має бути досить великою.

Мета вибіркового спостереження полягає в тому, що на основі відібраних з генеральної сукупності для обстеження одиниць необхідно дати оцінку невідомим параметрам генеральної сукупності.

Теоретичною основою вибіркового методу є математичні теореми закону великих чисел, викладення яких дається в курсі математичної статистики, і теорія ймовірності.

Великий внесок в розробку теоретичних основ вибіркового методу внесли Я.Бернуллі, С.Пуассон, К.Пірсон, Р.Фішер, В.Госсет, П.Л.Чебишев, О.М.Ляпунов та ін. Нерівність П.Л.Чебишева стосовно вибіркового методу може бути сформульована так: при необмеженому збільшенні числа незалежних спостережень ($n \to \infty$) в генеральній сукупності з обмеженою дисперсією з імовірністю як завгодно близькою до одиниці можна очікувати, що відхилення вибіркової середньої від генеральної середньої буде як завгодно мале, тобто

$$P\left(\left|\tilde{x}-\bar{x}\right|\leq\varepsilon\right)\to 1 \quad \text{при} \quad n\to\infty,$$

де P – імовірність нерівності, що стоїть у дужках; ε – будь-яке завгодно мале додатне число; $\tilde{x}$ і $\bar{x}$ – вибіркова і генеральна середня.

На основі теорем закону великих чисел розв'язуються два взаємопов'язаних і важливих в практичному відношенні питання вибіркового спостереження: розрахунок необхідної чисельності вибірки і визначення помилок вибірки при заданому рівні довірчої імовірності.

Результати вибіркового спостереження характеризуються середніми і відносними узагальнюючими показниками. Узагальнюючі показники генеральної сукупності (середня, частка, дисперсія та ін.) називають **генеральними**, а відповідні узагальнюючі показники вибіркової сукупності – **вибірковими.**

У вибірковому спостереженні прийняті такі позначення. Обсяг генеральної сукупності позначають через N, а вибіркової – через n. Середню величину і дисперсію ознаки в генеральній сукупності називають **генеральною середньою** і **генеральною дисперсією.** Генеральну середню позначають через $\bar{x}$, а генеральну дисперсію – через σ_0^2.

Середню величину і дисперсію ознаки у вибірковій сукупності називають **вибірковою середньою і вибірковою дисперсією.** Вибіркову середню позначають через $\tilde{x}$, а вибіркову дисперсію – через σ^2.

Частку одиниць, що володіють даною ознакою в генеральній сукупності, називають **генеральною часткою.** Відносна величина частки, що одержана в результаті вибіркового спостереження, називається **вибірковою часткою**, або **частістю.** Частість показує, яка частка одиниць вибіркової сукупності володіє досліджуваною ознакою. Генеральну частку позначають через p, а частість – через ω.

Наведемо приклад розрахунку зазначених показників для генеральної і вибіркової сукупності. Припустимо, що з 300 однакових за площею ділянок посіву, але з різною урожайністю ячменю, що представляють генеральну сукупність, відібрано у випадковому порядку 30 ділянок (10% всіх ділянок), що становлять вибіркову сукупність (табл. 6.1).

Визначимо за цими даними середню урожайність, дисперсію і частку ділянок з урожайністю 25 ц/га і більше для генеральної і вибіркової сукупностей:

а) **для генеральної сукупності**

середня урожайність

$$x=\frac{\Sigma xf}{\Sigma f}=\frac{24\cdot 70+25\cdot 150+26\cdot 80}{70+150+80}=\frac{7510}{300}=25{,}0\ \text{ц/га};$$

Таблиця 6.1

Урожайність ячменю і число ділянок генеральної і вибіркової сукупностей

Групи ділянок за урожайністю ячменю, ц/га x	Число ділянок	
	всього (генеральна сукупність) N	в тому числі відібрано (вибіркова сукупність) n
24	70	6
25	150	15
26	80	9
Разом	**300**	**30**

дисперсія

$$\sigma_0^2 = \frac{\Sigma(x-\overline{x})^2 f}{\Sigma f} = \frac{(24-25)^2 70 + (25-25)^2 150 + (26-25)^2 80}{70+150+80} = \frac{150}{300} = 0{,}50;$$

частка ділянок з урожайністю 25 ц/га і більше

$$p = \frac{230}{300} = 0{,}77, \text{ або } 77\%;$$

б) для вибіркової сукупності

середня урожайність

$$\widetilde{x} = \frac{\Sigma xf}{\Sigma f} = \frac{24 \cdot 6 + 25 \cdot 15 + 26 \cdot 9}{6+15+9} = \frac{753}{30} = \mathbf{25{,}1\ ц/га;}$$

дисперсія

$$\sigma^2 = \frac{\Sigma(x-\widetilde{x})^2 f}{\Sigma f} = \frac{(24-25{,}1)^2 6 + (25-25{,}1)^2 15 + (26-25{,}1)^2 9}{6+15+9} = \frac{14{,}70}{30} = 0{,}49;$$

частка ділянок з урожайністю 25 ц/га і більше

$$\omega = \frac{24}{30} = 0{,}80, \text{ або } 80\%.$$

Отже, генеральна середня становить 25,0 ц/га, генеральна дисперсія – 0,50, а генеральна частка – 0,77, відповідно вибіркова середня – 25,1 ц/га, вибіркова дисперсія – 0,49, вибіркова частість – 0,80. Як видно, ті самі показники генеральної і вибіркової сукупностей не збігаються. В абсолютному виразі різниця між середніми становить 0,1 ц/га, між частками – 0,03. Незначні розбіжності між середніми і відносними показниками свідчать про те, що вибіркова сукупність досить точно репрезентує генеральну сукупність.

6.2. Помилки вибірки

Між показниками вибіркової сукупності і шуканими показниками (параметрами) генеральної сукупності, як правило, існують деякі розбіжності, які називають **помилками вибірки.** Загальна помилка вибіркової характеристики складається з помилок двох родів: помилок реєстрації і помилок репрезентативності.

Помилки реєстрації властиві будь-якому статистичному спостереженню і поява їх може бути викликана неуважністю реєстратора, неточністю підрахунків, недосконалістю вимірювальних приладів тощо.

Помилки репрезентативності притаманні тільки вибірковому спостереженню і зумовлені самою його природою оскільки як би ретельно і правильно не проводився відбір одиниць, середні і відносні показники вибіркової сукупності завжди будуть якоюсь мірою відрізнятися від відповідних показників генеральної сукупності.

Розрізняють систематичні та випадкові помилки репрезентативності. **Систематичні помилки репрезентативності** – це неточності, які виникають внаслідок недотримання умов відбору одиниць у вибіркову сукупність, не надання рівної можливості кожній одиниці генеральної сукупності потрапити у вибірку. **Випадкові помилки репрезентативності** – це похибки, які виникають внаслідок того, що вибіркова сукупність точно не відтворює *характеристики генеральної сукупності (середню, частку, дисперсію та ін.)* через несуцільний характер обстеження.

При дотриманні принципу випадкового відбору розмір помилки вибірки *залежить насамперед від чисельності вибірки. Чим більше чисельність* вибірки при інших рівних умовах, тим меншою є величина помилки вибірки. При великій чисельності вибірки виразніше проявляється дія закону великих чисел, згідно з яким: з імовірністю, скільки завгодно близькою до одиниці, можна стверджувати, що при досить великому обсязі вибірки та обмеженій дисперсії вибіркові характеристики (середня, частка) будуть скільки завгодно мало відрізнятися від відповідних генеральних характеристик.

Розміри помилки вибірки також безпосередньо пов'язані зі ступенем варіювання досліджуваної ознаки, а ступінь варіювання, як зазначалося вище, в статистиці характеризується розміром дисперсії (розсіяння): чим менша дисперсія, тим меншою є помилка вибірки, тим надійніші статистичні висновки. Тому на практиці дисперсію ототожнюють з помилкою вибірки.

Оскільки параметр генеральної сукупності є шуканою величиною і він невідомий, потрібно орієнтуватися не на конкретну помилку, а середню з усіх можливих вибірок.

Якщо з генеральної сукупності відібрати кілька вибіркових сукупностей, то кожна із отриманих вибірок дасть різне значення конкретної помилки.

Середня квадратична величина μ, обчислена з усіх можливих значень конкретних помилок (ε_i), становитиме:

$$\mu = \sqrt{\frac{\Sigma \varepsilon_i^2 f_i}{\Sigma f_i}} = \sqrt{\frac{\Sigma(\tilde{x}_i - \bar{x})^2 f_i}{\Sigma f_i}},$$

де $\tilde{x}_i$ – вибіркові середні; $\bar{x}$ – генеральна середня; f_i – чисельність вибірок з величиною $\varepsilon_i = \tilde{x}_i - \bar{x}$.

Середнє квадратичне відхилення вибіркових середніх від генеральної середньої називають **середньою помилкою вибірки.**

Залежність величини помилки вибірки від її чисельності та від ступеня варіювання ознаки знаходить вираження у формулі середньої помилки вибірки μ.

Квадрат середньої помилки (дисперсія вибіркових середніх) прямо пропорційний дисперсії σ_0^2 і обернено пропорційний чисельності вибірки n:

$$\mu^2 = \frac{\sigma_0^2}{n},$$

де σ_0^2 – дисперсія ознаки у генеральній сукупності.

Звідси середню помилку в загальному вигляді визначають за формулою:

$$\mu = \frac{\sigma_0}{\sqrt{n}}.$$

Отже, визначивши за вибіркою середнє квадратичне відхилення, можна встановити значення середньої помилки вибірки, величина якої, як випливає з формули, тим більша, чим більшою є варіація випадкової величини і тим менша, чим більшою є чисельність вибірки.

Тому з мірою зростання обсягу вибірки розмір середньої помилки зменшується. Якщо, наприклад, потрібно зменшити середню помилку вибірки в два рази, то чисельність вибірки слід збільшити в чотири рази, якщо треба зменшити помилку вибірки в три рази, то обсяг вибірки слід збільшити в дев'ять разів і т. д.

У практичних розрахунках застосовують дві формули середньої помилки вибірки: для середньої і для частки.

При вибірковому вивченні середніх показників формула середньої помилки така:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}}.$$

При вивченні відносних показників (часток ознак) формула середньої помилки має вигляд:

$$\mu_p = \sqrt{\frac{p(1-p)}{n}},$$

де p – частка ознаки в генеральній сукупності.

Застосування наведених формул середньої помилки передбачає, що відомі генеральна дисперсія та генеральна частка. Проте в дійсності ці показники невідомі і обчислити їх неможливо через відсутність даних щодо генеральної сукупності. Тому виникає потреба заміни генеральної дисперсії та генеральної частки іншими, близькими до них величинами.

В математичній статистиці доведено, що такими величинами можуть бути вибіркова дисперсія (σ^2) та вибіркова частка (ω).

З урахуванням сказаного формули середньої помилки можуть бути записані так:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}};\ \mu_\omega = \sqrt{\frac{\omega(1-\omega)}{n}}.$$

Ці формули дають змогу визначити середню помилку при повторній вибірці. *Застосування простої випадкової повторної вибірки у практиці* є обмеженим. Насамперед практично недоцільно, а інколи неможливо повторне обстеження тих самих одиниць. Застосування безповторного відбору замість повторного диктується також вимогою підвищення ступеня точності і надійності вибірки. Тому на практиці найчастіше використовують спосіб безповторного випадкового відбору. За цим способом відбору одиниця сукупності, що відібрана у вибірку, в подальшому відборі участі не бере. Одиниці відбирають із генеральної сукупності, зменшеної на кількість раніше відібраних одиниць. Тому в зв'язку зі зміною чисельності генеральної сукупності після кожного відбору та ймовірності відбору для одиниць, що залишились, у формули середньої помилки вибірки вводиться поправочний множник

$$\frac{N-n}{N-1},$$

де N – чисельність генеральної сукупності; n – чисельність вибірки. При досить великому значенні N можна одиницею в знаменнику знехтувати. Тоді

$$\frac{N-n}{N-1} \approx \frac{N-n}{N} \approx 1-\frac{n}{N}.$$

Відтак формули середньої помилки вибірки для безповторного відбору для середньої і для частки відповідно мають вигляд:

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}; \qquad \mu_{\omega} = \sqrt{\frac{\omega(1-\omega)}{n}\left(1 - \frac{n}{N}\right)}.$$

Оскільки n завжди менше N, то додатковий множник $1 - \frac{n}{N}$ завжди менше одиниці. Отже, абсолютне значення помилки вибірки при безповторному відборі завжди менше, чим при повторному.

Якщо чисельність вибірки досить велика, то величина $1 - \frac{n}{N}$ близька до одиниці, а тому нею можна знехтувати. Тоді середню помилку випадкового безповторного відбору визначають за формулою власне випадкової повторної вибірки.

Розрахуємо для нашого прикладу середню помилку для урожайності і для частки ділянок з урожайністю 25 ц/га і більше.

Середня помилка вибірки:

а) середньої урожайності ячменю

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}50}{30}\left(1 - \frac{30}{300}\right)} = \sqrt{0{,}015} = \pm 0{,}12 \text{ ц/га;}$$

б) частки ділянок з урожайністю 25 ц/га і більше

$$\mu_p = \sqrt{\frac{p(1-p)}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}77 \cdot 0{,}23}{30}\left(1 - \frac{30}{300}\right)} = \sqrt{0{,}0053} = \pm 0{,}07.$$

Середня урожайність ячменю в генеральній сукупності $\tilde{x} \pm \mu_{\tilde{x}} = 25{,}1 \pm$ $\pm\, 0{,}12$ ц/га, тобто знаходиться в межах від 24,98 до 25,22 ц/га.

Частка ділянок з урожайністю 25 ц/га і більше в генеральній сукупності $p = \omega \pm \mu_p = 0{,}80 \pm 0{,}07$, тобто знаходиться в межах від 73 до 87%.

Середня помилка вибірки показує можливі відхилення характеристик вибіркової сукупності від характеристик генеральної сукупності. Разом з тим при проведенні вибіркового спостереження перед дослідниками часто стоїть завдання розрахунку не тільки середньої помилки, але і визначення граничної можливої помилки вибірки. Знаючи середню помилку, можна визначити межі, за які не вийде величина помилки вибірки. Однак стверджувати, що ці відхилення не перевищать заданої величини, можна не з абсолютною вірогідністю, а лише з певним ступенем імовірності. Рівень імовірності, що приймається при визначенні можливих меж, в яких містяться значення параметрів генеральної сукупності, називається **довірчим рівнем імовірності.**

Довірча імовірність – це досить висока і така, що практично вважається здійсненою в кожному конкретному випадку, імовірність, що гарантує отримання надійних статистичних висновків. Позначимо її через P, а імовірність перевищити цей рівень – α. Отже, $\alpha = 1 - P$. Імовірність α називають **рівнем значущості** (істотності), який характеризує відносне число помилкових висновків у загальному числі висновків і визначається як різниця між одиницею і довірчою імовірністю, що приймається.

Рівень довірчої імовірності встановлює дослідник, виходячи зі ступеня відповідальності і характеру завдань, що розв'язуються. У статистичних дослідах в сільському господарстві найчастіше приймається рівень довірчої імовірності $P = 0{,}95$; $P = 0{,}99$ (відповідно рівень значущості $\alpha = 0{,}05$; $\alpha = 0{,}01$) рідше $P = 0{,}999$. Наприклад, довірча імовірність $P = 0{,}99$ означає, що помилка оцінки у 99 випадках із 100 не перевищить встановленої величини і тільки в одному випадку із 100 може досягти обчисленого значення або перевищити його.

Помилка вибірки, що обчислена із заданим ступенем надійної імовірності, називається **граничною помилкою вибірки ε_p**.

Розглянемо, як встановлюється величина можливої граничної помилки вибірки. Величина ε_p пов'язана з нормованим відхиленням t, яке визначається як відношення граничної помилки вибірки ε_p до середньої помилки μ:

$$t = \frac{\varepsilon_p}{\mu}.$$

Для зручності розрахунків відхилення випадкової величини від її середнього значення звичайно виражають в одиницях середнього квадратичного відхилення. Вираз

$$\frac{x_i - \bar{x}}{\sigma}$$

називають **нормованим відхиленням.** В статистичній літературі t ще називають **коефіцієнтом довіри,** або **коефіцієнтом кратності середньої помилки вибірки.**

Так, нормоване відхилення для вибіркової середньої можна визначити за формулою:

$$t = \frac{\varepsilon_p}{\mu_{\tilde{x}}} = \frac{\tilde{x} - \bar{x}}{\mu_{\bar{x}}} = \frac{\tilde{x} - \bar{x}}{\sigma}\sqrt{n},$$

де $\mu_{\tilde{x}} = \dfrac{\sigma}{\sqrt{x}}$.

Із виразу $t = \dfrac{\varepsilon_p}{\mu}$ *можна знайти можливу граничну помилку, вибірки* $\varepsilon_p = t\mu$.

Підставивши замість μ її значення, наведемо формули граничних помилок вибірки для середньої і для частки при безповторному випадковому відборі:

Отже, гранична помилка вибірки залежить від величини середньої помилки і нормованого відхилення і дорівнює ± кратному числу середніх помилок вибірки.

Середня і гранична помилки вибірки – величини іменовані і виражаються в тих самих одиницях, що й середня арифметична і середнє квадратичне відхилення.

Нормоване відхилення функціонально пов'язане з імовірністю. Для знаходження значень t складені спеціальні таблиці (дод. 2), за якими можна знайти значення t при заданому рівні довірчої імовірності і значення імовірності при відомому t.

Наведемо значення t та відповідні до них імовірності для вибірок з чисельністю $n \geq 30$, що найчастіше використовується в практичних розрахунках:

t	1,00	1,96	2,00	2,58	3,00
P	0,6827	0,9500	0,9545	0,9901	0,9973

Отже, при $t = 1$ імовірність відхилення вибіркових характеристик від генеральних на величину однократної середньої помилки вибірки дорівнює 0,6827. Це означає, що в середньому із кожної 1000 вибірок 683 дадуть узагальнені характеристики, які відрізнятимуться від генеральних узагальнених характеристик не більше, ніж на величину однократної середньої помилки. При $t = 2$ імовірність дорівнює 0,9545. Це означає, що із кожної 1000 вибірок 954 дадуть узагальнені характеристики, які відрізнятимуться від генеральних узагальнених характеристик не більше, ніж на двократну середню помилку вибірки і т.д.

Однак в зв'язку з тим, що, як правило, проводиться тільки одна вибірка, то ми кажемо, що, наприклад, із імовірністю 0,9545 можна гарантувати, що розміри граничної помилки не перевищать двократну середню помилку вибірки.

Математично доведено, що відношення помилки вибірки до середньої помилки, як правило, не перевищує $\pm 3\mu$ при досить великій чисельності n, незважаючи на те, що помилка вибірки може набувати будь-яких значень. Іншими словами можна сказати, що при досить високій імовірності судження ($P = 0,9973$) гранична помилка вибірки, як правило, не перевищує трьох середніх помилок вибірки. Тому величину $\varepsilon_p = 3\mu$ можна прийняти за межу можливої помилки вибірки.

Визначимо для нашого прикладу граничну помилку вибірки для середньої урожайності і для частки ділянок з урожайністю 25 ц/га і більше. Довірчий рівень імовірності приймемо рівним $P = 0,9545$. За таблицею (дод. 2) знайдемо значення $t = 2$. Середні помилки вибірки для урожайності і частки ділянок з урожайністю 25 ц/га і більше були знайдені раніше і відповідно становили: $\mu_{\tilde{x}} = \pm 0{,}12$ ц/га; $\mu_p = \pm 0{,}07$.

Гранична помилка середньої урожайності ячменю:

$$\varepsilon_{\tilde{x}} = t\mu_{\tilde{x}} = 2 \cdot 0{,}12 = \pm 0{,}24 \text{ ц/га.}$$

Отже, різниця між вибірковою середньою урожайністю і генеральною середньою буде не більше 0,24 ц/га. Межі середньої урожайності в генеральній сукупності: $\bar{x} = \tilde{x} \pm \varepsilon_{\tilde{x}} = 25{,}1 + 0{,}24$, тобто від 24,86 до 25,34 ц/га.

Гранична помилка частки ділянок з урожайністю 25 ц/га і більше:

$$\varepsilon_p = t\mu_p = 2 \cdot 0{,}07 = \pm 0{,}14.$$

Отже, гранична помилка у визначенні частки ділянок з урожайністю 25 ц/га і більше не перевищить 14%, тобто питома вага ділянок із зазначеною урожайністю в генеральній сукупності знаходиться в межах: $p = \omega \pm \varepsilon_p = 0{,}80 \pm 0{,}14$, тобто від 66 до 94%.

6.3. Способи формування вибіркових сукупностей

Результати вибіркового спостереження багато в чому залежать від способів формування та відбору одиниць у вибіркову сукупність. Основним принципом правильності відбору одиниць є строго об'єктивний підхід до відбору одиниць для спостереження. Дотримання цього принципу дає змогу запобігти систематичних (тенденційних) помилок і найбільш точно і повно представити генеральну сукупність.

Попередження систематичних помилок досягається в результаті застосування науково обґрунтованих способів формування вибіркової сукупності.

При формуванні вибіркової сукупності мають бути забезпечені дві умови:

1) *рівні можливості для кожної одиниці генеральної сукупності потрапити* у вибірку (так званий **принцип рівноможливості**);

2) досить представницька чисельність вибіркової сукупності.

В статистиці застосовуються різні види і способи формування вибіркової сукупності. В кожному конкретному випадку залежно від цілого ряду умов, а саме: завдань дослідження, суті досліджуваного явища, специфіки об'єкта, обсягу сукупності, коливання ознаки, наявності матеріальних і трудових ресурсів вибирають найбільш оптимальну систему формування вибіркової сукупності, яка визначається видом і способом відбору.

За видами розрізняють: 1) **індивідуальний відбір** – у вибірку потрапляють окремі одиниці генеральної сукупності; 2) **груповий відбір** – у вибірку потрапляють якісно однорідні групи або серії досліджуваних одиниць; 3) **комбінований відбір** як комбінація індивідуального і групового відбору.

Спосіб відбору визначає конкретний механізм або процедуру вибірки одиниць з генеральної сукупності. В практиці застосування вибіркового спостереження найбільше поширення одержали такі види вибірки: власне-випадкова, механічна, типічна, серійна (гніздова) і комбінована. Ці способи можуть бути застосовані і в поєднанні один з одним.

При **власне-випадковій вибірці** відбір одиниць з генеральної сукупності проводиться без попереднього розчленування її на будь-які групи і одиниця спостереження збігається з обліковою одиницею.

Суть **випадкового відбору** полягає в тому, що кожна одиниця спостереження потрапляє у вибірку випадково – за жеребом. Залежно від способу відбору одиниць розрізняють повторний і безповторний відбір.

При **повторному відборі** (за схемою поверненого шару) кожна одиниця після її реєстрації повертається до генеральної сукупності і знову може бути відібраною. Цей спосіб відбору забезпечує постійність складу генеральної сукупності. Імовірність потрапляння кожної одиниці до вибірки залишається постійною, отже, зберігається незалежність наступного витягання одиниць від попередніх.

При **безповторному відборі** (за схемою неповерненого шару) кожна одиниця після її реєстрації до генеральної сукупності не повертається і в подальшому відборі участі не бере, тобто та сама одиниця не може двічі потрапити до вибірки. Тому безповторна вибірка краще репрезентує генеральну сукупність і, отже, дає меншу помилку, ніж повторна.

На відміну від повторного відбору при безповторному відборі не зберігається постійність генеральної сукупності, а імовірність потрапляння окремих одиниць до вибірки весь час змінюється (для одиниць, що залишилися, вона зростає). У зв'язку з цією особливістю до формул середньої та граничної помилок вибірки вводиться поправочний коефіцієнт, про що вже згадувалось вище.

Щоб позбутися елементів суб'єктивності при відборі одиниць з генеральної сукупності, можна користуватися таблицею випадкових чисел (дод. 7)

Випадковий відбір дає добрі результати в однорідних сукупностях, тобто в тих, де варіація ознаки є незначною. Якщо ж сукупність неоднорідна і складається з різних типів явищ, то необхідно застосувати типову вибірку.

Механічний відбір – це різновидність випадкового відбору. Суть його полягає в тому, що всі одиниці генеральної сукупності розташовуються в певному порядку (за зростанням або зменшенням, за алфавітом, географічним положенням тощо), а потім суто механічно через певний інтервал відбираються одиниці у вибіркову сукупність.

Наприклад, якщо треба відібрати 100 об'єктів із генеральної сукупності чисельністю 1000 одиниць, то величина інтервалу становитиме $h = N : n = 1000 : 100 = 10$, тобто слід відібрати по одній одиниці з кожного десятка. Щоб забезпечити випадковість відбору, доцільно першу вибірку з першого інтервалу провести за жеребом. Якщо відбір починають з третього об'єкта, то до вибірки потраплять 3, 13, 23 і т.д. об'єкти.

Середню і граничні помилки вибірки при механічному відборі розраховують за тими самими формулами, що і для випадкового безповторного відбору, оскільки механічний відбір, як правило, проводиться безповторно.

При **типовому відборі** всю генеральну сукупність попередньо поділяють на типові групи за досліджуваною ознакою, а потім із кожної групи випадковим або механічним способом відбирають необхідну кількість одиниць. При цьому до початку відбору необхідно забезпечити принцип пропорційного представництва кожної групи відповідно до їх чисельності або їх середніх квадратичних відхилень, або дисперсій. Можливий також відбір, пропорційний обом показникам – чисельності одиниць в типових групах і ступеню варіації ознаки. Такий відбір називається **оптимальним**. На практиці найчастіше застосовують вибірку, пропорційну чисельності типових груп.

Розчленування сукупності на типові групи дає можливість усунути вплив міжгрупової (систематичної) варіації на результати вибірки, оскільки у вибірці забезпечується представництво всіх груп, що може не мати місця при випадковому відборі. Отже, самою вибіркою відтворюється (відображується) міжгрупова варіація. Залишається варіація вибіркових даних навколо середніх (внутрішньогрупова або залишкова варіація), яку визначають по кожній із виділених типових груп.

Тому розмір середньої помилки вибірки буде визначатись тільки величиною внутрішньогрупової (залишкової) варіації ($\tilde{\sigma}^2_{\text{в гр}}$) ознаки, яка менша від загальної на величину міжгрупової варіації. Залишкову дисперсію обчислюють як середню арифметичну зважену із середніх квадратів відхилень, які вимірюють залишкову варіацію в кожній групі.

Середня помилка типової вибірки розраховується за формулою:

$$\mu_{\text{типов}} = \sqrt{\frac{\tilde{\sigma}^2_{\text{в гр}}}{n}}.$$

Наведемо формули граничних помилок вибірки для вибіркової середньої і частки для типової вибірки:

для середньої

$$\varepsilon_{\tilde{x}} = t\sqrt{\frac{\sigma^2_{\text{в гр}}}{n}\left(1-\frac{n}{N}\right)};$$

для частки

$$\varepsilon_{\omega} = t\sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}.$$

Типовий відбір дає точніший результат порівняно з випадковим або механічним відбором, оскільки розчленуванням сукупності на типові групи забезпечується потрапляння до вибірки одиниць від усіх виділених груп і типів.

Суть **серійного відбору** полягає в тому, що відбору підлягають не окремі одиниці генеральної сукупності, а цілі серії таких одиниць. У відібраних методом випадкового безповторного або механічного відбору серіях проводять суцільний опис усіх одиниць, що до них увійшли. При цьому загальне число серій, які складають генеральну сукупність, розглядають як її загальну чисельність N_c, а кількість відібраних – як обсяг вибірки n_c.

Оскільки при серійному способі формування вибіркової сукупності кожна серія виступає як самостійна одиниця спостереження, то **варіація** усередині серій (внутрішньосерійна $\bar{\sigma}^2_{вс}$) **при розрахунку середньої помилки має бути виключена. Отже, середня помилка вибірки у цьому випадку залежить тільки від міжсерійної** варіації $\bar{\sigma}^2_{мс}$:

$$\mu_c = \sqrt{\frac{\sigma^2_{мс}}{n}}.$$

Формули граничних помилок вибірки для середньої і для частки при серійному відборі матимуть вигляд:

для середньої **для частки**

$$\varepsilon_{\tilde{x}} = t\sqrt{\frac{\bar{\sigma}^2_{мс}}{n_c}\left(1-\frac{n_c}{N_c}\right)}; \qquad \varepsilon_{\omega} = t\sqrt{\frac{\omega_{мс}(1-\omega_{мс})}{n_c}\left(1-\frac{n_c}{N_c}\right)}.$$

Комбінований відбір. Розглянуті способи вибірки на практиці застосовуються не тільки самостійно, але і в комбінуванні в різних поєднаннях і з різною послідовністю. Так, наприклад, можна комбінувати серійний відбір з власне випадковою вибіркою. При цьому генеральна сукупність спочатку поділяється на серії і відбирається необхідне число серій, а далі у відібраних серіях проводиться випадковий відбір одиниць у вибіркову сукупність. Можлива також комбінація типової і серійної вибірки, коли серії відбираються в установленому порядку з кількох типових груп.

Розрізняють також одноступінчастий і багатоступінчастий способи відбору одиниць у вибіркову сукупність.

При одноступінчастій вибірці кожна відібрана одиниця зразу ж підлягає вивченню. Так обстежують одиниці вибіркової сукупності при власне-випадковій вибірці.

При багатоступінчастій вибірці спочатку проводять відбір з генеральної сукупності окремих груп, а потім з відібраних груп формують вибірку другого, третього і т.д. порядку, яку й аналізують.

В статистичній практиці найбільш широке застосування одержали двоступінчаста і триступінчаста вибірки. Прикладом двоступінчастої вибірки може бути аналіз якості насіння, при якому спочатку відбирають проби з партії насіння, а потім з відібраних проб виділяють наважку для визначення його якості (схожості, чистоти та інших посівних якостей насіння). Прикладом **триступінчастої** вибірки є відбір сімей для бюджетного обстеження, при якому на першій стадії з врахуванням виробничого напрямку відбирають райони, на другій – підприємства, і, насамкінець, на третій – окремі сім'ї.

Багатоступінчаста вибірка дає, як правило, менш точні результати порівняно з одноступінчастою, оскільки її помилки складаються з помилок на окремих ступенях відбору. Однак на практиці, якщо одноступінчасту вибірку організувати складно, використовують багатоступінчасту вибірку.

Кінцевою метою вибіркового спостереження є характеристика генеральної сукупності на основі даних, одержаних за вибіркою, і поширення цих даних на всю сукупність. При цьому вибіркові середні і відносні показники мають бути поширені на генеральну сукупність з урахуванням граничних можливих помилок вибірки.

Існує два способи такого поширення: 1) спосіб прямого перерахування, 2) спосіб поправочних коефіцієнтів.

Суть **способу прямого перерахування** полягає в перемноженні середнього значення ознаки знайденого в результаті вибіркового спостереження на число одиниць генеральної сукупності. Припустимо, що при вивченні шерстної продуктивності овець вибірковим спостереженням в особистих підсобних господарствах населення було встановлено, що середній настриг на вівцю становив 3,0 кг. Знаючи, що в господарствах населення є 2000 овець, можна способом прямого перерахування дістати величину валового настригу вовни: 3,0 · 2000 = 6000 кг, або 6,0 т. Якщо при цьому відомо, що гранична помилка вибірки з імовірністю $P = 0,95$ дорівнює 0,1 кг, і що генеральна середня з такою ж імовірністю коливається в межах від 2,9 до 3,1 кг, то загальний валовий настриг вовни буде коливатись від 5,8 до 6,2 т.

Спосіб поправочних коефіцієнтів застосовується у випадках, коли метою вибіркового спостереження є уточнення результатів суцільного спостереження. При уточненні даних суцільного спостереження на основі вибіркового спостереження визначається так званий поправочний коефіцієнт, яким і користуються для внесення поправок в дані суцільного спостереження. Припустимо, що за даними суцільного спостереження в господарствах населення зареєстровано 1500 корів. Контрольними обходами було охоплено 10% дворів, де зареєстровано 160 корів, а за даними суцільного спостереження числилось 155 корів. Поправочний коефіцієнт становитиме (160 : 155) · 100 = 1,0322, або 103,22%. Таким чином, недооблік корів при суцільному спостереженні становив 3,22%, а фактична чисельність корів в господарствах населення з поправкою на недооблік становить. 1500 · ·1,0322 = 1548 корів.

6.4. Визначення необхідної чисельності вибірки

При організації вибіркового спостереження виникає питання про те, якою повинна бути чисельність вибіркової сукупності, при якій межі можливої помилки не перевищать деякої заздалегідь заданої дослідником величини. Необхідно встановити таку чисельність вибірки, яка з довірчим рівнем імовірності P забезпечувала б одержання даних, що достатньо повно відображають узагальнюючі характеристики генеральної сукупності.

Надто велика вибірка призведе до нераціональних витрат трудових і матеріальних коштів, а недостатня – до великих помилок. Отже, треба встановити оптимальну чисельність вибірки, яка б гарантувала потрібну точність результатів і надійність висновків спостереження.

Необхідна чисельність вибірки залежить від таких факторів.

1. Розміру граничної помилки вибірки ε_p, тобто величини можливих відхилень показників генеральної сукупності від показників вибіркової сукупності. Чим менше розмір заданої граничної помилки, тим більшою має бути чисельність вибірки.

При визначенні необхідної чисельності вибірки гранична помилка вибірки заздалегідь задається самим дослідником залежно від характеру вирішуваних завдань і потрібної точності висновків. На практиці звичайно виходять з того, що гранична помилка вибірки по відношенню до середньої помилки не перевищує 1-5% Іншими словами, цей процент не повинен перевищувати прийнятий довірчий рівень значущості α

2 Ступеня варіації досліджуваної ознаки. Чим більше варіація (дисперсія, коефіцієнт варіації та ін.), тим більшою має бути чисельність вибірки.

3 Рівня довірчої імовірності P, з яким потрібно гарантувати припустимі розміри граничної помилки вибірки. Імовірність у свою чергу пов'язана з нормованим відхиленням t Чим більшим є заданий рівень довірчої імовірності P, тим більше нормоване відхилення t, тим більшою має бути чисельність вибіркової сукупності.

4. Способу відбору одиниць у вибіркову сукупність (повторний або безповторний відбір).

Отже, при визначенні необхідної чисельності вибірки мають бути задані такі умови: а) розмір граничної помилки; б) рівень варіації (дисперсія, коефіцієнт варіації та ін.), в) рівень довірчої імовірності і значення нормованого відхилення, що відповідає їй.

Формули для розрахунку необхідної чисельності вибірки виводяться з формул граничних помилок для середньої і для частки шляхом відповідних алгебраїчних перетворень.

Наведемо формули необхідної чисельності вибірки для різних способів відбору:

а) при визначенні середнього розміру ознаки

$$n=\frac{t^2\sigma^2}{\varepsilon_{\bar{x}}^2}$$ – власне випадкова і механічна повторна вибірка;

$$n=\frac{t^2\sigma^2 N}{\varepsilon_{\bar{x}}^2 N+t^2\sigma^2}$$ – власне випадкова і механічна безповторна вибірка;

$$n=\frac{t^2\tilde{\sigma}^2 N}{\varepsilon_{\bar{x}}^2 N+t^2\tilde{\sigma}^2}$$ – типова безповторна вибірка;

$$n=\frac{t^2\sigma_{м.с}^2 N_c}{\varepsilon_{\bar{x}}^2 N_c+t^2\sigma_{м.с}^2}$$ – серійна безповторна вибірка;

б) при визначенні частки ознаки

$$n=\frac{t^2\omega(1-\omega)}{\varepsilon_{\omega}^2}$$ – власне випадкова і механічна повторна вибірка;

$$n=\frac{t^2\omega(1-\omega)N}{\varepsilon_{\omega}^2 N+t^2\omega(1-\omega)}$$ – власне випадкова і механічна безповторна вибірка;

$$n=\frac{t^2\overline{\omega(1-\omega)}N}{\varepsilon_{\omega}^2 N+t^2\overline{\omega(1-\omega)}}$$ – типова безповторна вибірка;

$$n=\frac{t^2\omega_{м.с}(1-\omega_{м.с})N_c}{\varepsilon_{\omega}^2 N_c+t^2\omega_{м.с}(1-\omega_{м.с})}$$ – серійна безповторна вибірка.

При обчисленні необхідної чисельності вибірки потрібно знати міру коливання досліджуваної ознаки. Однак, дисперсія ознаки або її частка p у генеральній сукупності, як правило, невідомі і визначатимуться лише після проведення вибіркового спостереження. Не знаючи цих величин, не можна визначити необхідну чисельність вибірки.

Труднощі, що виникають, можна розв'язати такими шляхами:

1. Зміст фактичного значення σ_0^2 або p підставляють дані попередніх вибіркових спостережень, які проводилися в аналогічних цілях.

2. Можна провести пробні обстеження на невеликому обсязі вибірки і за даними кількох таких обстежень взяти найбільше значення дисперсії або частки.

3. Невідому величину середнього квадратичного відхилення можна знайти приблизно за величиною розмаху передбачуваної варіації ($R = x_{max} - x_{min}$). Доведено, що з імовірністю P = 0,997 можна стверджувати, що розмах варіації в нормальному розподілі ознаки укладається в 6σ (крайні значення знаходяться на відстані в той або інший бік від середньої величини на 3σ), тобто R = 6σ, звідси σ = 1/6 R.

4. Якщо розрахунок необхідної чисельності вибірки проводиться для альтернативної ознаки і її частка невідома хоча б приблизно, то вона приймається рівною своєму максимальному значенню 0,5 і дає величину дисперсії, яка дорівнює 0,25 ($pq = 0{,}5 \cdot 0{,}5$).

Тоді формули для визначення необхідної чисельності вибірки при повторному і безповторному відборі набудуть відповідно такого вигляду:

$$n = \frac{0{,}25t^2}{\varepsilon_\omega^2} \text{ і } n = \frac{0{,}25t^2N}{\varepsilon_\omega^2 N + 0{,}25t^2}.$$

Нерідко на практиці при визначенні необхідної чисельності вибірки гранична помилка вибірки задається не абсолютною величиною, а величиною відносної помилки, яка виражається в процентах. У цьому випадку і варіація ознаки має бути виражена в процентах, тобто дисперсію ознаки замінюють на коефіцієнт варіації (V).

Чисельність вибірки для випадку, коли гранична помилка задається в процентах, визначається за такими формулами:

повторний відбір — безповторний відбір

$$n = \frac{t^2V^2}{(\varepsilon_p\%)^2}; \qquad n = \frac{t^2V^2N}{(\varepsilon_p\%)^2 N + t^2V^2}.$$

Розглянемо приклади розрахунку необхідної чисельності вибірки при повторній і безповторній вибірці.

Повторний відбір. Приклад. В ТОВ проектується вибіркове визначення жирності молока, що приймається від населення. Загальне поголів'я корів в особистих підсобних господарствах населення становить 180 голів. За проведеними раніше дослідженнями встановлено, що середня жирність молока становить 3,6%, а середнє коливання жирності дорівнює 0,2% (σ = 0,2%).

Потрібно визначити, яке поголів'я корів слід піддати вибірковому обстеженню, щоб визначити середню жирність молока з граничною помилкою 0,1% ($\varepsilon_{\bar{x}}$ = 0,1%). Довірчий рівень імовірності P = 0,9545, якому відповідає нормоване відхилення t = 2 (дод. 2).

При випадковому повторному відборі чисельність вибірки визначається за формулою:

$$n = \frac{t^2\sigma^2}{\varepsilon_{\tilde{x}}^2} = \frac{2^2 \cdot 0{,}2^2}{0{,}1^2} = \frac{4 \cdot 0{,}04}{0{,}01} = 16 \text{ голів.}$$

Отже, вибірковому обстеженню досить піддати 16 корів, щоб з довірчою імовірністю $P = 0{,}9545$ (імовірність помилки в 5 випадках з 100) визначити середню жирність молока з помилкою, що не перевищує 0,1%.

Для одержання більш високої гарантії результатів вибіркового спостереження можна збільшити точність вибірки, тобто зменшити розміри припустимої граничної помилки. Так, якщо граничну помилку вибірки зменшити в два рази з 0,1 до 0,05%, то при тих самих умовах $\sigma = 0{,}2\%$ і $t = 2{,}0$ чисельність вибірки становитиме:

$$n = \frac{t^2\sigma^2}{\varepsilon_{\tilde{x}}^2} = \frac{2^2 \cdot 0{,}2^2}{0{,}05^2} = \frac{0{,}16}{0{,}0025} = 64 \text{ голови.}$$

Отже, збільшення точності вибірки у два рази призводить до зростання чисельності вибірки в 4 рази ($16 \cdot 4 = 64$).

Безповторний відбір. Приклад. На свинофермі із загальним поголів'ям поросят у віці 2-4 місяці $N = 2000$ голів необхідно визначити середню живу масу однієї голови в кінці місяця з точністю 1 кг ($\varepsilon_{\tilde{x}} = 1$ кг). Довірчий рівень імовірності $P = 0{,}9545$, якому відповідає $t = 2$. За даними зважування в минулому місяці (або пробному в даному) встановлено, що коливання живої маси в стаді становить 4 кг ($\sigma = 4$ кг).

Необхідно встановити, яку чисельність поросят слід відібрати для зважування, щоб середня жива маса була одержана з точністю до 1 кг.

Так як зважування проводиться безповторно, то

$$n = \frac{t^2\sigma^2 N}{\varepsilon_{\tilde{x}}^2 N + t^2\sigma^2} = \frac{2^2 \cdot 4^2 \cdot 2000}{1^2 \cdot 2000 + 2^2 \cdot 4^2} = \frac{128000}{2064} = 62 \text{ голови.}$$

Отже, досить зважити 62 голови, щоб без великих витрат і швидко одержати відповідь із заданою, практично вірогідною точністю.

6.5. Поняття статистичної оцінки. Точкова та інтервальна оцінка параметрів генеральної сукупності

Оскільки всі елементи генеральної сукупності для обчислення шуканого параметра, як правило, використати неможливо, то про цей параметр намагаються судити за даними однієї або кількох вибірок із генеральної сукупності.

Наближене значення шуканого параметра генеральної сукупності, встановлене за даними вибіркової сукупності, називають **вибірковою оцінкою параметра.**

Якщо шуканий параметр генеральної сукупності позначити через θ, а значення вибіркової характеристики – через $\tilde{\theta}$, то характеристика $\tilde{\theta}$ в даному випадку виступає як оцінка параметра генеральної сукупності θ.

У зв'язку з тим, що значення вибіркових характеристик встановлюються за даними випадкових вибірок, то і самі оцінки є випадковими величинами.

Оцінка параметрів є одним із центральних завдань математичної статистики і являє собою сукупність методів, які дозволяють робити науково обґрунтовані висновки щодо параметрів генеральної сукупності за даними випадкової вибірки з неї.

Оцінкою генеральної середньої (математичного сподівання) може виступати вибіркова середня, генеральної частки – вибіркова частка, генеральної дисперсії – вибіркова дисперсія тощо.

Для того щоб статистичні оцінки давали найкращі та добрі наближення оцінюваних параметрів, вони повинні володіти певними властивостями і задовольняти певним вимогам. Основними властивостями оцінок є властивості незміщеності, спроможності, ефективності і достатності.

Незміщеною називають статистичну оцінку $\tilde{\theta}$, математичне сподівання якої дорівнює оцінюваному параметру θ при будь-якому обсязі вибірки, тобто якщо вона задовольняє рівності

$$M(\tilde{\theta}) = \theta, \text{ або } M(\tilde{\theta}) - \theta = 0.$$

Оцінка називається **зміщеною**, якщо її математичне сподівання не дорівнює оцінюваному параметру, тобто $M(\tilde{\theta}) \neq \theta$.

Оцінка $\tilde{\theta}$ параметра θ називається **спроможною**, якщо вона підпорядковується закону великих чисел, тобто при $n \to \infty$ наближається за імовірністю до шуканого параметра:

$$\lim P\left(\left|\tilde{\theta} - \theta\right| < \varepsilon\right) = 1.$$

Спроможність оцінки означає, що чим більше обсяг вибірки, тим більша імовірність того, що помилка оцінки не перевищить скільки завгодно малого додатного числа ε.

Ефективною називають таку незміщену оцінку $\tilde{\theta}$, яка має найменшу дисперсію серед усіх можливих незміщених оцінок параметра θ, обчисленого за вибірками одного і того ж обсягу

Оцінка $\tilde{\theta}$ називається **достатньою (вичерпною)**, якщо вона включає всю інформацію, яка міститься у вибірці відносно шуканого параметра $\tilde{\theta}$.

Основними методами отримання оцінок параметрів генеральної сукупності за даними вибірки є методи моментів, аналогії, найменших квадратів, максимальної правдоподібності та ін.

З усіх перерахованих методів найбільш широко застосовується **метод максимальної (найбільшої) правдоподібності.** Суть цього методу, розробленого видатним англійським математиком-статистиком Р.Й. Фішером, полягає в тому, що із можливих оцінок параметра θ вибирається та, якій відповідає найбільша імовірність, тобто те значення, яке обертає функцію в максимум. Вибіркова оцінка $\tilde{\theta}$, яка обертає в максимум функцію правдоподібності, називається **оцінкою максимуму правдоподібності.**

Метод максимуму правдоподібності дозволяє одержати спроможні, ефективні, достатні та незначно зміщені оцінки.

Використовуючи метод максимуму правдоподібності, можна довести, що вибіркова середня арифметична є незміщеною, спроможною, ефективною і достатньою оцінкою генеральної середньої, а вибіркова дисперсія є найкращою оцінкою генеральної дисперсії. Цим значною мірою і пояснюється перевага, яка надається даним характеристикам порівняно з усіма іншими вибірковими характеристиками.

В теоретичному курсі математичної статистики доводиться, що математичне сподівання вибіркової дисперсії не дорівнює дисперсії генеральної сукупності. Тому вибіркова дисперсія є зміщеною оцінкою генеральної дисперсії.

Для отримання незміщеної оцінки дисперсії генеральної сукупності необхідно вибіркову дисперсію (σ^2) помножити на так звану поправку Бесселя $\frac{n}{n-1}$. Тоді виправлена, або скорегована, дисперсія (S^2) може бути визначена за формулою

$$S^2 = \sigma^2 \frac{n}{n-1} = \frac{\Sigma(x_i - \bar{x})^2}{n} \cdot \frac{n}{n-1} = \frac{\Sigma(x_i - \bar{x})^2}{n-1}.$$

Математичне сподівання такої виправленої вибіркової дисперсії при будь-якому обсязі вибірки дорівнює генеральній дисперсії. При достатньо великій чисельності вибірки поправка Бесселя $\frac{n}{n-1}$, яка корегує зміщення вибіркової дисперсії, практично дорівнює одиниці і тому нею можна знехтувати. При $n > 30$ (великі вибірки) практично немає різниці між оцінками σ^2 і S^2. При малих же значеннях ($n < 30$; малі вибірки) поправочний коефіцієнт значно відрізняється від одиниці. Тому при малому обсязі вибірки *завжди потрібно користуватися незміщеною оцінкою дисперсії S^2*. Можна довести, що оцінки σ^2 і S^2 є спроможними оцінками генеральної дисперсії σ_0^2.

Оцінка невідомого параметра генеральної сукупності може бути проведена двояко або одним числом (точкою) – **точкова оцінка**, або із зазначенням інтервалу, в якому із заданою імовірністю може знаходитись шуканий параметр, – **інтервальна оцінка.**

Суть точкової оцінки полягає в тому, що за найкращу оцінку шуканого параметра генеральної сукупності θ приймається знайдене за вибіркою його конкретне числове значення $\tilde{\theta}$, тобто приймається припущення, що $\theta = \tilde{\theta}$.

Оскільки сама вибіркова оцінка є випадковою величиною, а статистичні висновки в зв'язку з цим мають імовірнісний характер, то конкретна числова характеристика (точка) обов'язково повинна бути доповнена величиною середньої помилки (μ). Розміри помилки оцінки безпосередньо пов'язані з величиною її дисперсії (розсіювання)· чим менше дисперсія, тим менше помилка оцінки, тим надійніше статистичні висновки Тому дисперсію на практиці ототожнюють з помилкою оцінки, а середньоквадратичне відхилення вибіркової оцінки називають **середньою помилкою.**

Середню помилку оцінки в загальному вигляді визначають за формулою

$$\mu_{\tilde{\theta}} = \sqrt{D(\tilde{\theta})}$$

Квадрат середньої помилки (дисперсія вибіркових середніх) прямо пропорційний дисперсії σ^2 і обернено пропорційний чисельності вибірки n:

$$\mu_{\tilde{\theta}}^2 = \frac{\sigma^2}{n},$$

звідки формула для визначення середньої помилки оцінки прийме вигляд

$$\mu_{\tilde{\theta}} = \frac{\sigma}{\sqrt{n}}$$

Отже, визначивши за вибіркою середнє квадратичне відхилення, можна встановити значення середньої помилки оцінки, величина якої, як випливає із формули, тим більша, чим більша варіація випадкової величини і тим менша, чим більша чисельність вибірки

Результати точкової оцінки шуканого параметра генеральної сукупності можна записати таким чином за статистичну оцінку параметра генеральної сукупності приймається його вибіркове значення ($\theta = \tilde{\theta}$) з середньою помилкою $\pm\mu_{\tilde{\theta}}$

При невеликому обсязі вибірки точкова оцінка значною мірою випадкова і малоефективна і тому може істотно відрізнятися від параметра генеральної сукупності, тобто призводити до великих відхилень $\theta - \tilde{\theta}$ З цієї причини при невеликому обсязі вибірки доцільно користуватися інтервальною оцінкою.

Інтервальною називають оцінку, яка визначається двома числами – кінцями інтервалу, в якому із заданою імовірністю знаходиться шуканий параметр. Центром такого інтервалу, як правило, беруть знайдену вибіркову оцінку точки, а визначення самих кінців інтервалу пов'язується з середньою помилкою оцінки і довірчою імовірністю. Отже, інтервальна оцінка є подальшим доповненням і розширенням точкової оцінки параметра θ.

Встановивши довірчу імовірність, можна побудувати довірчий інтервал. **Довірчим інтервалом** для параметра θ називається такий інтервал, відносно якого можна із заздалегідь встановленою довірчою імовірністю $P = 1 - \alpha$, близькою до одиниці, стверджувати, що він містить невідоме значення параметра θ. Іншими словами, це інтервал, який покриває невідомий параметр θ із заданою імовірністю P.

Для побудови довірчого інтервалу необхідно вказати таке граничне значення помилки $\varepsilon_p = (\theta - \tilde{\theta})$, щоб імовірність її перевищення була не більше α, тобто

$$P(|\tilde{\theta} - \theta| < \varepsilon) = 1 - \alpha.$$

Отже, інтервальна оцінка параметра $\theta = \tilde{\theta} \pm \varepsilon_p$, а довірчий інтервал ($I_p$) має вигляд:

$$I_p = (\tilde{\theta} - \varepsilon_p;\ \tilde{\theta} + \varepsilon_p),$$

де $\tilde{\theta} - \varepsilon_p$ – початок інтервалу; $\tilde{\theta} + \varepsilon_p$ – кінець інтервалу.

Наприклад, шуканий довірчий інтервал для оцінки генеральної середньої матиме вигляд:

$$I_p(\tilde{x} - t\mu_{\tilde{x}};\ \tilde{x} + t\mu_{\tilde{x}}),$$

де $t\mu = \varepsilon_p$ – гранична помилка оцінки.

Для побудови надійного інтервалу спочатку необхідно визначити помилку вибірки, а потім за таблицями значень функції Лапласа (дод. 2) при заданому рівні імовірності знайти значення t.

Визначивши значення вибіркової оцінки і середньої помилки вибірки, можна при заданому рівні імовірності або відомому нормованому відхиленні (аналогічно побудові надійного інтервалу для оцінки генеральної середньої) побудувати довірчий інтервал і для оцінок інших вибіркових характеристик розподілів (наприклад, моди, медіани, дисперсії, частки та ін.).

Загальний вид довірчого інтервалу такий:

$$I_p(\tilde{\theta} - t\mu_{\tilde{\theta}};\ \tilde{\theta} + t\mu_{\tilde{\theta}}),$$

де $\mu_{\tilde{\theta}}$ – середня помилка оцінки, що використовується, і визначається як $\sqrt{D(\tilde{\theta})}$.

Ширина довірчого інтервалу безпосередньо залежить від величини граничної помилки, яка в свою чергу залежить від величини дисперсії (розсіювання) оцінки. Отже, чим менше дисперсія вибіркової оцінки, тим вужче довірчий інтервал, тим точніше і надійніше статистичні висновки.

В зв'язку з цим основна організаційна робота вибіркового спостереження полягає в тому, щоб прийняти міри, спрямовані на зменшення дисперсії. Дисперсія оцінки суттєво залежить від способів формування і відбору одиниць у вибіркову сукупність.

Проведемо точкову та інтервальну оцінку генеральної середньої за даними великої вибірки на такому прикладі. Є дані щодо стажу роботи 30 трактористів: 2? 5; 15? 7; 18; 20; 9; 6; 18; 15; 4; 16; 25; 8; 30; 1; 26; 20; 21; 6; 35; 30; 18; 26; 31; 3; 24; 32; 17; 22. Довірчий рівень імовірності $P = 0,9545$, якому відповідає $t = 2$.

Для визначення середньої помилки вибірки визначимо середню арифметичну і дисперсію.

Середній стаж роботи трактористів

$$\bar{x} = \frac{\Sigma x}{n} = \frac{510}{30} = 17 \text{ років.}$$

Вибіркова дисперсія

$$\sigma^2 = \frac{\Sigma x^2}{n} - \bar{x}^2 = \frac{11576}{30} - 17^2 = 385,87 - 289 = 96,87.$$

Скорегована дисперсія

$$S^2 = \sigma^2 \frac{n}{n-1} = 96,87 \frac{30}{30-1} = 100,21.$$

Незміщена оцінка дисперсії може бути визначена і за **іншою формулою:**

$$S^2 = \frac{\Sigma x^2 - n\bar{x}^2}{n-1} = \frac{11576 - 30 \cdot 17^2}{30-1} = 100,21.$$

Середня помилка вибіркової середньої

$$\mu_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{100,21}{30}} = \sqrt{3,34} = 1,83 \text{ року.}$$

Проведемо точкову оцінку середнього стажу роботи **в генеральній сукупності:**

$$\tilde{x} = \bar{x} = 17 \text{ років при } \mu_{\bar{x}} = 1,83 \text{ року.}$$

тобто точкова оцінка генеральної середньої може бути записана так:

$$\tilde{x} \pm \mu_{\tilde{x}} = 17 \pm 1{,}83 \text{ року.}$$

Це означає, що $\tilde{x} = 17$ років є оцінкою генеральної середньої з помилкою, що дорівнює 1,83 року.

Для проведення інтервальної оцінки і побудови довірчого інтервалу визначимо граничну помилку вибіркової середньої при P = 0,9545 і t = 2.

Гранична помилка вибіркової середньої

$$\varepsilon_{\tilde{x}} = t\mu_{\tilde{x}} = 2 \cdot 1{,}83 = 3{,}66 \text{ року.}$$

Побудуємо довірчий інтервал, в якому із заданим рівнем імовірності знаходиться середній стаж роботи трактористів в генеральній сукупності:

$$\tilde{x} - \varepsilon_{\tilde{x}} < \bar{x} < \tilde{x} + \varepsilon_{\tilde{x}};$$

$$17 - 3{,}66 < \bar{x} < 17 + 3{,}66;$$

$$13{,}34 < \bar{x} < 20{,}66.$$

Таким чином, довірчі межі інтервалу

$$I_{0{,}9545} = (13{,}34;\ 20{,}66),$$

що можна записати так: $\bar{x} = 17 \pm 3{,}66$ року.

Отже, з довірчою імовірністю P = 0,9545 можна стверджувати, що середній стаж роботи трактористів у генеральній сукупності знаходиться в інтервалі 13,34-20,66 року.

Точкова і інтервальна оцінка генеральної середньої в малих вибірках (n < 30) проводиться аналогічно оцінці у великих вибірках лише з тією різницею, що при визначенні граничної помилки замість t-критерію нормального розподілу використовується t-критерій Стьюдента (дод. 3).

6.6. Закони розподілу вибіркових характеристик

Під **законом розподілу** слід розуміти такий теоретичний розподіл, до якого прямує емпіричний розподіл при $n \to \infty$.

В статистиці широко використовуються різні види теоретичних розподілів, серед яких класичними вважаються нормальне, біноміальне і пуассонове. Серед названих законів розподілу, на якому грунтується більшість статистичних методів дослідження, є закон нормального розподілу.

Окремі закони пов'язані з характером розподілу окремих випадкових величин і застосовуються для розв'язання конкретних задач. Ці закони носять імена вчених, які їх відкрили. Серед них у статистичній науці і практиці найбільш широке застосування одержали закони розподілу Стьюдента, Пірсона і Фішера–Снедекора.

Кожний із законів розподілу має свою специфіку і область застосування в різних галузях знання.

Закони розподілу в основному використовуються для розв'язування задач, пов'язаних з оцінкою параметрів генеральної сукупності і перевіркою статистичних гіпотез.

Розглянемо закони розподілу, що одержали в статистичному аналізі найбільше застосування.

Нормальний розподіл. Більшість соціально-економічних і природних явищ підпорядковано закону нормального розподілу. Підпорядкованість закону нормального розподілу проявляється тим точніше, чим більше випадкових величин діє разом. Якщо жодна з випадково діючих причин за своєю дією не виявиться переважною над іншими, то закон розподілу дуже близько підходить до нормального.

Така закономірність проявляється, наприклад, в розподілі відхилень у виробничому процесі при нормальному рівні організації і технології, в розподілі населення певного віку за розміром взуття, одягу і в багатьох інших випадках.

Нормальний розподіл є симетричним розподілом, в якому більшість значень випадкової величини концентрується навколо середньої величини. Його особливістю є те, що чим більше значення окремих варіант відхиляються від середньої величини, тим рідше вони зустрічаються і тим менше імовірність їх появи. І навпаки, чим ближче варіанти до середнього значення, тим частіше вони зустрічаються і тим більше імовірність їх появи. Однакові за абсолютним значенням, але протилежні за знаком відхилення значень змінної x від середньої рівноімовірні.

Імовірність відхилень вибіркових середніх від генеральної середньої $(\tilde{x}-\bar{x})$ при великому числі спостережень $(n \to \infty)$ визначається законом нормального розподілу Лапласа–Гаусса.

Нормальним розподілом називають розподіл неперервної випадкової величини, який описується щільністю імовірності

$$\Phi(x)=\frac{1}{\sigma_0\sqrt{2\pi}}e^{-\frac{(x_i-\bar{x})^2}{2\sigma^2}},$$

де $\Phi(x)$ – щільність імовірності (ордината кривої); σ_0 – середнє квадратичне відхилення генеральної середньої, яке у практичних розрахунках заміняється вибірковим σ; π = 3,14 ... (постійна величина, яка характеризує відношення довжини кола до довжини його діаметра); e = 2,718... – основа натуральних логарифмів (число Ейлера).

Як видно, нормальний розподіл визначається двома параметрами: середньою арифметичною і середнім квадратичним відхиленням. Знаючи ці параметри, можна побудувати криву нормального розподілу.

Звичайно $\frac{x_i - \bar{x}}{\sigma}$ в даній формулі замінюється на *t*, де відхилення представлені в частках середнього квадратичного відхилення, прирівняного до одиниці. Завдяки нормуванню, дисперсія $t = 1$, а $\bar{x} = 0$.

Рівняння нормальної кривої при такій заміні приймає вигляд:

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

Його називають **стандартним рівнянням нормальної кривої**, а нормальну криву – **нормованою кривою.** При її побудові за емпіричними даними застосовують таку формулу:

$$y = \frac{hn}{\sigma} f(t),$$

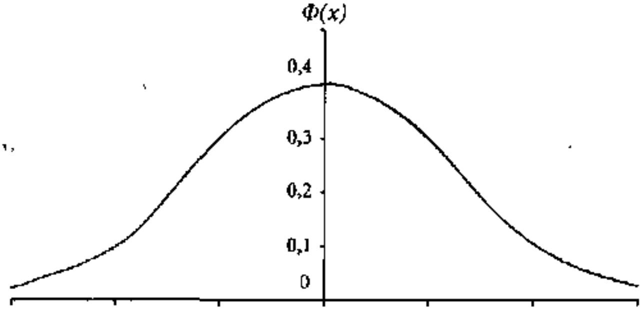
де *y* – ордината кривої (теоретична частота); *h* – величина інтервалу; *n* – чисельність сукупності; σ – середнє квадратичне відхилення; *f(t)* – функція щільності нормального розподілу.

Графік щільності нормального розподілу називається **нормальною кривою**, або кривою Гаусса (рис. 6.1).

Ця дзвоноподібна крива симетрична відносно осі ординат і асимптотично наближається до осі абсцис. Крива має точки перегину при $t = \pm 1$, тобто при таких відхиленнях значень ознаки від середньої арифметичної, які дорівнюють одному середньому квадратичному відхиленню. Площа, що обмежена кривою і віссю абсцис, дорівнює одиниці. Значення щільності імовірності $\Phi(x)$ залежить тільки від величини нормованого відхилення *t*, так як π і *e* – постійні величини. Так, при $t = 0$ співмножник $e^{-\frac{t^2}{2}} = 1$ і щільність імовірності максимальна $\Phi(0) = 0{,}3989$. По мірі зростання *t* щільність імовірності зменшується.

Для знаходження значень інтеграла імовірностей при заданому *t* складені спеціальні таблиці (дод. 2), за якими можна визначити значення *t* при заданому рівні імовірності *P* і значення *P* імовірності при відомому *t*.

Теоретичні значення *t* і *P*, що обчислені на основі стандартного рівняння нормальної кривої, використовуються в математичній статистиці, зокрема, у вибірковому методі, як нормативи (критерії), за допомогою яких проводиться оцінка вибіркових характеристик. В зв'язку з цим нормоване відхилення кривої нормального розподілу отримало назву *t*-критерію розподілу нормальної кривої.



Можливі відхилення вибіркової середньої від генеральної, що виражені в σ або t	-3σ	-2σ	$-\sigma$	$\bar{x}$	σ	2σ	3σ
	-3	-2	-1	0	1	2	3

Рис. 6.1. Крива нормального розподілу імовірностей

Розподіл Стьюдента. Теоретичні положення по оцінці вибіркових характеристик на основі малих вибірок ($n < 30$) вперше (1908 р.) розробив англійський математик-статистик В.Госсет (що друкував свої роботи під псевдонімом Стьюдент). Пізніше (1925 р.) Р.Фішер дав більш строге доведення цього розподілу, яке дістало назву t-розподілу Стьюдента.

Відхилення вибіркових середніх від генеральної середньої Стьюдент виразив в одиницях стандартного відхилення

$$t = \frac{\tilde{x}_i - \bar{x}}{S},$$

де в знаменнику використовується середнє квадратичне відхилення вибірки, тоді як в нормальному розподілі – середнє квадратичне відхилення генеральної сукупності (σ_0). Р.Фішер виразив ці відхилення в одиницях стандартної помилки

$$t = \frac{\bar{x}_i - \bar{x}}{S : \sqrt{n}},$$

де $\mu_{\bar{x}} = \dfrac{S}{\sqrt{n}}$ – середня помилка в малих вибірках.

Середнє квадратичне відхилення в малих вибірках визначається з врахуванням числа ступенів свободи варіації $(n-1)$:

$$S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}.$$

Теоретичний t-розподіл Стьюдента не залежить від параметрів генеральної сукупності, він пов'язаний тільки з величинами, що визначаються безпосередньо за даними вибірки.

В літературі з математичної статистики доводиться, що диференціальна функція t-розподілу Стьюдента (щільність розподілу імовірностей) має вигляд:

$$\varphi(t,n) = A\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}},$$

де A – величина, яка визначається з врахуванням числа ступенів свободи варіації $(k = n - 1)$ за допомогою гамма-функції (Γ-функції):

$$A = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)},$$

де $\Gamma(n) = \int\limits_0^{\infty} x^{n-1} e^{-x} dx$ – гамма-функція.

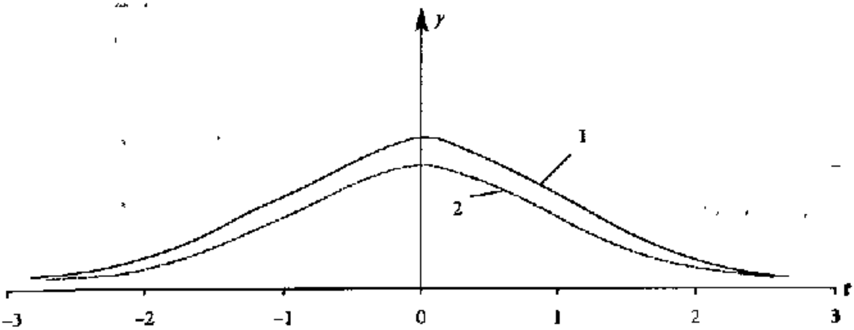
Як видно, величина A залежить тільки від обсягу вибірки і відповідає максимальній ординаті кривої розподілу при $t = 0$. Імовірність того, що помилка вибірки буде не більше заданої величини $\varepsilon_p = t\mu$, визначається інтегральною функцією

$$S(t,n) = \int\limits_{-\infty}^{t} \varphi(t,n)dt = A\int\limits_{-\infty}^{t}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt.$$

Інакше кажучи, $S(t, n) = P(t_{табл} < t_{факт})$, де $t_{табл}$ і $t_{факт}$ – табличне (теоретичне) і фактичне значення нормованого відхилення.

t-розподіл Стьюдента справедливий тільки для вибірок, взятих із генеральної сукупності з нормальним розподілом випадкової величини.

На рис. 6.2 порівнюється крива t-розподілу Стьюдента з кривою нормального розподілу.



1 – нормальний розподіл; 2 – t-розподіл Стьюдента

Рис. 6.2. Порівняння t-розподілу Стьюдента і нормального розподілу

Крива t-розподілу Стьюдента симетрична відносно осі ординат. На відміну від нормального розподілу під кінцями кривої t-розподілу Стьюдента при тих же значеннях t розміщена значно більша частина площі. Таким чином, *на частку більших відхилень від генеральної середньої припадає значна частина площі.* Це означає, що для малих вибірок імовірність допущення більших помилок суттєво підвищується.

При збільшенні обсягу вибірки t-розподіл Стьюдента наближається до нормального розподілу (практично вважається достатнім $n \geq 30$), а при $n \to \infty$ він стає нормальним.

Для визначення значень функції S (t, n) розподілу Стьюдента складено ряд спеціальних таблиць, в яких наводяться розрахункові значення S (t, n) при відповідному числі ступенів свободи варіації. За цими таблицями можна знайти імовірність помилки вибірки при заданому значенні нормованого відхилення t або значення t при заданому рівні імовірності судження P.

Наведемо витяг з таблиці імовірностей S (t, n) для значень n і t, які найбільш часто застосовуються (табл. 6.2).

Як видно з даних таблиці, імовірність розходження між вибірковою середньою малої вибірки і генеральною середньою залежить від двох величин: чисельності вибірки n і нормованого відхилення t. Можна побачити, що при збільшенні n цей розподіл прямує до нормального і при $n = 20$ вже мало від нього відрізняється. При $n \to \infty$ в таблиці наведені значення для функції нормального розподілу.

Таблиця 6.2

Витяг з таблиці значень функції $S(t, n)$ розподілу Стьюдента (імовірності помножені на 1000)

t \ n	4	5	6	7	8	9	10	12	15	20	∞
1	608	626	636	644	650	654	656	662	666	670	683
2	860	884	902	908	914	920	924	930	936	940	954
3	942	960	970	976	980	984	984	988	992	992	997

Інший аспект застосування розподілу Стьюдента наведений в додатку 3, в якому подано значення t-критерію Стьюдента при різному рівні значущості (α) і числі ступенів свободи варіації (k).

Розподіл Пірсона. Для оцінки розходжень між емпіричними і теоретичними частотами розроблено ряд критеріїв згоди, серед яких найбільш *широке застосування отримав критерій* χ^2 *– хі-квадрат.* На основі зіставлення фактичного і теоретичного (табличного) значення χ^2-критерію можна виявити належність даного емпіричного розподілу до деякого відомого теоретичного типу розподілу (наприклад, є або ні досліджуваний розподіл нормальним, біноміальним та ін.).

Крива, що характеризує розподіл χ^2, описується рівнянням

$$f(\chi^2) = \frac{1}{\Gamma\left(\frac{k}{2}\right)2^{\frac{k}{2}}}(\chi^2)^{\frac{k-2}{2}} \cdot e^{-\frac{\chi^2}{2}},$$

де k – число ступенів свободи варіації.

Враховуючи, що для цілих додатних чисел гамма-функція $\Gamma(n) = n - 1$, можна записати:

$$f(\chi^2) = \frac{1}{\left(\frac{k}{2} - 1\right)} \cdot \frac{1}{2^{\frac{k}{2}}}(\chi^2)^{\frac{k-2}{2}} \cdot e^{-\frac{\chi^2}{2}}.$$

З рівняння щільності імовірності видно, що розподіл χ^2 залежить тільки від числа ступенів свободи варіації ($k = n - 1$). Розподіл χ^2 не залежить від генеральної середньої і генеральної дисперсії. При великій чисельності вибірки (при $n = 30 - 40$) розподіл χ^2 практично стає нормальним.

Для χ^2-критерію складені спеціальні таблиці, в яких наведені його значення при певному числі ступенів свободи варіації і заданому рівні імовірності (дод. 6).

Викладення аспектів прикладного застосування χ^2-критерію дається в розділі, присвяченому питанням перевірки статистичних гіпотез (розділ 7).

Розподіл Фішера–Снедекора. При розв'язуванні ряду задач кореляційно-регресійного і дисперсійного аналізу використовується розподіл F, названий так по першій літері прізвища англійського математика-статистика Р.Фішера.

Якщо U і V – незалежні випадкові величини, розподілені за законом χ^2 зі ступенями свободи k_1 і k_2, то величина

$$F = \frac{U}{k_1} : \frac{V}{k_2}$$

підпорядковується розподілу F Фішера–Снедекора зі ступенями свободи k_1 і k_2. Приймаючи, що $U > V$, величина F буде мати значення не менше одиниці.

Щільність розподілу F має вигляд:

$$f(x) = \frac{\Gamma\left(\frac{k_1 + k_2}{2}\left(\frac{k_2}{k_1}\right)\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} x^{\frac{k_2 - 2}{2}} \left(1 + \frac{k_2}{k_1} x\right)^{-\frac{k_1 + k_2}{2}}.$$

З формули видно, що розподіл F визначається двома параметрами, тобто числами ступенів свободи варіації k_1 і k_2. Це дає змогу скласти таблиці розподілу випадкової величини F, в яких різним значенням рівня значущості і різним сполученням величин k_1 і k_2 відповідають певні значення F-критерію (дод. 4 і 5).

Використання F-критерію в статистичному аналізі детально розглянуто в розділі 8 «Дисперсійний аналіз».

6.7. Малі вибірки

Розглянуті вище прийоми розрахунку характеристик вибіркової сукупності (дисперсії, середньої і граничної помилок тощо) передбачають досить велику чисельність вибірки ($n > 30$). В той же час не завжди можливий і доцільний великий обсяг вибірки. У практиці виробничих спостережень та в науково-дослідній роботі часто доводиться користуватися невеликими за обсягом вибірками, чисельність яких не перевищує 30 одиниць (агрономічні і зоотехнічні досліди, перевірка якості продукції, пов'язана зі знищенням зразків тощо). В статистиці вони дістали назву **малих вибірок.** Відповідно вибірки з чисельністю більше 30 одиниць називають **великими вибірками.**

Невеликий обсяг вибірки зменшує її точність порівняно з великою вибіркою. Проте доведено, що результати, які отримані за малими вибірками, також можна поширювати на генеральну сукупність. Але тут необхідно враховувати деякі особливості, зокрема, при розрахунку середнього квадратичного відхилення. При малому обсязі вибірки слід користуватися незміщеною оцінкою дисперсії S^2.

Основи теорії малих вибірок розробив англійський математик-статистик В.Госсет (псевдонім Стьюдент). Дослідження Стьюдента показали, що при невеликій чисельності сукупності середнє квадратичне відхилення у вибірці значно відрізняється від середнього квадратичного відхилення в генеральній сукупності.

Оскільки середнє квадратичне відхилення генеральної сукупності є одним із параметрів кривої нормального розподілу, то використовувати функцію нормального розподілу для оцінки параметрів генеральної сукупності за даними малих вибірок в силу отримання великих помилок не правомірно.

При розрахунку середньої помилки по вибірках малої чисельності завжди треба користуватись незміщеною оцінкою дисперсії

$$S^2 = \frac{\Sigma(x_i - \tilde{x})^2}{n-1},$$

де $n - 1$ – **число ступенів свободи варіації** (k), під яким розуміють число одиниць, здатних приймати довільні значення, не змінюючи їх загальної характеристики (середньої).

Наприклад, проведено три спостереження: $x_1 = 4$; $x_2 = 2$; $x_3 = 6$. Середня величина

$$\bar{x} = \frac{x_1 + x_2 + x_3}{n} = \frac{4+2+6}{3} = \frac{12}{3} = 4.$$

Отже, вільно варіюючих величин залишається тільки дві, тому що третя може бути знайдена за відомими двома величинами і середньою:

$$x_3 = n - (x_1 + x_2) = 3 \cdot 4 - (4 + 2) = 6;$$

$$x_1 = n - (x_2 + x_3) = 3 \cdot 4 - (2 + 6) = 4 \text{ і т.д.}$$

Отже, для даного прикладу число ступенів свободи варіації дорівнює 2 ($k = n - 1 = 3 - 1 = 2$).

Стьюдент обґрунтував закон розподілу відхилень вибіркових середніх від генеральної середньої для малих вибірок. Згідно з розподілом Стьюдента імовірність того, що гранична помилка не перевищить t-кратну середню помилку в малих вибірках, залежить від величини t і чисельності вибірки.

Теоретичне нормоване відхилення для малих вибірок одержало назву t-критерію на відміну від t-критерію нормального розподілу, який застосовується у великих вибірках. Значення t-критерію Стьюдента наводяться в спеціальних таблицях (дод. 3).

Розглянемо порядок визначення середньої і граничної помилки для малої вибірки на такому прикладі. Припустимо, для визначення величини втрат при збиранні картоплі проведено перекопування п'яти випадково відібраних площадок по 4 м². Втрати по площадках становили (кг): 0,6; 0,2; 0,8; 0,4; 0,5.

Середня величина втрат

$$\bar{x} = \frac{\Sigma x}{n} = \frac{2,5}{5} = 0,5 \text{ кг.}$$

Спостерігаючи за окремими спостереженнями, величина втрат сильно варіює і середня лише по п'яти спостереженнях може мати велику помилку.

Для розрахунку помилок вибірки визначимо незміщену оцінку дисперсії

$$S^2 = \frac{\Sigma x^2 - n\tilde{x}^2}{n-1} = \frac{1,45 - 5 \cdot 0,5^2}{5-1} = \frac{1,45 - 1,25}{4} = \frac{0,20}{4} = 0,05.$$

Звідки $S = \sqrt{0,05} = 0,2236$ кг.

Розрахуємо середню помилку вибіркової середньої, де замість середнього квадратичного відхилення використовується його незміщена оцінка:

$$\mu_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{0,2236}{\sqrt{5}} = \frac{0,2236}{2,236} = 0,10 \text{ кг.}$$

За таблицями Стьюдента (дод. 3) встановимо, що при довірчій імовірності $P = 0,95$ (рівень значущості $\alpha = 0,05$) і при $k = n - 1 = 5 - 1 = 4$ ступенях свободи варіації $t = 2,78$. Тоді гранична помилка вибірки дорівнює

$$\varepsilon_{\bar{x}} = t\mu_{\bar{x}} = 2,78 \cdot 0,10 \simeq 0,28 \text{ кг.}$$

Отже, з імовірністю $P = 0,95$ можна стверджувати, що величина втрат на всьому полі становитиме $0,5 \pm 0,28$ кг, або від 0,22 до 0,78 кг з розрахунку на 4 м².

Як бачимо з прикладу, межі випадкових коливань при малих вибірках досить великі і можуть бути скорочені за рахунок збільшення чисельності вибірки і зменшення коливання (дисперсії) ознаки.

Якщо б ми використали для розрахунку довірчих меж генеральної середньої таблицю інтегралу імовірностей (дод. 2), то t було б рівним 1,96 і $\varepsilon_{\bar{x}} = t\mu_{\bar{x}} = 1,96 \cdot 0,10 = 0,20$ кг, тобто довірчий інтервал був би вужчим (від 0,30 до 0,70 кг).

Малі вибірки через свою невелику чисельність навіть при найретельнішій організації спостереження не відображають достатньо точно показники генеральної сукупності. Тому результати малих вибірок рідко використовуються для встановлення надійних меж, в яких знаходяться характеристики генеральної сукупності.

Критерій Стьюдента застосовується головним чином для перевірки статистичних гіпотез щодо істотності відмінностей між показниками двох або кількох малих вибірок (див. розділ 7).

Завдання для самоконтролю до розділу 6

1. У чому суть вибіркового спостереження, необхідність і доцільність його застосування?

2. У чому полягають переваги вибіркового спостереження порівняно з суцільним?

3. Які наукові умови застосування вибіркового спостереження?

4. Що означає репрезентативність вибірки?

5. Які є помилки вибірки?

6. Охарактеризуйте середню помилку вибірки, наведіть алгоритм її розрахунку.

7. Охарактеризуйте граничну помилку вибірки, наведіть алгоритм її розрахунку.

8. Що таке довірчий рівень імовірності і рівень значущості?

9. Які Ви знаєте способи відбору одиниць у вибіркову сукупність?

10. У чому полягає суть випадкового повторного і безповторного відбору?

11. У чому полягає суть механічного і типового відбору?

12. У чому полягає суть серійного і комбінованого відбору?

13. Як розраховується необхідна чисельність вибірки? Від яких факторів залежить чисельність вибірки?

14. У чому полягають особливості розрахунку необхідної чисельності вибірки при випадковому, типовому і серійному відборі?

15. Які є способи поширення даних вибіркового спостереження на генеральну сукупність?

16. Дайте поняття статистичної оцінки. Які вимоги ставляться до оцінок?

17. Що таке незміщена, спроможна, ефективна і достатня оцінка?

18. Що таке точкова і інтервальна оцінка генеральної середньої, довірчий інтервал?

19. Що таке закон розподілу? Дайте характеристику нормальному закону розподілу вибіркових характеристик.

20. У чому полягає суть розподілу Стьюдента і Пірсона?

21. Що таке мала вибірка? У чому полягають особливості визначення помилок вибірки в малих вибірках?

22. Для визначення середнього віку студентів факультету проведена випадкова безповторна вибірка 100 студентів із загального числа 1000 чол.

У результаті обстеження отримані такі дані:

Вік, років	До 18	19	20	21	22	Понад 22
Число студентів, чол	25	17	22	18	10	8

Визначте: а) середній вік студентів за вибіркою; б) середню і граничну помилку вибірки з імовірністю 0,954; в) з тією самою імовірністю довірчий інтервал для середньої і частки студентів, вік яких становить 21 рік і більше.

Розділ 7

Перевірка статистичних гіпотез

7.1. Поняття про статистичні гіпотези

В практичній і науковій діяльності часто для доведення справедливості того або іншого факту удаються до висловлювання гіпотез, які можуть бути перевірені на основі даних вибіркового спостереження.

Перевірка статистичних гіпотез має велике значення для практики. Зокрема, на ній ґрунтуються прийоми статистичного контролю якості продукції. Припустимо, що на підприємстві про якість продукції роблять висновки за результатами вибіркового контролю. Якщо вибіркова частка браку не перевищує заздалегідь встановленої (нормативної) величини, то партія продукції приймається. Однак висновок щодо відповідності якості продукції встановленим вимогам робиться на основі вибіркової перевірки і тому носить імовірнісний характер. Таким чином, судження про якість продукції не може розглядатися як категоричне. По суті, мова иде про припущення (гіпотезу), що частка браку у всій генеральній сукупності дорівнює або менше нормативної величини. Ця гіпотеза і має бути перевірена на основі результатів вибіркового спостереження.

Теорія перевірки статистичних гіпотез також має велике значення в статистичній обробці експериментальних даних. Так, якщо на основі експериментальних даних ставиться питання про заміну одного сорту пшениці іншим, то необхідно перевірити припущення (гіпотезу) про те, що новий сорт пшениці має більш високу врожайність порівняно зі старим сортом. Перевірку таких припущень на основі даних вибіркового спостереження називають **статистичною перевіркою гіпотез.** Перевірка гіпотез проводиться за допомогою методів математичної статистики.

Гіпотеза в широкому розумінні – це деяке наукове припущення щодо властивостей явищ, що їх вивчають, яке потребує перевірки та доведення.

Статистичною гіпотезою називається припущення відносно параметрів або форми розподілу генеральної сукупності, яке перевіряється на основі даних вибіркового спостереження. Позначається гіпотеза літерою H від латинського слова hypothesis.

Із визначення статистичної гіпотези випливає, що вона може стосуватися або окремих параметрів розподілу, або законів розподілу.

Прикладом статистичних гіпотез можуть бути припущення про те, що жива маса телят в господарствах району підпорядкована закону нормального розподілу; середня урожайність картоплі одного сорту перевищує середню урожайність другого сорту; середні розміри деталей, що виготовляються на однотипних верстатах в ремонтній майстерні підприємства, однакові.

В ході перевірки статистичної гіпотези необхідно встановити, чи узгоджуються дані спостереження з висунутою гіпотезою, чи можна відмінності між гіпотезою і результатами спостереження віднести до випадкових або ж ці відмінності зумовлені впливом яких-небудь систематично діючих причин. В результаті перевірки гіпотези слід прийняти рішення про вибір одного з можливих двох взаємовиключних висновків, які називають **альтернативними.** Наприклад, при випробуванні добавок мікроелементів у кормовий раціон тварин такими висновками будуть: а) добавка мікроелементів сприяє росту продуктивності тварин; б) добавка мікроелементів не сприяє росту, продуктивності тварин. За підсумками перевірки гіпотеза або приймається, або відхиляється.

Найбільш часто перевіряється припущення про те, що отримана за вибіркою величина незначно відрізняється від гіпотетичної (теоретично припустимої) або встановленої величини в генеральній сукупності. Для перевірки цього положення висувається гіпотеза про те, що істинна різниця між фактичними і гіпотетичними показниками дорівнює нулю. В зв'язку з цим гіпотезу, що перевіряється, називають **нульовою** і позначають H_0. Нульову гіпотезу ще називають **основною**, або **робочою, гіпотезою.**

У кожному випадку нульовій гіпотезі протиставляється **альтернативна (конкуруюча) гіпотеза** (H_a), яка заперечує нульову гіпотезу. Наприклад, якщо нульова гіпотеза полягає в припущенні, що середня урожайність зернових культур в генеральній сукупності дорівнює 35 ц/га, то альтернативна гіпотеза, зокрема, може полягати в тому, що середня урожайність в генеральній сукупності не дорівнює 35 ц/га. Зміст гіпотези записується після двокрапки, а сам запис має вигляд:

$$H_0 : \bar{x} = 35; \qquad H_a : \bar{x} \neq 35.$$

Щодо до нульової гіпотези, може бути вказана нескінченна множина альтернативних гіпотез. Зміст H_0 і H_a залежить від характеру вибіркової сукупності і конкретних завдань, які вирішуються за допомогою перевірки статистичних гіпотез. Формально будь-яка з двох протилежних гіпотез може виступати як нульова. Однак практичні наслідки будуть різними в зв'язку з різним характером можливих помилок, отриманих в результаті неправильного прийняття або заперечення нульової гіпотези. Тому вибір нульової гіпотези має принциповий характер і потребує спеціального обґрунтування. Про це докладно йтиметься при розгляді можливих помилок при перевірці гіпотез.

Гіпотези про параметри генеральної сукупності називають **параметричними**, про розподіли – **непараметричними**. За формою побудови гіпотези можуть бути **простими** і **складними**. **Простою** називають гіпотезу, яка містить тільки одне припущення, **складною** – гіпотезу, яка містить два і більше припущення. Наприклад, гіпотеза $H_0 : \sigma = 3$ є простою гіпотезою, а гіпотези $H_0 : \sigma > 3$ і $H_0 : \sigma < 3$, які включають деяку область імовірних значень досліджуваного параметра, є складними. Очевидно, що складні гіпотези включають множину простих гіпотез.

7.2. Помилки при перевірці статистичних гіпотез. Статистичні критерії і критична область

В результаті перевірки статистичної гіпотези, що ґрунтується на даних вибірки обмеженого обсягу, можна відхилити і прийняти нульову гіпотезу (відповідно вибіркові дані суперечать і узгоджуються з H_0). Звідси видно, що перевірка статистичних гіпотез пов'язана з ризиком прийняття помилкових рішень.

Неправильне рішення може бути прийняте у двох випадках. В зв'язку з цим розрізняють помилки двох родів.

Помилка першого роду полягає в тому, що нульова гіпотеза H_0 відхиляється, хоча в дійсності вона є правильною.

Помилка другого роду полягає в тому, що приймається нульова гіпотеза H_0, хоча насправді правильною є альтернативна гіпотеза H_a.

Якщо, наприклад, встановлено, що новий пестицид є кращим, хоча насправді його дія не відрізняється від старого, це помилка першого роду; якщо ми вирішили, що обидва види пестицидів однакові, тоді як насправді новий вид є кращим, то допущена помилка другого роду.

Правильні і неправильні рішення можуть бути отримані в двох випадках, що наочно ілюструє табл. 7.1.

Імовірність допустити помилку першого роду (невиправдане відхилення H_0) отримала назву **рівня значущості** і позначається α. Імовірність зробити помилку другого роду (прийняття неправильної гіпотези H_0) позначається β. Отже, можна сказати, що при великому числі вибірок частка хибних висновків дорівнює α, якщо правильна H_0, і дорівнює β, якщо правильна H_a.

Таблиця 7.1

Можливі результати перевірки нульової гіпотези

Результат перевірки H_0	Можливий стан гіпотези, що перевіряється	
	правильна гіпотеза H_0	правильна гіпотеза H_a
H_0 відхиляється	помилка першого роду α	правильне рішення
H_0 приймається	правильне рішення	помилка другого роду β

Помилки I і II роду за своїми наслідками нерівнозначні і ведуть до різних матеріальних втрат. Тому вибір рівня значущості повинен ґрунтуватись на обліку можливих втрат: чим більші ці втрати, тим меншим повинен бути рівень значущості. Однак, якщо знижується рівень значущості, збільшується імовірність появи помилок другого роду. В цьому розумінні помилки I і II роду є конкуруючими.

Оскільки помилки I і II роду практично виключити неможливо, то в кожному випадку необхідно прагнути до зменшення втрат від цих помилок. При практичній перевірці гіпотез прагнуть до того, щоб за помилку I роду прийняти ту із можливих помилок, яка спряжена з більш серйозними наслідками на практиці.

Рівень значущості встановлюється самим дослідником залежно від характеру і важливості задач, що їх розв'язують (за так званим принципом практичної впевненості). Рівень значущості являє собою ту мінімальну імовірність, починаючи з якої можна визнати подію практично неможливою. Можна користуватись стандартними значеннями α = 0,10; 0,05; 0,01; 0,001; 0,0001 та ін. Найчастіше α встановлюють на рівні 0,05 і 0,01. При більш відповідальних рішеннях α підвищують до 0,001. Рівень значущості, наприклад, α = 0,05 означає, що в середньому в 5 випадках із 100 є ризик допустити помилку I роду, тобто відхилити правильну гіпотезу (H_0).

Встановлюючи певний рівень значущості, дослідник контролює імовірність помилки I роду: чим він нижчий, тим частіше H_0 буде визнаватись правильною. Однак, як було зазначено вище, зниження рівня значущості веде до появи помилок другого роду. В більшості випадків єдиним шляхом одночасного зменшення імовірності появи помилок двох родів є збільшення чисельності вибірки.

Для перевірки нульової гіпотези і прийняття висновку щодо сумісності вибіркових даних з висунутою гіпотезою використовують спеціальні **статистичні критерії**, що є зведенням правил, за якими перевірювану гіпотезу або приймають, або відхиляють. Інакше кажучи, критерій визначає ті властивості, якими повинні володіти вибіркові дані, щоб гіпотеза могла бути *прийнята або відхилена*.

Для кожного виду гіпотез, що перевіряються, розроблені спеціальні критерії, серед яких найчастіше використовуються t-критерії нормального розподілу і розподілу Стьюдента, F-критерії Фішера–Снедекора, χ^2 (*хі*-квадрат) розподілу Пірсона та ін.

Статистичні критерії, які використовуються для перевірки статистичних гіпотез, бувають двох видів: параметричні і непараметричні.

Параметричними називають критерії, які ґрунтуються на припущенні, що розподіл випадкової величини в сукупності підпорядкований деякому відомому закону (наприклад, нормальному, біномінальному, Пуассона). До таких критеріїв відносяться критерії t, F, χ^2 та ін.

Непараметричними (порядковими) називають критерії, використання яких не пов'язане із знанням закону розподілу випадкової величини, їх можна застосовувати і тоді, коли досліджуваний розподіл значно відрізняється від нормального. До таких критеріїв належать, зокрема, критерій знаків, Вілкоксона, Уайта, Манна–Уїтні та ін.

Параметричні критерії більш ефективні порівняно з непараметричними. Проте вони можуть бути використані для сукупностей, які мають нормальний або близький до нормального розподіл. Непараметричні критерії можуть бути використані при будь-якій формі розподілу. Єдиною умовою їх застосування є взаємна незалежність даних спостереження.

У множині можливих значень вибраного критерію можна виділити дві підмножини, що не перетинаються, одна з яких містить значення критерію, а друга – ні. Перша підмножина називається **критичною областю**, а друга – **областю припустимих значень.**

Критичною областю називають ті значення критерію, при яких нульова гіпотеза відхиляється. **Областю припустимих значень** (областю прийняття H_0) називають сукупність значень використовуваного критерію, при яких нульова гіпотеза приймається.

Точки, які відділяють критичну область від області допустимих значень, називають **критичними точками.**

Розрізняють односторонню і двосторонню критичні області.

Односторонньою називають правосторонню або лівосторонню критичну область. Ці області визначаються такими нерівностями: для правосторонньої критичної області $k > a_{кр}$, де $a_{кр}$ – додатне число, для лівосторонньої $k < a_{кр}$, де $a_{кр}$ – від'ємне число.

Двостороння критична область визначається нерівностями $k < a_1$, $k > a_2$, де $a_2 > a_1$, або коротко $|k| > a_{кр}$, де $a_{кр} > 0$.

Вибір односторонньої або двосторонньої критичної області залежить від конкретних умов і мети задач, що розв'язуються. Наприклад, при альтернативній гіпотезі $H_a : \bar{x}_1 \neq \bar{x}_2$ слід користуватись двосторонньою критичною областю, а при гіпотезах $H_a : \bar{x}_1 > \bar{x}_2$ і – односторонньою (відповідно правосторонньою і лівосторонньою) критичною областю.

Критичну область доцільно будувати так, щоб вона найкращим чином відрізняла нульову від альтернативної гіпотези.

Критерій перевірки гіпотези повинен бути підібраний так, щоб ризик допущення помилок був мінімальним. При цьому дуже важливо визначити імовірність того, що не буде припущено помилку II роду. Ця імовірність характеризує чутливість критерію до помилок II роду і дістала назву потужності критерію.

Потужністю критерію називається імовірність відхилення випробуваної гіпотези H_0, коли правильною є альтернативна гіпотеза $H_a(1 - в)$. Отже, потужність критерію є імовірність того, що не буде припущено помилку II роду. Звичайно, бажано мати найпотужніший критерій, бо це забезпечить мінімальну імовірність припущення помилки II роду. Тому з усіх можливих критеріїв слід вибирати найпотужніши.

Потужність (чутливість) критерію може бути підвищена двома способами: а) збільшенням рівня значущості. Проте цей шлях не зовсім прийнятний, бо необґрунтовано підвищується імовірність помилок I роду; б) збільшенням чисельності вибірки.

При формулюванні висновків за результатами перевірки гіпотези керуються таким принципом (правилом): якщо фактичне значення критерію потрапляє в критичну область, то H_0 відхиляють, якщо ж фактичне значення критерію належить до області припустимих значень, то H_0 приймають.

Для кожного критерію складено спеціальні таблиці, за якими знаходять його табличне значення (критичні точки), що відокремлюють критичну область від області припустимих значень. Знайдене табличне значення критерію порівнюють з його фактичним значенням. Якщо фактичне значення критерію, визначене за даними вибірки, буде більшим від табличного значення, то нульову гіпотезу потрібно відхилити і прийняти альтернативну гіпотезу. Якщо ж фактичне значення критерію буде меншим або таким, що дорівнює табличному, то робиться висновок про згоду даних спостереження з нульовою гіпотезою, тобто підстави для відмови від H_0 немає і тому її треба прийняти.

Якщо, наприклад, у досліді перевіряють вплив будь-якого фактора на результативну ознаку за допомогою t-критерію Стьюдента, то висновки формулюються так. Якщо $t_{факт} > t_{\alpha}$, то нульову гіпотезу (H_0: фактор не впливає на результативну ознаку) відхиляють, а вплив фактора на результативну ознаку вірогідний, істотний. Якщо ж перевіряють вірогідність різниці між середніми двох або кількох малих вибірок, то у цьому разі ($t_{факт} > t_{\alpha}$) говорять, що відмінності між середніми настільки значні, що вони не можуть бути результатом випадкового варіювання вибіркових даних, тому вони повинні бути визнані істотними, вірогідними.

За ситуації, коли виявиться, що $t_{факт} < t_{\alpha}$, роблять зворотні висновки: нульова гіпотеза (H_0: фактор не впливає на результативну ознаку) приймається, вплив фактора на результативну ознаку неістотний, невірогідний, а сама різниця між середніми лежить у межах можливих випадкових коливань, а тому вона неістотна, невірогідна.

При цьому слід мати на увазі, що згода з нульовою гіпотезою не доводить її абсолютної справедливості. Це лише свідчення про необхідність подальшої її перевірки, зокрема шляхом збільшення обсягу вибірки або поки більш переконливі дослідження не дозволять зробити протилежний висновок. Тому при формулюванні остаточних висновків в цьому випадку більш правильно говорити про те, що дані спостереження не суперечать нульовій гіпотезі і, отже, не дають підстави для її відхилення.

7.3. Загальна схема перевірки статистичної гіпотези

Підсумовуючи, можна навести загальну схему (алгоритм) перевірки статистичної гіпотези. Ця перевірка, як зазначалося вище, може бути проведена з використанням параметричних і непараметричних критеріїв. Наведемо схему перевірки гіпотези, що передбачає знання закону розподілу генеральної сукупності, тобто для випадку застосування параметричних критеріїв.

Перевірка цієї статистичної гіпотези передбачає послідовне виконання таких етапів.

1. Оцінка вихідної інформації та описування статистичної моделі вибіркової сукупності.

2. Формулювання нульової і альтернативної гіпотез.

3. Встановлення рівня значущості, за допомогою якого контролюється помилка I роду.

4. Вибір найпотужнішого критерію для перевірки нульової гіпотези. Застосування найпотужнішого критерію дозволить контролювати ймовірність появи помилки II роду.

5. Обчислення за певним алгоритмом фактичного значення критерію.

6. Визначення критичної області та області згоди з нульовою гіпотезою, тобто встановлення табличного значення критерію.

7. Зіставлення фактичного і табличного значень критерію і формулювання висновків за результатами перевірки нульової гіпотези.

На практиці найчастіше доводиться розв'язувати задачі двох типів по перевірці гіпотез. Задачі першого типу пов'язані з перевіркою гіпотез про істотність відмінностей між параметрами статистичних сукупностей.

Прикладом таких задач може бути оцінка вірогідності відмінностей між середніми, дисперсіями, коефіцієнтами кореляції, регресії та ін.

Задачі другого типу пов'язані з перевіркою гіпотез про істотність відмінностей законів розподілу. До них відносять задачі по визначенню відповідності вибіркового розподілу теоретичному, частіше всього нормальному, оцінці близькості двох емпіричних розподілів, однорідності складу кількох сукупностей тощо.

Кожен з перелічених типів задач підрозділяють на різні види. Всі вони знаходять широке прикладне застосування в дослідженнях по галузях народного господарства.

Розглянемо деякі важливі передумови і особливості перевірки статистичних гіпотез задач першого типу, пов'язаних із застосуванням параметричних критеріїв і припущення нормального розподілу в генеральній сукупності.

Вибір конкретної схеми перевірки гіпотези залежить від характеру досліджуваної гіпотези, особливостей вихідної інформації та інших умов.

Розглянемо основні з них.

1. Обсяг вибіркової сукупності. При перевірці гіпотез за даними великих вибірок ($n > 30$) доцільно застосовувати t-критерій нормального розподілу, а за даними малих вибірок ($n < 30$) – t-критерій розподілу Стьюдента.

2. Рівність вибірок за чисельністю. Вибіркові сукупності за чисельністю можуть бути рівними і нерівними. Ці властивості необхідно враховувати при практичній перевірці гіпотез про істотність відмінностей між середніми, зокрема, при розрахунку середньої помилки двох вибіркових середніх.

3. Схема формування вибірок. Прийоми перевірки статистичних гіпотез залежать від характеру формування вибіркових сукупностей. Якщо спостереження однієї вибірки не протиставляються спостереженням другої вибірки, то такі вибірки називають **незалежними.** Якщо ж спостереження однієї вибірки деякою мірою пов'язані зі спостереженнями другої вибірки, то такі вибірки називаються **залежними.** Прикладом незалежних вибірок може бути дослід з двома групами тварин, одна з яких є контрольною, а на другій випробовують якийсь препарат. При цьому дослідна і контрольна група формуються випадково. Проте, якщо при проведенні цього досліду тварини будуть попередньо розподілені на групи за будь-якими ознаками (наприклад, за біологічними властивостями, продуктивністю тощо), а потім від кожної пари аналогів буде відібрано по одному представнику до контрольної і дослідної групи, то спостереження за такими вибірками вже *не можна розглядати як незалежні.*

Формування вибіркових сукупностей зумовлює різні прийоми оцінки вірогідності між середніми двох малих вибірок. Якщо вибірки незалежні, то статистичній оцінці підлягає різниця середніх, якщо залежні – середня різниця.

4. Рівність дисперсій. При перевірці гіпотез щодо середніх можливі два випадки щодо вибіркових дисперсій:

1) дисперсії рівні ($\sigma_1^2 = \sigma_2^2$);

2) дисперсії нерівні ($\sigma_1^2 \neq \sigma_2^2$).

В зв'язку з цим виникає спеціальне завдання перевірки гіпотези про істотність відмінностей двох дисперсій. Для перевірки гіпотези про рівність двох дисперсій в генеральних сукупностях використовується F-критерій розподілу Фішера, який ґрунтується на співвідношенні двох вибіркових скоригованих дисперсій: S_1^2 і S_2^2, що замінюють значення дисперсій в генеральних сукупностях, які, як правило, невідомі:

$$F = \frac{S_1^2}{S_2^2}, \text{ де } S_1^2 > S_2^2.$$

Критичні значення F-критерію Фішера F_α знаходять за спеціальними таблицями (дод. 4 і 5) при відповідному числі ступенів свободи і заданому рівні значущості α. Перевірка H_0 зводиться до такого. Спочатку визначають фактичне значення критерію $F_{факт}$. Потім за таблицями знаходять його критичне значення F_α. Якщо виявиться, що $F_{факт} > F_\alpha$, то H_0 відхиляється, якщо ж $F_{факт} < F_\alpha$, то H_0 приймається.

Залежно від того, рівні чи нерівні дисперсії в генеральних сукупностях, потрібно видозмінювати схему перевірки гіпотези.

Розв'язування задач другого типу пов'язане з перевіркою гіпотез відносно законів розподілу генеральних сукупностей. У практичних дослідженнях часто виникає необхідність встановити характер розподілу генеральних сукупностей. При цьому можуть виникати задачі трьох видів:

1) про узгодженість емпіричного (вибіркового) і теоретичного (генерального) розподілу;

2) про незалежність розподілу однієї ознаки від другої;

3) про однорідність двох та більше емпіричних розподілів.

Такі гіпотези, як і гіпотези відносно параметрів розподілу, перевіряють за допомогою спеціальних критеріїв згоди.

Критерієм згоди називають критерій перевірки гіпотези щодо передбачуваного закону невідомого розподілу в генеральній сукупності.

Є ряд критеріїв згоди: К.Пірсона, О.М.Колмогорова, М.В.Смирнова, Б.С.Ястремського та ін. Ці критерії дозволяють встановити, узгоджуються чи не узгоджуються досліджувані розподіли з теоретичними розподілами, а також те, наскільки істотними є розбіжності між цими розподілами.

7.4. Перевірка статистичних гіпотез щодо середніх величин

Серед найважливіших узагальнюючих характеристик, відносно яких найчастіше висуваються гіпотези, є середня величина. З метою перевірки гіпотези про рівність середніх в генеральній сукупності необхідно сформулювати нульову гіпотезу. При цьому, як правило, виходять з того, що обидві вибірки узяті з нормально розподіленої генеральної сукупності з математичним сподіванням, рівним $\overline{x}$ і з дисперсією, рівною σ_0^2. Якщо це припущення вірне, то $\overline{x}_1 = \overline{x}_2 = \overline{x}$. Фактично ж вибіркові середні $\widetilde{x}_1$ і $\widetilde{x}_2$ не будуть рівні через випадковості вибірки. Тому потрібно з'ясувати істотність розбіжностей між $\widetilde{x}_1$ і $\widetilde{x}_2$ – чи знаходиться їх різниця в межах можливої випадкової варіації, чи ж вона виходить за ці межі. Тоді задача перевірки гіпотези зводиться до перевірки істотності різниці

$$\left|\widetilde{x}_1 - \overline{x}\right| - \left|\widetilde{x}_2 - \overline{x}\right| = \left|\widetilde{x}_1 - \widetilde{x}_2\right| = \varepsilon_{\text{факт}}.$$

Кожна вибіркова середня має свою помилку μ:

$$\mu_1^2 = \frac{S_1^2}{n_1}; \qquad \mu_2^2 = \frac{S_2^2}{n_2};$$

де скориговані дисперсії

$$S_1^2 = \frac{\Sigma(x_{1i} - \widetilde{x}_1)^2}{n_1 - 1}; \qquad S_2^2 = \frac{\Sigma(x_{2i} - \widetilde{x}_2)^2}{n_2 - 1};$$

або за іншими формулами:

$$S_1^2 = \frac{\Sigma x_{1i}{}^2 - n_1 \widetilde{x}_1{}^2}{n_1 - 1}; \qquad S_2^2 = \frac{\Sigma x_{2i}{}^2 - n_2 \widetilde{x}_2{}^2}{n_2 - 1};$$

Узагальнена середня помилка двох вибіркових середніх

$$\overline{\mu}_{1-2} = \sqrt{\mu_1^2 + \mu_2^2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

Визначивши дисперсії і середню помилку вибіркових середніх, можна обчислити фактичне значення t-критерію і порівняти його з критичним (табличним) значенням при відповідному рівні значущості і числі ступенів свободи варіації (для вибірок з чисельністю $n \geq 30$ використовується t-критерій нормального розподілу, а для вибірок з чисельністю $n < 30$ – t-критерій Стьюдента).

Фактичне значення t-критерію визначається за формулою

$$t_{\text{факт}} = \frac{\left|\widetilde{x}_1 - \widetilde{x}_2\right|}{\overline{\mu}_{1-2}}.$$

Якщо вибіркове значення критерію потрапляє в критичну область ($t_{факт} > t_{\alpha}$), нульова гіпотеза про рівність середніх відхиляється; якщо ж вибіркове значення критерію потрапляє в область припустимих значень ($t_{факт} < t_{\alpha}$), нульова гіпотеза приймається.

Нульова гіпотеза про рівність середніх у двох генеральних сукупностях може бути також перевірена шляхом порівняння фактичної середньої різниці $\left(\varepsilon_{факт} = \tilde{x}_1 - \tilde{x}_2\right)$ з граничною випадковою помилкою при заданому рівні значущості(ε_{α}). Якщо фактична різниця між вибірковими середніми знаходиться в межах випадкової помилки ($\varepsilon_{факт} < \varepsilon_{\alpha}$), нульова гіпотеза приймається. Якщо ж фактична різниця між середніми виходить за межі випадкової помилки ($\varepsilon_{факт} > \varepsilon_{\alpha}$), нульова гіпотеза відхиляється.

При розв'язуванні конкретних задач по перевірці статистичних гіпотез відносно середніх необхідно враховувати такі моменти: 1) схему формування вибірок (вибірки незалежні і залежні); 2) рівність або нерівність обсягів вибірок; 3) рівність або нерівність дисперсій в генеральних сукупностях.

Алгоритм перевірки гіпотези відносно двох середніх дещо змінюється, якщо дисперсії по вибірках (S_1^2 і S_2^2) суттєво відрізняються. В цьому випадку при визначенні числа ступенів свободи вводиться поправка:

$$k = (n_1 - 1) + (n_2 - 1)\left(0{,}5 + \frac{S_1^2\ S_2^2}{S_1^4 + S_2^4}\right).$$

Коли ж при нерівних дисперсіях по вибірках нерівними є і їх чисельності (n_1 і n_2), табличне значення t-критерію Стьюдента слід розрахувати за формулою

$$t_{табл} = \frac{t_1 \frac{S_1^2}{n_1} + t_2 \frac{S_2^2}{n_2}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},$$

де t_1 і t_2 – табличні значення t-критерію Стьюдента, які беруться відповідно з $n_1 - 1$ і $n_2 - 1$ ступенями свободи.

Розглянемо приклад перевірки статистичної гіпотези про рівність двох середніх незалежних вибірок рівної чисельності ($n_1 = n_2$) і рівними дисперсіями ($\sigma_1^2 = \sigma_2^2$).

Нехай є дані щодо живої маси телят при народжені по двох групах корів чорно-рябої породи (корови одного віку). Перша група корів мала нормальну тривалість лактації (305 днів), а друга група доїлася протягом 320 днів. У кожну групу відібрано по 5 корів. Дані спостереження наведено в табл. 7.2.

Таблиця 7.2

Жива маса телят при народжені по групах корів з різною тривалістю лактації

Тварини	Жива маса телят, кг		Квадрати живої маси	
	І група корів	ІІ група корів	І група корів	ІІ група корів
	x_{1i}	x_{2i}	x_{1i}^2	x_{2i}^2
1	36	33	1296	1089
2	40	35	1600	1225
3	45	40	2025	1600
4	40	37	1600	1369
5	39	35	1521	1225
Разом	**200**	**180**	**8042**	**6508**

Співставлення живих мас телят по двох групах корів показує, що більш висока жива маса телят спостерігається у корів І групи, які мали нормальну тривалість лактації. Однак, в зв'язку з тим, що чисельність вибірок невелика ($n = 5$), не виключена можливість, що розбіжності між живими масами отримані в результаті дії випадкових причин.

Потрібно статистично оцінити різницю між середніми по двох групах корів.

За результатами перевірки гіпотези зробити висновок про те, що різниця між середніми лежить в межах випадкових коливань, або ж ця різниця настільки значна, що не узгоджується з нульовою гіпотезою про випадковий характер відмінностей між середніми.

Якщо буде доведено друге положення і відхилене перше, можна стверджувати, що тривалість лактації впливає на живу масу телят.

Умова задачі передбачає, що обидві вибірки взяті із нормально розподіленої генеральної сукупності. Формування груп випадкове (незалежне), тому оцінюватись повинна різниця між середніми.

Визначимо середню живу масу телят по двох групах корів:

$$\tilde{x}_1 = \frac{\Sigma x_{1i}}{n_1} = \frac{200}{5} = 40 \text{ кг}; \ \tilde{x}_2 = \frac{\Sigma x_{2i}}{n_2} = \frac{180}{5} = 36 \text{ кг}.$$

Фактична різниця між середніми становить:

$$\varepsilon_{\text{факт}} = \left|\tilde{x}_1 - \tilde{x}_2\right| = 40 - 36 = 4 \text{ кг}.$$

Істотність цієї різниці повинна бути оцінена. Для цього **необхідно перевірити** гіпотезу про рівність двох середніх.

Розглянемо докладно всі етапи схеми перевірки гіпотези.

1. Сформулюємо нульову H_0 і альтернативну H_a гіпотези:

$$H_0 : \overline{x}_1 = \overline{x}_2;\ H_a : \overline{x}_1 \neq \overline{x}_2.$$

2. Приймемо рівень значущості $\alpha = 0{,}05$, гарантуючи прийняття гіпотези або відмови від неї з імовірністю помилки тільки в 5 випадках із 100.

3. Найпотужнішим критерієм для перевірки такого роду гіпотези H_0 є t-критерій Стьюдента.

4. Сформулюємо правило прийняття рішення за результатами перевірки H_0. Оскільки за альтернативною гіпотезою $\overline{x}_1$ може бути або менше або більше $\overline{x}_2$, то критична область повинна бути встановлена з двох сторін: $t \leq -t_\alpha$ і $t \geq t_\alpha$, або коротше: $|t| \geq t_\alpha$.

Така форма завдання критерію називається **двосторонньою критичною областю.** Критична область при $\alpha = 0{,}05$ буде міститись в межах – всі значення вище, ніж верхня 2,5% і нижче, ніж 2,5% точки розподілу t-критерію Стьюдента.

З урахування сказаного висновки по перевірці H_0 можна сформулювати так: гіпотеза H_0 відхидяться, якщо фактичне значення t-критерію виявиться більшим за табличне значення, тобто якщо $t_{\text{факт}} > t_\alpha$. В протилежному випадку H_0 повинна бути прийнята.

5. Щоб перевірити H_0 потрібно визначити фактичне значення t-критерію Стьюдента і порівняти його з табличним значенням.

Для визначення фактичного значення t-критерію Стьюдента виконаємо такі обчислення.

6. Обчислимо по кожній вибірці скориговані на втрату ступенів свободи варіації дисперсії. Для цього попередньо піднесемо до квадрату значення x_{1i} і x_{2i}:

$$S_1^2 = \frac{\Sigma x_{1i}^2 - n_1 \tilde{x}_1^2}{n_1 - 1} = \frac{8042 - 5 \cdot 40^2}{5-1} = 10{,}5 \text{ кг};$$

$$S_2^2 = \frac{\Sigma x_{2i}^2 - n_1 \tilde{x}_1^2}{n_2 - 1} = \frac{6508 - 5 \cdot 36^2}{5-1} = 7{,}0 \text{ кг}.$$

7. Розрахуємо квадрати середніх помилок по кожній вибірці і узагальнену середню помилку різниці середніх:

$$\mu_1^2 = \frac{S_1^2}{n_1} = \frac{10{,}5}{5} = 2{,}1; \qquad \mu_2^2 = \frac{S_2^2}{n_2} = \frac{7{,}0}{5} = 1{,}4;$$

$$\overline{\mu}_{1-2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n}} = \sqrt{\mu_1^2 + \mu_2^2} = \sqrt{2{,}1 + 1{,}4} = \sqrt{3{,}5} = 1{,}87 \text{ кг}.$$

8. Розрахуємо фактичне значення t-критерію Стьюдента:

$$t_{факт} = \frac{\varepsilon_{факт}}{\overline{\mu}_{1-2}} = \frac{40-36}{1,87} = 2,14.$$

9. Встановимо табличне значення t-критерію Стьюдента, виходячи із рівня значущості α = 0,05 і загального числа ступенів свободи для двох вибірок:

$$k = (n_1 - 1) + (n_2 - 1) = (5 - 1) + (5 - 1) = 8.$$

За таблицею «Критичні точки розподілу Стьюдента» (дод. 3) знайдемо t при $\alpha = 0,05$ і $k = 8$: $t_{0,05} = 2,31$.

10. Співставимо фактичне і табличне значення t-критерію Стьюдента:

$$t_{факт} < t_{0,05};\ 2,14 < 2,31.$$

Оскільки $t_{факт} < t_{0,05}$ (вибіркове значення критерію знаходиться в області припустимих значень), нульова гіпотеза про рівність середніх в генеральних сукупностях приймається.

Отже, вплив тривалості лактації на живу масу телят при народженні виявляється недоведеним.

Однак слід звернути увагу на такий суттєвий момент: жива маса телят при народженні по всіх спостереженнях досліду вище в першій групі корів, які мають нормальну тривалість лактації. Тому замість альтернативної гіпотези $H_a : \overline{x}_1 \neq \overline{x}_2$ може бути взята інша. Оскільки немає підстав вважати, що при нормальній тривалості лактації жива маса телят буде нижчою, то очевидно, що більш доцільною формою альтернативної гіпотези є: $H_a : \overline{x}_1 > \overline{x}_2$.

Тоді критична область, що становить 0,05 всієї площі під кривою розподілу, буде розташована тільки з одного (правого) боку, так як від'ємні значення живих мас вважаються несумісними з умовами задачі. В зв'язку з цим табличне значення t-критерію слід визначати при подвоєному значенні рівня значущості (тобто при 2α; $t_\alpha = 2 \cdot 0,05 = 0,10$). Критерій перевірки гіпотези формулюється так: нульова гіпотеза відхиляється, якщо $t_{факт} > t_{2\alpha}$.

Така форма завдання критичної області називається **односторонньою.** Односторонній критерій більш чутливий до помилок другого роду, але його застосування припустиме лише у випадку, якщо доведена правомірність даної альтернативної гіпотези.

Встановимо за таблицями (дод. 3) табличне значення t-критерію при α=0,10 і $k = 8$, $t_{0,10} = 1,86$.

Отже, при використанні одностороннього критерію нульова гіпотеза відхиляється, тобто критерій виявився в критичній області ($t_{факт} > t_{0,10}$; 2,14 > 1,86). Таким чином, жива маса телят при народженні в групі корів з нормальною тривалістю лактації суттєво вище. Цей висновок точніший, ніж отриманий на основі двостороннього критерію, бо тут використана додаткова інформація для обґрунтування правильності застосування одностороннього критерію.

Такий самий висновок одержимо і шляхом порівняння можливої граничної *помилки двох вибірок* ε_α *з фактичною різницею середніх.*

Обчислимо можливу граничну помилку різниці середніх по двох вибірках:

$\varepsilon_{0,10} = t_{0,10} \cdot \overline{\mu}_{1-2} = 1,86 \cdot 1,87 = 3,48$ кг і порівняємо її з фактичною різницею середніх:

$$\varepsilon_{\text{факт}} > \varepsilon_{0,10};\ 4,00 > 3,48.$$

Зіставляючи граничну можливу помилку з фактичною різницею середніх, можна зробити аналогічний висновок про те, що висунута гіпотеза про рівність середніх не узгоджується з одержаними результатами.

Перевірку гіпотези для випадку залежних вибірок з рівними чисельностями і рівними дисперсіями розглянемо на такому прикладі.

Нехай є дані вибіркового спостереження щодо продуктивності корів-матерів і корів-дочок (табл. 7.3).

Таблиця 7.3

Продуктивність корів-матерів і корів-дочок

Номер пари	Надій на корову за рік, ц		Різниця надоїв	Квадрат різниці
	дочки	матері		
	x_{1i}	x_{2i}	$d_i = x_{1i} - x_{2i}$	d_i^2
1	50,3	44,1	6,2	38,44
2	43,4	42,2	1,2	1,44
3	50,3	47,6	2,7	7,29
4	41,6	38,1	3,5	12,25
5	51,9	50,0	1,9	3,61
Сума	**237,5**	**222,0**	**15,5**	**63,03**
Середні	**47,5**	**44,1**	**3,1**	–

Необхідно перевірити статистичну гіпотезу відносно середньої різниці між парами взаємопов'язаних спостережень в генеральній сукупності.

Оскільки спостереження двох вибірок попарно взаємопов'язані (залежні вибірки), то необхідно порівнювати не різницю між середніми, а середнє значення різниць між парами спостережень ($\overline{d}$).

Розглянемо всі етапи процедури перевірки гіпотези.

1. Сформулюємо нульову і альтернативну гіпотези:

$$H_0 : \overline{d} = 0;\ H_a : \overline{d} \neq 0.$$

При такій альтернативі необхідно застосувати двосторонній критерій.

2. Рівень значущості приймемо рівним $\alpha = 0,05$.

3. Найпотужнішим критерієм перевірки H_0 є t-критерій Стьюдента.

4. Обчислимо середню різницю

$$\tilde{d} = \frac{\Sigma d_i}{n} = \frac{15,5}{5} = 3,1 \text{ ц},$$

або $\tilde{d} = \left|\tilde{x}_1 - \tilde{x}_2\right| = 47,5 - 44,1 - 3,1$ ц.

5. *Розрахуємо скореговану дисперсію середньої різниці:*

$$S^2_{\tilde{d}} = \frac{\Sigma d_i^2 - n\tilde{d}^2}{n-1} = \frac{63,03 - 5 \cdot 3,1^2}{5-1} = 3,745.$$

6. Визначимо середню помилку середньої різниці:

$$\mu_{\tilde{d}} = \sqrt{\frac{S^2_{\tilde{d}}}{n}} = \sqrt{\frac{3,745}{5}} = 0,865 \text{ ц}.$$

7. Обчислимо фактичне значення t-критерію Стьюдента:

$$t_{\text{факт}} = \frac{\left|\tilde{d}\right|}{\mu_{\tilde{d}}} = \frac{3,1}{0,865} = 3,58.$$

8. Встановимо число ступенів свободи, виходячи із чисельності пар взаємопов'язаних різниць:

$$k = n - 1 = 5 - 1 = 4.$$

9. Знайдемо табличне значення t-критерію Стьюдента при $k = 4$ і $\alpha=0,05$; $t_{0,05} = 2,78$ (дод. 3).

10. Порівняємо фактичне і табличне значення критерію:

$$t_{\text{факт}} > t_{0,05}\,;\ 3,58 > 2,78.$$

Фактичне значення критерію вище за табличне. Отже, величина середньої різниці між надоями двох вибірок істотна і нульова гіпотеза відхиляється.

Такі самі висновки дістанемо, порівнюючи можливу граничну помилку з фактичною середньою різницею:

$$\varepsilon_{0,05} = t\mu_{\tilde{d}} = 2,78 \cdot 0,865 = 2,4 \text{ ц}.$$

Гранична помилка показує, що в результаті випадкового варіювання середня різниця може досягати 2,4 ц. Фактична середня різниця вище:

$$\varepsilon_{\text{факт}} > \varepsilon_{0,05};\ 3,1 > 2,4.$$

Отже, за результатами досліду можна з високим ступенем імовірності стверджувати, що відмінності в значеннях середніх надоїв корів-матерів і корів-дочок вірогідні.

7.5. Перевірка статистичних гіпотез щодо розподілів

Поряд з перевіркою статистичних гіпотез щодо середніх інколи потрібно перевірити гіпотези щодо характеру розподілу. Гіпотези про розподіли полягають в тому, що розподіл в генеральній сукупності підпорядковується якому-небудь певному закону. Перевірка гіпотези полягає в тому, щоб на основі порівняння фактичних (емпіричних) частот з передбачуваними (теоретичними) частотами зробити висновок про відповідність фактичного розподілу гіпотетичному розподілу.

Процедура перевірки гіпотези про відповідність емпіричного розподілу теоретичному складається з таких етапів:

1. Обчислення оцінок параметрів передбачуваного розподілу на основі даних вибіркового спостереження.

2. Визначення теоретичних частот на основі одержаних оцінок параметрів і виходячи з теоретичної функції частот.

3. Оцінка близькості емпіричного розподілу теоретичному на основі певного критерію згоди.

Оскільки нормальний розподіл зустрічається досить часто, то найчастіше перевіряють гіпотези про відповідність вибіркового розподілу нормальному. Однак, поряд з нормальним розподілом генеральні сукупності можуть бути розподілені і за іншими законами. Тому вибір теоретичного закону розподілу повинен базуватись на глибокому розумінні характеру формування досліджуваного явища або процесу. Певну роль у вирішенні цього питання може відігравати розрахунок статистичних характеристик вибіркових розподілів і побудова графіків (гістограми, полігону, кумуляти тощо). Так, про форму розподілу роблять висновок за вибірковими коефіцієнтами скошеності і ексцесу: якщо вони рівні нулю або близькі до нуля, то можна припустити, що досліджуваний розподіл належить до нормального; якщо середня арифметична та дисперсія рівні або дуже близькі одна до одної, то можна припустити, що вибірковий розподіл відповідає розподілу Пуассона.

Для перевірки гіпотези щодо відповідності вибраних законів розподілу (нормальне, біноміальне, Пуассона тощо) в генеральній сукупності в більшості випадків при розрахунку критеріїв згоди використовуються відхилення емпіричних частот від теоретичних. Чим менше це відхилення, тим точніше теоретичний розподіл відтворює вибірковий та навпаки.

При перевірці статистичних гіпотез відносно розподілів може бути використаний ряд критеріїв. З множини критеріїв згоди, які використовуються при перевірці гіпотез щодо розподілів найчастіше за інші застосовують найпотужніший параметричний критерій Пірсона. (χ^2 – *хі*-квадрат). Його обчислюють як суму частки від ділення квадрату різниці між емпіричними і теоретичними частотами на теоретичні частоти:

$$\chi^2 = \sum_{i=1}^{l} \frac{(n_i - \widetilde{n}_i)^2}{\widetilde{n}_i},$$

де l – число інтервалів (класів, груп) на які розбито вибірковий розподіл; n_i – частоти емпіричного розподілу; $\widetilde{n}_i$ – частоти теоретичного розподілу.

З формули випливає, що чим менше розбіжність між n_i і $\widetilde{n}_i$, тим ближче за значенням один до одного емпіричні та теоретичні частоти, тим менше χ^2. При повному збігу теоретичних і вибіркових частот $\chi^2 = 0$, у протилежному випадку $\chi^2 > 0$. Область зміни χ^2 від 0 до ∞. При великому числі ступенів свободи ($k \to \infty$) розподіл χ^2 набуває форми, близької до нормального розподілу.

Щоб оцінити близькість емпіричного і теоретичного розподілів, необхідно розрахувати фактичне значення χ^2 і порівняти його з табличним значенням при заданому рівні значущості (α) і відповідному числі ступенів свободи k.

Число ступенів свободи визначають по-різному залежно від характеру гіпотези, що її перевіряють, та особливостей вихідної інформації. Так, якщо перевіряється гіпотеза про узгодженість вибіркового і теоретичного розподілів, то число ступенів свободи визначають за формулою

$$k = l - 1 - S,$$

де l – число інтервалів (класів, груп) вибірки; S – число параметрів генерального розподілу, які оцінюються за даними вибірки.

При оцінці відповідності емпіричного розподілу нормальному число ступенів свободи $k = l - 1 - 2 = l - 3$, оскільки для побудови кривої нормального розподілу оцінюються два параметри: середня арифметична і середнє квадратичне відхилення. Якщо перевіряється відповідність вибіркового розподілу Пуассона, то оцінюється один параметр λ. Тоді число ступенів свободи $k = l - 1 - 1 = l - 2$.

Якщо вихідні дані подано у вигляді таблиці розподілу частот і необхідно перевірити гіпотезу щодо незалежності розподілу двох ознак, то число ступенів свободи визначають за формулою:

$$k = (a - 1) \cdot (b - 1),$$

де a – число рядків; b – число стовпців.

Так, число ступенів свободи $k = (a - 1) \cdot (b - 1) = (2 - 1) \times (2 - 1) = 1$ для таблиці 2×2, $k = 4$ для таблиці 3×3, $k = 2$ для таблиці 3×2 і т.д.

Якщо перевіряється гіпотеза щодо однорідності двох сукупностей, то число ступенів свободи визначають за формулою $k = n - 1$, де n – число інтервалів (класів, груп).

Як видно, для всіх випадків число ступенів свободи, крім обов'язкових обмежень, завжди зменшується на одиницю, тобто має місце один лінійний обмежуючий зв'язок – рівність сум емпіричних і теоретичних частот.

Якщо отримане за вибіркою значення $\chi^2_{\text{факт}} \le \chi^2_{\alpha}$, то нульова гіпотеза приймається. Якщо ж $\chi^2_{\text{факт}} > \chi^2_{\alpha}$, то нульова гіпотеза відхиляється.

Фактичне значення можна обчислити і за іншою формулою, яка випливає з вищенаведеної:

$$\chi^2 = \sum_{i=1}^{l} \frac{n_i^2}{\tilde{n}_i} - n.$$

Ця формула не потребує обчислення квадратів відхилень (в чому полягає її простота), її можна використовувати і для перевірки правильності обчислень.

Якщо вихідні дані подано у вигляді чотирьохклітинної таблиці розподілу частот за двома ознаками (2 × 2) з чисельностями n_i:

n_1	n_2
n_3	n_4

то фактичне значення χ^2 може бути визначене за формулою:

$$\chi^2 = \frac{(n_1 n_4 - n_2 n_3)(n_1 + n_2 + n_3 + n_4)}{(n_1 + n_2)(n_3 + n_4)(n_1 + n_3)(n_2 + n_4)}.$$

Критерій χ^2 Пірсона використовується для розв'язування ряду задач, зокрема, при перевірці гіпотез про згоду (відповідність) вибіркового і теоретичного розподілів, про незалежність розподілів, про однорідність сукупностей. Стосовно цих задач критерій χ^2 називають критерієм згоди, незалежності і однорідності.

Застосування критерію χ^2 вимагає дотримання ряду умов, найважливішими серед яких є:

1) обсяг вибірки повинен бути досить великим (при $n < 50$ потужність критерію χ^2 значно знижується);

2) чисельність окремих інтервалів (класів) має бути не менше п'яти одиниць. Якщо ця умова не виконується, то проводиться об'єднання малочисельних інтервалів з числом одиниць менше 5 (як виняток таких інтервалів може бути не більше 20% від загальної кількості їх);

3) частоти не можна перетворювати в частки, оскільки це може призвести до збільшення величини відхилень $n - \tilde{n}$.

Розглянемо приклад перевірки статистичної гіпотези про відповідність емпіричного розподілу нормальному. Для цього використаємо дані ряду розподілу 100 господарств за надоєм молока на корову (див. табл. 4.2).

Розрахунок характеристик варіаційного ряду розподілу 100 господарств за надоєм молока (моди, медіани, коефіцієнтів скошеності та ексцесу) показав, що емпіричний розподіл дуже близький до симетричного і характеризується такими параметрами:

середній надій на корову $\widetilde{x} = 32{,}6$ ц;

вибіркове середнє квадратичне відхилення $\sigma = 3{,}2$ ц;

величина інтервалу $h = 2{,}0$ ц;

чисельність вибіркової сукупності $n = 100$.

Розглянемо всі етапи процедури перевірки гіпотези.

1. Сформулюємо нульову і альтернативну гіпотези: H_0 – емпіричний розподіл відповідає нормальному; H_a – емпіричний розподіл не відповідає нормальному.

2. Приймемо рівень значущості $\alpha = 0{,}05$.

3. Найпотужнішим критерієм перевірки цієї гіпотези є χ^2-критерій згоди Пірсона.

4. Для перевірки H_0 необхідно розрахувати χ^2 і порівняти його з табличним значенням χ^2_α.

Спочатку необхідно, виходячи із припущення про відповідність емпіричного розподілу нормальному, побудувати теоретичний розподіл (криву нормального розподілу), для побудови якого використаємо параметри вибіркового розподілу $\widetilde{x} = 32{,}6$ ц; $S = 3{,}2$ ц.

5. Порядок розрахунку теоретичних частот нормального розподілу і критерію χ^2 наведено в табл. 7.4.

6. Дамо деякі пояснення до розрахунків. Значення нормованого відхилення t_i визначається як $\frac{x_i - \widetilde{x}}{S}$, де за x_i приймається серединне значення інтервалу. Виражаючи довжину інтервалу h також в одиницях середнього квадратичного відхилення як $\frac{h}{S}$ при відомому обсязі вибірки n, можна розрахувати теоретичні (очікувані) частоти для будь-якого інтервалу, використовуючи таке співвідношення:

$$\widetilde{n} = n\frac{h}{S}f(t) = n\frac{h}{S}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}.$$

Значення функції нормального розподілу $f(t)$ знайдемо за таблицею (дод. 1)

Таблиця 7.4

Розрахунок теоретичних частот нормального розподілу і критерію χ^2

Середнє значення інтервалу	Фактичне число господарств	Нормоване відхилення	Значення функції нормального розподілу	Теоретичні (очікувані) частоти (із заокругленням)	Різниці частот	Квадрат різниці	Зважені квадрати різниці
x_i	n_i	$t = \frac{x_i - \bar{x}}{S}$	$f(t)$	$\tilde{n}_i$	$n_i - \tilde{n}_i$	$(n_i - \tilde{n}_i)^2$	$\frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$
27	8	– 1,75	0,0863	6	2	4	0,667
29	16	– 1,12	0,2131	14	2	4	0,286
31	17	– 0,50	0,3521	22	– 5	25	1,136
33	25	0,12	0,3961	25	0	0	0,000
35	18	0,75	0,3011	19	– 1	1	0,053
37 39	11 } 16 5	1,37 2,00	0,1561 0,0540	10 } 14 4	2	4	0,268
Разом	**100**	–	–	**100**	**0**	–	**2,428**

Значення постійного виразу $\frac{h}{S}$ для даної задачі становитиме

$$\tilde{n} = n\frac{h}{S} = 100\frac{2}{3,2} = 62,5.$$

Тоді теоретичні частоти $\tilde{n}_i$ нормального розподілу становитимуть: для першого інтервалу $62,5 \cdot 0,0863 \approx 6$; для другого інтервалу $62,5 \cdot 0,2131 \approx$ ≈ 14 і т.д.

Інтервали з числом одиниць менше 5 необхідно об'єднати. Тому два останніх інтервали укрупнимо. Підрахуємо суми теоретичних і фактичних частот і перевіримо їх рівність загальному підсумку ($n = 100$).

7. Обчислимо фактичне значення критерію χ^2. Послідовність обчислення χ^2 наведена в трьох останніх колонках (табл. 7.4).

$$\chi^2_{\text{факт}} = \sum_{i=1}^{l} \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i} = 2,428.$$

Такий самий результат дістанемо за іншою формулою

$$\chi^2_{\text{факт}} = \sum_{i=1}^{l} \frac{n_i^2}{\tilde{n}_i} - n = \frac{8^2}{6} + \frac{16^2}{14} + \frac{17^2}{22} + \frac{25^2}{25} + \frac{18^2}{19} + \frac{16^2}{14} - 100 = 102,428 - 100 = 2,428.$$

8. Для встановлення табличного значення χ^2 необхідно визначити число ступенів свободи *k*. Визначимо його як число інтервалів (груп, класів, з врахуванням укрупнення) без одиниці і мінус дві постійні величини, які описують криву нормального розподілу ($\tilde{x}$ і σ):

$$k = l - 1 - S = 6 - 1 - 2 = 3.$$

За таблицею (дод. 6) при $\alpha = 0,05$ встановимо табличне значення $\chi^2_{0,05} = 7,815$.

9. Порівняємо фактичне значення χ^2, розраховане за даними вибірки, з табличним:

$$\chi^2_{\text{факт}} < \chi^2_{0,05};\ \ 2,428 < 7,815.$$

Оскільки, $\chi^2_{0,05} > \chi^2_{\text{факт}}$, нульова гіпотеза про нормальний розподіл господарств за надоями молока на корову приймається. Іншими словами, фактичні дані узгоджуються з нульовою гіпотезою.

7.6. Перевірка статистичної гіпотези про істотність розбіжностей між дисперсіями

Задача перевірки гіпотези про рівність дисперсій виникає досить часто. Наприклад, при аналізі стабільності виробничого процесу до і після впровадження нової техніки (коливання у випуску продукції вимірюється за допомогою середнього квадратичного відхилення), при вивченні ступеня однорідності двох сукупностей відносно будь-якої ознаки (стажу роботи, рівня продуктивності праці, продуктивності тварин тощо). Найчастіше необхідність перевірки гіпотези про рівність дисперсій виникає при порівнянні середніх величин сукупностей, оскільки при цьому в більшості випадків передбачається, що генеральні дисперсії рівні. Так як вибіркові дисперсії, як правило, нерівні, в ході перевірки статистичної гіпотези про рівність середніх необхідно перевірити гіпотезу про істотність різниці дисперсій.

Гіпотеза про рівність дисперсій перевіряється за допомогою *F*-критерію Фішера, який являє собою відношення двох вибіркових дисперсій S_1^2 і S_2^2 при відповідних ступенях свободи варіації k_1 і k_2:

$$F = \frac{S_1^2}{S_2^2},\ \text{де } S_1^2 > S_2^2.$$

Цей критерій докладно розглядається при вивченні дисперсійного аналізу (розділ 8). Тому ми тут обмежимося лише розглядом схеми застосування критерію.

Критичне значення F_α знаходиться в спеціальних таблицях (дод. 4 і 5) з урахуванням зазначених ступенів свободи при заданому рівні значущості α. Для зручності обчислень ці таблиці складені для відношень більшої дисперсії до меншої. Тому при визначенні вибіркового значення $F_{факт}$ завжди більшу вибіркову дисперсію необхідно ділити на меншу. Отже, завжди $F > 1$.

Для перевірки нульової гіпотези про рівність дисперсій в генеральних сукупностях $(H_0 : \sigma_1^2 = \sigma_2^2;\ H_a : \sigma_1^2 \neq \sigma_2^2)$ слід обчислити фактичне значення F-критерію і порівняти його з табличним значенням. Оскільки дисперсії генеральних сукупностей невідомі, то перевірку висунутої гіпотези здійснюють на основі порівняння вибіркових дисперсій S_1^2 і S_2^2. При цьому слід мати на увазі, що якщо відношення $S_1^2 : S_2^2$ близьке до одиниці, то, очевидно, немає підстав для відхилення нульової гіпотези. Якщо це відношення значно відрізняється від одиниці, то даний факт може бути підставою для її відхилення.

Перевірка гіпотези H_0 зводиться до такого правила: якщо фактичне значення критерію потрапляє в область припустимих значень ($F_{факт} < F_\alpha$), нульова гіпотеза про рівність дисперсій в генеральних сукупностях приймається; якщо ж фактичне значення критерію потрапляє в критичну область ($F_{факт} > F_\alpha$), то від нульової гіпотези слід відмовитися.

Якщо альтернативна гіпотеза формулюється так: , то $H_a : \sigma_1^2 \neq \sigma_2^2$ використовується двосторонній критерій. Якщо ж відомо, що одна з дисперсій передбачувано більше другої, то альтернативна гіпотеза $H_a : \sigma_1^2 > \sigma_2^2$ і використовується односторонній критерій.

Порядок перевірки гіпотези про істотність різниці між дисперсіями розглянемо на такому прикладі. Нехай шляхом направленого відбору з місцевого стада овець виведена нова лінія, що характеризується більшою плодючістю на одну матку. Коротко вихідну сукупність назвемо групою І, а нову лінію – групою ІІ.

Припустимо, встановлено, що дисперсія в групі ІІ менша (група більш однорідна).

Для перевірки істотності різниці дисперсій відібрано таку кількість тварин:

$$n_1 = 25;\ n_2 = 18,$$

при цьому $S_1^2 = 12{,}8;\ S_2^2 = 4{,}6.$

Сформулюємо нульову і альтернативну гіпотези:

$$\left(H_0 : \sigma_1^2 = \sigma_2^2;\ H_a : \sigma_1^2 > \sigma_2^2\right).$$

Рівень значущості приймемо рівним $\alpha = 0{,}05$.

Визначимо фактичне значення F-критерію:

$$F_{\text{факт}} = \frac{S_1^2}{S_2^2} = \frac{12{,}8}{4{,}6} = 2{,}78.$$

Табличне значення F-критерію при $k_1 = n_1 - 1 = 25 - 1 = 24$ і $k_2 = n_2 - 1 = 18 - 1 = 17$ і $\alpha = 0{,}05$ становить $F_{0,05} = 2{,}19$ (дод. 4).

Оскільки $F_{\text{факт}} > F_{0,05}$; $2{,}78 > 2{,}19$ нульова гіпотеза про рівність дисперсій відхиляється. Це означає, що дисперсія ознаки в II групі суттєво нижче.

Для великих вибірок перевірка гіпотези про рівність дисперсій може бути проведена на основі t-критерію нормального розподілу:

$$t = \frac{S_1 - S_2}{\overline{\mu}_{1-2}},$$

де $\overline{\mu}_{1-2} = \sqrt{\mu_1^2 + \mu_2^2}$; $\mu_1 = \dfrac{S_1}{\sqrt{2(n_1 - 1)}}$; $\mu_2 = \dfrac{S_1}{\sqrt{2(n_2 - 1)}}$.

Тут ми розглянули лише випадок перевірки гіпотези про рівність двох дисперсій. Якщо ж потрібна оцінка істотності відмінностей ряду дисперсій, то застосовуються інші критерії:

а) критерій Кохрана, якщо чисельності вибірок рівні;

б) критерій Бартлетта, якщо чисельності вибірок нерівні.

Завдання для самоконтролю до розділу 7

1. Що називають статистичною гіпотезою?
2. Дайте визначення нульової і альтернативної гіпотези.
3. Охарактеризуйте помилки першого і другого роду, їх взаємозв'язок.
4. Що називається статистичним критерієм? Назвіть їх види і охарактеризуйте.
5. Що таке критична область і область припустимих значень?
6. Що називається потужністю критерію?
7. Наведіть загальну схему перевірки статистичної гіпотези.
8. Наведіть алгоритм перевірки статистичних гіпотез щодо середніх величин.
9. Назвіть етапи перевірки статистичних гіпотез щодо розподілів.

10. Які висновки треба зробити щодо прийняття або відхилення нульової гіпотези, якщо фактичне значення критерію більше фактичного значення і навпаки?

11. По двох сортах капусти урожайність досліджувалась 6 років. По першому сорту середня урожайність за 6 років становила 300 ц/га при дисперсії 36, по другому – відповідно 350 і 42 ц/га. Встановіть істотність різниці в урожайності капусти при рівні значущості $\alpha = 0{,}05$.

Розділ 8

Дисперсійний аналіз

8.1. Теоретичні основи і принципова схема дисперсійного аналізу

Розглянуті вище прийоми перевірки статистичних гіпотез щодо істотності відмінностей між двома середніми на практиці мають обмежене застосування. Це пов'язано з тим, що для виявлення дії всіх можливих умов і факторів на результативну ознаку польові та лабораторні досліди, як правило, проводять із застосуванням не двох, а більшого числа вибірок (12-20 та більше).

Часто дослідники порівнюють середні кількох вибірок, об'єднаних у єдиний комплекс. Наприклад, вивчаючи вплив різних видів і доз добрив на урожайність сільськогосподарських культур досліди повторюють у різних варіантах. У цих випадках попарні порівняння стають громіздкими, а статистичний аналіз усього комплексу потребує застосування особливого методу. Такий метод, розроблений в математичній статистиці, дістав назву дисперсійного аналізу. Вперше його застосував англійський статистик Р.Фішер при обробці результатів агрономічних дослідів (1938 р.).

Дисперсійний аналіз – це метод статистичної оцінки надійності прояву залежності результативної ознаки від одного або кількох факторів. За допомогою методу дисперсійного аналізу проводиться перевірка статистичних гіпотез відносно середніх в кількох генеральних сукупностях, які мають нормальний розподіл.

Дисперсійний аналіз є одним з основних методів статистичної оцінки результатів експерименту. Все більш широке застосування отримує він і в аналізі економічної інформації. Дисперсійний аналіз дає змогу встановити, наскільки вибіркові показники зв'язку результативних і факторних ознак достатні для поширення одержаних за вибіркою даних на генеральну сукупність. Достоїнством цього методу є те, що він дає досить надійні висновки по вибірках невеликої чисельності.

Досліджуючи варіацію результативної ознаки під впливом одного або кількох факторів, за допомогою дисперсійного аналізу можна одержати, крім загальних оцінок істотності залежностей, також і оцінку відмінностей у величині середніх, що формуються при різних рівнях факторів, та істотності взаємодії факторів. Дисперсійний аналіз застосовується для вивчення залежностей як кількісних, так і якісних ознак, а також при їх поєднанні.

Суть цього методу полягає в статистичному вивченні вірогідності впливу одного або кількох факторів, а також їх взаємодії на результативну ознаку. Відповідно до цього за допомогою дисперсійного аналізу вирішуються три основних завдання: 1) загальна оцінка істотності відмінностей між груповими середніми; 2) оцінка вірогідності взаємодії факторів; 3) оцінка істотності відмінностей між парами середніх. Найчастіше такі завдання доводиться вирішувати дослідникам при проведенні польових і зоотехнічних дослідів, коли вивчається вплив кількох факторів на результативну ознаку одночасно.

Принципова схема дисперсійного аналізу включає встановлення основних джерел варіювання результативної ознаки і визначення обсягів варіації (сум квадратів відхилень) за джерелами її утворення; визначення числа ступенів свободи, що відповідають компонентам загальної варіації; обчислення дисперсій як відношення відповідних обсягів варіації до їх числа ступенів свободи; аналіз співвідношень між дисперсіями; оцінка вірогідності різниці між середніми і формулювання висновків.

Зазначена схема зберігається як при простих моделях дисперсійного аналізу, коли дані групуються за однією ознакою, так і при складних моделях, коли дані групуються за двома і більшим числом ознак. Однак із збільшенням числа групувальних ознак ускладнюється процес розкладання загальної варіації за джерелами її утворення.

Відповідно до принципової схеми дисперсійний аналіз можна подати у вигляді п'яти послідовно виконуваних етапів:

1) визначення і розкладання варіації;
2) визначення числа ступенів свободи варіації;
3) обчислення дисперсій та їх співвідношень;
4) аналіз дисперсій та їх співвідношень;
5) оцінка вірогідності різниці між середніми і формулювання висновків з перевірки нульової гіпотези.

Найбільш трудомісткою частиною дисперсійного аналізу є перший етап – визначення і розкладання варіації за джерелами її утворення. Порядок розкладання загального обсягу варіації докладно розглядався в розділі 5.

В основі розв'язування задач дисперсійного аналізу лежить закон розкладання (додавання) варіації, відповідно до якого загальна варіація (коливання) результативної ознаки поділяється на дві: варіацію, зумовлену дією досліджуваного фактора (факторів), і варіацію, викликану дією випадкових причин, тобто

$$\sigma^2_{заг} = \sigma^2_{м.гр} + \sigma^2_{в.гр}.$$

Припустимо, що досліджувана сукупність поділена за факторною ознакою на кілька груп, кожна з яких характеризується своєю середньою величиною результативної ознаки. При цьому варіацію цих величин можна пояснити двома видами причин: такими, які діють на результативну ознаку систематично і піддаються регулюванню в ході здійснюваного експерименту, і такими, які регулюванню не піддаються. Очевидно, що міжгрупова (факторна або систематична) варіація залежить переважно від дії досліджуваного фактора, а внутрішньогрупова (залишкова або випадкова) – від дії випадкових факторів.

Щоб оцінити вірогідність відмінностей між груповими середніми, необхідно визначити міжгрупову та внутрішньогрупову варіації. Якщо міжгрупова (факторна) варіація значно перевищує внутрішньогрупову (залишкову) варіацію, то фактор впливає на результативну ознаку, істотно змінюючи значення групових середніх величин. Але виникає питання, яке співвідношення між міжгруповою і внутрішньогруповою варіаціями можна розглядати як достатнє для висновку про вірогідність (істотність) відмінностей між груповими середніми.

Для оцінки істотності відмінностей між середніми і формулювання висновків з перевірки нульової гіпотези ($H_0 : \bar{x}_1 = \bar{x}_2 = \ldots = \bar{x}_n$) в дисперсійному аналізі використовується своєрідний норматив – F-критерій, закон розподілу якого встановив Р.Фішер. Цей критерій являє собою відношення двох дисперсій: факторної, породжуваної дією досліджуваного фактора, та залишкової, зумовленої дією випадкових причин:

$$F = \frac{S_1^2}{S_2^2}, \text{ де } S_1^2 > S_2^2.$$

Дисперсійне відношення $F = s_1^2 : s_2^2$ американським статистиком Снедекором запропоновано позначати літерою F на честь винахідника дисперсійного аналізу Р.Фішера.

Дисперсії s_1^2 і s_2^2 є оцінками дисперсії генеральної сукупності. Якщо вибірки з дисперсіями s_1^2 і s_2^2 зроблені з однієї і тієї самої генеральної сукупності, де варіація величин мала випадковий характер, то розбіжність у величинах s_1^2 і s_2^2 також випадкова.

Якщо в експерименті перевіряють вплив кількох факторів (A, B, C і т.д.) на результативну ознаку одночасно, то дисперсія, що обумовлена дією кожного з них, має бути порівняна з , тобто

$$F_{\text{факт}}(A)=\frac{S_A^2}{S_{\text{в.гр}}^2};\quad F_{\text{факт}}(B)=\frac{S_B^2}{S_{\text{в.гр}}^2};\quad F_{\text{факт}}(C)=\frac{S_C^2}{S_{\text{в.гр}}^2}.$$

Якщо значення факторної дисперсії значно більше залишкової, то фактор істотно впливає на результативну ознаку і навпаки.

В багатофакторних експериментах крім варіації, зумовленої дією кожного фактора, практично завжди є варіація, яка зумовлена взаємодією факторів $\left(S_{AB}^2,\ S_{AC}^2,\ S_{BC}^2,\ S_{ABC}^2\right)$. Суть взаємодії полягає в тому, що ефект одного фактора істотно змінюється на різних рівнях другого (наприклад, ефективність якості ґрунту при різних дозах добрив).

Взаємодія факторів також має бути оцінена шляхом порівняння відповідних дисперсій з $S_{\text{в.гр}}^2$:

$$F_{\text{факт}}(AB)=\frac{S_{AB}^2}{S_{\text{в.гр}}^2};\qquad F_{\text{факт}}(AC)=\frac{S_{AC}^2}{S_{\text{в.гр}}^2};$$

$$F_{\text{факт}}(BC)=\frac{S_{BC}^2}{S_{\text{в.гр}}^2};\qquad F_{\text{факт}}(ABC)=\frac{S_{ABC}^2}{S_{\text{в.гр}}^2}.$$

При обчисленні фактичного значення F-критерію в чисельнику береться більша з дисперсій, тому $F > 1$. Очевидно, що чим більше критерій F, тим значнішим є розбіжності між дисперсіями. Якщо $F = 1$, то питання про оцінку істотності відмінностей дисперсій знімається.

Для визначення меж випадкових коливань відношення дисперсій Р.Фішер розробив спеціальні таблиці F-розподілу (дод. 4 і 5). Критерій F функціонально зв'язаний з імовірністю і залежить від числа ступенів свободи варіації k_1 і k_2 двох порівнюваних дисперсій. Звичайно використовуються дві таблиці, що дозволяють робити висновки про гранично високе значення критерію F для рівнів значущості 0,05 і 0,01. Рівень значущості 0,05 (або 5%) означає, що тільки в 5 випадках із 100 критерій F може приймати значення, що дорівнює вказаному в таблиці або вище його. Зниження рівня значущості з 0,05 до 0,01 призводить до збільшення значення критерію F між двома дисперсіями в силу дії тільки випадкових причин.

Значення критерію F також залежить безпосередньо від числа ступенів свободи двох порівнюваних дисперсій. Якщо число ступенів свободи прямує до нескінченості ($k \to \infty$), то відношення F для двох дисперсій прямує до одиниці.

Табличне значення критерію F показує можливу випадкову величину відношення двох дисперсій при заданому рівні значущості і відповідному числі ступенів свободи для кожної з порівнюваних дисперсій. В зазначених таблицях наводиться величина F для вибірок, зроблених з однієї і тієї самої генеральної сукупності, де причини зміни величин тільки випадкові.

Значення F знаходять за таблицями (дод. 4 і 5) на перетині відповідного стовпця (число ступенів свободи для більшої дисперсії – k_1) і рядка (число ступенів свободи для меншої дисперсії – k_2). Так, якщо більшій дисперсії (чисельник F) $k_1 = 3$, а меншій (знаменник F) $k_2 = 9$, то $F_б$ при рівні значущості $\alpha = 0{,}05$ становитиме 3,63 (дод. 4). Отже, в результаті дії випадкових причин, оскільки вибірки малочисельні, дисперсія однієї вибірки може при 5%-му рівні значущості перевищувати дисперсію другої вибірки в 3,63 рази. При зниженні рівня значущості з 0,05 до 0,01 табличне значення критерію F, як зазначалося вище, буде збільшуватися. Так, при тих самих ступенях свободи $k_1 = 3$ і $k_2 = 9$ і $\alpha = 0{,}01$ табличне значення критерію F становитиме 6,99 (дод. 5).

Розглянемо порядок визначення числа ступенів свободи в дисперсійному аналізі. Число ступенів свободи, що відповідає загальній сумі квадратів відхилень, розкладається на відповідні компоненти аналогічно розкладанню сум квадратів відхилень ($W_{заг} = W_{м.гр} + W_{в.гр}$), тобто загальне число ступенів свободи (k_0) розкладається на число ступенів свободи для міжгрупової (k_1) і внутрішньогрупової (k_2) варіацій.

Так, якщо вибіркова сукупність, що складається з N спостережень, поділена на m груп (число варіантів досліду) і n підгруп (кількість повторностей), то число ступенів свободи k відповідно становитиме:

а) для загальної суми квадратів відхилень $(W_{заг})$

$$k_0 = m \cdot n - 1 = N - 1;$$

б) для міжгрупової суми квадратів відхилень $(W_{м.гр})$

$$k_1 = m - 1;$$

в) для внутрішньогрупової суми квадратів відхилень $(W_{в.гр})$

$$k_2 = k_0 - k_1 = (N - 1) - (m - 1) = N - m,$$

або $k_2 = m\,(n - 1)$.

Згідно з правилом додавання варіації:

$$k_0 = k_1 + k_2;\; N - 1 = (m - 1) + (N - 1) - (m - 1) = N - 1.$$

Наприклад, якщо в досліді сформовано чотири варіанти досліду ($m = 4$) у п'яти повторностях кожен ($n = 5$), і загальна кількість спостережень $N = m \cdot n = 4 \cdot 5 = 20$, то число ступенів свободи відповідно дорівнює:

$k_0 = N - 1 = 20 - 1 = 19$; $k_1 = m - 1 = 4 - 1 = 3$;

$k_2 = k_0 - k_1 = 19 - 3 = 16$; або $k_2 = N - m = 20 - 4 = 16$;

або $k_2 = m\,(n - 1) = 4\,(5 - 1) = 16$.

Перевіримо розрахунок: $k_0 = k_1 + k_2$; $19 = 3 + 16$.

Знаючи суми квадратів відхилень і число ступенів свободи, можна визначити незміщені (скориговані) оцінки для трьох дисперсій:

$$S^2_{заг} = \frac{W_0}{k_0} = \frac{W_0}{N-1}; \qquad S^2_{м.гр} = \frac{W_{м.гр}}{k_1} = \frac{W_{м.гр}}{m-1};$$

$$S^2_{в.гр} = \frac{W_{в.гр}}{k_2} = \frac{W_{в.гр}}{N-m}.$$

Нульову гіпотезу H_0 за критерієм F перевіряють так само, як і за t-критерієм Стьюдента. Щоб прийняти рішення з перевірки H_0, необхідно розрахувати фактичне значення критерію $F_{факт}$ і порівняти його з табличним значенням F_α для прийнятого рівня значущості α і відповідного числа ступенів свободи k_1 і k_2 для двох дисперсій.

Якщо $F_{факт} > F_\alpha$, то згідно з прийнятим рівнем значущості можна зробити висновок, що відмінності вибіркових дисперсій визначаються не лише випадковими факторами; вони істотні. Нульову гіпотезу в цьому випадку відхиляють і є підстава стверджувати, що фактор істотно впливає на результативну ознаку. Якщо ж $F_{факт} < F_\alpha$, то нульову гіпотезу приймають і є підстава стверджувати, що відмінності між порівнюваними дисперсіями знаходяться в межах можливих випадкових коливань: дія фактора на результативну ознаку не є істотною.

Застосування тієї або іншої моделі дисперсійного аналізу залежить як від кількості досліджуваних факторів, так і від способу формування вибірок.

Залежно від кількості факторів, що визначають варіацію результативної ознаки, вибірки можуть бути сформовані за однією, двома і більшим числом факторів. Відповідно до цього дисперсійний аналіз поділяється на однофакторний і багатофакторний. Інакше його ще називають однофакторним і багатофакторним дисперсійним комплексом.

Схема розкладання загальної варіації залежить від формування груп. Воно може бути випадковим (спостереження однієї групи не пов'язані із спостереженнями другої групи) і невипадковим (спостереження двох вибірок пов'язані між собою спільністю умов експерименту). Відповідно дістають незалежні і залежні вибірки. Незалежні вибірки можуть бути сформовані як з рівною, так і з нерівною чисельністю. Формування залежних вибірок передбачає їх рівну чисельність.

Якщо групи сформовані у невипадковому порядку, то загальний обсяг варіації результативної ознаки включає в себе поряд з факторною (міжгруповою) і залишковою варіацією варіацію повторностей, тобто $W_{заг} = W_{м.гр} + W_{в.гр} + W_{повт}$.

На практиці в більшості випадків доводиться розглядати залежні вибірки, коли умови для груп і підгруп вирівнюються. Так, у польовому досліді всю ділянку розбивають на блоки з максимально вирівняннями умовами. При цьому кожен варіант досліду отримує рівні можливості бути представленим в усіх блоках, чим досягається вирівнювання умов для всіх перевірюваних варіантів досліду. Такий метод побудови досліду дістав назву методу **рендомізованих блоків.** Аналогічно проводяться і досліди з тваринами.

При обробці методом дисперсійного аналізу соціально-економічних даних необхідно мати на увазі, що через багаточисельність факторів та їх взаємозв'язок важко навіть при самому ретельному вирівнюванні умов встановити ступінь об'єктивного впливу кожного окремого фактора на результативну ознаку. Тому рівень залишкової варіації зумовлюється не тільки випадковими причинами, але й суттєвими факторами, які не були враховані при побудові моделі дисперсійного аналізу. Внаслідок цього залишкова дисперсія, як база порівняння, інколи стає неадекватною своєму призначенню, вона явно завищується за величиною і не може виступати як критерій істотності впливу факторів. В зв'язку з цим при побудові моделей дисперсійного аналізу стає актуальною проблема відбору найважливіших факторів і вирівнювання умов для проявлення дії кожного з них. Крім того, застосування дисперсійного аналізу передбачає нормальний або близький до нормального розподіл досліджуваних статистичних сукупностей. Якщо ця умова не витримується, то оцінки, одержані в дисперсійному аналізі, виявляться перебільшеними.

8.2. Дисперсійний аналіз при групуванні даних за однією ознакою

Порядок проведення дисперсійного аналізу при групуванні даних за однією ознакою розглянемо на такому прикладі.

В досліді вивчався вплив нових комбікормів на прирости живої маси бройлерів (табл. 8.1).

Аналіз даних таблиці показує, що середні прирости живої маси вище по групах бройлерів, що одержували дослідний комбікорм і дослідний комбікорм з добавкою вітамінів. Отже, варіація середніх по варіантах досліду зумовлювалась різним складом раціонів. Але прирости живої маси варіювали і всередині груп, тобто мала місце внутрішньогрупова (залишкова) варіація, викликана рештою неврахованих факторів. Про це свідчать середні прирости живої маси по повторностях: вони коливаються від 34 до 38 г.

Таблиця 8.1

Середньодобові прирости живої маси бройлерів, г

Варіант досліду	Повторність					Сума Σx_j	Середні $\bar{x}_j$
	1	2	3	4	5		
Стандартний комбікорм	30	34	31	33	32	160	32
Дослідний комбікорм	32	36	33	34	35	170	34
Дослідний комбікорм + + добавка вітамінів	39	43	37	36	40	195	39
Сума Σx_i	**101**	**113**	**101**	**103**	**107**	**525**	**35**
Середні $\bar{x}_i$	**34**	**38**	**34**	**34**	**36**	**35**	*x*

На основі вихідних даних методом дисперсійного аналізу перевіримо вірогідність впливу різних за якістю раціонів на прирости живої маси бройлерів. Для цього висунемо і перевіримо статистичну гіпотезу відносно середніх в генеральних сукупностях.

Сформулюємо нульову і альтернативну гіпотези:

$$H_0 : \bar{x}_1 = \bar{x}_2 = \bar{x}_3; \;\; H_a : \bar{x}_1 \neq \bar{x}_2 \neq \bar{x}_3.$$

Рівень значущості приймемо рівним $\alpha = 0{,}05$. Найпотужнішим критерієм перевірки H_0 є F-критерій Фішера. Для перевірки H_0 і формулювання висновків за результатами дисперсійного аналізу необхідно обчислити фактичне значення F-критерію Фішера і порівняти його з табличним значенням F_α.

Для розрахунку фактичного значення F-критерію виконаємо всі необхідні операції по його обчисленню відповідно до етапів дисперсійного аналізу.

При групуванні даних за однією ознакою загальний обсяг варіації можна розкласти на варіацію, пов'язану з дією групувальної ознаки (міжгрупову або систематичну), і варіацію внутрішньогрупову (залишкову):

$$W_{заг} = W_{м\,гр} + W_{в.гр}$$

Для визначення сум квадратів відхилень піднесемо до квадрату середньодобові прирости і суми приростів (табл. 8.2).

Позначимо загальне число спостережень $N = 15$, число варіантів досліду $m = 3$, число повторностей $n = 5$, $N = m \cdot n = 3 \cdot 5 = 15$.

На підставі одержаних в табл. 8.2 даних обчислимо:

$$W_{заг} = \Sigma x_{ij}^2 - \frac{(\Sigma x_{ij})^2}{N} = 18555 - \frac{525^2}{15} = 180;$$

Таблиця 8.2

Квадрати середньодобових приростів живої маси

Варіант досліду	Повторність					Сума Σx_j^2	Середні $\left(\Sigma x_j\right)^2$
	1	2	3	4	5		
Стандартний комбікорм	900	1156	961	1089	1024	5130	25600
Дослідний комбікорм	1024	1296	1089	1156	1225	5790	28900
Дослідний комбікорм + + добавка вітамінів	1521	1849	1369	1296	1600	7635	38025
Сума Σx_i^2	**3445**	**4301**	**3419**	**3541**	**3849**	**18555**	**92525**
Середні $\left(\Sigma x_i\right)^2$	**10201**	**12769**	**10201**	**10609**	**11449**	**55229**	**–**

$$W_{\text{рац}} = \frac{\Sigma(\Sigma x_{ij})^2}{n} - \frac{(\Sigma x_{ij})^2}{N} = \frac{92525}{5} - \frac{525^2}{15} = 130;$$

$$W_{\text{зал}} = W_{\text{заг}} - W_{\text{рац}} = 180 - 130 = 50.$$

Визначення і розкладання варіації дало такі результати:

$$W_{\text{заг}} = W_{\text{рац}} + W_{\text{зал}};$$

$$180 = 130 + 50\,;$$

$$100{,}0\% = 72{,}2\% + 27{,}8\%.$$

Отже, 72,2% загальної варіації середньодобових приростів бройлерів у досліді припадає на варіацію досліджуваного фактора (раціонів), а 27,8% варіації зумовлено неврахованими факторами.

Встановимо число ступенів свободи варіації для кожної суми квадратів відхилень при $N = 15$, $m = 3$, $n = 5$.

Тоді число ступенів свободи варіації для загальної суми квадратів відхилень

$$k_0 = N - 1 = 15 - 1 = 14;$$

для міжгрупової (раціонів)

$$k_1 = m - 1 = 3 - 1 = 2;$$

для залишкової визначається за різницею так само, як і залишкова варіація:

$$k_2 = k_0 - k_1 = (N - 1) - (m - 1) = 14 - 2 = 12.$$

Знаючи суми квадратів відхилень і ступені свободи, визначимо дисперсії, як відношення суми квадратів відхилень до відповідного числа ступенів свободи варіації.

Для дисперсійного аналізу викликає інтерес міжгрупова і залишкова дисперсії, а загальна дисперсія в аналізі участі не бере, тому її не обчислюємо:

$$S^2_{\text{рац}} = \frac{W_{\text{рац}}}{k_1} = \frac{130}{2} = 65{,}00; \qquad S^2_{\text{зал}} = \frac{W_{\text{зал}}}{k_2} = \frac{50}{12} = 4{,}17.$$

Зіставимо дисперсії, тобто знайдемо фактичне значення F-критерію Фішера:

$$F_{\text{факт}} = \frac{S^2_{\text{рац}}}{S^2_{\text{зал}}} = \frac{65{,}00}{4{,}17} = 15{,}59.$$

Для перевірки нульової гіпотези необхідно визначити табличне значення F-критерію Фішера (дод. 4) і порівняти його з одержаним фактичним значенням.

Більшій дисперсії $S^2_{\text{рац}}$ відповідає число ступенів свободи варіації $k_1 = 2$ (чисельник відношення), меншій дисперсії $S^2_{\text{зал}}$ число ступенів свободи варіації $k_2 = 12$ (знаменник відношення). Отже, згідно з дод. 4 теоретичне (табличне) значення F-критерію знаходиться на перетині другого стовпця і дванадцятого рядка: $F_{0,05} = 3{,}88$.

Порівняємо фактичне і табличне значення F-критерію:

$$F_{\text{факт}} > F_{0,05}\,; \; 15{,}59 > 3{,}88.$$

Оскільки $F_{\text{факт}} > F_{0,05}$, висунута нульова гіпотеза щодо випадкових розбіжностей в групових середніх має бути відхилена і прийнята альтернативна гіпотеза: значення генеральних середніх істотно відрізняються. Іншими словами, фактичні дані не узгоджуються з нульовою гіпотезою. Отже, вплив різних за якістю раціонів на середньодобові прирости живої маси бройлерів вірогідний та істотний.

В дисперсійному аналізі при розкладанні загальної варіації результативної ознаки за компонентами необхідно враховувати порядок формування вибірок: залежні чи незалежні. Якщо в прикладі, що розглядується, вибірки сформовані як незалежні (тобто в кожну групу відібрано по п'ять тварин у випадковому порядку), то загальна сума квадратів відхилень поділяється на два компонента: $W_{\text{рац}}$ і $W_{\text{зал}}$. Однак частіше в дослідах з тваринами формуються групи тварин-аналогів відповідно до числа повторностей по кожному варіанту досліду. Кожен варіант досліду має за повторність – тварину-представника з групи аналогів.

При проведенні експерименту (розглянутий вище приклад) з перевірки впливу різних раціонів на прирости живої маси бройлерів могло бути сформовано п'ять груп тварин-аналогів, оскільки дослід мав проводиться у п'ятикратному повторенні. Групи складалися з трьох голів відповідно до варіантів раціону і кожний варіант досліду мав представника з однієї і тієї самої групи тварин-аналогів. Отже, середні по повторностях (по графах табл. 8.1) відбивають тоді відмінності між виділеними групами. Ці відмінності необхідно виключити із загального варіювання приростів, оскільки вони не є випадковими, не пов'язані із зміною раціонів і певною мірою можуть затушовувати вплив на прирости досліджуваного фактора.

Розкладання загальної варіації результативної ознаки при залежному формуванні вибірок ведеться за схемою:

$$W_{заг} = W_{рац} + W_{повт} + W_{зал},$$

тобто на відміну від незалежного формування вибірок включає ще й варіацію повторностей.

За даними табл. 8.2 обчислимо суму квадратів відхилень повторностей:

$$W_{повт} = \frac{\Sigma(\Sigma x_j)^2}{m} - \frac{(\Sigma x_{ij})^2}{N} = \frac{55229}{3} - \frac{525^2}{15} = 34{,}7.$$

Цій сумі квадратів відхилень відповідає $k_{повт} = n - 1 = 5 - 1 = 4$, тоді дисперсія повторностей становитиме:

$$S^2_{повт} = \frac{W_{повт}}{k_{повт}} = \frac{34{,}7}{4} = 8{,}68.$$

Виключення із загальної суми квадратів відхилень $W_{повт}$ призводить і до зміни $W_{зал}$ відповідних їй ступенів свободи варіації:

$$W_{зал} = W_{заг} - W_{рац} - W_{повт} = 180 - 130 - 34{,}7 = 15{,}3;$$

$$k_{зал} = (N - 1) - (m - 1) - (n - 1) = 14 - 2 - 4 = 8.$$

а отже, і залишкової дисперсії

$$S^2_{зал} = \frac{W_{зал}}{k_{зал}} = \frac{15{,}3}{8} = 1{,}91.$$

Зіставимо дисперсію раціонів і повторностей із залишковою дисперсією. Результати розрахунків оформимо в табл. 8.3.

Таблиця 8.3

Аналіз дисперсій

Джерело варіації	Сума квадратів відхилень	Ступінь свободи варіації	Дисперсія	Відношення дисперсій	
	W_i	k_i	S_i^2	$F_{факт}$	$F_{0,05}$
Раціони	130,0	2	65,00	34,03	4,46
Повторності	34,7	4	8,68	4,54	3,84
Залишкова	15,3	8	1,91	1	–
Загальна	180,0	14	×	–	–

7*

Табличне значення F-критерію для оцінки відношення дисперсії раціонів до залишкової дисперсії знаходимо на перетині другого стовпця і восьмого рядка – $F_{0,05} = 4,46$, а для відношення дисперсії повторностей до залишкової – на перетині четвертого стовпця і восьмого рядка – $F_{0,05} = 3,84$ (дод. 4). Порівняння $F_{факт}$ з табличними значеннями призводить до відмови від нульової гіпотези. Отже, відмінності між середніми за раціонами і повторностями не можуть бути віднесені на рахунок випадкового варіювання.

8.3. Застосування дисперсійного аналізу для оцінки вірогідності різниці двох середніх

Критерій F дозволяє встановити наявність або відсутність істотних зв'язків між груповими середніми в цілому, однак він не показує, між якими середніми різниця істотна, а між якими не істотна.

Тому, якщо проведений дисперсійний аналіз призвів до відмови від нульової гіпотези, що передбачає рівність середніх, та показав істотність впливу фактора, що перевіряється, на результативну ознаку, то цей загальний висновок може бути конкретизований, доповнений перевіркою істотності відмінностей між парами середніх.

Для перевірки нульової гіпотези про вірогідність відмінностей між парами середніх використовується ряд методик і критеріїв.

Для випадку, коли чисельності вибірок (груп) між собою рівні ($n_1 = n_2 = \ldots = n_k$) і групування проведене за якісною ознакою, найточніша оцінка вірогідності відмінностей між середніми може бути здійснена при використанні методики Хартлі. В свою чергу ця методика ґрунтується на використанні критерію Q Тьюкі. При нерівних чисельностях вибірок доцільно застосовувати метод контрастів Шеффе. При невеликій кількості варіантів досліду (не більше трьох) можливе використання t-критерію Стьюдента. При багатоваріантному досліді (більше трьох) застосування цього критерію дає збільшену кількість вірогідних зв'язків.

Методика застосування Q-критерію Тьюкі така:

1) будують ранжирований ряд групових середніх;

2) визначають попарні різниці між середніми, спочатку між середніми, що стоять поряд, потім між середніми, віддаленими одна від одної на дві, три позиції і т. д.;

3) обчислюють фактичне значення Q-критерію Тьюкі як відношення знайдених різниць до середньої помилки вибірки:

$$Q_{\text{факт}} = \frac{\left|\tilde{x}_i - \tilde{x}_j\right|}{\sqrt{\frac{S^2_{\text{зал}}}{n}}},$$

де $S^2_{\text{зал}}$ – дисперсія, обчислена у дисперсійному аналізі, n – кількість спостережень у кожному варіанті досліду (кількість повторностей);

4) визначають табличне значення Q-критерію Тьюкі (дод. 7); при цьому значення Q_α різні для середніх, віддалених одна від одної на одну, дві, три позиції і т.д. і залежать від рівня значущості (його треба взяти таким самим, як і в дисперсійному аналізі), числа ступенів свободи і порядку різниць l ($l =$ $= 2$ для сусідніх в ранжированому ряду середніх, $l = 3$ для середніх, які відстоять одна від одної на дві позиції і т.д.);

5) зіставляють $Q_{\text{факт}}$ і Q_α; якщо $Q_{\text{факт}} > Q_\alpha$, то нульову гіпотезу про рівність двох середніх відхиляють, а різницю між середніми вважають істотною (вірогідною), якщо ж $Q_{\text{факт}} < Q_\alpha$, то нульову гіпотезу приймають, а різницю між середніми визнають неістотною, оскільки вона знаходиться в межах можливих випадкових коливань.

Вірогідність різниці між парами середніх може бути оцінена також шляхом співставлення її з гранично можливою помилкою вибірки ε_p, яка вказує межі граничних випадкових коливань і дістала назву **найменшої істотної різниці (НІР).** Можлива гранична помилка при використанні t-критерію Стьюдента розраховується за формулою $\varepsilon_p = t\overline{\mu}_{1-2}$, при використанні більш строгого Q-критерію Тьюкі – за формулою $\varepsilon_p = Q\overline{\mu}_{1-2}$.

Якщо різниця між двома порівнюваними середніми більша за абсолютною величиною, ніж можлива гранична помилка (НІР), то роблять висновок про істотність різниці цих двох середніх. Якщо ж можлива гранична помилка буде більшою від фактичної різниці, то різниця між двома середніми лежить у межах можливих випадкових коливань, тобто вона є невірогідною.

Проведемо оцінку істотності різниці між парами середніх по розглянутому прикладу, оскільки був одержаний загальний висновок щодо вірогідності впливу раціонів на прирости живої маси бройлерів.

Нагадаємо, що за даними досліду, що аналізується, вже розраховані три групові середні (табл. 8.1): $\tilde{x}_1 = 32$; $\tilde{x}_2 = 34$; $\tilde{x}_3 = 39$ (г). Залишкова дисперсія, обчислена в дисперсійному аналізі, становила $S^2_{\text{зал}} = 1{,}91$, їй відповідало число ступенів свободи $k_{\text{зал}} = 8$, кількість повторностей $n = 5$.

Сформулюємо H_0 і H_a. За даними досліду можливі такі три порівняння:

$$H_0 : \bar{x}_1 = \bar{x}_2; \qquad H_a : \bar{x}_1 \neq \bar{x}_2;$$

$$H_0 : \bar{x}_2 = \bar{x}_3; \qquad H_a : \bar{x}_2 \neq \bar{x}_3;$$

$$H_0 : \bar{x}_1 = \bar{x}_3; \qquad H_a : \bar{x}_1 \neq \bar{x}_3;$$

Рівень значущості $\alpha = 0{,}05$. Найбільш потужним критерієм перевірки H_0 такого роду є Q-критерій Тьюкі.

Для перевірки H_0 необхідно розрахувати фактичні значення критерію $Q_{факт}$ і порівняти їх з табличними значеннями $Q_{0,05}$.

Побудуємо ранжирований ряд групових середніх за величиною середніх приростів живої маси:

$$\tilde{x}_1 = 32; \qquad \tilde{x}_2 = 34; \qquad \tilde{x}_3 = 39.$$

Визначимо середню помилку вибірок:

$$\bar{\mu}_{1-2} = \sqrt{\frac{S^2_{зал}}{n}} = \sqrt{\frac{1{,}91}{5}} = \sqrt{0{,}38} = 0{,}6 \text{ г}.$$

Обчислимо фактичне значення Q-критерію Тьюкі для кожної пари середніх:

$$Q_1 = \frac{|\tilde{x}_1 - \tilde{x}_2|}{\bar{\mu}_{1-2}} = \frac{|32 - 34|}{0{,}6} = 3{,}33;$$

$$Q_2 = \frac{|\tilde{x}_2 - \tilde{x}_3|}{\bar{\mu}_{1-2}} = \frac{|34 - 39|}{0{,}6} = 8{,}33;$$

$$Q_3 = \frac{|\tilde{x}_1 - \tilde{x}_3|}{\bar{\mu}_{1-2}} = \frac{|32 - 39|}{0{,}6} = 11{,}67.$$

Встановимо критичні значення Q-критерію Тьюкі (дод. 7) при $= 0{,}05$; $k_{зал} = 8$:

для різниць першого ($l = 2$) порядку $Q_{0,05} = 3{,}261$;

для різниць другого ($l = 3$) порядку $Q_{0,05} = 4{,}041$.

Порівняємо фактичні і табличні значення Q-критерію (табл. 8.4).

Оскільки для всіх трьох порівнянь $Q_{факт} > Q_{0,05}$, нульові гіпотези про рівність середніх в генеральних сукупностях відхиляються.

Ті самі висновки матимемо, порівнюючи можливу граничну помилку (НІР) з фактичною різницею між парами середніх.

Визначимо ε_p:

а) для різниць першого порядку

$$\varepsilon_{0,05} = Q_{0,05} \cdot \bar{\mu}_{1-2} = 3{,}261 \cdot 0{,}6 = 1{,}96 \text{ г};$$

Таблиця 8.4

Дані для аналізу статистичної оцінки різниці між парами середніх

Фактична різниця між середніми, г	Значення Q-критерію Тьюкі		Можлива гранична помилка	
	$Q_{факт}$	$Q_{0,05}$	$E_{факт}$	$F_{0,05}$
Першого порядку				
$\tilde{x}_2 - \tilde{x}_1 = 2$	3,33	3,261	2	1,96
$\tilde{x}_3 - \tilde{x}_2 = 5$	8,33	3,261	5	1,96
Другого порядку				
$\tilde{x}_3 - \tilde{x}_1 = 7$	11,67	4,041	7	2,42

б) для різниць другого порядку

$$\varepsilon_{0,05} = Q_{0,05} \cdot \overline{\mu}_{1-2} = 4,041 \cdot 0,6 = 2,42 \text{ г};$$

Можлива гранична помилка (НІР) показує, що внаслідок випадкового варіювання різниця між парами середніх першого порядку може досягати 1,96 г, другого порядку – 2,42 г.

Порівняння фактичних і гранично можливих значень різниць між парами середніх (табл. 8.4) показує, що всі три фактичні різниці перевищують межі можливих випадкових коливань.

Отже, з імовірністю помилки тільки в 5 випадках із 100 можна стверджувати, що різні за якістю раціони істотно впливають на прирости живої маси бройлерів.

8.4. Дисперсійний аналіз при групуванні даних за двома ознаками

В статистичній практиці частіше мають справу з багатофакторними дослідами, в яких вивчають вплив на результативну ознаку двох і більше факторів одночасно.

За аналогією з комбінаційними групуваннями багатофакторні моделі дисперсійного аналізу мають незаперечну перевагу порівняно з однофакторними моделями: вони дають змогу виявити ступінь впливу не тільки кожного фактора окремо, а й їхню взаємодію. Наприклад, ефективність добрив підвищується при покращанні догляду за рослинами, переваги високоурожайних сортів повністю проявляються при високій агротехніці їх вирощування. Ці завдання розв'язуються за допомогою побудови комбінаційних групувань і таблиць. Методи ж розкладання варіації дають числові характеристики взаємодії факторів, а використання імовірнісних оцінок дає змогу зробити висновок про її вірогідність.

Дисперсійний аналіз при групуванні даних за двома факторами ведеться за тією самою принциповою схемою, що й при групуванні даних за одним фактором. При цьому також необхідно враховувати порядок формування груп: випадкове чи невипадкове (незалежні чи залежні вибірки). При залежному формуванні вибірок схема розкладання суми квадратів відхилень *ускладнюється в зв'язку з виділенням суми* квадратів відхилень повторень.

Відмінність багатофакторного аналізу від однофакторного полягає в тому, що загальний обсяг варіації розкладається на більше число компонентів. В міру розкладання сукупності на групи і підгрупи ускладнюються розрахунки з розкладання загального обсягу варіації на складові частини, а також аналіз дисперсій.

Розглянемо схему розкладання загальної варіації на складові частини для випадку з двома факторами (A і B). Джерелами варіації при групуванні даних за двома факторами будуть: перший фактор – A, другий фактор – B, їх взаємодія – AB, залишкове варіювання.

Тоді загальну суму квадратів відхилень можна подати в такому вигляді: $W_{заг} = W_A + W_B + W_{AB} + W_{зал}$.

Розкладання загальної варіації доцільно здійснити в два етапи. Для цього потрібно побудувати дві таблиці. На першому етапі із загального варіювання слід виділити варіацію, пов'язану з двома факторами і залишкову варіацію:

$$W_{заг} = W_{A+B} + W_{зал},$$

а на другому етапі – відповідно:

$$W_{A+B} = W_A + W_B + W_{AB}.$$

Такий порядок розкладання загальної варіації справедливий для незалежних вибірок. Якщо ж вибірки залежні (наприклад, рендомізовані блоки, латинський квадрат тощо), то з'являється новий компонент варіації, пов'язаний з повторностями.

Тоді схема розкладання загальної варіації набуває такого вигляду:

I етап $W_{заг} = W_{A+B} + W_{повт} + W_{зал}$;

II етап $W_{A+B} = W_A + W_B + W_{AB}$;

і в цілому $W_{заг} = W_A + W_B + W_{AB} + W_{повт} + W_{зал}$.

Для випадку з трьома факторами і залежними вибірками схема розкладення загальної варіації ускладнюється:

$$W_{заг} = W_A + W_B + W_C + W_{AB} + W_{AC} + W_{BC} + W_{ABC} + W_{повт} + W_{заг}.$$

Факторні моделі з великою кількістю факторів (трьома і більше) доцільно *досліджувати кореляційним методом з використанням ЕОМ*.

Використання дисперсійного аналізу при групуванні даних за двома факторами розглянемо на такому прикладі. В польовому досліді вивчався вплив різних доз добрив на урожайність озимої пшениці, висіяної на ділянках з різними попередниками (табл. 8.5).

Таблиця 8.5

Урожайність озимої пшениці, ц/га

Попередник (група)	Добрива (підгрупа)	Повторність				Сума x_j	Середня $\bar{x}_j$
		1	2	3	4		
Горох	$P_{50}K_{30}$ (фон)	48,6	43,1	46,0	44,8	182,5	45,6
	Фон + N_{45}	51,5	47,2	50,1	48,6	197,4	49,4
	Фон + N_{90}	55,9	53,0	56,1	53,2	218,2	54,5
Багаторічні трави	$P_{50}K_{30}$ (фон)	50,5	46,5	46,7	48,6	192,3	48,1
	Фон + N_{45}	56,8	48,4	52,7	52,0	209,9	52,5
	Фон + N_{90}	64,0	60,4	62,1	61,6	248,1	62,0
Сума	$\sum x_i$	**327,3**	**298,6**	**313,7**	**308,8**	**1248,4**	**52,0**
Середня	$\bar{x}_i$	**54,5**	**49,8**	**52,3**	**51,5**	**52,0**	×

Схема побудови досліду така: ділянку, на якій проводився дослід, було розбито на чотири блоки, що відрізняються між собою рельєфом і механічним складом ґрунтів. Кожен з варіантів досліду у випадковому порядку було розподілено в усіх чотирьох блоках, чим зумовлювалося значне вирівнювання умов в усіх варіантах досліду, які перевірялись. Отже, дослід проведено в *чотирьохкратній повторності і побудовано за методом* рендомізованих блоків (залежні вибірки).

Потрібно методом дисперсійного аналізу перевірити статистичну гіпотезу відносно середніх у генеральних сукупностях.

Порівнюючи середню урожайність по групах (попередниках) і підгрупах (добривах), *можна помітити, що урожайність озимої пшениці* закономірно зростає зі зміною попередника і збільшенням доз добрив. Разом з тим висунемо нульові гіпотези про випадковість відмінностей середньої урожайності за варіантами досліду, тобто про те, що фактори не впливають на рівень урожайності озимої пшениці:

а) $H_0 : \bar{x}_1 = \bar{x}_2$; $H_a : \bar{x}_1 \neq \bar{x}_2$;

б) $H_0 : \bar{y}_1 = \bar{y}_2 = \bar{y}_3$; $H_a : \bar{y}_1 \neq \bar{y}_2 \neq \bar{y}_3$;

в) H_0 – ефективність взаємодії факторів **у генеральних сукупностях** однакова;

H_a – ефективність взаємодії факторів у генеральних сукупностях неоднакова;

Тут $\tilde{x}_i$ – середні по групах; $\tilde{x}_j$ – середні по підгрупах.

Рівень значущості візьмемо таким, що дорівнює $\alpha = 0{,}05$. Для перевірки H_0 використаємо критерій F.

Схема побудови досліду показує, що загальну варіацію урожайності озимої пшениці можна розкласти на 5 компонентів:

$$W_{заг} = W_п + W_д + W_{вз} + W_{повт} + W_{зал},$$

де $W_{заг}$ - варіація урожайності за рахунок впливу: попередників $W_п$, добрив $W_д$, взаємодії факторів (попередників і добрив) $W_{вз}$, повторностей $W_{повт}$; $W_{зал}$ – залишкова варіація.

Розкладання загальної варіації проведемо в два етапи. На першому етапі виділимо із загальної варіації

$$W_{заг} = W_{п+д} + W_{повт} + W_{зал},$$

тобто варіацію, що створюється спільним впливом попередників і добрив, варіацію по блоках (повторностях) і залишкову (відхилення індивідуальних спостережень від середніх по кожному фактору окремо).

На другому етапі виділимо варіацію (суму квадратів відхилень), пов'язану з дією факторів (кожного окремо і їхню спільну взаємодію):

$$W_{п+д} = W_п + W_д + W_{вз}.$$

Позначимо число спостережень в досліді $N = 24$, кількість груп за попередниками $m = 2$, кількість підгруп за добривами $l = 3$ і за блоками $n = 4$ ($N = mln = 2 \cdot 3 \cdot 4 = 24$).

Для спрощення обчислень сум квадратів відхилень зменшимо всі вихідні дані на постійну величину, близьку до середньої величини ($a = 50$) і виразимо дані досліду у відхиленнях від постійної величини.

Результати обчислень запишемо в табл. 8.6.

Таблиця 8.6

Відхилення від умовного початку ($y = x - a$; $a = 50$)

Попередник (група)	Добрива (підгрупа)	Повторність				Сума Σx_j
		1	2	3	4	
Горох	$P_{50}K_{30}$ (фон)	1,4	– 6,9	– 4,0	5,2	17,5
	Фон + N_{45}	1,5	– 2,8	0,1	– 1,4	– 2,6
	Фон + N_{90}	5,9	3,0	6,1	3,2	18,2
Багаторічні трави	$P_{50}K_{30}$ (фон)	0,5	– 3,5	– 3,3	– 1,4	– 7,7
	Фон + N_{45}	6,8	– 1,6	2,7	2,0	9,9
	Фон + N_{90}	14,0	10,4	12,1	11,6	48,1
Сума	Σy_t	**27,3**	**– 1,4**	**13,7**	**8,8**	**48,4**

Перевіримо правильність розрахунків: загальна сума урожайності $\left(\Sigma x_{ij} = 1248{,}4\right)$ має дорівнювати сумі відхилень від умовного початку $\left(\Sigma y_{ij} = = 48{,}4\right)$ плюс добуток умовного початку на кількість спостережень (aN):

$$\Sigma x_{ij} = \Sigma y_{ij} + aN = 48{,}4 + 50 \cdot 24 = 1248{,}4.$$

Для визначення сум квадратів відхилень $W_{заг}$; $W_{п+д}$; $W_{повт}$ і $W_{зал}$ піднесемо до квадрата всі індивідуальні відхилення, їх суми по графі і рядку і загальний підсумок (табл. 8.7).

Таблиця 8.7

Квадрати відхилень

Попередник (група)	Добрива (підгрупа)	Повторність				Сума квадратів Σy_j^2	Квадрат суми $(\Sigma y_j)^2$
		1	2	3	4		
Горох	$P_{50}K_{30}$ (фон) Фон + N_{45} Фон + N_{90}	1,96 2,25 34,81	47,6 17,84 9,00	16,00 0,01 37,21	27,04 1,96 10,24	92,61 12,06 91,26	306,25 6,76 331,24
Багаторічні трави	$P_{50}K_{30}$ (фон) Фон + N_{45} Фон + N_{90}	0,25 46,24 196,00	12,25 2,56 108,16	10,8 97,29 146,41	1,96 4,00 134,56	25,35 60,09 585,13	59,29 98,01 2313,61
Сума квадратів	Σy_i^2	**281,51**	**187,42**	**217,81**	**179,76**	**866,50**	**3115,16**
Квадрат суми	$(\Sigma y_i)^2$	**745,29**	**1,96**	**187,69**	**77,44**	**1012,38**	**2342,56**

В результаті піднесення до квадрата відхилень отримаємо всі потрібні дані для визначення сум квадратів відхилень на першому етапі:

$$\Sigma y_{ij}^2 = 866{,}50; \qquad \Sigma(\Sigma y_j)^2 = 3115{,}16; \qquad \Sigma(\Sigma y_i)^2 = 1012{,}38;$$

$$(\Sigma y_{ij})^2 = 2342{,}56.$$

Користуючись одержаними сумами, визначимо необхідні суми квадратів відхилень:

$$W_{заг} = \Sigma y_{ij}^2 - \frac{(\Sigma y_{ij})^2}{N} = 866{,}50 - \frac{2342{,}56}{24} = 768{,}89;$$

$$W_{п+д} = \frac{\Sigma(\Sigma y_j)^2}{n} - \frac{(\Sigma y_{ij})^2}{N} = \frac{3115{,}16}{4} - \frac{2342{,}56}{24} = 681{,}18;$$

$$W_{повт} = \frac{\Sigma(\Sigma y_i)^2}{ml} - \frac{(\Sigma y_{ij})^2}{N} = \frac{1012,38}{2 \cdot 3} - \frac{2342,56}{24} = 71,12;$$

$$W_{зал} = W_{заг} - W_{п+д} - W_{повт} = 768,89 - 681,18 - 71,12 = 16,59.$$

Для кожної суми квадратів відхилень визначимо число ступенів свободи варіації:

$k_0 = N - 1 = 24 - 1 = 23;$ $k_{п+д} = m \cdot l - 1 = 2 \cdot 3 - 1 = 5;$

$k_{повт} = n - 1 = 4 - 1 = 3;$ $k_{зал} = k_0 - k_{п+д} - k_{повт} = 23 - 5 - 3 = 15.$

На другому етапі проведемо розкладання варіації, пов'язаної з дією факторів (за рахунок кожного фактора та їх взаємодії):

$$W_{п+д} = W_п + W_д + W_{вз}.$$

Для розрахунку необхідних сум квадратів відхилень на підставі даних останнього стовпця табл. 8.6 складемо нову таблицю відхилень (табл. 8.8), в якій назви рядків містять градації за попередниками, а стовпців – за дозами добрив. Позначимо суми відхилень за попередниками – *A*, а за дозами добрив – *B* і піднесемо до квадрата відхилення (табл. 8.9).

Таблиця 8.8

Відхилення по варіантах досліду

Попередник	Добрива			Сума *A*
	$P_{50}K_{30}$ (фон)	фон + N_{45}	фон + N_{90}	
Горох	– 17,5	– 2,6	18,2	– 1,9
Багаторічні трави	– 7,7	9,9	48,1	50,3
Сума *B*	**– 25,2**	**7,3**	**66,3**	**48,4**

Таблиця 8.9

Квадрати відхилень по варіантах досліду

Попередник	Добрива			Сума квадратів	Квадрат сум A^2
	$P_{50}K_{30}$ (фон)	фон + N_{45}	фон + N_{90}		
Горох	306,25	6,76	331,24	644,25	3,61
Багаторічні трави	59,29	98,01	2313,61	2470,91	2530,09
Сума квадратів	**365,54**	**104,77**	**2644,16**	**3115,16**	**2533,70**
Квадрат сум B^2	**635,04**	**53,29**	**4395,69**	**5084,02**	**2342,56**

Сума квадратів відхилень в табл. 8.9 (3115,16) дорівнює підсумку останньої графи табл. 8.7. На підставі цієї суми було визначено варіацію урожайності, зумовленої спільною дією двох факторів:

$W_{п+д} = 681,18$ з числом ступенів свободи $k_{п+д} = 5$.

Ця сума квадратів відхилень складається із таких складових компонентів:

$$W_{n+д} = W_n + W_д + W_{вз}.$$

Обчислимо потрібні суми квадратів відхилень:

$$W_n = \frac{\Sigma A^2}{n \cdot l} - \frac{\Sigma(\Sigma y_{ij})^2}{N} = \frac{2533{,}70}{4 \cdot 3} - \frac{2342{,}56}{24} = 113{,}53;$$

$$W_д = \frac{\Sigma B^2}{m \cdot n} - \frac{\Sigma(\Sigma y_{ij})^2}{N} = \frac{5084{,}02}{2 \cdot 4} - \frac{2342{,}56}{24} = 537{,}89;$$

$$W_{вз} = W_{n+д} - W_n - W_д = 681{,}18 - 113{,}53 - 537{,}89 = 29{,}76.$$

Визначимо число ступенів свободи варіації для кожної з обчислених сум квадратів відхилень. Для цього необхідно ступені свободи варіації двох факторів ($k_{n+д} = 5$) розподілити між трьома компонентами, що складають цю суму квадратів відхилень (k_n, $k_д$ і $k_{вз}$):

$k_n = m - 1 = 2 - 1 = 1;$ $\quad$ $k_д = l - 1 = 3 - 1 = 2;$

$k_{вз} = k_{n+д} - k_n - k_д = 5 - 1 - 2 = 2.$

Об'єднаємо результати обчислень двох етапів, розрахуємо і проаналізуємо дисперсії (табл. 8.10).

Таблиця 8.10

Розрахунок і аналіз дисперсій

Джерело варіації	Сума квадратів відхилень	Ступінь свободи варіації	Дисперсія	Відношення дисперсій	
	W_i	k_i	S_i^2	$F_{факт}$	$F_{0,05}$
Попередники	113,53	1	113,53	102,28	4,54
Добрива	537,89	2	268,94	242,29	3,68
Взаємодія	29,76	2	14,88	13,40	3,68
Повторності	71,12	3	23,71	21,36	3,29
Залишкова	16,59	15	1,11	1,00	–
Загальна	768,89	23	–	–	–

Фактичне дисперсійне відношення знайдемо за формулою

$$F_{факт} = \frac{S^2_{фактор}}{S^2_{зал}}.$$

Наприклад, $F_{факт}$ по добривах становитиме $S_д^2 : S_{зал}^2 = 268{,}94 : 1{,}11 = 242{,}29$, по взаємодії факторів $S_{вз}^2 : S_{зал}^2 = 14{,}88 : 1{,}11 = 13{,}40$ і т.д. Табличні значення F при рівні значущості $\alpha = 0{,}05$ визначимо за додатком 4 для оцінки відношення дисперсії попередників до дисперсії залишкової – на перетині 1-го стовпця і 15-го рядка ($F_{0,05}$ =4,54); для відношення дисперсії добрив до дисперсії залишкової – 2-го стовпця і 15-го рядка ($F_{0,05}$ = 3,68) і т.д.

Порівняння фактичних і табличних значень F-критерію Фішера при заданому рівні значущості α = 0,05 показує, що в усіх випадках $F_{факт} > F_{0,05}$. Отже, висунуті нульові гіпотези не узгоджуються з фактичними даними і тому їх потрібно відхилити. Відмінності в урожайності за факторами і повторностями є істотними, вірогідними, вірогідним є проявлення ефекту взаємодії факторів.

Оскільки за результатами дисперсійного аналізу нульові гіпотези відхилено і доведено істотність відмінностей між середніми, можна оцінити вірогідність різниць між парами середніх. Для цього потрібно обчислити середню і можливу граничну помилку вибірок (ε_p) і аналогічно тому, як це було зроблено в попередньому прикладі, знайти фактичні різниці між парами середніх і порівняти їх з можливими граничними різницями (НІР).

Завдання для самоконтролю до розділу 8

1. У чому полягає суть і призначення дисперсійного аналізу?

2. Наведіть принципову схему дисперсійного аналізу.

3. Які критерії застосовуються в дисперсійному аналізі для оцінки різниць кількох середніх?

4. Розкрийте смисл нульової гіпотези в дисперсійному аналізі.

5. Що показує фактичне і табличне значення F-критерію Фішера?

6. Наведіть і поясніть суть формули фактичного дисперсійного відношення

7. Наведіть алгоритм дисперсійного аналізу при групуванні даних за однією ознакою.

8. Наведіть алгоритм дисперсійного аналізу при групуванні даних за двома і більше ознаками.

9. За яких обставин і чому проводиться попарне порівняння середніх?

10. Що таке найменша істотна різниця (НІР)?

11. За наведеними даними щодо середньодобового приросту живої маси молодняку великої рогатої худоби визначити вірогідність впливу типу годівлі. Дослід проведено в 5-кратній повторності. Загальний обсяг варіації – 36 000. Рівень значущості α = 0,05.

Тип годівлі	1	2	3
Середньодобовий приріст живої маси, г	600	680	730

Розділ 9

Кореляційний аналіз

9.1. Поняття про кореляційний аналіз

Вивчення реальної дійсності показує, що практично кожне суспільне явище знаходиться в тісному зв'язку і взаємодії з іншими явищами, якими б випадковими вони не здавалися на перший погляд. Так, наприклад, рівень урожайності сільськогосподарських культур залежить від множини природних і економічних факторів, тісно пов'язаних між собою.

Дослідження і вимірювання взаємозв'язків і взаємозалежностей соціально-економічних явищ є одним з найважливіших завдань статистики.

Для дослідження взаємозв'язків між явищами статистика використовує ряд методів і прийомів: статистичні групування (прості і комбінаційні), індексний, кореляційний і дисперсійний аналіз, балансовий, табличний, графічний та ін. Зміст, специфіка і можливості застосування деяких з перелічених методів уже були розглянуті в попередніх розділах підручника. Індексний і графічний методи розглядаються відповідно в 11 і 12 розділах.

Поряд з уже розглянутими методами вивчення взаємозв'язків особливе місце займає метод кореляції, який є логічним продовженням таких методів, як аналітичне групування, дисперсійний аналіз і зіставлення паралельних рядів. В поєднанні з цими методами він надає статистичному аналізу закінчений, завершений характер.

Засновниками теорії кореляції є англійські статистики Ф. Гальтон (1822–1911) і К. Пірсон (1857–1936).

Термін кореляція походить від англійського слова correlation – співвідношення, відповідність (взаємозв'язок, взаємозалежність) між ознаками, що виявляється при масовому спостереженні зміни середньої величини однієї ознаки залежно від значення іншої. Ознаки, що пов'язані між собою кореляційним зв'язком, називають **корельованими.**

Кореляційний аналіз дає змогу виміряти ступінь впливу факторних ознак на результативні, встановити єдину міру тісноти зв'язку і роль досліджуваного фактора (факторів) у загальній зміні результативної ознаки. Кореляційний метод дозволяє одержати кількісні характеристики ступеня зв'язку між двома і більшим числом ознак, а тому на відміну від розглянутих вище методів дає більш широке уявлення про зв'язок між ними.

Зв'язки між факторами досить різноманітні. При цьому одні ознаки виступають в ролі факторів, що діють на інші, зумовлюючи їх зміну, другі – в ролі дії цих факторів. Перші з них називають **факторними** ознаками, другі – **результативними.**

Досліджуючи зв'язки між ознаками, необхідно виділити насамперед два види зв'язків: функціональний (повний) і кореляційний (статистичний) зв'язок.

Функціональним називають такий зв'язок між ознаками, при якому кожному значенню однієї змінної (аргументу) відповідає строго визначене значення другої змінної (функції). Такі зв'язки спостерігаються в математиці, фізиці, хімії, астрономії та інших науках.

Наприклад, площа круга ($S = \pi R^2$) і довжина кола ($C = 2\pi R$) повністю визначається величиною радіуса, площі трикутника і прямокутника – довжиною їх сторін тощо. Так, із збільшенням радіуса кола на 1 см його довжина збільшується на 6,28 см, на 2 см – на 12,56 см і т.д.

У сільськогосподарському виробництві прикладом функціонального зв'язку може бути зв'язок між виручкою від продажу продукції, ціною реалізації 1 ц і кількістю реалізованої продукції; валовим збором, урожайністю і розміром посівної площі, фондовіддачею, вартістю валової продукції і основних фондів, заробітною платою і кількістю відпрацьованого часу при погодинній оплаті тощо.

Функціональний зв'язок виявляється як у сукупності в цілому, так і в кожній її одиниці абсолютно точно і виражається за допомогою аналітичних формул.

В соціально-економічних явищах функціональні зв'язки між ознаками трапляються рідко. Тут дуже часто мають місце такі зв'язки між змінними величинами, при яких чисельному значенню однієї з них відповідає кілька значень інших. Такий зв'язок між ознаками дістав назву **кореляційного (статистичного) зв'язку.** Наприклад, відомо, що із збільшенням доз мінеральних добрив і поліпшенням їхньої структури (співвідношення), як правило, урожайність сільськогосподарських культур підвищується, але добре відомо, що приріст урожайності у кожному окремому випадку буде різним при однакових нормах внесення добрив. Крім того, одні і ті самі норми добрив навіть при дуже вирівняних умовах часто по-різному впливають на урожайність. Крім самих добрив, на величину формування урожайності впливають також інші фактори, насамперед такі, як якість ґрунту, опади, строки і способи сівби та збирання тощо. Відома закономірність між урожайністю і добривами проявиться при досить великій кількості спостережень і при порівнянні досить великої кількості середніх значень результативної і факторної ознак.

Прикладом кореляційного зв'язку в сільськогосподарському виробництві може бути зв'язок між продуктивністю тварин і рівнем годівлі, якістю кормів, породністю худоби, між стажем роботи і продуктивністю праці робітників тощо.

Кореляційний зв'язок є неповним, він проявляється при великій кількості спостережень, при порівнянні середніх значень результативної і факторної ознак. У цьому відношенні виявлення кореляційних залежностей пов'язано з дією закону великих чисел: тільки при досить великій кількості спостережень індивідуальні особливості і другорядні фактори згладяться і залежність між результативною і факторною ознаками, якщо вона має місце, виявиться досить виразно.

За допомогою кореляційного аналізу вирішують такі основні завдання:

а) визначення середньої зміни результативної ознаки під впливом одного або кількох факторів (в абсолютному або відносному вимірі);

б) характеристика ступеня залежності результативної ознаки від одного з факторів при фіксованому значенні інших факторів, включених до кореляційної моделі;

в) визначення тісноти зв'язку між результативними і факторними ознаками (як з усіма факторами, так і з кожним фактором окремо при виключенні впливу інших);

г) визначення і розкладання загального обсягу варіації результативної ознаки на відповідні частини і встановлення ролі кожного окремого фактора в цій варіації;

д) статистична оцінка вибіркових показників кореляційного зв'язку.

Кореляційний зв'язок виражається відповідними математичними рівняннями. За напрямом зв'язок між корелюючими ознаками може бути прямим і оберненим. При **прямому зв'язку** обидві ознаки змінюються в одному напрямі, тобто із збільшенням факторної ознаки зростає результативна і навпаки (наприклад, зв'язок між якістю ґрунту і врожайністю, рівнем годівлі і продуктивністю тварин, стажем роботи і продуктивністю праці). **При оберненому зв'язку** обидві ознаки змінюються в різних напрямах (наприклад, зв'язок між урожайністю і собівартістю продукції, продуктивністю праці і собівартістю продукції).

За **формою**, або **аналітичним виразом**, розрізняють зв'язки прямолінійні (або просто лінійні) і нелінійні (або криволінійні). Якщо зв'язок між ознаками виражається рівнянням прямої лінії, то його називають **лінійним зв'язком**, якщо ж він виражається рівнянням будь-якої кривої (параболи, гіперболи, показникової, степеневої і т.д.), то такий зв'язок називають **нелінійним**, або **криволінійним**.

Залежно від **кількості досліджуваних ознак** розрізняють парну (просту) і множинну кореляцію. При **парній кореляції** вивчають зв'язок між двома ознаками (результативною і факторною), при **множинній кореляції** – зв'язок між трьома і більшим числом ознак (результативною і двома та більшим числом факторів).

За допомогою методу кореляційного аналізу вирішується два головних завдання: 1) визначення форми і параметрів рівняння зв'язку; 2) вимірювання тісноти зв'язку.

Перше завдання вирішується знаходженням рівняння зв'язку і визначенням його параметрів. Друге – за допомогою розрахунку різних показників тісноти зв'язку (коефіцієнта кореляції, кореляційного відношення, індексу кореляції та ін.).

Схематично кореляційний аналіз можна поділити на п'ять етапів:

1) постановка завдання, встановлення наявності зв'язку між досліджуваними ознаками;

2) відбір найістотніших факторів для аналізу;

3) визначення характеру зв'язку, його напряму і форми, вибір математичного рівняння для вираження існуючих зв'язків;

4) розрахунок числових характеристик кореляційного зв'язку (визначення параметрів рівняння і показників тісноти зв'язку);

5) статистична оцінка вибіркових показників зв'язку.

Науково обґрунтоване застосування кореляційного методу потребує передусім глибокого розуміння суті взаємозв'язків соціально-економічних явищ. Сам метод не встановлює наявність і причини виникнення зв'язків між досліджуваними явищами, його призначення полягає в їх кількісному вимірюванні. На першому етапі кореляційного аналізу здійснюється загальне ознайомлення з досліджуваним об'єктом і явищами, уточнюються мета і завдання дослідження, встановлюється теоретична можливість причинно-наслідкового зв'язку між ознаками.

Встановлення причинних залежностей в досліджуваному явищі передує власне кореляційному аналізу. Тому застосуванню методів кореляції повинен передувати глибокий теоретичний аналіз, який охарактеризує основний процес, що протікає в досліджуваному явищі, визначить суттєві зв'язки між окремими його сторонами і характер їх взаємодії.

Попередній аналіз даних створює основу для формулювання конкретного завдання дослідження зв'язків, відбору найважливіших факторів, встановлення можливої форми взаємозв'язку ознак і тим самим приводить до математичної формалізації – до вибору математичного рівняння, яке найбільш повно відтворить існуючі зв'язки.

Одним із найважливіших питань кореляційного аналізу є відбір результативної і факторної (факторних) ознак. Факторні і результативні ознаки, що відбираються для кореляційного аналізу, повинні бути суттєвими, перші повинні безпосередньо впливати на другі. Відбір факторів для включення їх в кореляційну модель повинен базуватися передусім на теоретичних основах і практичному досвіді аналізу досліджуваного соціально-економічного явища. Велику допомогу в розв'язанні цього завдання можуть надати такі статистичні прийоми і методи, як зіставлення паралельних рядів, побудова таблиць розподілу чисельностей за двома ознаками (кореляційних таблиць), побудова статистичних групувань як за результативною ознакою з аналізом взаємопов'язаних з нею факторів, так і за факторною ознакою (або комбінацією факторних ознак) з аналізом їх впливу на результативну ознаку.

Відбір факторів для парних кореляційних моделей не складний: з множини факторів, що впливають на результативну ознаку, відбирається один із найважливіших факторів, який в основному визначає варіацію результативної ознаки або ж фактор, істотність впливу якого на результативну ознаку передбачається вивчити або перевірити. Відбір факторів для множинних кореляційних моделей має ряд особливостей і обмежень. Вони будуть розглянуті при викладенні питань множинної кореляції.

Однією з головних проблем побудови кореляційної моделі є визначення форми зв'язку і на цій основі встановлення типу аналітичної функції, що відображає механізм зв'язку результативної ознаки з факторною (факторними). Під **формою кореляційного зв'язку** розумають тип аналітичного рівняння, що виражає залежність між досліджуваними ознаками.

Вибір того або іншого рівняння для дослідження зв'язків між ознаками є найбільш важким і відповідальним завданням, від якого залежать результати кореляційного аналізу. Всі подальші найретельніші розрахунки можуть бути знецінені, якщо форма зв'язку вибрана невірно. Важливість цього етапу полягає в тому, що правильно встановлена форма зв'язку дає змогу підібрати і побудувати найбільш адекватну модель і на основі її розв'язання одержати статистично вірогідні і надійні характеристики.

Встановлення форми зв'язку між ознаками в більшості випадків обґрунтовується теорією або практичним досвідом попередніх досліджень. Якщо форма зв'язку невідома, то при парній кореляції математичне рівняння може бути встановлено за допомогою складання кореляційних таблиць, побудови статистичних групувань, перегляду різних функцій на ЕОМ і вибір такого рівняння, яке дає найменшу суму квадратів відхилень фактичних даних від вирівняних (теоретичних) значень та ін.

Залежно від вихідних даних теоретичною лінією регресії можуть бути різні типи кривих або пряма лінія. Так, якщо зміна результативної ознаки під впливом фактора характеризується постійними приростами, то це вказує на лінійний характер зв'язку, якщо ж зміни результативної ознаки під впливом фактора характеризується постійними коефіцієнтами зростання, то є підстава припустити криволінійний зв'язок.

Особливе місце в обґрунтуванні форми зв'язку при проведенні кореляційного аналізу належить графікам, побудованим у системі прямокутних координат на основі емпіричних даних. Графічне зображення фактичних даних дає наочне уявлення про наявність і форму зв'язку між досліджуваними ознаками.

Згідно з правилами математики при побудові графіка на осі абсцис відкладають значення факторної ознаки, а на осі ординат – значення результативної ознаки. Відклавши на перетині відповідних значень двох ознак точки, одержимо точковий графік, який називають **кореляційним полем.** За характером розміщення точок на кореляційному полі роблять висновок про напрям і форму зв'язку. Достатньо подивитися на графік, щоб прийти до висновку про наявність і форму зв'язку між ознаками. Якщо точки концентруються навколо уявної осі, напрямленої зліва, знизу, направо, вгору, то зв'язок прямий, якщо ж навпаки зліва, зверху, направо, вниз – зв'язок обернений. Якщо точки розкидані по всьому полю, то це свідчить про те, що зв'язок між ознаками відсутній або дуже слабкий. Характер розміщення точок на кореляційному полі вказує також і на наявність прямолінійного або криволінійного зв'язку між досліджуваними ознаками.

За допомогою графіка добирають відповідне математичне рівняння для кількісної оцінки зв'язку між результативною і факторною ознаками. Рівняння, що відображає зв'язок між ознаками, називають **рівнянням регресії**, або **кореляційним рівнянням.** Якщо рівняння регресії зв'язує лише дві ознаки, то воно називається **рівнянням парної регресії.** Якщо рівняння зв'язку відображає залежність результативної ознаки від двох і більше факторних ознак, воно називається **рівнянням множинної регресії.** Криві, побудовані на основі рівнянь регресії, називають **кривими регресії**, або **лініями регресії**

Розрізняють емпіричну і теоретичну лінії регресії. Якщо на кореляційному полі з'єднати точки відрізками прямої лінії, то одержимо ламану лінію з деякою тенденцією, яка називається **емпіричною лінією регресії**. **Теоретичною лінією регресії** називається та лінія, навколо якої концентруються точки кореляційного поля і яка вказує основний напрям, основну тенденцію зв'язку. Теоретична лінія регресії повинна відображати зміну середніх величин результативної ознаки в міру зміни величин факторної ознаки при умові повного взаємопогашення всіх інших – випадкових по відношенню до фактора – причин. Отже, ця лінія має бути проведена так, щоб сума відхилень точок кореляційного поля від відповідних точок теоретичної лінії дорівнювала нулю, а сума квадратів відхилень була б мінімальною величиною. Пошук, побудова, аналіз і практичне застосування теоретичної лінії регресії називають **регресійним аналізом.**

За емпіричною лінією регресії не завжди вдається встановити форму зв'язку і добрати рівняння регресії. В таких випадках будують і розв'язують різні рівняння регресії. Потім оцінюють їх адекватність і добирають таке рівняння, яке забезпечує найкращу апроксимацію (наближення) фактичних даних до теоретичних і достатню статистичну вірогідність і надійність.

Якщо підходити строго, регресійно-кореляційний аналіз слід розчленувати на **регресійний** і **кореляційний.** Регресійний аналіз вирішує питання побудови, розв'язання і оцінки рівнянь регресії, а при кореляційному аналізі до цих питань приєднується ще коло питань, пов'язаних із визначенням тісноти зв'язку між результативною і факторною (факторними) ознаками. В подальшому викладенні регресійно-кореляційний аналіз розглядається як єдине ціле і називається просто – **кореляційний аналіз.**

Щоб результати кореляційного аналізу знайшли практичне застосування і дали науково обґрунтовані результати, повинні виконуватись певні вимоги відносно об'єкта дослідження і якості вихідної статистичної інформації. Основні з цих вимог такі:

– якісна однорідність досліджуваної сукупності, що передбачає близькість формування результативних і факторних ознак. Необхідність виконання цієї умови випливає із змісту параметрів рівняння зв'язку. З математичної статистики відомо, що параметри є середніми величинами. В якісно однорідній сукупності вони будуть типовими характеристиками, в якісно різнорідній – спотвореними, що перекручують характер зв'язку. Кількісна однорідність сукупності полягає у відсутності одиниць спостереження, які за своїми числовими характеристиками суттєво відрізняються від основної маси даних. Такі одиниці спостереження слід виключати із сукупності і вивчати окремо;

– досить велике число спостережень, оскільки зв'язки між ознаками виявляються тільки внаслідок дії закону великих чисел. Кількість одиниць спостереження повинна в 6-8 разів перевищувати кількість включених у модель факторів;

– випадковість і незалежність окремих одиниць сукупності одна від одної. Це означає, що значення ознак у одних одиницях сукупності не повинні залежати від значень у інших одиницях даної сукупності;

– стійкість і незалежність дії окремих факторів;

– сталість дисперсії результативної ознаки при зміні факторних ознак;

– нормальний розподіл ознак.

9.2. Парна (проста) лінійна кореляція

Найпростішим видом кореляційного зв'язку є зв'язок між двома ознаками: результативною і факторною. Такий зв'язок називають **парною кореляцією**, або **простою кореляцією**.

В економічних дослідженнях взаємозв'язку двох факторів серед множини функцій часто розглядається прямолінійна форма зв'язку, яка виражається рівнянням прямої лінії

$$\widetilde{y}_x = a + bx,$$

де $\widetilde{y}_x$ – вирівняне значення результативної ознаки (залежна змінна); x – значення факторної ознаки (незалежна змінна); a – початок відліку, або значення при $b = 0$ (економічного змісту не має); b – коефіцієнт регресії, який показує середню змінну залежної змінної при зміні незалежної змінної на одиницю (одне своє значення).

Коефіцієнти регресії є величинами іменованими і мають одиниці вимірювання, що відповідають змінним, між якими вони характеризують зв'язок.

Якщо $b > 0$, то зв'язок прямий, якщо $b < 0$, то зв'язок обернений, якщо $b = 0$, то зв'язок відсутній.

Параметри рівняння a і b визначають способом найменших квадратів. Він дає можливість знайти ту криву, яка порівняно з іншими кривими проходить найближче до точок кореляційного поля, що відображають фактичні дані, тобто дає найменшу суму квадратів відхилень фактичних значень результативної ознаки від вирівняних (теоретичних) значень:

$$\sum (y_i - \tilde{y}_x)^2 = \min .$$

Порядок одержання системи нормальних рівнянь при парній кореляції такий. Для одержання першого рівняння системи необхідно всі члени вихідного рівняння кореляційного зв'язку помножити на коефіцієнти при першому невідомому (a) і одержані добутки підсумувати. Потім для отримання другого рівняння необхідно всі члени вихідного рівняння помножити на коефіцієнт при другому невідомому (b) і також всі добутки підсумувати.

Техніка одержання системи нормальних рівнянь залишається аналогічною і для побудови системи рівнянь з більшим числом змінних. Так, для парного лінійного зв'язку система нормальних рівнянь має вигляд:

$$\begin{cases} \sum y = an + b\sum x; \\ \sum yx = a\sum x + b\sum x^2. \end{cases}$$

Параметри a і b рівняння прямої лінії можна визначити за іншими робочими формулами:

$$a = \frac{\sum y \sum x^2 - \sum yx \sum x}{n\sum x^2 - \sum x \sum x}; \qquad b = \frac{n\sum xy - \sum y \sum x}{n\sum x^2 - \sum x \sum x};$$

або

$$a = \bar{y} - b\bar{x}; \qquad b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}.$$

Рівняння кореляційного зв'язку мають як пізнавальне, так і практичне значення, їх використовують для обчислення теоретичної лінії регресії, очікуваних (теоретичних, вирівняних) і прогнозованих значень залежної змінної при тих або інших значеннях фактора (факторів). При цьому слід мати на увазі, що рівняння дає середнє співвідношення між результативною і факторною ознаками, тому найбільшу точність збігання мають розрахункові значення результативної ознаки при величині фактора, близького до середнього його рівня.

Ступінь наближення розрахункових значень результативної ознаки до її фактичного значення залежить від того, наскільки досконала кореляційна модель. Якщо вона включає всі основні фактори, що визначають варіацію результативної ознаки, то точність буде досить високою.

Розглянемо приклад аналізу кореляційного зв'язку між двома ознаками (парна кореляція): продуктивністю корів – надоєм молока на середньорічну корову (ц) і рівнем годівлі – витратами кормів на одну корову за рік (ц кормових одиниць; табл. 9.1).

Таблиця 9.1

Розрахунок даних для визначення показників кореляційного зв'язку

Господарство	Надій на корову за рік, ц	Витрати кормів на корову, ц кормових одиниць	Розрахункові дані			
	y	x	yx	y^2	x^2	Очікуване (розрахункове значення надою), ц $\tilde{y}_x$
1	35,0	42,4	1484,0	1225,0	1797,8	37,7
2	39,8	41,0	1631,8	1584,0	1681,0	36,6
3	30,2	35,0	1057,0	912,0	1225,0	31,7
4	42,4	47,4	2009,8	1797,8	2246,8	41,7
5	32,5	35,6	1157,0	1056,2	1267,4	32,2
6	34,1	38,8	1323,1	1162,8	1505,4	34,7
7	43,6	50,0	2180,0	1901,0	2500,0	43,9
8	40,1	46,2	1852,6	1608,0	2134,4	40,8
9	33,4	36,7	1225,8	1115,6	1346,9	33,0
10	36,9	39,9	1472,3	1361,6	1592,0	35,7
Разом	**368,0**	**413,0**	**15393,4**	**13724,0**	**17296,7**	**368,0**
У середньому	36,8	41,3	1539,3	1372,4	1729,7	36,8

Результативною ознакою в даному прикладі є продуктивність корів (y), а факторною – рівень годівлі (x).

Для визначення форми зв'язку між продуктивністю корів і рівнем годівлі побудуємо графік – кореляційне поле (рис. 9.1). На осі абсцис відкладемо значення факторної ознаки (незалежної змінної – рівня годівлі, а на осі ординат – результативної ознаки (залежної змінної – продуктивності корів).

Рис. 9.1. Кореляційне поле залежності надою на корову від витрат кормів

Графік показує, що в даному випадку зв'язок близький до прямолінійного і його можна виразити рівнянням прямої лінії

$$\tilde{y}_x = a + bx$$

Розв'язання цього рівняння регресії покаже зміну продуктивності корів під впливом рівня годівлі при виключенні випадкових коливань ознаки.

Параметри рівняння прямої лінії a і b знайдемо із системи нормальних рівнянь:

$$\begin{cases} \sum y = an + b\sum x; \\ \sum yx = a\sum x + b\sum x^2. \end{cases}$$

Усі необхідні для розв'язання системи рівнянь дані розрахуємо в табл. 9.1.

Одержані дані підставимо в систему рівнянь:

$$\begin{cases} 368 = 10a + 413b & \Big| : 10 \\ 15393{,}4 = 413a + 17296{,}7b & \Big| : 413 \end{cases}$$

Поділимо рівняння на коефіцієнти при a, тобто перше рівняння на 10, а друге – на 413:

$$\begin{cases} 36{,}80 = a + 41{,}30b; \\ 37{,}2722 = a + 41{,}8806b. \end{cases}$$

Віднімемо перше рівняння із другого:

$$0{,}4722 = 0{,}5806b,$$

звідси $b = 0{,}4722 : 0{,}5806 = 0{,}813$ ц на 1 ц кормових одиниць.

Підставимо значення $b = 0{,}813$ в перше рівняння і знайдемо a:

$$368 = 10\,a + 413 \cdot 0{,}813$$

$$368 = 10\,a + 335{,}8$$

$$a = \frac{368 - 335{,}8}{10} = 3{,}22 \text{ ц}.$$

Рівняння регресії (кореляційне рівняння), яке виражає зв'язок між продуктивністю корів і рівнем годівлі, буде мати вигляд:

$$\tilde{y}_x = 3{,}22 + 0{,}813x.$$

Коефіцієнт регресії $b = 0{,}813$ показує, що з підвищенням рівня годівлі на 1 ц кормових одиниць продуктивність корів в середньому по даній сукупності господарств зростає на 0,813 ц.

Параметри рівняння регресії можна визначити і за іншими формулами:

$$a = \frac{\sum y \sum x^2 - \sum yx \sum x}{n\sum x^2 - \sum x \sum x} = \frac{368 \cdot 17296{,}7 - 15393{,}4 \cdot 413}{10 \cdot 17296{,}7 - 413 \cdot 413} = 3{,}22;$$

$$b = \frac{n\sum yx - \sum y \sum x}{n\sum x^2 - \sum x \sum x} = \frac{10 \cdot 15393{,}4 - 368 \cdot 413}{10 \cdot 17296{,}7 - 413 \cdot 413} = 0{,}813,$$

або $a = \overline{y} - b\overline{x} = 36{,}8 - 0{,}813 \cdot 41{,}3 = 36{,}8 - 33{,}58 = 3{,}22;$

$$b = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2} = \frac{1539{,}34 - 36{,}8 \cdot 41{,}3}{1729{,}7 - 41{,}3^2} = 0{,}813.$$

Перевіримо правильність розв'язання системи рівнянь, виходячи із рівності $\overline{y} = a + b\overline{x}$:

$$36{,}80 = 3{,}22 + 0{,}813 \cdot 41{,}3;$$

$$36{,}80 = 3{,}22 + 33{,}58; \; 36{,}80 = 36{,}80.$$

За рівнянням регресії можна розрахувати очікувані (розрахункові або теоретичні) значення продуктивності корів $(\tilde{y}_x)$ при різних значеннях витрат кормів на корову (x). Для цього замість x підставимо його конкретні значення:

$$\widetilde{y}_x = 42{,}4 = 3{,}22 + 0{,}813 \cdot 42{,}4 - 37{,}7 \text{ ц};$$

$$\widetilde{y}_x = 41{,}0 = 3{,}22 + 0{,}813 \cdot 41{,}0 - 36{,}6 \text{ ц} \quad \text{і т.д.}$$

Усі обчислені дані запишемо в останню графу таблиці 9.1. За цими даними на рис. 9.1 побудуємо теоретичну лінію регресії.

Перевіримо правильність усіх розрахунків, зіставивши суми фактичного і розрахункового надою молока на корову:

$$\Sigma y = \Sigma \widetilde{y}_x; \qquad 368{,}0 = 368{,}0.$$

9.3. Показники тісноти зв'язку

При кореляційному зв'язку разом з досліджуваним фактором або кількома факторами при множинній кореляції на результативну ознаку впливають і інші фактори, які не враховуються або не можуть бути точно враховані. При цьому дія їх може бути направлена як в сторону підвищення результативної ознаки, так і в сторону її зниження. Отже, дослідження зв'язку відбувається в умовах, коли цей зв'язок більшою або меншою мірою затушовується суперечливою дією інших причин. Тому одне із завдань кореляційного аналізу полягає у визначенні тісноти зв'язку між ознаками, у визначенні сили дії досліджуваного фактора (факторів) на результативну ознаку.

Тіснота зв'язку у кореляційному аналізі характеризується за допомогою спеціального відносного показника, який отримав назву **коефіцієнта кореляції.**

При парній лінійній залежності тіснота зв'язку визначається за допомогою **лінійного коефіцієнта кореляції**

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y},$$

де $\overline{xy} = \dfrac{\Sigma xy}{n}$; $\quad \overline{y} = \dfrac{\Sigma y}{n}$; $\quad \overline{x} = \dfrac{\Sigma x}{n}$;

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - (\overline{x})^2}; \qquad \sigma_y = \sqrt{\frac{\Sigma y^2}{n} - (\overline{y})^2}.$$

Він може бути обчислений за іншими формулами:

$$r = \frac{\Sigma(y - \overline{y})(x - \overline{x})}{n\sigma_x \cdot \sigma_y}; \qquad r = \frac{n\Sigma yx - \Sigma y \Sigma x}{\sqrt{[n\Sigma y^2 - (\Sigma y)^2][n\Sigma x^2 - (\Sigma x)^2]}};$$

$$r = \frac{\Sigma(y-\bar{y})(x-\bar{x})}{\sqrt{\Sigma(y-\bar{y})^2(x-\bar{x})^2}}; \qquad r = b\frac{\sigma_x}{\sigma_y}.$$

Коефіцієнт кореляції знаходиться в межах від 0 до ± 1. Якщо коефіцієнт кореляції дорівнює нулю, то зв'язок відсутній, а якщо одиниці, то зв'язок функціональний. Знак при коефіцієнті кореляції вказує на напрям зв'язку («+» – прямий, «–» – обернений). Чим ближче коефіцієнт кореляції до одиниці, тим зв'язок між ознаками тісніший.

Квадрат коефіцієнта кореляції називається **коефіцієнтом детермінації** (r^2). Він показує, яка частка загальної варіації результативної ознаки визначається досліджуваним фактором. Якщо коефіцієнт детермінації виражений в процентах, то його слід читати так: варіація (коливання) залежної змінної на стільки-то процентів зумовлена варіацією фактора.

Між лінійним коефіцієнтом кореляції (r) і коефіцієнтом повної регресії (b) є такий зв'язок:

$$r = b\frac{\sigma_x}{\sigma_y}. \text{ Звідси } b = r\frac{\sigma_y}{\sigma_x}.$$

Отже, знаючи коефіцієнт кореляції (r) і значення середніх квадратичних відхилень по x і y, можна визначити коефіцієнт регресії (b) і навпаки, знаючи коефіцієнт регресії (b) і відповідні середні квадратичні відхилення, можна обчислити коефіцієнт кореляції (r).

При парній лінійній залежності коефіцієнт кореляції і коефіцієнт повної регресії мають однакові знаки (плюс, мінус).

Лінійний коефіцієнт кореляції призначений для оцінки ступеня тісноти зв'язку при лінійній залежності. Для випадків нелінійного зв'язку між ознаками використовується інша формула коефіцієнта кореляції, яка випливає з правила додавання дисперсій:

$$\sigma^2_{\text{заг}} = \sigma^2_{\text{м гр}} + \sigma^2_{\text{в гр}}.$$

Із наведеної рівності видно, що чим більшим є вплив фактора на результативну ознаку, тим більшою мірою значення її дисперсії ($\sigma^2_{\text{м гр}}$) наближається до значення загальної дисперсії результативної ознаки. Відповідно, чим більше $\sigma^2_{\text{м гр}}$ і менша $\sigma^2_{\text{в.гр}}$, тим зв'язок між ознаками буде тіснішим, і навпаки. Відтак, відношення міжгрупової (факторної) і загальної дисперсій використовується для оцінки тісноти зв'язку між ознаками. Формула коефіцієнта кореляції має вигляд:

$$r = \sqrt{\frac{\sigma^2_{\text{м гр}}}{\sigma^2_{\text{заг}}}}.$$

Враховуючи, що $\sigma^2_{\text{м.гр}} = \sigma^2_{\text{заг}} + \sigma^2_{\text{в гр}}$, формулу коефіцієнта кореляції можна подати як

$$r = \sqrt{1 - \frac{\sigma^2_{\text{в.гр}}}{\sigma^2_{\text{заг}}}}.$$

Обидві формули коефіцієнта кореляції застосовуються для обчислення тісноти зв'язку при будь-якій формі зв'язку.

Із правила додавання дисперсій видно, що значення коефіцієнта кореляції перебуває в межах від 0 до 1. Знак коефіцієнта кореляції з формули не виводиться. Якщо вивчається зв'язок між двома ознаками (парна проста кореляція), то напрямок зв'язку (знак перед r) визначається безпосередньо за знаком перед коефіцієнтом регресії лінійного рівняння.

При парній криволінійній залежності тіснота зв'язку, як і при лінійній залежності, визначається за допомогою спеціального показника, аналогічного розглянутому вище коефіцієнту кореляції r.

Цей показник (щоб підкреслити його належність до криволінійного зв'язку) позначають символом i_r і називають індексом кореляції:

$$i_r = \sqrt{\frac{\sigma^2_{\text{м.гр}}}{\sigma^2_{\text{заг}}}}, \text{ або } i_r = \sqrt{1 - \frac{\sigma^2_{\text{в гр}}}{\sigma^2_{\text{заг}}}}.$$

Числове значення індексу кореляції аналогічне коефіцієнту кореляції: якщо $i_r = 1$ – зв'язок функціональний, якщо $i_r = 0$ – зв'язок відсутній; чим i_r ближче до одиниці, тим зв'язок між ознаками тісніший.

Якщо відомі коефіцієнти регресії рівняння зв'язку, то індекс кореляції можна визначити за іншою, простішою формулою. Так, при параболічній залежності формула індексу кореляції може бути подана як

$$i_r = \sqrt{\frac{\sigma^2_{\text{м.гр}}}{\sigma^2_{\text{заг}}}} = \sqrt{\frac{a\Sigma y + b\Sigma yx + c\Sigma yx^2 - n\bar{y}^2}{\Sigma y^2 - n\bar{y}^2}}.$$

або

$$i_r = \sqrt{1 - \frac{\sigma^2_{\text{в.гр}}}{\sigma^2_{\text{заг}}}} = \sqrt{1 - \frac{\Sigma y^2 - a\Sigma y - b\Sigma yx - c\Sigma yx^2}{\Sigma y^2 - n\bar{y}^2}}.$$

Тіснота зв'язку при множинній кореляції визначається за допомогою **коефіцієнта множинної кореляції** (R) і **коефіцієнта множинної детермінації** (R^2). За змістом вони аналогічні коефіцієнтам кореляції і детермінації при парному зв'язку. Їх обчислення ґрунтується на порівнянні міжгрупової (факторної) і загальної дисперсій:

$$R = \sqrt{\frac{\sigma^2_{\text{м.гр}}}{\sigma^2_{\text{заг}}}}, \text{ або } R = \sqrt{1 - \frac{\sigma^2_{\text{в.гр}}}{\sigma^2_{\text{заг}}}}.$$

Ця формула може бути застосована для визначення тісноти зв'язку при будь-якій формі зв'язку.

Величина R змінюється від 0 до 1 і розглядається як додатна, оскільки при множинних залежностях зв'язок результативної ознаки з одними факторами може бути додатним, а з іншими – від'ємним.

Для випадку залежності результативної ознаки від двох факторів формула коефіцієнта множинної кореляції має вигляд

$$R = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}},$$

де r_r – парні лінійні коефіцієнти кореляції.

Наведена формула застосовується для визначення тісноти зв'язку при лінійній залежності.

Для визначення тісноти зв'язку між результативною ознакою і кожним фактором при виключені впливу інших факторів визначають **часткові коефіцієнти кореляції,** які характеризують «чистий» вплив фактора на результативну ознаку. Для їх розрахунку використовують парні коефіцієнти кореляції.

У випадку залежності результативної ознаки від двох факторів (x_1 і x_2) можна розрахувати три коефіцієнта часткової кореляції:

1) між y і x_1 при виключенні впливу x_2:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}};$$

2) між y і x_2 при виключенні впливу x_1:

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})}};$$

3) між x_1 і x_2 при виключенні впливу y :

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{yx_2})}}.$$

Коефіцієнти кореляції при парних і множинних зв'язках, а також індекс кореляції – це відносні величини, тому вони можуть бути використані для зіставлення тісноти зв'язку по кількох явищах, що аналізуються.

Слід мати на увазі, що показники зв'язку залежать від розмаху варіювання досліджуваних ознак. Чим більшою є варіація змінних, тим вищою буде величина показників тісноти зв'язку.

Визначимо тісноту зв'язку між досліджуваними ознаками для нашого прикладу. Оскільки між продуктивністю корів і рівнем годівлі має місце лінійний зв'язок, тісноту зв'язку визначимо за допомогою лінійного коефіцієнта кореляції

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y},$$

де $\overline{xy} = \frac{\Sigma xy}{n} = \frac{15393{,}4}{10} = 1539{,}34;$ $\quad \bar{x} = \frac{\Sigma x}{n} = \frac{413}{10} =$

$= 41{,}3$ ц кормових одиниць $\quad y = \frac{\Sigma y}{n} = \frac{368}{10} = 3{,}68$ ц;

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2} = \sqrt{\frac{17296{,}7}{10} - 41{,}3^2} = 4{,}90 \text{ ц кормових одиниць;}$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - (\bar{y})^2} = \sqrt{\frac{13724{,}0}{10} - 36{,}8^2} = 4{,}26 \text{ ц.}$$

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{1539.34 - 41{,}3 \cdot 36{,}8}{4{,}90 \cdot 4{,}26} = \frac{19{,}50}{20{,}87} = 0{,}9344$$

Коефіцієнт кореляції показує, що між продуктивністю корів і рівнем годівлі має місце тісний (сильний) зв'язок.

Коефіцієнт детермінації $r^2 = 0{,}9344^2 = 0{,}8731$ показує, що 87,31% загального коливання продуктивності корів зумовлено відмінностями у рівні годівлі, а решта 12,69% (100 – 87,31) – іншими факторами, які в даному випадку не було враховано.

Коефіцієнт кореляції можна знайти і за іншими формулами:

$$1)\ r = \frac{n\Sigma yx - \Sigma y \Sigma x}{\sqrt{[n\Sigma y^2 - (\Sigma y)^2][n\Sigma x^2 - (\Sigma x)^2]}} =$$

$$= \frac{10 \cdot 15393{,}4 - 368 \cdot 413}{\sqrt{(10 \cdot 13724{,}0 - 368^2)(10 \cdot 17296{,}7 - 413^2)}} = \frac{1950{,}0}{2086{,}8} = 0{,}9344;$$

$$2)\ r = b\frac{\sigma_x}{\sigma_y} = 0{,}813\frac{4{,}8969}{4{,}2614} = 0{,}9344.$$

9.4. Криволінійна кореляція

Дослідження форми зв'язку іноді зумовлює потребу використання нелінійних (криволінійних) рівнянь регресії. Це пояснюється тим, що взаємодія між ознаками, що характеризують окремі явища і процеси, нерідко має більш складний характер, ніж просто пропорційні залежності.

Характерною особливістю цього зв'язку є те, що рівномірна зміна однієї ознаки супроводжується нерівномірною зміною (збільшенням або зменшенням) значення іншої ознаки.

Нелінійні форми зв'язку притаманні багатьом процесам у сільському господарстві. Так, ріст і розвиток рослин, накопичення ними продуктивної маси, як правило, в часі розвивається нелінійно. Відомо також, що якщо ґрунти насичені вологою більше певної норми, то урожайність сільськогосподарських культур починає знижуватися. Продуктивність корів залежно від числа отелень (віку корів) спочатку має тенденцію до зростання, досягаючи максимуму в 5-7 отеленні, а потім починає закономірно знижуватися.

При дослідженні криволінійних зв'язків, так само як і при вивченні лінійних зв'язків, принципове значення має вибір форми і рівняння зв'язку, яке найточніше виявить наявний зв'язок. Для розв'язання цього завдання використовуються ті самі прийоми, що й при обґрунтуванні лінійного зв'язку. При цьому особлива увага належить графічному методу.

Криволінійні форми зв'язку досить різноманітні. В статистичному аналізі найчастіше використовують параболу другого порядку, гіперболу і степеневу функцію.

При криволінійній залежності система рівнянь будується так само, як і для лінійного зв'язку: вихідне рівняння множиться на коефіцієнти при невідомих і добутки підсумовуються почленно. Так, система рівнянь для параболи другого порядку $\tilde{y}_x = a + bc + cx^2$ має вигляд:

$$\Sigma y = an + b\Sigma x + c\Sigma x^2;$$

$$\Sigma yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3;$$

$$\Sigma yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4.$$

Однією з особливостей параболи другого порядку є те, що вона завжди має точку перегину (критичну точку), яка характеризує оптимальний варіант розміру величини результативної ознаки і змінює свій напрям тільки один раз. Якщо в рівнянні параметр a_1 виражений додатним числом, а параметр – a_2 від'ємним, то крива змінює напрям із зростання на зниження.

Система рівнянь для гіперболи $\tilde{y}_x = a + b\frac{1}{x}$

має вигляд

$$\begin{cases} \Sigma y = an + b\Sigma\frac{1}{x}; \\ \Sigma y\frac{1}{x} = a\Sigma\frac{1}{x} + b\Sigma\frac{1}{x^2}. \end{cases}$$

Формули, які випливають із розв'язання цієї системи **рівнянь, для** визначення параметрів гіперболи мають вигляд:

$$a = \frac{\Sigma y\Sigma\frac{1}{x^2} - \Sigma\frac{y}{x}\Sigma\frac{1}{x}}{n\Sigma\frac{1}{x^2} - \Sigma\frac{1}{x}\Sigma\frac{1}{x}}; \qquad b = \frac{n\Sigma\frac{y}{x} - \Sigma y\Sigma\frac{1}{x}}{n\Sigma\frac{1}{x^2} - \Sigma\frac{1}{x}\Sigma\frac{1}{x}}.$$

Щоб полегшити обчислення параметрів рівнянь регресії способом найменших квадратів при криволінійній залежності, вибране рівняння регресії доцільно звести до лінійного вигляду відповідними перетвореннями.

Процес перетворень нелінійних рівнянь регресії в лінійні називають **лінеаризацією.**

Покажемо на прикладі трьох нелінійних функцій, найчастіше застосовуваних при вивченні взаємозв'язків, перетворення до лінійного вигляду.

1. Гіперболу $\tilde{y}_x = a + \frac{b}{x}$ зводять до лінійного вигляду, замінивши x новою змінною (її зворотним значенням $z = \frac{1}{x}$);

$$\tilde{y}_x = a + bz;$$

2. Параболу другого порядку $\tilde{y}_x = a + bx + cx^2$ перетворюють, замінивши квадрат значень факторної ознаки ($z = x^2$). Одержимо лінійну функцію з двох змінних:

$$\tilde{y}_x = a + bx + cz;$$

3. Степеневу $\tilde{y}_x = ax^b$ зводять до лінійного вигляду логарифмуванням

8

$$\lg y = \lg a + b\Sigma \lg x.$$

Подальші розрахунки аналогічні розрахункам лінійної функції. Система рівнянь має вигляд:

$$\begin{cases} \Sigma \lg y = n \lg a + b\Sigma \lg x; \\ \Sigma \lg y \lg x = \lg a\Sigma \lg x + b\Sigma(\lg x)^2. \end{cases}$$

Формули для визначення параметрів степеневої функції

$$\lg a = \frac{\Sigma \lg y(\Sigma \lg x)^2 - \Sigma \lg y \lg x\Sigma \lg x}{n\Sigma(\lg x)^2 - \Sigma \lg x\Sigma \lg x};$$

$$\lg b = \frac{n\Sigma \lg y \lg x - \Sigma \lg y\Sigma \lg x}{n\Sigma(\lg x)^2 - \Sigma \lg x\Sigma \lg x}.$$

На відміну від прямолінійної залежності коефіцієнти криволінійної регресії не можна інтерпретувати однозначно, оскільки швидкість зміни результативної ознаки при різному значенні фактора буде неоднаковою. Наприклад, якщо залежність добових надоїв від віку корів, яка характеризується тим, що із зміною віку спочатку продуктивність зростає, а потім поступово знижується, виразити рівнянням параболи другого порядку $\widetilde{y}_x = a + bx + cx^2$, то коефіцієнт a_1 покаже швидкість приросту продуктивності корів, а a_2 – її уповільнення.

Порядок визначення показників зв'язку при криволінійній залежності розглянемо на такому прикладі (табл. 9.2).

Таблиця 9.2

Дані для розрахунку показників кореляційного зв'язку

Порядковий номер	Добовий надій, кг	Вік корів, років	Розрахункові дані							Очікуване значення добових надоїв, кг
	y	x	x^2	x^3	x^4	yx	yx^2	y^2		$\widetilde{y}_x$
1	2	3	4	5	6	7	8	9		10
1	8	3	9	27	81	24	72	64		6,85
2	7	3	9	27	81	21	63	49		6,85
3	8	3	9	27	81	24	72	64		6,85
4	9	4	16	64	256	36	144	81		9,19
5	8	4	16	64	256	32	128	64		9,19
6	9	4	16	64	256	36	144	81		9,19
7	10	5	25	125	625	50	250	100		10,83
8	9	5	25	125	625	45	225	81		10,83
9	11	5	25	125	625	55	275	121		10,83
10	12	6	36	216	1296	72	432	144		11,78
11	10	6	36	216	1296	60	360	100		11,78
12	12	7	49	343	2401	84	588	144		12,01

1	2	3	4	5	6	7	8	9	10
13	13	7	49	343	2401	91	637	169	12,01
14	12	7	49	343	2401	84	588	144	12,01
15	15	8	64	512	4096	120	960	225	11,54
16	12	8	64	512	4096	96	768	144	11,54
17	11	9	81	729	6561	99	891	121	10,37
18	10	9	81	729	6561	90	810	100	10,37
19	8	10	100	1000	10000	80	800	64	8,49
20	7	10	100	1000	10000	70	700	49	8,49
Разом	**201**	**123**	**859**	**6591**	**53995**	**1269**	**8907**	**2109**	**201,00**

Для визначення форми зв'язку між добовими надоями (y) і віком корів (x) *побудуємо графік – кореляційне поле* (рис. 9.2).

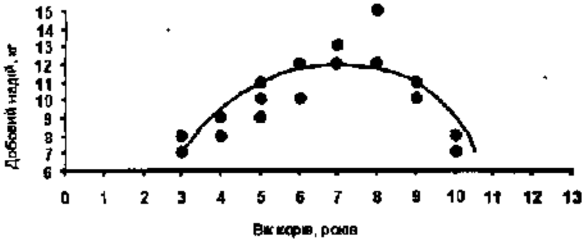


Рис. 9.2. Кореляційне поле залежності добових надоїв від віку корів

З графіка видно, що між добовим надоєм і віком корів зв'язок нелінійний. Добовий надій зростає у міру роздоювання до 6-7 отелення, а потім знижується. Розташування точок на кореляційному полі показує, що зв'язок між надоєм і віком корів можна виразити рівнянням параболи другого порядку:

$$\tilde{y}_x = a + bx + cx^2,$$

де $\tilde{y}_x$ – добовий надій, кг; x – вік корів, років; a, b, c – параметри рівняння.

Для визначення параметрів рівняння регресії a, b, c складемо систему рівнянь, для чого послідовно перемножимо всі члени вихідного рівняння на коефіцієнти при невідомих, а одержані добутки підсумуємо:

$$\begin{cases} \sum y = an + b\sum x + c\sum x^2; \\ \sum yx = a\sum x + b\sum x^2 + c\sum x^3; \\ \sum yx^2 = a\sum x^2 + b\sum x^3 + c\sum x^4. \end{cases}$$

8*

Усі потрібні для розв'язання системи нормальних рівнянь дані (Σy; Σx; Σx^2; Σx^3; Σx^4; Σyx; Σyx^2; Σy^2) розрахуємо в табл. 9.2.

Підставимо одержані дані в систему рівнянь:

$$\begin{cases} 201 = 20a + 123b + 859c; & (1) \\ 1269 = 123a + 859b + 6591c; & (2) \\ 8907 = 859a + 6591b + 53995c. & (3) \end{cases}$$

Розв'яжемо систему рівнянь і знайдемо коефіцієнти регресії a, b, c:

а) поділимо всі члени рівняння на коефіцієнти при a (перше на 20, друге – на 123, третє – на 859):

$$\begin{cases} 10{,}0500 = a + 6{,}1500b + 42{,}9500c; & (4) \\ 10{,}3171 = a + 6{,}9837b + 53{,}5854c; & (5) \\ 10{,}3690 = a + 7{,}6729b + 62{,}8580c. & (6) \end{cases}$$

б) віднімемо з 5-го рівняння 4-е і із 6-го рівняння 5-е, в результаті одержимо систему рівнянь з двома невідомими:

$$\begin{cases} 0{,}2671 = 0{,}8337b + 10{,}6354c; & (7) \\ 0{,}0519 = 0{,}6892b + 9{,}2726c; & (8) \end{cases}$$

в) поділимо обидва рівняння на коефіцієнти при b:

$$\begin{cases} 0{,}3204 = b + 12{,}7569c; & (9) \\ 0{,}0753 = b + 13{,}4541c; & (10) \end{cases}$$

г) віднімемо з 9-го рівняння 10-е:

$0{,}2451 = -\ 0{,}6972\ c$,

звідси $c = -\ 0{,}3515$;

д) підставимо значення c в рівняння 9 і знайдемо коефіцієнт регресії b:

$$0{,}3204 = b + 12{,}7569 \cdot (-\ 0{,}3515)\ ;\ b = 4{,}8044;$$

е) визначимо коефіцієнт регресії a, підставивши значення b і c в перше рівняння:

$$201 = 20a = +\ 123 \cdot 4{,}8044 + 859\ (-\ 0{,}3515);$$

$$a = -\ 4{,}4001.$$

Перевіримо правильність обчислення коефіцієнтів регресії за такою формулою:

$$\bar{y} = a + b\bar{x} + c\bar{x}^2,$$

де $\bar{y} = \dfrac{\Sigma y}{n} = \dfrac{201}{20} = 10{,}05$; $\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{123}{20} = 6{,}15$;

$$\bar{x}^2 = \frac{\Sigma x^2}{n} = \frac{859}{20} = 42{,}95.$$

$$10{,}05 = -4{,}4001 + 4{,}8044 \cdot 6{,}15 + (-0{,}3515) \cdot 42{,}95 = 10{,}05.$$

Отже, *рівняння регресії, яке характеризує зв'язок між добовим* надоєм і віком корів, має вигляд:

$$\tilde{y}_x = -4{,}4001 + 4{,}8044x - 0{,}3515x^2.$$

Коефіцієнт регресії $b = 4{,}8044$ показує, що у міру зростання віку корів до 7 років (див. графік і очікуване значення надоїв – $\tilde{y}_x$) добові надої збільшуються на 4,8044 кг, а потім із збільшенням віку продуктивність корів зменшується. Про це свідчить коефіцієнт регресії $c = -0{,}3515$ кг, який показує уповільнення приростів продуктивності корів.

Оптимальне значення фактора можна розрахувати за формулою

$$-\frac{b}{2c} = -\frac{4{,}8044}{2(-0{,}3515)} = 7 \text{ років}.$$

Визначимо очікувані (розрахункові) значення добових надоїв для різного віку корів ($\tilde{y}_x$).

Для цього в рівняння регресії замість x (вік корів) підставимо його конкретні значення $x = 3, 4, 5, ...10$. Так, очікуване значення добового надою для корів у віці 3-х років становить

$$\tilde{y}_{x=3} = -4{,}4001 + 4{,}8044 \cdot 3 - 0{,}3515 \cdot 3^2 = 6{,}85 \text{ кг};$$

для корів у віці 4-х років

$$\tilde{y}_{x=4} = -4{,}4001 + 4{,}8044 \cdot 4 - 0{,}3515 \cdot 4^2 = 9{,}19 \text{ кг і т.д.}$$

Результати розрахунків запишемо в останню колонку табл. 9.2.

Перевіримо правильність розрахунків:

$$\Sigma y = \Sigma \tilde{y}_x; \quad 201 = 201.$$

За очікуваними значеннями добових надоїв на рис. 9.2 побудуємо теоретичну лінів регресії.

$$i_r = \sqrt{\frac{\sigma^2_{\text{м гр}}}{\sigma^2_{\text{заг}}}} = \sqrt{\frac{a\Sigma y + b\Sigma yx + c\Sigma yx^2 - n\bar{y}^2}{\Sigma y^2 - n\bar{y}^2}} =$$

$$= \sqrt{\frac{-4{,}4001 \cdot 201 + 4{,}8044 \cdot 1269 + (-0{,}3515) \cdot 8907 - 20 \cdot 10{,}05^2}{2109 - 20 \cdot 10{,}05^2}} =$$

$$= \sqrt{\frac{61{,}503}{88{,}950}} = \sqrt{0{,}6914} = 0{,}8315.$$

Визначимо тісноту зв'язку між добовими надоями і віком корів, для чого розрахуємо індекс кореляції

Такий самий результат може бути одержаний і за іншою формулою (через залишкову дисперсію):

$$i_r = \sqrt{1 - \frac{\sigma^2_{\text{в.гр.}}}{\sigma^2_{\text{заг}}}} = \sqrt{1 - \frac{\Sigma y^2 - a\Sigma y - b\Sigma yx - c\Sigma yx^2}{\Sigma y^2 - n\bar{y}^2}} =$$

$$= \sqrt{1 - \frac{2109 - (-4{,}4001) \cdot 201 - 4{,}8044 \cdot 1269 - (-0{,}3515) \cdot 8907}{2109 - 20 \cdot 10{,}05^2}} =$$

$$= \sqrt{1 - \frac{27{,}447}{88{,}950}} = \sqrt{1 - 0{,}3086} = \sqrt{0{,}6914} = 0{,}8315.$$

Коефіцієнт кореляції показує, що між продуктивністю і віком корів має місце тісний зв'язок. Коефіцієнт детермінації ($i_r^2 = 0{,}8315^2 = 0{,}6914$) показує, що 69,14% відмінностей в добових надоях пов'язані з віком корів, а решта 30,86% – з іншими факторами, дію яких у даному випадку не було враховано.

9.5. Множинна кореляція

Рівень результативних показників сільськогосподарського виробництва формується під впливом цілого комплексу взаємопов'язаних між собою факторів, які діють з різною силою і з різною спрямованістю. Тому на практиці найчастіше доводиться вивчати взаємозв'язки між кількома ознаками одночасно.

Особливе значення у вивченні взаємозв'язків між ознаками в сільськогосподарському виробництві належить багатофакторному кореляційно-регресійному аналізу, при якому визначається залежність результативної ознаки від кількох факторів одночасно.

Використання ЕОМ і типових програм кореляційно-регресійного аналізу дає змогу розв'язувати кореляційні моделі різних залежностей і вибрати з цієї множини таке рівняння, яке найточніше описує ступінь наближення фактичних даних до теоретичних і відповідно дає найменшу суму квадратів відхилень фактичних даних від розрахованих за рівнянням зв'язку.

Багатофакторний кореляційно-регресійний аналіз може бути застосований для:

1) розрахунку очікуваних (теоретичних) значень результативної ознаки;

2) зіставлення і оцінки фактичного і розрахункового значень результативної ознаки;

3) порівняльного аналізу різних сукупностей;

4) об'єктивної оцінки результатів роботи підприємств;

5) виявлення резервів виробництва;

6) розроблення нормативів;

7) прогнозування суспільних явищ тощо.

Парна кореляція через те, що разом з досліджуваним фактором на *результативну ознаку впливають й інші фактори, не завжди дає правильне* уявлення про зв'язок між результативною і факторною ознакою (перебільшує або применшує міру залежності). Перевага багатофакторного кореляційно-регресійного аналізу порівняно з простою кореляцією полягає в тому, що він дає змогу оцінити ступінь впливу на результативну ознаку кожного з включених у модель (рівняння) факторів при фіксованому положенні (звичайно на середньому рівні) решти факторів.

Методологія множинної кореляції ґрунтується на загальних принципах кореляційного аналізу. Водночас в ній ускладнюється змістовний аналіз, зростає складність математичного апарату.

При формуванні множинної кореляційної моделі необхідно враховувати ряд обмежень, пов'язаних з відбором, кількістю і взаємозв'язком факторів, вибором форми зв'язку (рівняння регресії).

Відбір найістотніших факторів до кореляційної моделі є одним з найважливіших і принципових завдань багатофакторного кореляційно-регресійного аналізу. Природно, що всі фактори, які впливають на досліджувану результативну ознаку, до рівняння регресії включити не можна. З усього комплексу таких факторів необхідно відібрати найбільш важливі, істотні. Захоплення великою кількістю факторів при відносно невеликій чисельності сукупності може призвести до неякісних результатів. Крім того, із збільшенням в рівнянні регресії кількості параметрів значно утруднюється інтерпретація одержаних результатів.

Велику роль у відборі факторів відіграють завчасно побудовані і проаналізовані факторні групування. Дуже важливого значення тут набувають комбінаційні групування, які дозволяють визначити вплив на результативну ознаку фактора, що цікавить дослідника, при фіксованих значеннях інших факторів. Можна зробити безперечний висновок про те, що **статистичні групування становлять основу для кореляційного і дисперсійного аналізу і найбільшої ефективності останні досягають в поєднанні з методом групувань.**

Практичні розрахунки показують, що для забезпечення стійкості параметрів рівняння зв'язку кількість факторів, включених до моделі, має бути в 6-8 разів меншою від чисельності досліджуваної сукупності. При цьому сукупність, з якої відбирають фактори, повинна бути якісно однорідною.

Відбираючи фактори, потрібно виключати ті, що взаємно дублюють один одного і перебувають у функціональному зв'язку. Функціональний або близький до нього зв'язок між самими факторами вказує на мультико-лінеарність (для двох – колінеарність). Наявність мультиколінеарності свідчить про те, що ці фактори відображають ту саму сторону впливу на результативну ознаку.

При високій корельованості факторів (тіснота зв'язку між двома факторами перевищує $r > 0{,}8$) вплив одного з них акумулює і вплив другого. Одержані при цьому кореляційні моделі стають нестійкими.

При формуванні кореляційної моделі до неї потрібно включити один з цих факторів, який істотніше впливає на результативну ознаку. При мультиколінеарності включення до кореляційної моделі взаємопов'язаних факторів можливе тоді, коли тіснота зв'язку між ними менша, ніж тіснота зв'язку результативної ознаки з кожним фактором. Потрібно, щоб кореляційна модель містила незалежні і такі, що не дублюють один одного, фактори. Небажаним є включення до однієї моделі часткових і загальних факторів. Повністю слід виключити фактори, функціонально пов'язані з результативною ознакою.

Важкою і складною проблемою побудови рівняння множинної регресії є також вибір функції зв'язку, тобто вибір математичного рівняння, яке найповніше проявляє характер взаємозв'язку між результативною ознакою і включеними до рівняння регресії факторами.

Одна із складностей полягає у взаємозв'язку і взаємодії факторів між собою та з результативною ознакою. Тому звичайні прийоми, використовувані при виборі форми зв'язку при парній кореляції (графічний та ін.), тут малоприйнятні.

Вибір рівняння регресії може спиратися на положення теорії досліджуваного явища або практичний досвід попередніх досліджень. Якщо таких даних немає, то допомогти у вирішенні цього питання може побудова комбінаційних групувань, таблиць розподілу чисельностей, експертні оцінки, вивчення парних зв'язків між результативною ознакою і кожним фактором, графіки, перебирання функцій різних типів (при розв'язанні задач на ЕОМ), послідовний перехід від лінійних рівнянь зв'язку до більш складних видів тощо.

Виконання усіх цих прийомів пов'язане із значною кількістю зайвих підрахунків. Тому, приймаючи до уваги, що кореляційні зв'язки в більшості випадків відображаються функціями лінійного типу або степеневими, які шляхом логарифмування або заміни змінних можна звести до лінійного вигляду, рівняння множинної регресії можна будувати у лінійній формі.

При n змінних лінійне рівняння має вигляд:

$$\tilde{y}_x = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n,$$

де $\tilde{y}_x$ – залежна змінна (результативна ознака);

a_1 – незалежні змінні (фактори);

a_0 – початок відліку, який економічного смислу не має;

$a_1, a_2, ..., a_n$ – коефіцієнти регресії.

Рівняння, за допомогою якого виражається кореляційний зв'язок між кількома ознаками, називають **рівнянням множинної регресії**. Параметри рівняння регресії, так само як і у випадку парної кореляції, знаходять способом найменших квадратів.

Коефіцієнти множинної регресії показують ступінь середньої зміни результативної ознаки при зміні відповідної факторної ознаки на одиницю (одне своє значення) за умови, що всі інші фактори, які включені до рівняння регресії, залишаються постійними (фіксованими) на одному (звичайно середньому) рівні.

Коефіцієнти множинної регресії, які характеризують зв'язок між результативною ознакою і фактором при фіксованому значенні інших факторів, називаються **коефіцієнтами чистої регресії,** а коефіцієнти парної регресії – **коефіцієнтами повної регресії.**

Коефіцієнти чистої регресії, що мають різний фізичний смисл і одиниці вимірювання, не дають чіткого уявлення про те, які саме фактори найістотніше впливають на результативну ознаку. Крім того, величина коефіцієнтів регресії залежить від ступеня варіації ознаки.

Щоб привести коефіцієнти чистої регресії до порівнянного вигляду, їх виражають у стандартизованій формі у вигляді коефіцієнтів еластичності (E) і бета-коефіцієнтів (β).

Коефіцієнти еластичності показують, на скільки процентів змінюється величина результативної ознаки при зміні відповідного фактора на один процент при фіксованому значенні інших факторів.

Коефіцієнти еластичності і коефіцієнти чистої регресії зв'язані між собою таким відношенням:

$$E_i = a_i \frac{\bar{x}_i}{\bar{y}},$$

де a_i – коефіцієнт чистої регресії при i-му факторі;

$\bar{x}_i$ і $\bar{y}$ – середні значення відповідно i-го фактора і результативної ознаки.

Бета-коефіцієнти показують, на скільки середньоквадратичних відхилень σ_y зміниться результативна ознака при зміні відповідного фактора на одне значення середньоквадратичного відхилення σ_x (при постійності інших факторів, включених до рівняння регресії).

Бета-коефіцієнти обчислюються за формулою:

$$\beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y},$$

де a_i – коефіцієнт чистої регресії при i-му факторі;

σ_{x_i} і σ_y – середні квадратичні відхилення відповідно по i-му фактору і результативній ознаці.

З наведеної формули випливає, що бета-коефіцієнти мають той самий знак (плюс, мінус), що й коефіцієнти чистої регресії.

По суті, бета-коефіцієнти характеризують фактори, у розвитку яких приховуються найбільші резерви поліпшення результативної ознаки.

При парному лінійному зв'язку коефіцієнт кореляції являє собою бета-коефіцієнт:

$$r = b \frac{\sigma_x}{\sigma_y} = \beta.$$

Як зазначалося вище, коефіцієнт множинної детермінації (R^2) показує, яка частина загальної варіації результативної ознаки визначається варіацією факторів, включених до кореляційної моделі. Щоб визначити частку впливу кожного фактора у загальній варіації, треба коефіцієнт множинної детермінації розкласти на складові частини:

$$R^2 = d_1^2 + d_1^2 + ... + d_n^2,$$

де d_i^2 – коефіцієнти окремого визначення, які можна знайти за формулою:

$$d_i^2 = \frac{a_i C_{yx_i}}{\sigma_y^2}.$$

де a_i – коефіцієнти чистої регресії;

σ_y^2 – дисперсія результативної ознаки;

C_{yx_i} – спряжена варіація результативної ознаки і однієї з факторних ознак.

Величину C_{yx_i} визначають за формулою:

$$C_{yx_i} = \frac{\sum yx_i}{n} - \bar{y} \cdot \bar{x}_i.$$

Розкладання загального обсягу варіації результативної ознаки на складові частини можна здійснити й іншим способом. Для цього потрібно знайти добуток парних коефіцієнтів кореляції (r_{yx_i}) на відповідні бета-коефіцієнти (β_i), а одержані по всіх факторах результати підсумувати ($\sum r_{yx_i}\beta_i$).

Щоб визначити частку впливу кожного фактора в сумарному впливі факторів, включених до рівняння регресії, розраховують коефіцієнти окремого визначення (d_i^2):

$$d_i^2 = \frac{r_{yx_i}\beta_i}{\sum r_{yx}\beta_i} = \frac{r_{yx}\beta_i}{R^2}.$$

Якщо потрібно частку впливу кожного фактора визначити у процентах, то знайдені коефіцієнти множать на сто процентів.

Порядок визначення і аналізу показників зв'язку при множинній кореляції розглянемо на прикладі залежності урожайності зернових культур (y) від чотирьох факторів: якості ґрунту (x_1); кількості мінеральних добрив, внесених на 1 га зернових культур, ц діючої речовини (x_2); вартості силових і робочих машин на 100 га ріллі, тис. грн. (x_3) трудозабезпеченості (кількість середньорічних працівників на 100 га сільськогосподарських угідь, чол. (x_4).

Вихідні дані подамо у вигляді матриці (табл. 9.3).

Таблиця 9.3

Матриця вихідних даних для багатофакторного кореляційно-регресійного аналізу урожайності зернових культур

Господарство	Урожайність, ц/га	Якість ґрунту, балів	Кількість мінеральних добрив на 1 га зернових культур, ц діючої речовини	Вартість силових і робочих машин на 100 га ріллі, тис. грн.	Трудозабезпеченість, чол.
	y	x_1	x_2	x_3	x_4
1	2	3	4	5	6
1	18,9	46	1,2	17,9	15,6
2	19,1	48	1,2	19,4	13,7

3	19,2	50	1,3	18,9	17,0
4	19,6	56	1,3	19,7	15,8
5	20,4	52	1,6	20,0	17,3
6	20,6	65	2,2	18,1	16,4
7	21,0	60	1,6	19,0	16,1
8	21,5	55	2,3	18,8	16,6
9	21,8	58	1,6	20,5	17,5
10	22,0	78	1,4	21,8	18,0
11	22,1	60	1,7	20,3	17,4
12	23,3	76	1,9	21,3	16,0
13	23,8	71	2,1	22,2	18,2
14	23,9	59	2,1	20,4	17,5
15	24,3	65	2,1	21,3	16,3
16	25,0	66	2,1	21,8	17,1
17	25,2	68	1,8	18,1	16,8
18	25,3	67	2,2	21,2	18,8
19	25,8	71	1,9	22,8	17,3
20	26,2	73	2,0	20,8	16,5
21	26,4	75	2,5	21,8	17,1
22	26,7	76	1,6	22,0	18,9
23	28,0	59	2,6	22,3	19,8
24	28,2	79	2,7	20,4	20,8
25	28,5	57	2,7	22,6	20,0
26	28,6	65	2,5	23,0	20,4
27	28,7	61	2,8	24,0	19,9
28	29,0	80	2,7	20,2	19,4
29	29,2	89	3,0	21,9	19,8
30	29,7	86	2,8	21,5	19,0

Попереднє вивчення форми залежності між вказаними ознаками показало, що зв'язок може бути виражений за допомогою лінійного рівняння регресії:

$$\tilde{y}_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4.$$

Розв'язавши рівняння множинної регресії і розрахувавши інші показники кореляційного зв'язку на ЕОМ, одержимо таку машинограму:

1. Коефіцієнт рівняння регресії, a_i	2. Середня помилка коефіцієнта регресії, μ_{a_i}	3. t-коефіцієнт
-4,8215		2,7728
0,0730	0,0263	4,0710
2,8948	0,7110	2,4036
0,4753	0,1977	2,0837
0,4860	0,2332	
4. Парний коефіцієнт кореляції, r_{yx_i}	**5. β-коефіцієнт β_i**	**6. Коефіцієнт еластичності $Е_i$**
0,6740	0,2338	0,1964
0,8577	0,4453	0,2434
0,6895	0,2163	0,4052
0,8077	0,2356	0,3525

Продовження

7. *Середні значення* $\bar{y}$ і $\bar{x}$	8. *Середнє квадратичне відхилення*, i	9. Кореляційна матриця
24,40 65,70 2,05 20,80 17,70	3,40 10,89 0,52 1,55 1,65	1 0,6740 0,8577 0,6895 0,8077 1 0,5395 0,4143 0,4869 1 0,5186 0,7390 1 0,6167 1
10. Загальний коефіцієнт кореляції, R	**11. Коефіцієнт детермінації R^2**	**12. Різниця $y - \bar{y}_x$**
0,9375	0,8790	1. 18,90 18,098 0,802 2. 19,10 18,033 1,067 3. 19,20 19,835 – 0,035 ... ... 30. 29,70 29,011 0,689

На ЕОМ одержана така кореляційна залежність урожайності від включених до моделі факторів (1-й стовпець машинограми):

$$\tilde{y}_x = -4{,}8215 + 0{,}0730x_i + 2{,}8948x_2 + 0{,}4753x_3 + 0{,}4860x_4.$$

Подальший аналіз пов'язаний з перевіркою значущості коефіцієнтів регресії. Для цього визначимо табличне значення t-критерію нормального розподілу ($n > 30$) і порівняємо його з фактичними значеннями (3-й стовпець машинограми).

Табличне значення t-критерію нормального розподілу при заданому рівні довірчої імовірності $P = 0{,}95$ становитиме $t = 1{,}96$ (дод. 2).

Відповідні фактичні значення *нормованих відхилень для коефіцієнтів* регресії такі:

$$t_{a_1} = 2{,}7728; \quad t_{a_3} = 2{,}4036;$$

$$t_{a_2} = 4{,}0710; \quad t_{a_4} = 2{,}0837.$$

Фактичні значення коефіцієнтів t вище табличного значення (t = 1,96). Тому наведене вище рівняння регресії можна використати для подальшого аналізу.

Коефіцієнти регресії показують, на скільки зміниться урожайність зернових культур при зміні кожного фактора на одиницю його виміру при фіксованих значеннях інших факторів, включених до рівняння. Так, поліпшення якості ґрунту на один бал збільшує урожайність на 0,0730 ц/га, збільшення дози внесення добрив на 1 ц діючої речовини – на 2,8948 ц/га, збільшення вартості силових і робочих машин на 100 га ріллі на одну тисячу гривень – на 0,4753 ц/га, збільшення трудозабезпеченості працівниками на 100 га сільськогосподарських угідь на одного чоловіка – на 0,4860 ц/га.

Коефіцієнт множинної кореляції (10-й стовпець машинограми), який характеризує тісноту зв'язку між урожайністю та її факторами, дорівнює $R = 0{,}9375$.

Коефіцієнт множинної детермінації (11-й стовпець машинограми) $R^2 = 0{,}9375^2 = 0{,}8790$ показує, що варіація урожайності в зв'язку із зміною розглядуваних факторів становить 87,90%.

Тісноту зв'язку між ознаками, включеними в рівняння регресії, характеризують 4-й і 9-й стовпці машинограми і на їх основі складена така матриця парних коефіцієнтів кореляції:

	y	x_1	x_2	x_3	x_4
y	1	0,6740	0,8577	0,6895	0,8077
x_1		1	0,5395	0,4143	0,4689
x_2			1	0,5186	0,7390
x_3				1	0,6167
x_4					1

З даних матриці видно, що урожайність знаходиться в досить тісному зв'язку з включеними до моделі факторами. Так, тіснота зв'язку між урожайністю і якістю ґрунту становить $r_{yx_1} = 0{,}6740$, між дозами добрив $r_{yx_2} = 0{,}8577$, між забезпеченістю технікою і робочою силою – відповідно $r_{yx_3} = 0{,}6895$ і $r_{yx_4} = 0{,}8077$. Значна тіснота зв'язку спостерігається і між факторами.

Найбільший вплив на урожайність зернових культур, якщо робити висновки за наведеним рівнянням регресії, мають дози внесених добрив і трудозабезпеченість, тому що коефіцієнти регресії при цих коефіцієнтах найбільші ($a_2 = 2{,}8948$, $a_4 = 0{,}4880$); потім забезпеченість технікою ($a_3 = 0{,}4753$) і якість ґрунту ($a_1 = 0{,}0730$).

Однак, коефіцієнти регресії, що мають різний фізичний смисл і одиниці вимірювання, не дають чіткого уявлення про те, які фактори найістотніше впливають на урожайність. Для проведення такого аналізу на ЕОМ розраховані коефіцієнти еластичності, які показують, на скільки процентів зміниться величина результативної ознаки у разі зміни величини фактора на 1% при фіксованому значенні інших факторів (6-й стовпець машинограми).

На підставі обчислених коефіцієнтів еластичності

$$E_1 = 0{,}1964; \quad E_3 = 0{,}4052;$$

$$E_2 = 0{,}2434; \quad E_4 = 0{,}3525$$

можна зробити висновок, що збільшення на 1% забезпеченості технікою веде до збільшення урожайності відповідно на 0,4052 %, трудозабезпеченості – на 0,3525 %, добрив – на 0,2434 і якості ґрунту – на 0,1964 %.

Таким чином, найбільший вплив на урожайність мають забезпеченість технікою і робочою силою.

Проте і цих даних недостатньо, щоб скласти об'єктивне уявлення про те, як по групі досліджуваних господарств розподіляються фактори за їх впливом на резерви зростання урожайності зернових культур.

З цією метою на ЕОМ обчислюють β-коефіцієнти, які показують, на скільки середньоквадратичних відхилень σ_y зміниться результативна ознака (урожайність) при зміні відповідного фактора на одне значення свого середньоквадратичного відхилення σ_y. По суті, β-коефіцієнти характеризують фактори, в розвитку яких приховано найбільші резерви збільшення результативної ознаки (урожайності).

Фактичні значення β-коефіцієнтів (5-й стовпець машинограми) такі:

$$\beta_1 = 0{,}2338;\ \beta_3 = 0{,}2163;$$
$$\beta_2 = 0{,}4453;\ \beta_4 = 0{,}2356.$$

У розрахованій нами моделі найбільші можливості збільшення урожайності закладено в добривах (β_2 = 0,4453), тому що при зміні на одне середнє квадратичне відхилення доз добрив урожайність змінюється на 0,4453 свого середнього квадратичного відхилення. Далі за ступенем впливу йдуть такі фактори: забезпеченість робочою силою (β_4 = 0,2356), якість грунту (β_1 = 0,2338) і забезпеченість технікою (β_3 = 0,2163). Сила впливу на урожайність останніх трьох факторів практично є однаковою.

Коефіцієнт множинної детермінації, який дорівнює R^2 = 0,8790, свідчить про те, що коливання урожайності, які пояснюються варіацією включених до рівняння регресії факторів, дорівнює 87,9 %. Викликає інтерес розкладання загального обсягу варіації урожайності на варіацію за рахунок кожного включеного в рівняння регресії фактора. Для цього розрахуємо коефіцієнти детермінації, які визначаються як добуток парних коефіцієнтів кореляції на β-коефіцієнти за відповідними факторами (4-й і 5-й стовпці машинограми).

Усі розрахунки зведемо в табл. 9.4.

Таблиця 9.4

Розкладання загального обсягу варіації за факторами

Номер	Фактор	Парний коефіцієнт кореляції	β-коефіцієнт	Добуток, %
	x_i	r_{yx_i}	β_i	$r_{yx_i}\beta_i$
1	x_1 – якість ґрунту, балів	0,6740	0,2338	15,8
2	x_2 – добрива, ц діючої речовини	0,8577	0,4453	38,2
3	x_3 – силові і робочі машини, тис. грн.	0,6895	0,2163	14,9
4	x_4 – трудозабезпеченість, чол.	0,8077	0,2356	19,0
	Разом	–	–	87,9

Таким чином, із 87,9% загального коливання урожайності зернових культур 15,8% пояснюється варіацією якості грунту, 38,2% – кількістю внесених добрив, 14,9% – забезпеченістю технікою і 19,0% – трудозабезпеченістю. Найвпливовішим фактором, як показали розрахунки, є добрива.

Заключним етапом багатофакторного кореляційно-регресійного аналізу є оцінка результатів роботи кожного господарства за досягнутим рівнем урожайності зернових культур. Для цього потрібно порівняти розрахунковий (теоретичний) і фактичний рівні урожайності (12-й стовпець машинограми). З метою економії місця тут наведено дані лише для перших трьох і останнього господарства.

Плюсова різниця свідчить про те, що в цих господарствах фактична урожайність виявилася вище розрахункової, а мінусова різниця навпаки.

Порівняння фактичного і теоретичного рівнів урожайності в окремих господарствах показує, що в 14 господарствах з 30 урожайність виявилася вище розрахункової. В інших 16 господарствах вона не перевищувала теоретичного рівня. Це говорить про те, що ще в багатьох господарствах не використано достатньою мірою наявні резерви збільшення урожайності зернових культур.

9.6. Статистична оцінка вибіркових показників зв'язку

У тих випадках, коли вивчення кореляційної залежності базується на вибіркових даних, виникає потреба оцінки вибіркових показників кореляції (коефіцієнтів регресії і кореляції).

Статистична оцінка виб!ркових показниюв кореляцп дае змогу зробити висновок про те, наскільки виб1рков1 статистичш показники вщповщають показникам генеральнсм сукупносп. Однак така оцінка проводиться у випадках, коли: 1) ВН6ірКа сформована у випадковому порядку; 2) виб1рка зроблена з нормально розподітено1 сукупносп; 3) вщхилення фактичних значень результативно:! ознаки вщ її теоретичних значень, обчислених за р!внянням> також розподігеш нормально.

Розглянемо порядок статистично1 оцінки виб1ркових показниюв зв'язку при парній лінійній регресії.

В кореляційному анал1з1 середня помилка виб1рки обчислюеться на основ! залишковО1 дисперсй, оск1льки ця величина характеризуе то-4Нісгь піщбору кривО1 до фактичних даних. Проте залишкова дисперсія, розрахована за виб!рковими даними, зменшуе дійсну величину залишково1 дисперсп в генеральнШ сукупносі, тобто е зміщеною оцінкою. Це зміщення коригуеться внесениям в знаменник формули залишково"1 дисперси поправки на втрату ступен1В свободи. При парнш лшшнш залежносп втрачаються вщповщно числу параметр1в р1вняння (я і *Ъ)* два ступеня свободи, при кореляцй трьох зм1нних з параметрами я, *Ь'*\ с-три ступеня свободи і т.д.

Квадрат середньо1 помилки виб!ркового коефіц1ента регресії являе собою вiщношення залишково'1 дисперсй, скоригованоі на втрату числа *ступетв* свободи вар!ацй до суми квадра^в віщхилень незалежно! змішно!.

Позначаючи залишкову дисперс1ю через S^, а квадрат середньо*1 помилки виб!ркового коеф!ц!ента регресії через *fi*,* одержимо

$$\mu^2_{в} = \frac{S^2_{yx}}{\sum(x_i - \bar{x})^2}; \quad де \quad S^2_{yx} = S^2_{зал}\frac{n}{n-m} = \frac{\sum(y_i - \tilde{y}_x)^2}{n-m};$$

Де *n* - чисельн1сть виб)рки; *m* - юльюсть параметр1в р1вняння **регресії,** яке дор1внюе двом при парши лЫйшй залежност1.

Вщпов1дно середня помилка коефщента регресії:

$$\mu_{в} = \sqrt{\frac{S^2_{yx}}{\sum(x_i - \bar{x})^2}}.$$

Гранична помилка виб!ркового коефнйента рарєсй визначаеться за формулою:

$$\varepsilon_{в} = t\mu_{в},$$

де t – значення нормованого відхилення, величина якого встановлюється за таблицями. Для великих вибірок ($n > 30$) значення t знаходять за дод. 2, для малих вибірок ($n < 30$) – за дод. 3.

Довірчі межі коефіцієнта регресії у генеральній сукупності (b_0) становитимуть:

$$b_0 = в \pm t\mu_в.$$

Вірогідність вибіркового коефіцієнта регресії визначається як відношення:

$$t = \frac{b}{\mu_в}$$

Якщо $t_{факт} > t_{табл}$ при заданому рівні значущості і відповідному числі ступенів свободи варіації, то нульова гіпотеза про рівність коефіцієнта регресії у генеральній сукупності нулю ($b_0 = 0$) відхиляється і робиться висновок про те, що вибірковий коефіцієнт регресії є вірогідним. Якщо ж $t_{факт} < t_{табл}$, то нульова гіпотеза приймається і робиться висновок про те, що значення b у вибірці є неістотним, випадковим.

Обчислимо середню і граничну помилку для коефіцієнта регресії, що характеризує залежність продуктивності корів від рівня годівлі ($b = 0{,}813$ ц).

Визначимо залишкову дисперсію, використовуючи коефіцієнти рівняння регресії $\tilde{y}_x = 3{,}22 + 0{,}813\ x$ і дані табл. 9.1.

$$\sigma^2_{зал} = \frac{\Sigma y^2 - a\Sigma y - в\Sigma yx}{n} =$$

$$= \frac{1374{,}0 - 3{,}22 \cdot 368 - 0{,}813 \cdot 15393{,}4}{10} = \frac{24{,}2}{10} = 2{,}42.$$

Обчислимо скориговану залишкову дисперсію

$$S^2_{yx} = \sigma^2_{зал}\frac{n}{n-m} = 2{,}42\frac{10}{10-2} = 3{,}025,$$

де m – число параметрів рівняння регресії ($m = 2$).

Визначимо середню помилку параметра b:

$$\mu_в = \sqrt{\frac{S^2_{yx}}{\Sigma(x_i - \tilde{x})^2}} = \sqrt{\frac{3{,}025}{239{,}8}} = \sqrt{0{,}0126} = 0{,}1123\ \text{ц},$$

де $\Sigma(x_i - \tilde{x})^2 = n\sigma^2_x = 10 \cdot 4{,}8969 = 239{,}8$,

або $\Sigma(x_i - \tilde{x})^2 = \Sigma x^2 - n\tilde{x}^2 = 17296{,}7 - 10 \cdot 41{,}3^2 = 239{,}8.$

Перевіримо вірогідність вибіркового коефіцієнта регресії b, висунувши нульову гіпотезу, а саме: коефіцієнт регресії в генеральній сукупності дорівнює нулю: $H_0 : b_0 = 0$; $\quad H_a : b_0 \neq 0$.

Розрахуємо фактичне значення t-критерію Стьюдента:

$$t = \frac{b}{\mu_в} = \frac{0{,}813}{0{,}1123} = 7{,}240.$$

За таблицею (дод. 3) при $\alpha = 0{,}05$ і числі ступенів свободи $k = n - m = 10 - 2 = 8$ знайдемо $t_{0,05} = 2{,}307$.

Оскільки $t_{факт} > t_{0,05}$ $(7{,}240 > 2{,}307)$, від нульової гіпотези, яка передбачає відсутність зв'язку між урожайністю і якістю ґрунту в генеральній сукупності $(b_0 = 0)$, слід відмовитись. Вибірковий коефіцієнт регресії $b = 0{,}813$ є вірогідним, істотним.

Обчислимо граничну помилку вибіркового коефіцієнта регресії:

$$\varepsilon_в = t_{0,05}\mu_в = 2{,}307 \cdot 0{,}1123 = 0{,}26 \text{ ц.}$$

Визначимо інтервал, в якому із заданим рівнем значущості знаходиться коефіцієнт регресії в генеральній сукупності:

$$b_0 = b \pm \varepsilon_в = 0{,}813 \pm 0{,}26, \text{ або } 0{,}553 \leq b_0 \leq 1{,}073.$$

Отже з рівнем значущості $\alpha = 0{,}05$ (з імовірністю помилитись в 5 випадках із 100) можна стверджувати, що величина коефіцієнта регресії, який характеризує зв'язок між продуктивністю корів і рівнем годівлі в генеральній сукупності, перебуває в інтервалі від 0,553 до 1,073 ц на 1 ц кормових одиниць.

Для перевірки вірогідності вибіркового коефіцієнта кореляції визначають його середню і граничну помилки вибірки.

Середня помилка вибіркового коефіцієнта кореляції визначається за формулою:

$$\mu_r = \frac{1 - r_0^2}{n - m},$$

де r_0 - значення коефіцієнта кореляції в генеральній сукупності, яке наближено замінюється значенням вибіркового коефіцієнта кореляції; n - чисельність вибірки; m - число параметрів рівняння регресії.

Застосування цієї формули пов'язано з двома обмеженнями: 1) чисельність вибірки повинна бути достатньо великою; 2) вибірковий коефіцієнт кореляції не повинен бути близьким до одиниці. У тих випадках, коли вибірковий коефіцієнт кореляції близький до одиниці $(r \geq 0{,}8)$, для перевірки гіпотези про його вірогідність рекомендується застосовувати метод перетвореної кореляції, запропонований Р.Фішером.

Гранична помилка вибіркового коефіцієнта кореляції визначається за формулою:

$$\varepsilon_r = t\mu_r.$$

Довірчі межі, в яких знаходиться значення коефіцієнта кореляції в генеральній сукупності, становитимуть

$$r_0 = r \pm t\mu_r.$$

Вірогідність вибіркового коефіцієнта кореляції визначається як відношення

$$t_{\text{факт}} = \frac{r}{\mu_r}.$$

Якщо $t_{\text{факт}} > t_{\text{табл}}$, то нульова гіпотеза про рівність нулю коефіцієнта кореляції у генеральній сукупності відхиляється і робиться висновок про вірогідність одержаного за вибіркою коефіцієнта кореляції. Якщо ж $t_{\text{факт}} < t_{\text{табл}}$, то нульова гіпотеза приймається і робиться висновок про те, що отримане за вибіркою значення коефіцієнта кореляції неістотне, має випадковий характер.

Крім розглянутого прийому, оцінка вибіркового коефіцієнта кореляції може бути проведена більш простими способами. Для визначення вірогідності вибіркового коефіцієнта кореляції можна користуватися спеціальними таблицями (дод. 9), в яких наводяться критичні значення коефіцієнта кореляції при заданому рівні значущості (α) і відповідному числі ступенів волі варіації ($k = n - m$).

Якщо при заданому α і відповідному k фактичне значення вибіркового коефіцієнта кореляції більше табличного значення ($r_{\text{табл}}$), то тіснота зв'язку між досліджуваними ознаками вважається вірогідною і навпаки.

Проведемо оцінку вірогідності вибіркового коефіцієнта кореляції, що характеризує тісноту зв'язку між продуктивністю корів і рівнем годівлі ($r = 0{,}9344$).

Оскільки у прикладі чисельність вибірки невелика ($n = 10$), а вибірковий коефіцієнт кореляції близький до одиниці ($r = 0{,}9344$), оцінку його вірогідності проведемо за допомогою методу Р.Фішера, який дістав назву перетвореної кореляції.

Р.Фішер довів, що розподіл логарифмічної функції вибіркового лінійного коефіцієнта кореляції (z) наближається до кривої нормального розподілу навіть при невеликому обсязі вибірки і високому значенні r.

Величина z визначається за формулою

$$z = \frac{1}{2}\ln\frac{1+r}{1-r}.$$

Перехід від r до z і назад здійснюється за допомогою спеціальних таблиць, що виключають потребу логарифмування.

Середня квадратична помилка z-розподілу залежить тільки від обсягу вибірки і визначається за формулою:

$$\mu_z = \frac{1}{\sqrt{n-3}}.$$

Обчислимо середню помилку z-розподілу для нашого прикладу

$$\mu_z = \frac{1}{\sqrt{10-3}} = \frac{1}{7} = 0{,}378.$$

За таблицею (дод. 8) знайдемо, що коефіцієнту кореляції r = 0,9344 відповідає z = 1,5890.

Визначимо відношення z до середньої помилки вибіркового коефіцієнта кореляції:

$$t_{\text{факт}} = \frac{z}{\mu_z} = \frac{1{,}658}{0{,}378} = 4{,}386.$$

Знайдемо табличне значення t-критерію Стьюдента (дод. 3) при α = 0,05 і $k = 10 - 2 = 8$; $t_{0,05} = 2{,}307$.

Оскільки фактичне відношення t виявилося більше табличного $t_{0,05}$ ($4{,}386 > 2{,}307$), то можна зробити висновок про те, що висунута гіпотеза про рівність нулю коефіцієнта кореляції у генеральній сукупності не узгоджується з фактичними даними, у зв'язку з чим вона повинна бути відхилена. Вибірковий коефіцієнт кореляції є вірогідним, істотним.

Побудуємо довірчий інтервал, в якому із заданим рівнем значущості знаходиться коефіцієнт кореляції в генеральній сукупності:

$r_0 = z \pm t\mu_z = 1{,}658 + 2{,}307 \cdot 0{,}378 = 1{,}658 \pm 0{,}872$, тобто від 0,786 до 2,530.

Користуючись таблицею значень z у зворотному порядку, знайдемо границі довірчого інтервалу для коефіцієнта кореляції в генеральній сукупності:

$$0{,}66 < r_0 < 0{,}99.$$

Отже, із заданим рівнем значущості = 0,05 можна стверджувати, що тіснота зв'язку між продуктивністю корів і рівнем годівлі в генеральній сукупності знаходиться в межах від 0,66 до 0,99.

Вірогідність вибіркового коефіцієнта кореляції може бути встановлена і без обчислень за таблицею Р.Фішера (дод. 9).

Для нашого прикладу табличне значення коефіцієнта кореляції при $\alpha = 0,05$ і $k = 8$ становитиме $r_{0,05} = 0,632$.

Оскільки $r_{факт} > r_{0,05}$ (0,9344 > 0,632), можна підтвердити попередній висновок про те, що вибірковий коефіцієнт кореляції є вірогідним. Це дає підставу для висновку про дійсний зв'язок між продуктивністю корів і рівнем годівлі в генеральній сукупності

9.7. Непараметричні критерії оцінки кореляційного зв'язку

Наведені вище формули для визначення тісноти зв'язку між ознаками передбачають, що сукупності, до яких вони застосовуються, мають нормальний або близький до нормального розподіл. Якщо ж характер розподілу досліджуваної сукупності навіть передбачувано невідомий, то тісноту зв'язку можна обчислити за допомогою **непараметричних критеріїв** визначення тісноти зв'язку.

Особливістю цих критеріїв є те, що тіснота зв'язку між ознаками визначається не за кількісними значеннями варіантів, а за допомогою порівняння їх рангів. Під **рангом** розуміють порядковий номер одиниці сукупності в ранжированому ряду розподілу. Чим менші розбіжності між рангами, тим тісніший зв'язок між ознаками

До непараметричних критеріїв показників тісноти зв'язку відносяться коефіцієнти: кореляції рангів, знаків Фехнера, асоціації, контингенції та ін

Коефіцієнт кореляції рангів – це один з найпростіших показників тісноти зв'язку (його ще називають ранговим коефіцієнтом кореляції Спірмена). Суть його розрахунку полягає в такому Парні спостереження двох взаємопов'язаних ознак (результативної і факторної) ранжируються, а потім відповідно величині ознаки їм надається ранг від 1 до n. Тіснота зв'язку визначається на основі близькості рангів, і формула коефіцієнта кореляції рангів буде мати вигляд:

$$r_p = 1 - \frac{6\Sigma d^2}{n(n-1)}.$$

де d – різниці між величинами рангів в порівнюваних рядах; n – число спостережень.

Смисл його такий самий, як і лінійного коефіцієнта кореляції. Коефіцієнт кореляції рангів, як і лінійний коефіцієнт кореляції, може приймати значення від – 1 до + 1. Якщо ранги двох паралельних рядів повністю співпадають, то $\Sigma d^2 = 0$, і тоді має місце прямий функціональний зв'язок, а $r_p = 1$. При повному зворотному зв'язку (ранги розміщуються в зворотному порядку) $r_p = -1$. Ранжирувати обидві ознаки потрібно в одному і тому самому порядку: або від менших значень ознаки до більших, або навпаки.

Методику розрахунку коефіцієнта кореляції рангів покажемо на прикладі визначення тісноти зв'язку між урожайністю хмелю і кількістю внесених органічних добрив на 1 га хмелю (табл. 9.5).

$$r_p = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \quad 16}{10(10^2-1)} = 1 - 0{,}097 = 0{,}903$$

Таблиця 9.5

Дані для розрахунку коефіцієнта кореляції рангів

Господарство	Урожайність хмелю, ц/га	Внесено органічних добрив на 1 га хмелю, т	Ранги		Різниця рангів	Квадрат різниці рангів
			за урожайністю	за добривами		
	y	x	R_y	R_x	$d = R_y - R_x$	d^2
1	9,5	42	1	2	–1	1
2	10,1	45	2	4	–2	4
3	10,6	40	3	1	2	4
4	11,5	46	4	5	–1	1
5	12,7	44	5	3	2	4
6	12,9	49	6	6	0	0
7	13,5	51	7	7	0	0
8	15,6	60	8	9	–1	1
9	17,0	58	9	8	1	1
10	18,2	70	10	10	0	0
Разом	–	–	–	–	–	**16**

Розрахований коефіцієнт кореляції рангів свідчить про наявність прямого тісного зв'язку між урожайністю хмелю і кількістю внесених органічних добрив.

Вірогідність коефіцієнта кореляції рангів можна перевірити за таблицею Фішера (дод. 9). Табличне значення коефіцієнта кореляції при $\alpha = 0{,}05$ і $k = n - m = 10 - 2 = 8$ становить $r_p = 0{,}632$. Оскільки $r_{факт} > r_{005}$ (0,903 > 0,632), можна зробити висновок про те, що вибірковий коефіцієнт кореляції рангів є вірогідним.

Недоліком коефіцієнта кореляції рангів є те, що однаковим різницям можуть відповідати зовсім відмінні різниці значень ознак (у випадку кількісних ознак). Тому для останніх слід вважати кореляцію рангів приблизною мірою оцінки тісноти зв'язку.

Коефіцієнт кореляції рангів може бути також використаний для визначення тісноти зв'язку між якісними (атрибутивними) ознаками, яким може бути надана рангова оцінка.

Коефіцієнт Фехнера застосовується для оцінки тісноти зв'язку на основі порівнянь знаків відхилень значень результативної і факторної ознак від їх середніх, його обчислюють за формулою

$$k = \frac{\Sigma a - \Sigma b}{\Sigma a + \Sigma b},$$

де Σa – сума збігів знаків; Σb – сума незбігів знаків.

Коефіцієнт Фехнера змінюється від 0 до ± 1. Якщо знаки всіх відхилень збігаються, то $\Sigma b = 0$, а коефіцієнт Фехнера дорівнює одиниці, що свідчить про наявність прямого зв'язку. Якщо і знаки всіх відхилень будуть різними, то $\Sigma a = 0$, а коефіцієнт Фехнера дорівнює – 1, що вказує на наявність оберненого зв'язку.

Розглянемо порядок обчислення коефіцієнта Фехнера на прикладі табл. 9.6.

Таблиця 9.6

Дані для розрахунку коефіцієнта Фехнера

Господарство	Щільність корів на 100 га сільськогосподарських угідь, гол.	Виробництво молока на 100 га сільськогосподарських угідь, ц	Різниця щодо середньої (+, –)		Збіг (a), незбіг (b) знаків
			для x	для y	
	x	y			
1	14	356	–	–	a
2	15	320	–	–	a
3	15	366	–	–	a
4	17	425	–	+	b
5	19	382	+	–	b
6	21	450	+	+	a
7	25	480	+	+	a
Разом	126	2779	–	–	–
У середньому	18	397	–	–	–

Знак мінус означає, що значення ознаки менше середньої, знак плюс – більше середньої. Збіг знаків по обох ознаках означає узгоджену варіацію, незбіг – порушення узгодженості.

Коефіцієнт Фехнера для нашого прикладу становитиме

$$k_{\phi} = \frac{\Sigma a - \Sigma b}{\Sigma a + \Sigma b} = \frac{5-2}{5+2} = \frac{3}{7} = 0{,}4286.$$

Одержана додатна величина коефіцієнта Фехнера свідчить про те, що між виробництвом молока на 100 га сільськогосподарських угідь і чисельністю корів є прямий кореляційний зв'язок.

Слід мати на увазі, що коефіцієнт Фехнера тільки констатує наявність і напрям кореляційного зв'язку і не залежить від величини відхилень результативної і факторної ознак від відповідних середніх, у зв'язку з чим оцінка тісноти зв'язку є наближеною. Коефіцієнт Фехнера може бути деяким орієнтиром в оцінці інтенсивності зв'язку.

Тісноту зв'язку між атрибутивними (якісними) ознаками можна **виміряти** за допомогою спеціальних коефіцієнтів **асоціації і контингенції**, запропонованих відповідно Д.Юлом і К.Пірсоном.

Для їх обчислення будується чотириклітинна таблиця, яка показує зв'язок між двома ознаками, кожна з яких повинна бути альтернативною, тобто такою, що складається з двох якісно відмінних один від одного значень (наприклад, стан посівів задовільний або незадовільний, землі удобрені або не удобрені та ін.).

Загальна схема чотириклітинної таблиці має вигляд (табл. 9.7).

Таблиця 9.7

Чотириклітинна таблиця для розрахунку коефіцієнтів асоціації і контингенції

Ознаки	A	Не A	ΣB
B	a	b	$a+b$
Не B	c	d	$c+d$
ΣA	$a+c$	$b+d$	N

В цій таблиці A і B ознаки, між якими вивчається зв'язок; не A і не B – протилежні (альтернативні) ознаки; a, b, c, d – частоти відповідних комбінацій ознак; N – загальне число спостережень.

Коефіцієнти обчислюються за формулами:

асоціації $r = \dfrac{ad - bc}{ad + bc}$;

контингенції $r = \dfrac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$.

Методику розрахунку коефіцієнтів асоціації і контингенції розглянемо на прикладі визначення тісноти зв'язку між двома якісними ознаками: термінами обробки хмільників пестицидами і ступенем їх ураженості хворобами (табл. 9.8).

Таблиця 9.8

Розподіл ділянок хмільників, уражених і не уражених хворобами

Обробка хмільників	Ділянки		Разом
	не уражені	уражені	
Своєчасна	60	15	75
Із запізненням	18	47	65
Разом	78	62	140

У таблиці $a = 60$; $b = 15$; $c = 18$; $d = 47$.

Розрахуємо коефіцієнти

асоціації $r = \dfrac{ad - bc}{ad + bc} = \dfrac{60 \cdot 47 - 15 \cdot 18}{60 \cdot 47 + 15 \cdot 18} = \dfrac{2550}{3090} = \mathbf{0{,}825;}$

контингенції $r = \dfrac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} =$

$$= \frac{60 \cdot 47 - 15 \cdot 18}{\sqrt{(60 + 15)(18 + 47)(60 + 18)(15 + 47)}} = \frac{2550}{4855{,}5} = 0{,}525.$$

Одержані коефіцієнти асоціації і контингенції вказують на досить тісний зв'язок між термінами обробки і ураженістю хворобами хмільників. При цьому коефіцієнт контингенції дає більш обережну оцінку тісноти зв'язку між ознаками.

Коефіцієнти асоціації і контингенції можуть приймати будь-які значення від -1 до $+1$. Коефіцієнт контингенції завжди менше коефіцієнта асоціації. Для великих вибірок ($n \geq 30$) зв'язок практично вважається значущим, якщо $r \geq 0{,}5$ або $r \geq 0{,}3$. Величини коефіцієнтів асоціації і контингенції, як показників тісноти зв'язку, тлумачаться так само, як і величина коефіцієнта кореляції.

9.8. Особливості кореляційного аналізу в рядах динаміки

Наведені вище приклади кореляційного аналізу обчислені на матеріалах просторових статистичних сукупностей. Однак при вивченні зміни явищ у часі часто виникає потреба оцінювати ступінь взаємозв'язку рівнів яких-небудь рядів динаміки різного змісту, але пов'язаних між собою. Наприклад, можна поставити запитання: якою мірою зміна в динаміці забезпеченості виробництва добривами, технікою впли-ває на зміну урожайності, або рівня годівлі тварин на їх продуктивність і т.д. У цьому випадку застосовують кореляцію рядів динаміки.

Кореляційний аналіз динамічних рядів проводиться за тією самою схемою, що й аналіз взаємозв'язків у просторі. Разом з тим є деякі особливості застосування методів кореляції щодо аналізу рядів динаміки. Однією з таких особливостей є наявність для більшості рядів динаміки основної тенденції (тренда) в зміні їх рівнів, тоді як однією з умов застосування теорії кореляції є незалежність окремих спостережень. В динамічних же рядах фактором зміни рівнів виступає, крім інших, і час, тому, як правило, кожний наступний часовий рівень пов'язаний з попереднім. Наприклад, обсяги виробництва сільськогосподарської продукції безумовно величини, які залежать від рівня їх виробництва в попередні роки. Те ж саме можна сказати і про урожайність сільськогосподарських культур, продуктивність тварин тощо. Така залежність між кожним наступним і попереднім рівнями ряду динаміки одержала назву **автокореляції.**

Кореляція між рівнями динамічних рядів правильно показує тісноту зв'язку між ознаками лише в тому випадку, якщо в кожному з них відсутня автокореляція. Наявність чи відсутність автокореляції завжди потрібно перевіряти, і однією з умов наукового застосування методів кореляційного аналізу стосовно рядів динаміки є виключення трендів з обох рядів. У загальному випадку можна передбачати, що в рядах динаміки, які складаються з відхилень від тренда, автокореляції немає.

Другою особливістю, яку необхідно брати до уваги при зіставленні рівнів рядів динаміки, є наявність часового лагу. Під **часовим лагом** розуміють, з одного боку, період, через який спостерігається сильна кореляція між рівнями одного і того самого ряду, і, з іншого боку, період відставання у розвитку двох взаємопов'язаних рядів. Найтісніший зв'язок у двох взаємопов'язаних рядах виникає, якщо зсунути один ряд відносно другого на період лагу. Тому при наявності відставання в розвитку двох взаємопов'язаних показників необхідно зсунути рівні одного ряду відносного другого на деякий проміжок часу, що дасть змогу одержати більш правильну оцінку ступеня тісноти кореляційного зв'язку.

Третьою особливістю кореляції рядів динаміки є можливість змінної кореляції. **Змінною кореляцією** називають зміну показників тісноти зв'язку (коефіцієнтів кореляції) протягом часу. Тому показник тісноти зв'язку в динамічних рядах можна подати як серію коефіцієнтів, розрахованих подібно ковзній середній. За цією серією коефіцієнтів кореляції, обчислених у ковзному порядку з виключенням одного члена ряду і включенням наступного через певний інтервал ковзання (триріччя, п'ятиріччя і т.д.), можна одержати більш повні відомості про те, як змінювалась тіснота зв'язку в часі. Це дає змогу вияснити ті передумови, які призвели до зміни тісноти зв'язку між відображуваними в рядах динаміки явищами.

Насамкінець, четвертою особливістю кореляції в рядах динаміки є те, що при їх аналізі, як правило, є досить обмежена кількість спостережень, особливо коли вивчається зміна будь-якого явища по роках. Крім того, залучення до аналізу даних за тривалий період часу спряжено з непорівнянністю за територією, часом обліку, методикою обчислення показників і т. ін.

При кореляційному аналізі динамічних рядів доводиться розв'язувати два завдання: 1) виміряти зв'язок послідовних рівнів одного і того самого ряду динаміки; 2) виміряти зв'язок між змінами ознак двох динамічних рядів різного змісту, але пов'язаних між собою. Для розв'язання першого завдання обчислюються коефіцієнти автокореляції і авторегресії, які показують залежність між послідовними рівнями ряду динаміки, для розв'язання другого завдання – коефіцієнти кореляції і регресії.

Коефіцієнт автокореляції обчислюється за безпосередніми даними рядів динаміки, коли фактичні рівні одного ряду розглядаються як значення факторної ознаки, а рівні цього ж самого ряду зі зсувом на один період приймаються за результативну ознаку.

Коефіцієнт автокореляції обчислюється за формулою коефіцієнта кореляції для парної залежності

$$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}.$$

Коефіцієнт кореляції розраховується за відхиленнями від вирівняних значень в обох рядах динаміки, що корелюються. Як правило, визначають відхилення фактичних рівнів від тренда, що характеризує основну тенденцію розвитку кожного ряду динаміки.

Коефіцієнт кореляції визначається за формулою

$$r_{d_x d_y} = \frac{\Sigma d_x d_y}{\sqrt{\Sigma d_x^2 \Sigma d_y^2}},$$

де $d_x = x_i - \widetilde{x}_t$; $d_y = y_i - \widetilde{y}_t$.

Як зазначалося вище, корелювання рівнів рядів динаміки правильно показує тісноту зв'язку між рядами лише в тому випадку, якщо в кожному з них відсутня автокореляція. Якщо ж за результатами розрахунків коефіцієнтів автокореляції буде доведено наявність автокореляції рівнів вихідних рядів динаміки, то не слід корелювати безпосередньо рівні порівнюваних динамічних рядів, а попередньо необхідно виключити автокореляцію рівнів у динамічних рядах.

Є декілька способів виключення автокореляції. Один із них пов'язаний з кореляцією відхилень фактичних рівнів від тренда. Суть цього способу полягає в тому, що корелюють не самі рівні, а відхилення фактичних рівнів від вирівняних, що відображають тренд, тобто корелюють залишкові величини. Виключення трендів дасть змогу послабити автокореляцію і привести вихідні дані до такого вигляду, який найбільш придатний для застосування класичних методів теорії кореляції.

При корелюванні відхилень фактичних рівнів ряду динаміки від вирівняних необхідно виконати такі операції:

1) здійснити аналітичне вирівнювання порівнюваних рядів динаміки за певним, характерним для них аналітичним рівнянням;

2) обчислити величину відхилення кожного фактичного рівня ряду динаміки від відповідного йому вирівняного значення, тобто знаходять $d_x = x_i - \widetilde{x}_t$; $d_y = y_i - \widetilde{y}_t$.

3) визначити тісноту зв'язку між розрахованими відхиленнями за допомогою вищенаведеної формули коефіцієнта кореляції.

Другий спосіб виключення автокореляції базується на корелюванні послідовних різниць між кожним наступним і попереднім рівнями, тобто величин $\Delta_y = y_t - y_{i-1}$; $\Delta_x = x_t - x_{t-1}$, де Δ_{y_t} і Δ_{x_t} – абсолютні прирости (зниження) із змінною базою в рядах динаміки показників y і x.

Формула коефіцієнта кореляції перших різниць, що використовується для визначення тісноти зв'язку між досліджуваними рядами динаміки, має вигляд:

$$r_{\Delta_x \Delta_y} = \frac{\Sigma \Delta_x \Delta_y}{\sqrt{\Sigma \Delta_x^2 \Delta_y^2}}.$$

При цьому слід мати на увазі, що різниця першого порядку виключає автокореляцію лише в тих рядах динаміки, в яких зміна в часі відбувається за прямою лінією. Якщо ж зміна рівнів рядів динаміки відбувається за параболою другого порядку, то виключати автокореляцію можна за допомогою корелювання других різниць (різниць між першими різницями).

Застосування кореляційного аналізу в динамічних рядах покажемо на прикладі взаємозв'язку в динаміці урожайності соняшнику і кількості внесених мінеральних добрив на 1 га соняшнику (табл. 9.9).

Таблиця 9.9

Дані для розрахунку коефіцієнта кореляції в рядах динаміки

Порядковий номер року	Урожайність соняшнику, ц/га	Внесено на 1 га соняшнику мінеральних добрив, ц діючої речовини	Розрахункові величини		
	y	x	yx	y^2	x^2
1	10,7	0,8	8,56	114,49	0,64
2	12,3	1,1	13,53	151,29	1,21
3	11,4	1,2	13,68	129,96	1,44
4	15,6	1,6	24,96	243,36	2,56
5	13,0	1,4	18,20	169,00	1,96
6	17,5	1,5	26,25	306,25	2,25
7	16,8	1,4	23,52	282,24	1,96
8	19,5	1,7	33,15	308,25	2,89
9	18,4	2,0	36,80	338,56	4,00
10	19,6	1,8	35,28	384,16	3,24
Разом	**154,8**	**14,5**	**233,93**	**2499,56**	**22,15**

Якщо корелювати рівні цих двох рядів динаміки, приймаючи урожайність за результативну ознаку (y), а кількість внесених добрив – за факторну ознаку (x), то розрахований коефіцієнт кореляції покаже, в якому ступені обидва корельованих ряди динаміки схильні до схожого впливу протягом даного періоду часу.

Величина коефіцієта кореляції становитиме

$$r_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}} =$$

$$= \frac{233{,}93 - \frac{14{,}5 \cdot 154{,}8}{10}}{\sqrt{\left(22{,}15 - \frac{14{,}5^2}{10}\right)\left(2499{,}56 - \frac{154{,}8^2}{10}\right)}} = \frac{9{,}47}{10{,}80} = \mathbf{0{,}8768}.$$

Одержаний коефіцієнт кореляції свідчить про наявність досить тісної кореляційної залежності між урожайністю соняшнику і кількістю внесених мінеральних добрив. Така тіснота зв'язку зумовлена тим, що в двох порівнюваних рядах динаміки короткочасні коливання мають однакову тенденцію, отже, і тренди відображають однаковий напрям змін у часі як урожайності, так і кількості внесених добрив.

У зв'язку з цим перед тим, як робити висновок про дійсну тісноту зв'язку між досліджуваними ознаками, необхідно перевірити обидва ряда динаміки на автокореляцію. Найчіткіше автокореляція виявляється між рівнями рядів динаміки, розташованими рядом. Щоб оцінити ступінь залежності між сусідніми рівнями динамічного ряду, обчислимо коефіцієнти автокореляції для першого ряду, що характеризує динаміку урожайності, і для другого ряду, що характеризує динаміку кількості внесених добрив. Автокореляція визначається зіставленням даних, які відносяться до двох суміжних років, тобто величин $x_i(x)$ і $x_{i+1}(y)$. Для розрахунку коефіцієнтів автокореляції побудуємо дві допоміжні таблиці (табл. 9.10 і 9.11).

Таблиця 9.10

Дані для розрахунку коефіцієнта автокореляції по ряду динаміки урожайності соняшнику

x	y	xy	x^2	y^2
10,7	12,3	131,61	114,49	151,29
12,3	11,4	140,22	151,29	129,96
11,4	15,6	177,84	129,96	243,36
15,6	13,0	202,80	243,36	169,00
13,0	17,5	227,50	169,00	306,25
17,5	16,8	294,00	306,25	282,24
16,8	19,5	327,60	282,24	380,25
19,5	18,4	358,80	380,25	338,56
18,4	19,6	360,64	338,56	384,16
Разом 135,2	**144,1**	**2221,01**	**2115,40**	**2385,07**

Таблиця 9.11

Дані для розрахунку коефіцієнта автокореляції по ряду динаміки внесення добрив

x	y	xy	x^2	y^2
0,8	1,1	0,88	0,64	1,21
1,1	1,2	1,32	1,21	1,44
1,2	1,6	1,92	1,44	2,56
1,6	1,4	2,24	2,56	1,96
1,4	1,5	2,10	1,96	2,25
1,5	1,4	2,10	2,25	1,96
1,4	1,7	2,38	1,96	2,89
1,7	2,0	3,40	2,89	4,00
2,0	1,8	3,60	4,00	3,24
Разом 12,7	13,7	19,94	18,91	21,51

Розрахуємо коефіцієнти автокореляції по кожному ряду динаміки:
а) для урожайності соняшнику

$$r_{xy} = \frac{\Sigma xy - \dfrac{\Sigma x \Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \dfrac{(\Sigma y)^2}{n}\right]}} =$$

$$= \frac{2221{,}01 - \dfrac{135{,}2 \cdot 144{,}1}{9}}{\sqrt{\left(2115{,}40 - \dfrac{135{,}2^2}{9}\right)\left(2385{,}07 - \dfrac{144{,}1^2}{9}\right)}} = \frac{56{,}31}{81{,}07} = 0{,}6946.$$

б) для кількості внесених добрив

$$r_{xy} = \frac{\Sigma xy - \dfrac{\Sigma x \Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \dfrac{(\Sigma y)^2}{n}\right]}} =$$

$$= \frac{19{,}94 - \dfrac{12{,}7 \cdot 13{,}7}{9}}{\sqrt{\left(18{,}91 - \dfrac{12{,}7^2}{9}\right)\left(21{,}51 - \dfrac{13{,}7^2}{9}\right)}} = \frac{0{,}61}{0{,}8083} = \mathbf{0{,}7547}.$$

Як свідчать числові значення коефіцієнтів кореляції, у взаємопов'язаних рядах динаміки існує досить значна автокореляція і наш висновок про тісноту зв'язку між рівнями динамічних рядів (урожайності і добрив) перекручується автокореляцією. В зв'язку з цим не слід корелювати безпосередньо рівні рядів динаміки, а спочатку необхідно *виключити основну тенденцію* зміни рівнів і корелювати вже відхилення від тренда.

З цією метою визначимо лінії трендів, щоб потім їх виключити з аналізу. Вирівнювання рядів динаміки здійснимо за рівнянням прямої лінії $y_t = a_0 + a_1 t$, де t – порядковий номер року; a_0 і a_1 – параметри рівняння.

Для вирівнювання ряду динаміки урожайності складемо допоміжну табл. 9.12.

Таблиця 9.12

Дані для розрахунку лінії тренда ряду динаміки урожайності

Порядковий номер року, t	Урожайність, ц/га y	t	t^2	Розрахункові значення урожайності, ц/га $\widetilde{y}_t$
1	10,7	10,7	1	11,97
2	12,3	24,6	4	12,75
3	11,4	34,2	9	13,53
4	15,6	62,4	16	14,31
5	13,0	65,0	25	15,09
6	17,5	105,0	36	15,87
7	16,8	117,6	49	16,65
8	19,5	156,0	64	17,43
9	18,4	165,6	81	18,21
10	19,6	196,0	100	18,99
Разом 55	**154,8**	**937,1**	**385**	**154,80**

Параметри рівняння a_0 і a_1 знайдемо з такої системи нормальних рівнянь:

$$\begin{cases} \Sigma y = a_0 n + a_1 \Sigma t; \\ \Sigma yt = a_0 \Sigma t + a_1 \Sigma t^2; \end{cases} \qquad \begin{cases} 154{,}8 = 10a_0 + 55a_1 & |:10 \\ 937{,}1 = 55a_0 + 385a_1 & |:55 \end{cases}$$

Поділимо перше рівнянняна 10, а друге на 55:

$$\begin{cases} 15{,}48 = a_0 + 5{,}50a_1; \\ 17{,}04 = a_0 + 7{,}00a_1. \end{cases}$$

Віднімемо з другого рівняння перше

$$1{,}56 = 2{,}00\ a_1.$$

Звідси $a_1 = 1{,}56 : 2{,}00 = 0{,}78$ ц/га.

Визначимо a_0

$$154{,}8 = 10\ a_0 + 55 \cdot 0{,}78$$

$$a_0 = \frac{154{,}8 - 55 \cdot 0{,}78}{10} = \frac{111{,}9}{10} = 11{,}19;$$

Рівняння лінійного тренду має вигляд

$$\tilde{y}_t = 11{,}19 + 0{,}78t.$$

Коефіцієнт регресії $a_1 = 0{,}78$ ц/га показує, що в середньому за досліджуваний період урожайність щорічно зростала на 0,78 ц/га.

Підставляючи в одержане рівняння значення t ($t = 1, 2, ..., 10$), дістанемо вирівняні (розрахункові) значення урожайності.

Проведемо вирівнювання ряду динаміки кількості внесених добрив, для чого складемо допоміжну табл. 9.13.

Таблиця 9.13

Дані для розрахунку лінії тренда ряду динаміки добрив

Порядковий номер року, t	Добрива, ц діючої речовини y	t	t^2	Розрахункові значення добрив, ц діючої речовини $\tilde{y}_t$
1	0,8	0,8	1	1,09
2	1,1	2,2	4	1,17
3	1,2	3,6	9	1,25
4	1,6	6,4	16	1,33
5	1,4	7,0	25	1,41
6	1,5	9,0	36	1,49
7	1,4	9,8	49	1,57
8	1,7	13,6	64	1,65
9	2,0	18,0	81	1,73
10	1,8	18,0	100	1,81
Разом 55	**14,5**	**88,4**	**385**	**14,50**

Розв'язавши систему рівнянь

$$\begin{cases} 14{,}5 = 10a_0 + 55a_1 & \Big| : 10; \\ 88{,}4 = 55a_0 + 385a_1 & \Big| : 55, \end{cases}$$

дістанемо $a_0 = 1{,}01$; $a_1 = 0{,}08$ ц діючої речовини.

Рівняння лінійного тренду має вигляд:

$$\tilde{y}_t = 1{,}01 + 0{,}08\ t.$$

Коефіцієнт регресії $a_1 = 0{,}08$ ц діючої речовини показує, що в середньому за досліджуваний період дози внесених добрив щорічно збільшувались на 0,08 ц діючої речовини.

Знайдені тренди виявляться виключеними, якщо будуть обчислені відхилення від них фактичних даних. Значення фактичних даних і вирівняних рівнів та їх відхилення для рядів динаміки урожайності і добрив подані в табл. 9.14.

Таблиця 9.14

Відхилення від трендів

Порядковий номер року, t	Урожайність, ц/га y	$\tilde{y}_t$	$d_y = y_t - \tilde{y}_t$	Добрива, ц діючої речовини x	$\tilde{x}_t$	$d_x = x_t - \tilde{x}_t$	$d_x d_y$	d_x^2	d_y^2
1	10,7	11,97	– 1,27	0,8	1,09	– 0,29	0,3683	0,0841	0,1629
2	12,3	12,75	– 0,45	1,1	1,17	– 0,07	0,0315	0,0049	0,2025
3	11,4	13,53	– 2,13	1,2	1,25	– 0,05	0,1065	0,0025	4,5369
4	15,6	14,51	1,29	1,6	1,33	0,27	0,3483	0,0729	1,6641
5	13,0	15,09	– 2,09	1,4	1,41	– 0,01	0,0209	0,0001	4,3681
6	17,5	15,87	1,63	1,5	1,49	0,01	0,0163	0,0001	2,6569
7	16,8	16,65	0,15	1,4	1,57	– 0,17	– 0,0255	0,0289	0,05225
8	19,5	17,43	2,07	1,7	1,65	0,05	0,1035	0,0025	4,2849
9	18,4	18,21	0,19	2,0	1,73	0,27	0,0513	0,0729	0,0361
10	19,6	18,99	0,61	1,8	1,81	– 0,01	– 0,0061	0,0001	0,3721
Разом	**154,8**	**154,80**	**0**	**14,5**	**14,50**	**0**	**1,0150**	**0,2690**	**19,7570**

Розрахунок коефіцієнта кореляції здійснимо за відхиленнями фактичних рівнів від трендів:

$$r_{d_x d_y} = \frac{\Sigma d_x d_y}{\sqrt{\Sigma d_x^2 \Sigma d_y^2}} = \frac{1,0150}{\sqrt{0,2690 \cdot 19,7570}} = \frac{1,0150}{2,3053} = 0,4403.$$

Величина розрахованого коефіцієнта кореляції найбільш точно характеризує тісноту зв'язку в рядах динаміки урожайності і добрив. Наявність високої автокореляції в рядах (r_{xy} = 0,6946 і r_{xy} = 0,7547) певною мірою підсилювало високий рівень коефіцієнта кореляції (r = 0,8768) в досліджуваних рядах динаміки.

До аналогічного висновку ми прийдемо, якщо будемо корелювати різниці між *наступним* і *попереднім рівнями* обох рядів *динаміки*.

Для оцінки суттєвості вибіркового коефіцієнта автокореляції використовують спеціальні таблиці з критичними значеннями коефіцієнта автокореляції при різних рівнях значущості (дод. 10). Фактичне значення коефіцієнта автокореляції ($r_{a_{факт}}$) зіставляється з табличним ($r_{a_{табл}}$) при довірчому рівні імовірності судження і відповідній чисельності вибірки. Якщо фактичне значення буде більшим від його критичного значення, вказаного в таблиці, то роблять висновок про наявність автокореляції в генеральній сукупності. Якщо ж фактичне значення коефіцієнта автокореляції буде менше його табличного значення, тобто $r_{a_{факт}} < r_{a_{табл}}$, то потрібно відмовитись від гіпотези про наявність автокореляції в генеральній сукупності.

Проведемо оцінку вірогідності коефіцієнтів автокореляції. Для цього порівняємо їх фактичні значення з табличним при рівні значущості $\alpha = 0{,}05$ і $n = 10$ (число років).

За дод. 10 встановимо, що табличне значення коефіцієнта автокореляції становить $r_{0{,}05} = 0{,}360$.

Оскільки $r_{a_{факт}} > r_{0{,}05}$ (0,6946 > 0,360; 0,7547 > 0,360), можна зробити висновок про те, що одержані коефіцієнти автокореляції є вірогідними, а досліджуваним рядам динаміки притаманна висока автокореляція його рівнів в генеральних сукупностях.

Завдання для самоконтролю до розділу 9

1. Дайте поняття функціонального і кореляційного зв'язку. Як проявляється кореляційний зв'язок?

2. Які Ви знаєте форми кореляційного зв'язку? Наведіть приклади.

3. Які задачі розв'язуються за допомогою кореляційно-регресійного аналізу?

4. Назвіть основні етапи кореляційного аналізу і розкрийте їх суть.

5. Які основні прийоми встановлення форми зв'язку між ознаками? Що таке кореляційне поле?

6. Як визначаються параметри рівняння регресії при лінійній і криволінійній залежності?

7. Що характеризує коефіцієнт регресії?

8. Охарактеризуйте рівняння регресії.

9. Які показники використовують для вимірювання тісноти зв'язку між ознаками при лінійній і криволінійній залежності?

10. Що характеризує коефіцієнт кореляції і коефіцієнт детермінації?

11. На що вказує знак лінійного коефіцієнта кореляції?

12. Як розраховується кореляційне відношення і що воно показує?

13. Дайте визначення множинної кореляції.

14. Що характеризують коефіцієнти регресії в рівнянні множинної регресії?

15. Який зміст коефіцієнтів повної і чистої регресії?

16. У чому суть коефіцієнтів еластичності і -коефіцієнта?

17. Яка методика перевірки істотності коефіцієнта регресії?

18. Як оцінюється істотність лінійного коефіцієнта кореляції?

19. Охарактеризуйте коефіцієнт кореляції рангів. Наведіть алгоритм його розрахунку.

20. За допомогою яких показників вивчається і вимірюється кореляційна залежність між якісними показниками?

21. Охарактеризуйте коефіцієнти асоціації і контингенції.

22. У чому полягають особливості кореляційного аналізу в рядах динаміки? Що таке автокореляція?

23. За наведеними даними розрахуйте коефіцієнти регресії, кореляції і детермінації. Зробіть висновки.

№ підприємства	1	2	3	4	5
Річна продуктивність праці робітника, тис. грн.	22	10	15	30	12
Фондоозброєність, тис. грн.	70	41	65	90	50

24. Лінійний коефіцієнт кореляції між добовою переробкою сировини і вартістю основних виробничих фондів становить 0,8. Яка частка варіації добової переробки сировини зумовлена варіацією вартості основних виробничих фондів?

Розділ 10

Ряди динаміки

10.1. Поняття про ряди динаміки та їх види. Наукові умови побудови рядів динаміки

Соціально-економічні явища, які вивчаються статистикою, постійно змінюються і розвиваються як у просторі, так і в часі. З часом – від місяця до місяця, від року до року – змінюється чисельність і склад населення, обсяг і структура виробленої продукції, рівень продуктивності праці, урожайності сільськогосподарських культур і т.д. Тому одним з важливих завдань статистики є вивчення суспільних явищ в безперервному розвитку і динаміці. **Динамікою** в статистиці прийнято називати процес розвитку, руху соціально-економічних явищ у часі. Для відображення і аналізу динаміки будують динамічні (хронологічні, часові) ряди. Дослідження динаміки дає змогу охарактеризувати процес розвитку явищ, розкрити основні шляхи, тенденції і темпи цього розвитку.

Рядом динаміки називають ряд статистичних показників, які характеризують зміну суспільних явищ у часі. Наприклад, чисельність населення країни на певні дати (дати перепису або дати обліку), урожайність зернових культур у господарствах області за 1991-2002 рр., поголів'я корів в держгоспі на початок кожного місяця тощо.

Кожний ряд динаміки складається з двох обов'язкових елементів: періодів часу (t) і рівнів (y). Показниками часу в рядах динаміки можуть бути або певні дати (моменти) часу, або окремі періоди (роки, квартали, місяці, декади, доба).

Рівнем ряду динаміки називають статистичний показник, який характеризує величину суспільного явища на даний момент або за певний період часу. Вони відображають кількісну оцінку (міру) розвитку досліджуваного суспільного явища.

Рівні динамічного ряду можуть бути виражені абсолютними, відносними і середніми величинами. При аналізі рядів динаміки всі ці величини необхідно використовувати в комплексі, вони мають доповнювати один одного. Рівні ряду динаміки можуть характеризувати величину статистичного показника на певний момент (яку-небудь дату) і за відповідний період часу (рік, місяць, день, годину тощо). В зв'язку з цим розрізняють *моментні* та *інтервальні ряди динаміки*.

Моментними називають ряди динаміки, які характеризують розмір явища на певний період часу. Прикладом моментного ряду динаміки є така інформація про спискову чисельність працівників підприємства в 2002 р. (табл. 10.1).

Таблиця 10.1

Чисельність працівників підприємства в 2002 р.

Дата	1.01.2002 р.	1.04.2002 р.	1.07.2002 р	1.10.2002 р.	1.01.2003 р.
Чисельність працівників, чол.	250	254	260	263	260

За допомогою моментних рядів динаміки характеризується найчастіше стан умов і факторів виробництва. Наприклад, динамічний ряд наявності кормів і поголів'я худоби на початок кожного місяця, потужність тракторного парку на кінець року і т.д.

В моментному ряду динаміки одні й ті самі одиниці сукупності входять до складу кількох рівнів. Тому підсумовування рівнів моментного ряду динаміки не має смислу, оскільки при цьому підсумки позбавлені економічного змісту. Так, сума чисельності працівників підприємства на 1.01 і 1.04.2002 р. (250 + 254 = 504) не має реального смислу. Проте визначення різниці між рівнями моментного динамічного ряду має певний смисл. Так, різниця між чисельністю працівників підприємства на 1.04 і 1.01.2002 р. (254 – 250) характеризує абсолютний приріст чисельності працівників за цей період.

Інтервальними називають ряди динаміки, які характеризують розмір явищ за певний період часу. Прикладом інтервального ряду динаміки можуть бути дані, наведені в табл. 10.2.

За допомогою інтервальних динамічних рядів, як правило, характеризуються підсумки виробничого процесу (обсяги виробленої продукції, виконаних робіт, затрат праці, кількості внесених добрив тощо). Рівні інтервального ряду динаміки абсолютних показників на відміну від рівнів моментного ряду не містяться в попередніх або наступних показниках. Тому важливе економічне значення має підсумовування цих рівнів, сума рівнів інтервального ряду динаміки характеризує обсяг досліджуваного явища за більш довгий період. Наприклад, підсумовування валового збору цукрових буряків в господарстві за досліджуваний період (1998–2002 рр.) дає уяву про обсяг її виробництва за 5 років (44465 т). Для виявлення тенденції зміни досліджуваного явища рівні інтервального ряду динаміки можна укрупнювати.

Таблиця 10.2

Динаміка валового збору цукрових буряків у господарстві за 1998–2002 рр.

Рік	1998	1999	2000	2001	2002
Валовий збір цукрових буряків, т	8120	8444	9103	8567	10231

При вивченні динаміки соціально-економічних явищ вирішується цілий ряд завдань, основними з яких є такі: 1) характеристика за допомогою системи показників динаміки інтенсивності зміни рівнів ряду від періоду до періоду або від дати до дати; 2) визначення середніх значень динамічного ряду за той або інший період; 3) виявлення і кількісна оцінка основної тенденції розвитку (тренда) досліджуваного явища; 4) прогнозування розвитку явища на перспективу; 5) виявлення факторів, що зумовили зміну досліджуваного суспільного явища в часі; 6) аналіз сезонних коливань.

Однією з важливих вимог правильного обчислення і аналізу показників динаміки є дотримання умов зіставлення порівнюваних між собою рівнів ряду динаміки. Проблема порівнянності даних особливо гостро стоїть в динамічних рядах, оскільки вони, як правило, охоплюють значні періоди часу, за які могли виникнути зміни, що призводять до непорівнянності статистичних даних.

При побудові й аналізі рядів динаміки необхідно забезпечити порівнянність рівнів ряду передусім за територією, методикою розрахунку показників, періодом, або моментом часу, об'єктом і одиницею спостереження, ступенем охоплення одиниць досліджуваної сукупності, одиницями вимірювання тощо.

Розглянемо основні наукові умови порівнянності рівнів ряду динаміки.

Непорівнянність даних, що виникає внаслідок адміністративно-територіальних змін, часто виявляється в статистичній практиці. Це зумовлено тим, що межі територій господарств, районів, областей і т.д. протягом досліджуваного періоду змінюються внаслідок приєднання до них нових територій або від'єднання окремих частин їхньої території. Для приведення даних до порівнянного виду необхідно виконати перерахунок даних за попередні роки (до зміни території) з урахуванням нових меж.

Найбільш суттєвою вимогою при побудові ряду динаміки є єдина методика обчислення рівня за кожний з періодів, що розглядається. Завдяки цьому забезпечується порівнянність статистичних показників за змістом. Наприклад, при вивченні динаміки урожайності сільськогосподарських культур показник урожайності повинен відноситись до однієї і тієї самої посівної площі (весняної продуктивної, фактично зібраної і т.д.). При дослідженні динаміки вартісних показників обсягу продукції необхідно усунути вплив зміни цін. На практиці для вирішення цього завдання кількість продукції, виробленої в різні періоди, оцінюють в цінах одного періоду, які називають фіксованими, або порівнянними. Якщо ряд динаміки подано узагальнюючими показниками в умовно-натуральних одиницях вимірювання, коефіцієнти сумірництва для всіх рівнів мають бути єдиними.

Порівнянність рівнів ряду динаміки за періодом, або моментом спостереження, означає, по-перше, що всі показники характеризують явище або за певний період часу, або на певний момент часу. В зв'язку з цим неправомірно порівнювати, наприклад, середньорічне число тракторів з числом тракторів на початок або кінець року, по-друге, в інтервальних динамічних рядах рівні повинні відноситись до рівних періодів часу, а в моментних – повинні бути, як правило, рівні відрізки часу між моментами (датами) спостереження. Крім того, не можна поєднувати в одному ряду динаміки періоди і моменти часу.

Порівнянність за об'єктом спостереження означає, що всі рівні ряду динаміки відносяться до одного і того самого об'єкта спостереження. Наприклад, при дослідженні динаміки продуктивності корів об'єктом спостереження можуть бути державні, колективні, фермерські господарства, особисті підсобні господарства населення або всі категорії в цілому. Для одержання порівнянної в динаміці продуктивності корів показник повинен розраховуватись по одній і тій самій категорії господарства або по їх сукупності.

Порівнянність за одиницями спостереження передбачає, що всі рівні одержані по одних і тих самих одиницях спостереження. Одиницями спостереження можуть бути окремі підприємства або їхні підрозділи. Тому, наприклад, при вивченні динаміки урожайності сільськогосподарських культур показник урожайності повинен визначитися по одних і тих самих сільськогосподарських підприємствах, держгоспах, фермах і т.д.

Крім перелічених вимог, без урахування яких неможливо побудувати ряд динаміки, потрібно дотримуватися одних і тих самих одиниць вимірювання. Так, якщо дані про валовий збір за одні роки наводяться в тонах, а за інші – в центнерах, то необхідно перерахувати весь ряд в одні і ті самі одиниці вимірювання.

Науково обгрунтоване формування рядів динаміки вимагає також виділення строго однорідних періодів (етапів) у розвитку досліджуваних соціально-економічних явищ, тому що всебічного аналізу динамічних процесів можна досягти лише в межах однорідних періодів. Періодизацію динамічних рядів слід проводити на основі глибокого теоретичного аналізу основних процесів і законів, які визначають розвиток досліджуваного явища.

10.2. Показники ряду динаміки

Одним з важливих завдань аналізу рядів динаміки є вивчення особливостей розвитку досліджуваних явищ за окремі періоди. Для виявлення напрямку та інтенсивності змін досліджуваних суспільних явищ за певні періоди часу визначають систему абсолютних і відносних показників динаміки. До таких показників відносяться: абсолютний приріст, темп (коефіцієнт) зростання, темп приросту, абсолютне значення одного процента приросту і середні показники ряду динаміки (середній рівень ряду динаміки, середній абсолютний приріст, середній темп зростання і приросту та ін.).

Показники абсолютного приросту, темпу зростання і приросту, а також абсолютного значення одного процента приросту отримують, порівнюючи між собою вихідні рівні ряду динаміки. При цьому рівень, з яким порівнюють, називають **базисним**, а порівнюваний – **поточним рівнем.**

Якщо порівнянню підлягають декілька послідовних рівнів, то можливі два варіанти порівняння:

1) кожен рівень вихідного ряду динаміки зіставляють з одним і тим самим рівнем, взятим за базу порівняння. Найчастіше за базу по-

рівняння береться або початковий (перший) рівень, або ж рівень, з якого починається якийсь новий етап розвитку явища. Вибір бази порівняння повинен бути обґрунтований історично і економічно. Таке порівняння дістало назву **порівняння з постійною базою**;

2) кожен рівень вихідного ряду динаміки порівнюють з безпосередньо попереднім рівнем. Таке порівняння називають **порівнянням зі змінною базою**.

Відповідно до цих двох варіантів порівняння отримують дві системи показників ряду динаміки. При порівнянні кожного рівня з одним і тим самим рівнем, взятим за базу порівняння, одержують **базисні показники**; при порівнянні кожного рівня з безпосередньо попереднім рівнем отримують **ланцюгові показники.**

Для характеристики абсолютної швидкості зростання (зниження) рівнів ряду динаміки обчислюють показник абсолютного приросту (A).

Абсолютний приріст являє собою різницю між двома рівнями, один з яких взято за базу порівняння.

Він показує, на скільки одиниць кожен даний рівень відрізняється від рівня, взятого за базу порівняння. Абсолютний приріст може мати додатний або від'ємний знак. Якщо наступний рівень ряду динаміки більший за попередній, то абсолютний приріст буде мати знак «+», якщо менше – знак «–».

Динамічний ряд абсолютних приростів дає змогу визначити напрям (зростання, зниження) динаміки досліджуваного явища. Крім того, порівнянням абсолютних приростів між собою можна встановити характер зростання або зниження в абсолютному вираженні (рівномірний, прискорений, стрибкоподібний та ін.).

Розглянемо порядок обчислення абсолютних приростів та інших показників динаміки за даними про динаміку поголів'я корів в господарствах району за 1997–2002 рр. (табл. 10.3).

Таблиця 10.3

Показники динаміки поголів'я корів в господарствах району за 1997–2002 рр.

Рік	Поголів'я корів, гол.	Абсолютний приріст		Темп (коефіцієнт) зростання		Темп приросту, %		Абсолютне значення одного процента приросту	Пункти зростання, %
		базисний	ланцюговий	базисний	ланцюговий	базисний	ланцюговий		
1	2	3	4	5	6	7	8	9	10
1997	16200	–	–	–	–	–	–	–	–
1998	16875	675	675	1,0417	1,0417	4,17	4,17	162,0	4,17

Продовження табл. 10.3

1	2	3	4	5	6	7	8	9	10
1999	17121	921	246	1,0568	1,0146	5,68	1,46	168,7	1,51
2000	17566	1366	445	1,0843	1,0260	8,43	2,60	171,2	2,75
2001	17083	883	– 483	1,0545	0,9725	5,45	– 2,75	175,7	– 2,98
2002	18360	2160	1277	1,1333	1,0747	13,33	7,47	170,8	7,88

Абсолютні прирости становитимуть:

базисні (до 1997 р.)

$$A_1 = y_1 - y_0 = 16875 - 16200 = 675 \text{ гол.};$$

$$A = y_2 - y_0 = 17121 - 16200 = 921 \text{ гол. і т.д.};$$

ланцюгові (до попереднього року)

$$A_1 = y_1 - y_0 = 16875 - 16200 - 675 \text{ гол.};$$

$$A_2 = y_2 - y_1 = 17121 - 16875 = 246 \text{ гол. і т.д.}$$

Ланцюгові і базисні абсолютні прирости пов'язані між собою такою рівністю: сума ланцюгових приростів дорівнює відповідному базисному приросту, тобто загальному приросту за весь відповідний період часу.

Для даного прикладу:

2160 = 675 + 246 + 445 + (– 483) + 1277.

Темп (коефіцієнт) зростання (K) – це відношення двох рівнів, один з яких взято за базу порівняння.

Темп зростання характеризує відносну швидкість зміни явища і показує у скільки разів кожний даний рівень більший або менший рівня, який взято за базу порівняння. Він може бути виражений у вигляді коефіцієнтів або процентів. Темп зростання, виражений у процентах, називають **процентом зростання.**

Величина темпу зростання буде більшою від одиниці, якщо рівень щодо бази порівняння зростає, і меншою за одиницю, якщо рівень щодо бази порівняння зменшується.

Темпи (коефіцієнти) зростання для нашого прикладу становитимуть:

базисні: $K_1 = y_1 : y_0 = 16875:16200 = 1{,}0417$;

$K_2 = y_2 : y_0 = 17121:16200 = 1{,}0568$ і т.д.;

ланцюгові: $K_1 = y_1 : y_0 = 16875:16200 = 1{,}0417$;

$K_2 = y_2 : y_1 = 17121:16875 = 1{,}0146$ і т.д.

Між ланцюговими і базисними темпами зростання, вираженими коефіцієнтами, є такий взаємозв'язок:

а) добуток ланцюгових коефіцієнтів зростання дорівнює базисному коефіцієнту зростання за відповідний період. Наприклад:

$$K_1 \cdot K_2 \cdot K_3 = \frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \cdot \frac{y_3}{y_2} = \frac{y_3}{y_0}.$$

Для нашого прикладу:

$$1{,}1333 = 1{,}0417 \cdot 1{,}0146 \cdot 1{,}0260 \cdot 0{,}9725 \cdot 1{,}0747.$$

б) частка від ділення двох сусідніх базисних коефіцієнтів зростання дорівнює відповідному ланцюговому коефіцієнту зростання.

Наприклад, $K_2 : K_1 = \frac{y_2}{y_0} : \frac{y_1}{y_0} = \frac{y_2}{y_1}$.

Цей взаємозв'язок дає змогу здійснювати перехід від ланцюгових коефіцієнтів зростання до базисних і навпаки.

Для нашого прикладу:

$$K_2 = 1{,}0568 : 1{,}0417 = 1{,}0146;$$

$$K_3 = 1{,}0843 : 1{,}0568 = 1{,}0260 \quad \text{і т.д.}$$

Поряд з темпами зростання відносна зміна явища у часі може бути також охарактеризована за допомогою **темпів приросту**, які являють собою відношення абсолютного приросту до рівня, взятого за базу порівняння (T).

Темп приросту, як і абсолютний приріст, може бути як додатним, так і від'ємним числом (відповідно при зростанні і зниженні рівня) і *виражається у вигляді коефіцієнтів або процентів. На практиці темпи* приросту найчастіше виражаються у формі процентів.

Вони показують, на скільки процентів збільшився або зменшився поточний рівень порівняно з базисним, взятим за 100%.

Визначимо для нашого прикладу темпи приросту (в %):

базисні:

$$T_1 = (A_1 : y_0) \cdot 100\% = (675 : 16200) \cdot 100 = 4{,}17;$$

$$T_2 = (A_2 : y_0) \cdot 100\% = (921 : 16200) \cdot 100 = 5{,}68 \quad \text{і т.д.}$$

ланцюгові

$$T_1 = (A_1 : y_0) \cdot 100\% = (675 : 16200) \cdot 100 = 4{,}17;$$

$$T_2 = (A_2 : y_0) \cdot 100\% = (246 : 16875) \cdot 100 = 1{,}46 \text{ і т. д.}$$

Між темпом зростання і темпом приросту існує такий зв'язок:

$$T_i = K_i - 1,$$

або

$$T_i = (K_i \cdot 100\%) - 100\%,$$

тобто темп приросту завжди на одиницю менше відповідного темпу зростання, вираженого у формі коефіцієнта, або на 100%, якщо його виражено в процентах.

Отже, щоб визначити темп приросту, потрібно від темпу зростання відняти одиницю, якщо його виражено коефіцієнтом, або 100%, коли він виражений у процентах.

Для динамічного ряду поголів'я корів темпи приросту, розраховані цим способом, становитимуть:

$T_1 = K_1 \cdot 100\% - 100\% = 1{,}0417 \cdot 100\% - 100\% = 4{,}17\%;$

$T_2 = K_1 \cdot 100\% - 1005 = 1{,}0568 \cdot 100\% - 100\% = 5{,}68\%$ і т.д.

Поряд з показниками темпів зростання і приросту в аналізі динамічних рядів викликає інтерес ще один відносний показник, який дає змогу визначити вагомість кожного процента приросту, і те, яка абсолютна величина приховується за цим процентом. Таким показником є **абсолютне значення одного процента приросту** (Π). Він обчислюється як відношення абсолютного приросту до відповідного темпу приросту, вираженого в процентах

$$\Pi_t = \frac{A_t}{T_t(\%)}.$$

Розрахунок цього показника має економічний зміст тільки на ланцюговій основі, оскільки на базисній основі по всіх часових відрізках буде отримано одне і те саме значення показника – сота частина початкового (першого) рівня.

Абсолютне значення одного процента приросту для нашого прикладу становитиме:

$$\Pi_1 = \frac{A_1}{T_1} = \frac{675}{4{,}1667} = 125{,}0 \text{ гол};$$

$$\Pi_2 = \frac{A_2}{T_2} = \frac{246}{1{,}4578} = 168{,}7 \text{ гол};$$

Цей показник можна обчислити значно простіше. Підставивши у формулу значення відповідних показників, дістанемо, що величина абсолютного значення одного процента приросту являє собою соту частину попереднього рівня:

$$\Pi_t = \frac{A_t}{T_t(\%)} = \frac{A_t}{\frac{A_t}{y_0} \cdot 100} = \frac{y_0}{100} = 0{,}01 y_0.$$

Таким чином, не вдаючись до спеціального розрахунку, можна визначити, що абсолютне значення одного процента приросту для відповідних часових відрізків становитиме:

$\Pi_1 = y_0 : 100 = 16200 : 100 = 162{,}0$ гол.;

$\Pi_2 = y_1 : 100 = 16875 : 100 = 168{,}7$ гол. і т.д.

Цей показник має важливе практичне значення в економічному аналізі, оскільки темпи зростання можуть сповільнюватись або залишатися на одному рівні, а абсолютне значення одного процента приросту зростати. Звичайно така закономірність спостерігається в динамічних рядах з рівнями, що постійно зростають.

Слід відмітити, що в динамічних рядах відносних величин (процентів зростання і приросту) їх безпосереднє порівняння можна здійснювати тільки шляхом визначення різниці рівнів. Ці різниці дістали назву **пунктів зростання.** Їх розраховують як різницю базисних процентів зростання або приросту двох суміжних періодів. На відміну від темпів приросту, які не можна підсумовувати та перемножувати, пункти зростання можна підсумовувати, в результаті чого дістанемо темп приросту відповідного періоду порівняно з базисним періодом.

Пункти зростання для нашого прикладу становитимуть:

для 1999 *р.* $ПР_1 = T_2 - T_1 = 5{,}68 - 4{,}17 = 1{,}51\%$;

для 2000 р. $ПР_2 = T_3 - T_2 = 8{,}43 - 5{,}88 = 2{,}75\%$ і т.д.

Контроль правильності розрахунків: сума пунктів зростання дорівнює загальному темпу приросту за весь період:

$$4{,}17 + 1{,}51 + 2{,}75 + (-2{,}98) + 7{,}88 = 13{,}33\%.$$

Це відповідає темпу приросту рівня 2002 р. порівняно з 1997 р.

Для отримання узагальнюючих показників динаміки соціально-економічних явищ визначаються різного роду середні величини: середній рівень динамічного ряду, середній абсолютний приріст, середній темп зростання і приросту та ін. Середню з рівнів динамічного ряду називають **хронологічною середньою**. Середній рівень ряду динаміки характеризує типовий розмір рівнів ряду.

Порядок розрахунку середнього рівня для інтервальних і моментних рядів динаміки відрізняється. В інтервальних рядах динаміки з рівновіддаленими один від одного рівнями середній рівень обчислюється за формулою середньої арифметичної простої:

$$\bar{y} = \frac{\sum y}{n},$$

де n – число рівнів ряду динаміки.

Середній рівень динамічного ряду для нашого прикладу (табл. 10.3) становитиме:

$$\bar{y} = \frac{\sum y}{n} = \frac{16200 + 16875 + 17121 + 17566 + 17083 + 18360}{6} =$$

$$= \frac{103205}{6} = 17201 \text{ гол.}$$

Якщо інтервальний ряд динаміки має нерівновіддалені один від одного рівні, то середній рівень розраховується за формулою середньої арифметичної зваженої:

$$\bar{y} = \frac{\sum yt}{\sum t},$$

де t – відрізок часу, протягом якого зберігалося дане значення рівня y.

Наприклад, середня чисельність працівників агрофірми за перше півріччя становила 500 чоловік, за третій квартал – 520, за четвертий – 515 чоловік. Звідси середньорічна чисельність працівників становитиме:

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{500 \cdot 6 + 520 \cdot 3 + 515 \cdot 3}{12} = \frac{6105}{12} = 509 \text{ чол.}$$

У моментних динамічних рядах з рівними проміжками між датами середній рівень обчислюється за формулою:

$$\bar{y}_{\text{хронол}} = \frac{0{,}5y_1 + y_2 + ... + 0{,}5y_n}{n-1},$$

де n – число рівнів ряду динаміки.

Застосування формули проілюструємо на даних про спискову чисельність працівників підприємства за 2002 р. (табл. 10.1). Середньорічна чисельність працівників підприємства становитиме:

$$\bar{y} = \frac{0{,}5y_1 + y_2 + ... + 0{,}5y_n}{n-1} =$$

$$= \frac{0{,}5 \cdot 250 + 254 + 260 + 263 + 0{,}5 \cdot 260}{5-1} = \frac{1032}{4} = 258 \text{ чол.}$$

У моментних динамічних рядах з нерівними проміжками між датами середній рівень розраховують за формулою середньої арифметичної зваженої, тобто шляхом зваження рівнів за кількістю рівних періодів:

$$\bar{y} = \frac{\sum yt}{\sum t}.$$

Порядок обчислення середнього рівня для моментного ряду динаміки показано в табл. 10.4.

Отже, середня чисельність працівників підприємства в липні 2002 р. становитиме:

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{1300 + 2610 + 1848 + 1310 + 1060}{5 + 10 + 7 + 5 + 4} = \frac{8128}{31} = 262 \text{ чол.}$$

Таблиця 10.4

Динаміка чисельності працівників підприємства в липні 2002 р.

Показник	Число місяця				
	1	6	16	23	28
Чисельність працівників, чол. (y)	260	261	264	262	265
Число днів перебування на підприємстві певної кількості працівників (t)	5	10	7	5	4
Кількість людино-днів (yt)	1300	2610	1848	1310	1060

Середній абсолютний приріст характеризує середню швидкість зростання (або зниження) рівня. Для моментних та інтервальних рядів динаміки з рівними проміжками між датами його обчислюють як середню арифметичну просту з ланцюгових абсолютних приростів або як різницю між кінцевим і початковим рівнем, поділену на кількість членів ряду, зменшених на одиницю:

$$\overline{A} = \frac{\sum A}{n}, \text{ або } \overline{A} = \frac{y_n - y_0}{n-1},$$

де n – число абсолютних приростів; $n - 1$ – кількість рівнів ряду динаміки, зменшених на одиницю.

Середній абсолютний приріст для ряду динаміки поголів'я корів (табл. 10.3) становитиме:

$$\overline{A} = \frac{\sum A}{n} = \frac{675 + 246 + 445 + (-483) + 1277}{5} = \frac{2160}{5} = 432 \text{ гол.},$$

$$\text{або } \overline{A} = \frac{y_n - y_0}{n-1} = \frac{18360 - 16200}{5} = \frac{2160}{5} = 432 \text{ гол.}$$

Для узагальнюючої характеристики темпів зростання за ряд років обчислюють середній темп (коефіцієнт) зростання ($\overline{K}$). Він показує, в скільки разів у середньому кожен даний рівень ряду більший (або менший) від попереднього рівня. Для динамічних рядів з рівними проміжками між датами середній темп зростання обчислюється за формулою середньої геометричної:

$$\overline{K} = \sqrt[n]{K_1 \cdot K_2 \cdot ... \cdot K_n},$$

де K – коефіцієнти зростання за окремі періоди часу; n – число коефіцієнтів зростання.

Середній коефіцієнт зростання може бути визначений за іншою формулою, що випливає з наведеної вище формули:

$$\overline{K} = \sqrt[n-1]{\frac{y_n}{y_0}},$$

де n – число рівнів ряду динаміки; y_0 і y_n – початковий і кінцевий рівні ряду динаміки.

На основі середнього темпу зростання можна визначити середній темп приросту ($\overline{T}$). Він показує, на скільки процентів у середньому збільшується (або зменшується) даний рівень порівняно з попереднім. Його розраховують як різницю між середнім темпом зростання, вираженим у процентах і 100%:

$$\overline{T}_t = (\overline{K}_t \cdot 100\%) - 100\%.$$

За даними табл. 10.3 розрахуємо середній коефіцієнт зростання поголів'я корів за 1997–2002 рр. Для цього використаємо формулу середньої геометричної

$$\overline{K} = \sqrt[n]{K_1 \cdot K_2 \cdot \ldots \cdot K_n} =$$

$$= \sqrt[5]{1{,}0417 \cdot 1{,}0146 \cdot 1{,}0260 \cdot 0{,}9725 \cdot 1{,}0747} = \sqrt[5]{1{,}1333}$$

Прологарифмуємо рівняння:

$$\lg \overline{K} = \frac{\lg 1{,}0417 + \lg 1{,}0146 + \lg 1{,}0260 + \lg 0{,}9725 + \lg 1{,}0747}{5} =$$

$$= \frac{0{,}01775 + 0{,}00629 + 0{,}01115 + (-0{,}01211) + 0{,}03129}{5} =$$

$$= \frac{0{,}05436}{5} = 0{,}0109.$$

За таблицями антилогарифмів знайдемо середній коефіцієнт зростання: $\overline{K} = 1{,}025$, або 102,5%.

Такий самий результат дістанемо і за іншою (зручнішою) формулою, що випливає із взаємозв'язку ланцюгових і базисних коефіцієнтів: добуток ланцюгових коефіцієнтів дорівнює базисному крайніх періодів

$$\overline{K} = \sqrt[n-1]{\frac{y_n}{y_0}} = \sqrt[6-1]{\frac{18360}{16200}} = \sqrt[5]{1{,}1333}.$$

Прологарифмуємо рівняння:

$$\lg \overline{K} = \frac{1}{5} \lg 1{,}1333 = \frac{1}{5} 0{,}05436 = 0{,}0109.$$

За таблицями антилогарифмів знайдемо середній коефіцієнт зростання: $\overline{K} = 1{,}025$, або 102,5%.

Такий самий результат матимемо і за іншою формулою:

$$\lg K = \frac{\lg 18360 - \lg 16200}{5} = \frac{4{,}26378 - 4{,}20951}{5} = \frac{0{,}05436}{5} = 0{,}0109.$$

$$\overline{K} = 1{,}025 ,$$

або 102,5%.

Обчисливши систему показників по ряду динаміки поголів'я корів (табл. 10.3), можна зробити такі висновки. Досліджуваний ряд динаміки є інтервальним. Середнє поголів'я корів за рік в господарствах району становить 17201 гол. Щорічно воно зростало в середньому на 432 голови. У цілому ж поголів'я корів за досліджуваний період (1997–2002 рр.) в районі зросло на 2160 гол., або на 13,33%. За середнім коефіцієнтом зростання можна встановити, що середній щорічний темп зростання поголів'я корів становить 2,5% (102,5 – 100). Із зростанням поголів'я корів збільшувалося абсолютне значення одного процента приросту з 162,0 гол. в 1998 р. до 170,8 гол. в 2002 р.

При порівняльному аналізі кількох рядів динаміки, що відображають різні економічні явища за однакові відрізки часу, визначають **коефіцієнт випередження.** Він показує, у скільки разів швидше зростає рівень одного динамічного ряду (i) порівняно з другим (j), тобто

$$K_{\text{вип}} = \frac{T_i}{T_j} .$$

Наприклад, при порівнянні темпів зростання продуктивності праці і заробітної плати, урожайності і посівних площ, продуктивності тварин і їхньої чисельності тощо.

Так, якщо за п'ятиріччя (1997–2002 рр.) базисний темп зростання продуктивності праці в господарствах району становив T_i = 115,6%, а темп зростання заробітної плати – T_j = 108,7%, коефіцієнт випередження темпу зростання продуктивності праці над темпом зростання заробітної плати дорівнює: $K_{\text{вип}} = T_i : T_j$ = 115,6 : 107,7 = 1,063, тобто темп зростання продуктивності праці випереджав тем зростання заробітної плати в 1,063 рази.

10.3. Прийоми виявлення основної тенденції розвитку в рядах динаміки

Для всебічної характеристики зміни соціально-економічних явищ у часі розрахунку *тільки одних показників динаміки* та *їхніх середніх* величин не досить. У зв'язку з цим статистика пропонує ряд спеціальних прийомів обробки й аналізу динамічних рядів.

Важливе місце у вивченні розвитку суспільних явищ належить порівняльному аналізу кількох рядів динаміки. При цьому можна порівнювати динамічні ряди як однойменних, так і різнойменних показників, що стосуються різних територій або є складовими частинами цілого. Абсолютні рівні таких рядів динаміки, як правило, внаслідок відмінностей методики обчислення показників, грошової оцінки продукції та інших причин безпосередньо непорівнянні. Тому доцільно порівнювати не абсолютні, а відносні показники і за ними робити висновки про те, яке явище і на якій території зростає (або знижується) швидше. Цей прийом дістав назву **приведення рядів динаміки до однієї основи**, тобто до загальної бази порівняння, яку беруть за одиницю або за сто процентів.

Суть цього прийому полягає в тому, що дані про величину показника, що вивчається, за рік (або інший відрізок часу), взятий за базу порівняння, беруть таким, що дорівнює 100%, рівні окремих років (або інших відрізків часу) порівнюють з ним, а частку виражають в процентах.

Практичне застосування прийому приведення рядів динаміки до однієї основи розглядається в наступному параграфі розділу підручника, присвяченому факторному аналізу рядів динаміки (табл. 10.13 і 10.14).

У тих випадках, коли рівні ряду динаміки за одні роки непорівнянні з рівнями за інші роки в зв'язку з територіальними, відомчими, організаційними змінами, зміною методики обчислення показників або за іншими причинами і виникає потреба забезпечити порівнянність рівнів, удаються до **змикання динамічних рядів**, тобто об'єднання двох і більше рядів в один зімкнутий ряд.

Суть цього прийому полягає в наступному. Рівні року, протягом якого відбулися зміни, як до змін, так і після змін, беруть за базу порівняння (звичайно за 100%), інші – порівнюють з ним і виражають у процентах. В результаті цього дістанемо єдиний ряд відносних величин, що характеризує зміну досліджуваного явища за весь період.

Припустимо, є дані за 1996–2002 рр. щодо посівної площі картоплі в господарствах району, в територіальних межах якого в 1999 р. відбулися зміни (табл. 10.5).

Аналіз таблиці показує, що в зв'язку зі зміною меж району в 1999 р. дані про посівну площу картоплі за 2000–2002 рр. непорівнянні з даними за 1996–1998 рр. Щоб мати порівнянні дані, виконаємо змикання цих рядів динаміки.

Таблиця 10.5

Динаміка посівної площі картоплі в господарствах району за 1996–2002 рр., га

Межі району	Роки						
	1996	1997	1998	1999	2000	2001	2002
До змін	2100	2208	2315	2430	–	–	–
Після змін	–	–	–	2686	2717	2804	2861

Змикання рядів динаміки і зведення їх до порівнянного вигляду здійснимо двома способами:

а) вираженням ряду динаміки у відносних показниках, прийнявши за базу порівняння один і той самий період;

б) перерахунком абсолютних показників.

Змикання рядів способом вираження рядів відносними показниками динаміки виконаємо так. Візьмемо рік, в якому відбулися територіальні зміни (в нашому прикладі це 1999 р.), за базу порівняння або 100%, а решту рівнів порівняємо з цим роком і отримані дані виразимо в процентах. Отже, за 100% для першого ряду динаміки (1996–1999 рр.) буде прийнята величина посівної площі картоплі, що дорівнює 2430 га, а для другого ряду динаміки (1999–2002 рр.) – 2686 га.

Так, наприклад, відносний показник динаміки посівної площі в 1996 р. порівняно з 1999 р. становитиме 86,4% [(2100 : 2430) · 100], в 2002 р. порівняно з 1999 р. – 106,5% [(2861 : 2686) · 100] і т.д.

Внаслідок отримаємо ряди відносних показників динаміки посівної площі картоплі з однаковою базою порівняння, які можна замінити одним зімкнутим рядом.

Змикання рядів динаміки способом перерахунку абсолютних показників здійснимо за допомогою коефіцієнта перерахунку ($K_п$), який визначимо як відношення двох рівнів посівної площі після зміни меж району до посівної площі перед цією зміною:

$$K_п = \frac{2686}{2430} = 1{,}105.$$

Перемноживши посівну площу картоплі першого ряду динаміки (1996–1998 рр.) на коефіцієнт перерахунку, матимемо дані, які порівняні з даними про посівну площу картоплі другого ряду динаміки (2000–2002 рр.). Так, у 1996 р. посівна площа картоплі в порівнянному показникові дорівнюватиме 2320 га (2100 · 1,105), в 1997 р. – 2440 га (2208 · 1,105) і т.д.

Усі розрахунки з приведення рядів динаміки до порівнянного вигляду зведемо в табл. 10.6.

Таблиця 10.6

Зімкнуті ряди відносних і абсолютних змін посівної площі картоплі в господарствах району за 1996–2002 рр.

Показник \ Рік	1996	1997	1998	1999	2000	2001	2002
Відносні ряди, % а) до зміни меж	86,4	90,9	95,3	100,0	–	–	–
б) після зміни меж	–	–	–	100,0	101,2	104,4	106,5
Зімкнутий ряд динаміки, одержаний способом розрахунку відносних показників динаміки, %	86,4	90,9	95,3	100,0	101,2	104,4	106,5
Зімкнутий ряд динаміки, одержаний способом перерахунку абсолютних показників, га	2320	2440	2558	2686	2717	2804	2861

Добуті зімкнуті ряди динаміки дають змогу визначити дані про посівну площу картоплі за різні роки порівнянними, з них видно, що посівна площа картоплі в районі як в абсолютних, так і у відносних показниках систематично зростала.

Під впливом випадкових факторів (у сільськогосподарському виробництві до них відносять передусім метеорологічні умови) рівні ряду динаміки часто-густо сильно коливаються по періодах часу, при цьому тенденція розвитку затушовується, наочно не проявляється. В зв'язку з цим одним із основних завдань аналізу рядів динаміки є виявлення основної тенденції розвитку соціально-економічних явищ. Під **загальною тенденцією динамічного ряду** розуміють тенденцію до зростання, зниження або стабілізації рівня будь-якого суспільного явища.

Виявлення тенденції в динамічних рядах дає змогу оцінити характер розвитку досліджуваного явища, визначити ефективність факторів, що формують основну тенденцію, встановити рівні досліджуваного явища на перспективу.

Виявлення основної тенденції зміни рівнів динамічного ряду передбачає її кількісне вираження, деякою мірою вільної від випадкових причин. Це досягається шляхом абстрагування від індивідуальних, випадкових змін ознаки. Виявлення основної тенденції розвитку (тренда) називається у статистиці також **вирівнюванням часового ряду**, а прийоми виявлення основної тенденції – **прийомами вирівнювання**. Вирівнювання дає змогу охарактеризувати особливості зміни у часі даного динамічного ряду в найбільш загальному вигляді як функцію часу, передбачаючи, що через час можна виразити вплив основних факторів.

У практиці економічного аналізу нерідкі випадки, коли загальна тенденція явища до зростання або зниження проявляється досить чітко. Наведені дані про динаміку поголів'я корів (табл. 10.3) показують, що в динамічному ряду має місце загальна тенденція до зростання поголів'я корів. Проте для виявлення тенденції в рядах динаміки не досить одного візуального аналізу ряду, якщо його рівні через будь-які об'єктивні або випадкові причини істотно коливаються, то зростаючи, то знижуючись. Це затушовує, наочно не проявляє основну тенденцію розвитку явища. Наприклад, якщо урожайність будь-якої сільськогосподарської культури під впливом метеорологічних умов, що діють в різних напрямках, дуже коливається по роках, то основна тенденція зміни урожайності може не проявлятися безпосередньо. В таких випадках для проявлення основної тенденції потрібно вдатися до спеціальних прийомів обробки динамічних рядів.

До таких прийомів відносяться укрупнення періодів, згладжування ряду динаміки способом ковзної середньої, вирівнювання ряду динаміки по середньому абсолютному приросту, середньому коефіцієнту зростання і способу найменших квадратів (аналітичне вирівнювання рядів динаміки).

Розглянемо на конкретних прикладах умови і техніку виявлення основної тенденції розвитку динамічних рядів кожним із названих прийомів.

Одним з найпростіших прийомів виявлення тенденції розвитку є прийом **укрупнення періодів**. Суть його полягає в тому, що абсолютні або середні рівні ряду динаміки за короткі інтервали (рік, місяць, декаду, день тощо), що зазнають випадкових коливань, замінюють узагальнюючим (звичайно середнім) значенням за триваліший період (трирічча, п'ятиріччя тощо).

По суті, спосіб укрупнення періодів являє собою типологічне групування рівнів ряду динаміки, тому при його застосуванні необхідно дотримуватись наукових основ побудови статистичних групувань.

При укрупненні періодів дуже важливо науково обґрунтовано і правильно виділити періоди часу для укрупнення. Періоди, що їх виділяють, мають бути однорідними в якісному відношенні і досить тривалими за часом, щоб відбулося погашення випадкових коливань явища.

ється в окремі роки головним чином метеорологічними умовами. Оскільки вплив метеорологічних умов для досліджуваної зони в більшості випадків вирівнюється по п'ятирічних періодах, укрупнення здійснено по п'ятиріччях. Завдяки такому укрупненню взаємопогасяться випадкові фактори і виявиться загальна тенденція зміни урожайності.

Щоб отримати середні рівні по п'ятиріччях, спочатку знайдемо суми урожайності за кожне п'ятиріччя (1988–1992 рр., 1993–1997 рр., 1998-2002 рр.), а потім добуті суми поділимо на кількість років в укрупненому періоді (п'ять).

Знайдені суми і середні запишемо центруючи їх на середину кожного п'ятиріччя (відповідно 1990, 1995 і 2000 рр.).

В результаті проведеного укрупнення періодів ряду динаміки чіткіше проявляється тенденція зростання урожайності за роки, що аналізуються. Так, добуті результати показують, що від п'ятиріччя до п'ятиріччя урожайність соняшнику в ТОВ району систематично зростала (з 12,2 ц/га в 1988–1992 рр. до 17,0 ц/га в 1998–2002 рр., тобто на 4,8 ц/га, або на 39,3%).

При укрупненні періодів число членів динамічного ряду дуже скорочується. Цей істотний недолік значною мірою усувається при використанні прийому вирівнювання динамічних рядів **способом ковзних середніх.**

Цей спосіб також ґрунтується на укрупненні періодів. Суть розрахунку ковзних середніх полягає в тому, що склад періоду безперервно і постійно змінюється – відбувається зсув на одну дату при збереженні постійного інтервалу періоду (триріччя, п'ятиріччя тощо).

Ковзна середня – це середня укрупнених періодів, створених послідовним виключенням кожного початкового рівня інтервалу і заміни його черговим наступним рівнем ряду. Таким чином, відбувається ніби ковзання періоду і отриманої середньої по динамічному ряду. Наприклад, при згладжувані по триріччях

$$\bar{y}_1 = \frac{y_0 + y_1 + y_2}{3}; \quad \bar{y}_2 = \frac{y_1 + y_2 + y_3}{3}; \quad \bar{y}_3 = \frac{y_2 + y_3 + y_4}{3} \text{ і т.д.}$$

Цей прийом, як і попередній, ґрунтується на відомому теоретичному положенні про те, що в середніх величинах взаємно погашаються випадкові відхилення і виявляється типове, закономірне.

При виявленні тенденції прийомом ковзних середніх, так само, як і при використанні прийому укрупнення періодів, одним з важливих питань є питання щодо тривалості періодів. Інтервал має бути досить великим і забезпечити взаємне погашення випадкових відхилень рівнів. Якщо в розвитку явища має місце циклічність (періодичність), то інтервал ковзання слід брати рівним тривалості циклу. Чим довше інтервал ковзання, тим більшою мірою вирівнюється ряд в результаті усереднення вихідних рівнів.

Покажемо порядок розрахунку ковзних середніх, використовуючи дані про урожайність соняшнику (табл. 10.7).

Ковзні середні розрахуємо також по п'ятирічних періодах. Для розрахунку ковзних середніх підсумуємо урожайність за перші п'ять років (1988–1992 рр.). а потім, опускаючи дані першого в ряду динаміки року, підсумуємо урожайність за наступне п'ятиріччя (1989–1993 рр.) і т.д. Відбувається мовби ковзання по ряду динаміки. Добуті суми поділимо на число років в періоді ковзання (п'ять), а обчислену середню віднесемо до середини періоду ковзання (в нашому прикладі – третій рік кожного п'ятирічного періоду ковзання).

Розраховані ковзні середні показують стійку тенденцію зростання урожайності соняшнику в ТОВ району.

Ковзна середня згладжує варіацію рівнів, але не дає ряду динаміки, в якому всі вихідні рівні були б замінені вирівняними. Це пояснюється недоліком вирівнювання ряду способом ковзної середньої, при якому вирівняний ряд «скорочується» порівняно з вихідним на $(n - 1) : 2$ члена з одного та другого кінця (під n розуміють число членів, з яких визначають ковзні середні).

Прагнення в процесі вирівнювання ряду замінити всі вихідні рівні вирівняними зумовлює застосування досконаліших прийомів вирівнювання рядів динаміки. До таких прийомів належать: вирівнювання по середньому абсолютному приросту, середньому коефіцієнту зростання і способу найменших квадратів.

В основі застосування способу вирівнювання ряду динаміки по **середньому абсолютному приросту** лежить припущення, що кожен наступний рівень змінюється порівняно з попереднім приблизно на однакову величину, що дорівнює середньому абсолютному приросту.

Рівняння, що відображує тенденцію розвитку явища за цим способом вирівнювання ряду динаміки, має вигляд:

$$\tilde{y}_t = y_0 + \overline{A}t,$$

де $\tilde{y}_t$ – вирівняні рівні ряду динаміки; y_0 – початковий рівень ряду динаміки; $\overline{A}$ – середній абсолютний приріст; t – порядковий номер дати $(t = 1, 2, 3, ..., n)$.

Техніку виявлення тенденції на основі середнього абсолютного приросту розглянемо на прикладі динамічного ряду посівної площі цукрових буряків (табл. 10.8).

Таблиця 10.8

Динаміка посівної площі цукрових буряків в господарстві за 1997–2002 рр.

Рік	Посівна площа, га	Порядковий номер року	Площа, вирівняна по середньому абсолютному приросту	Відхилення фактичних рівнів від вирівняних
	y	t	$\tilde{y}_t = y_0 + At$	$y - \tilde{y}_t$
1997	250	0	250	0
1998	261	1	260	0
1999	268	2	270	-2
2000	281	3	280	1
2001	290	4	290	3
2002	300	5	300	0

Аналіз ряду динаміки показує, що для нього характерне постійне збільшення посівної площі цукрових буряків. При цьому щорічні абсолютні прирости посівної площі стабільні і становлять близько 10 га. Найбільш прийнятним способом вирівнювання рядів динаміки, які мають сталі абсолютні прирости, є спосіб вирівнювання рядів по середньому абсолютному приросту.

Визначимо середній абсолютний приріст посівної площі:

$$\overline{A} = \frac{y_n - y_0}{n - 1} = \frac{300 - 250}{6 - 1} = \frac{50}{5} = 10 \text{ га},$$

де y_0 – початковий рівень ряду динаміки; y_n – кінцевий рівень ряду динаміки; n – число рівнів ряду динаміки (n = 6 років).

Отже, посівна площа цукрових буряків щорічно збільшувалась в середньому на 10 га.

Визначимо вирівняні по середньому абсолютному приросту значення посівної площі для кожного року, підставляючи у рівняння замість t його значення:

$$\tilde{y}_t = y_0 + \overline{A}t,$$

де t – порядковий номер року (t = 0, 1, 2, 3, 4, 5). Вирівняні значення посівної площі становитимуть:

в 1997 р. (при $t = 0$) $\tilde{y}_t = y_0 + \overline{A}t = 250 + 10 \cdot 0 = 250$ га;

в 1998 р. (при $t = 1$ $\tilde{y}_t = y_0 + \overline{A}t = 250 + 10 \cdot 1 = 260$ га і т.д.

Вирівняний по середньому абсолютному приросту ряд динаміки на графіку являє собою пряму лінію, яка з'єднує мінімальне і максимальне значення. Як видно з таблиці, відхилення фактичних рівнів від вирівняних незначні. Отже, вирівнювання ряду динаміки по середньому абсолютному приросту дало змогу точніше відобразити тенденцію зміни посівної площі цукрових буряків в господарстві.

Водночас необхідно відмітити, що теоретична лінія, яка вирівнює ряд динаміки, цілком залежить тільки від двох крайніх значень рівнів ряду динаміки (початкового і кінцевого), які можуть суттєво змінюватись під впливом випадкових коливань. Відповідно тенденція, яка дійсно має місце в досліджуваному явищі, буде спотворена. В зв'язку з цим прийом вирівнювання рядів динаміки по середньому абсолютному приросту доцільно використовувати лише для рядів, що мають стабільні щорічні абсолютні прирости. Практично цей прийом використовується в динамічних рядах, які охоплюють нетривалий період часу, протягом якого не відбувається суттєвих якісних змін у рівнях факторів, що визначають тенденцію, і ступенях їх впливу на досліджувану ознаку.

Вирівнювання ряду динаміки по **середньому коефіцієнту зростання** застосовується в тих випадках, коли в досліджуваному ряду кожен наступний рівень змінюється порівняно з попереднім приблизно в одну і ту саму кількість разів, що дорівнює величині середнього коефіцієнта зростання, тобто коли фактори, що визначають основну тенденцію, зумовлюють від періоду до періоду однакові коефіцієнти зростання досліджуваного явища.

Вирівняні значення рівнів ряду динаміки визначають за формулою:

$$\tilde{y}_t = y_0 \cdot \bar{k}^t,$$

де $\bar{k}$ – середній коефіцієнт зростання.

Таблиця 10.9

Динаміка вартості основних виробничих фондів тваринництва в ТОВ за 1997–2002 рр.

Рік	Вартість основних виробничих фондів на 100 га сільськогосподарських угідь (фондозабезпеченість), тис. грн	Порядковий номер року	Фондозабезпеченість, вирівняна по середньому коефіцієнту зростання	Відхилення фактичного рівня від вирівняного
	y	t	$\tilde{y}_t = y_0 \cdot \bar{k}^t$	$y - \tilde{y}_t$
1997	160	0	160	0
1998	169	1	170	–1
1999	181	2	180	1
2000	190	3	191	–1
2001	202	4	202	0
2002	214	5	214	0

Порядок вирівнювання по середньому коефіцієнту зростання розглянемо на прикладі динаміки фондозабезпеченості (табл. 10.9).

Аналіз динамічного ряду показує, що абсолютні прирости збільшуються від 9-10 тис. грн. у перші роки до 12 тис. грн. в останні роки, а коефіцієнти зростання залишаються приблизно однаковими і становлять 1,06-1,07. Отже, для даного динамічного ряду характерно збільшення кожного наступного рівня порівняно з попереднім в ту саму кількість разів, яка дорівнює величині середнього коефіцієнта зростання. Тому даний ряд динаміки доцільно вирівнювати по середньому коефіцієнту зростання.

Визначимо середній коефіцієнт зростання фондозабезпеченості за формулою:

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_0}} = \sqrt[6-1]{\frac{214}{160}}; \quad \lg \bar{k} = \frac{\lg y_n - \lg y_0}{n-1} = \frac{\lg 214 - \lg 160}{6-1} =$$

$$= \frac{2{,}3304 - 2{,}2041}{5} = \frac{0{,}1263}{5} = 0{,}02526.$$

За таблицями антилогарифмів встановимо значення $\bar{k} = 1{,}06$.

Отже, фондозабезпеченість господарства щороку в середньому зростала в 1,06 рази, або на 6,0%.

Обчислимо вирівняні по середньому коефіцієнту зростання значення фондозабезпеченості:

в 1997 р. (при $t = 0$) $\tilde{y}_t = y_0 \cdot \bar{k}^t = 160 \cdot 1{,}06^0 = 160$ тис. грн.;

в 1998 р. (при $t = 1$) $\tilde{y}_t = y_0 \cdot \bar{k}^t = 160 \cdot 1{,}06^1 = 170$ тис. грн. і т.д.

Слід мати на увазі, що при визначенні вирівняних значень рівнів ряду динаміки по середньому коефіцієнту зростання, так само як і при вирівнюванні по середньому абсолютному приросту, у вихідному і вирівняному рядах динаміки початкові і кінцеві рівні збігаються.

Вирівняний по середньому коефіцієнту зростання ряд динаміки являє собою показникову криву. Цьому способу вирівнювання рядів динаміки притаманний той самий недолік, що й вирівнюванню по середньому абсолютному приросту. Тут при визначенні вирівняних значень використовуються тільки два крайніх рівні динамічного ряду (початковий і кінцевий), які внаслідок впливу випадкових факторів можуть бути нехарактерними для досліджуваного суспільного явища.

Досконалішим і точнішим прийомом вирівнювання рядів динаміки, який враховує всі рівні вихідного ряду, є **аналітичне вирівнювання по способу найменших квадратів.**

Вирівнювання по цьому способу ґрунтується на припущенні, що зміни досліджуваного ряду динаміки можуть бути наближено виражені певним математичним рівнянням (апроксимуючою функцією), за яким і визначають вирівняні рівні динамічного ряду. Іншими словами, рівні ряду динаміки розглядаються як функція часу $\tilde{y}_t = f(t)$, де $\tilde{y}_t$ – рівні динамічного ряду, визначені за відповідним рівнянням на момент часу t.

Аналітичне вирівнювання можна провести з використанням різних типів функцій: прямої лінії, параболи другого порядку, показникової кривої (експоненти), гіперболи тощо.

Рівняння, що виражає рівні ряду динаміки як деяку функцію часу t, називають **трендом**. Поняття про рівняння тенденції було введене в статистику англійським вченим Гукером у 1902 р. Він запропонував називати таке рівняння трендом (the trend).

Суть аналітичного вирівнювання динамічних рядів полягає в тому, що фактичні рівні ряду замінюються рядом рівнів, які змінюються плавно (теоретичними рівнями), обчисленими на основі певної кривої, вибраної в припущенні, що вона найточніше відображає загальну тенденцію зміни досліджуваного соціально-економічного явища у часі.

Підбір найбільш придатної функції є важливим і відповідальним завданням, від якого в остаточному підсумку залежать результати вирівнювання. В основі вирішення його має бути змістовний теоретичний аналіз істотності досліджуваного явища і законів його розвитку. Треба підібрати таку криву, яка б максимально близько проходила до фактичних рівнів. Добитися цього можна за умови, що сума квадратів відхилень фактичних рівнів (y) від розрахованих за рівнянням ($\tilde{y}_t$) буде мінімальною $(y - \tilde{y}_t)^2 = \min$.

У практиці економічних досліджень найчастіше застосовують такий підхід: добирають кілька рівнянь, визначають їх параметри, а потім віддають перевагу тому, в якому $(y - \tilde{y}_t)^2$ і коефіцієнт варіації найменші.

Наближено обґрунтувати рівняння, що відображує основну тенденцію, можна за допомогою побудови графіка (лінійної діаграми).

Вирівнювання динамічних рядів способом найменших квадратів, як і вирівнювання за допомогою інших прийомів, має здійснюватись в межах однаякісних періодів. Якщо в динамічному ряду є якісно специфічні періоди, то виявляти тенденцію доцільно в межах кожного з них.

Залежно від вихідних даних для вирівнювання рядів динаміки можуть бути вибрані різні типи кривих або пряма лінія. Аналіз динаміки соціально-економічних явищ показує, що їхня зміна супроводжується постійними зростаючими і спадаючими абсолютними приростами, постійними темпами зростання і приросту, прискоренням або уповільненням, тобто їхнє вирівнювання слід здійснювати за рівнянням прямої лінії, параболи другого порядку або показникової кривої. Основна тенденція (тренд) показує, як виливають систематичні фактори на рівні ряду динаміки, а відхилення фактичних рівнів від вирівняних характеризує варіацію рівнів, викликану індивідуальними особливостями кожного періоду. Випадкова (залишкова) варіація в рядах динаміки може бути виміряна способами, якими вимірюється звичайна варіація, наприклад за допомогою залишкового середнього квадратичного відхилення

$$\sigma_{зал} = \sqrt{\frac{\sum (y - \tilde{y}_t)}{n}}$$

або коефіцієнта варіації $V = \frac{\sigma_{зал}}{\bar{y}} \cdot 100\%$.

Показники варіації рівнів динамічних рядів можуть бути використані для оцінки правильності вибору апроксимуючої функції (рівняння) для вирівнювання, а також оцінки порівняльної стійкості окремих динамічних рядів. Очевидно, що чим показники варіації менше, тим вирівнювання здійснене точніше, а ряди динаміки стійкіші.

Вирівнювання динамічних рядів за **рівнянням прямої лінії** доцільно проводити тоді, коли для емпіричного ряду характерні більш або менш постійні ланцюгові абсолютні прирости, тобто тоді, коли рівні ряду змінюються приблизно в арифметичній прогресії.

Стосовно рядів динаміки аналітичне рівняння прямої лінії має вигляд:

$$\tilde{y}_t = a_0 + a_1 t,$$

де $\tilde{y}_t$ – вирівняні значення рівнів динамічного ряду;

t – час, тобто порядкові номери періодів;

a_0 і a_1 – параметри рівняння шуканої прямої;

a_0 – початок відліку (економічного змісту не має);

a_1 – коефіцієнт регресії або пропорційності, який показує середній щорічний приріст (зниження) досліджуваного явища (тангенс кута нахилу прямої лінії до осі абсцис).

Параметри a_0 і a_1 шуканої прямої, що задовольняють вимозі способу найменших квадратів, знаходять, розв'язуючи таку систему нормальних рівнянь:

$$\begin{cases} \sum y = a_0 n + a_1 \sum t; \\ \sum yt = a_0 \sum t + a_1 \sum t^2, \end{cases}$$

де n – число рівнів ряду динаміки.

Техніка утворення системи рівнянь така. Перше рівняння дістають множенням всіх членів вихідного рівняння ($\tilde{y}_t = a_0 + a_1 t$) на коефіцієнт при a_0 (на одиницю) і підсумовуванням знайдених добутків. Щоб мати друге рівняння, всі члени вихідного рівняння необхідно помножити на коефіцієнт при a_1 (на t) і знайдені добутки підсумувати. Аналогічно будують систему нормальних рівнянь і для інших кривих (параболи другого порядку, гіперболи тощо).

Розрахунок параметрів рівняння (a_0 і a_1) можна значно спростити, якщо відлік часу (t) проводити так, щоб сума показників часу дорівнювала нулю ($t = 0$). Цього досягають так. Рівень, що знаходиться всередині ряду динаміки, беруть за умовний початок відліку, або нульове значення. Для того щоб сума показників часу дорівнювала нулю, умовні позначення дат потрібно давати так.

При непарному числі рівнів ряду динаміки для отримання $t = 0$ рівень, що знаходиться всередині ряду, прирівнюють до нуля, а рівні, розташовані вище його, позначають числами із знаком мінус: – 1, – 2, – 3 і т.д., а нижче – числами зі знаком плюс: + 1, + 2, + 3 і т.д.

При парному числі рівнів ряду динаміки рівні, що лежать вище серединного значення (воно знаходиться всередині між двома серединними датами), позначають натуральними числами із знаком мінус: – 1, – 3, – 5 і т.д., а рівні, що лежать нижче серединного значення, – натуральними числами зі знаком плюс: + 1, + 3, + 5 і т.д.

За умовою, що $t = 0$, система нормальних рівнянь спрощується і приймає вигляд:

$$\begin{cases} \sum y = a_0 n; \\ \sum yt = a_1 \sum t^2. \end{cases}$$

Звідки $a_0 = \dfrac{\sum y}{n}$; $a_1 = \dfrac{\sum yt}{\sum t^2}$.

Розв'язуючи вихідну систему нормальних рівнянь способом визначників, параметри a_0 і a_1 можна обчислити за іншими формулами, які дають змогу дістати точніші результати за рахунок зведення до мінімуму помилки через закруглення в обчисленнях параметрів:

$$a_0 = \frac{\sum y \sum t^2 - \sum yt \sum t}{n\sum t^2 - \sum t \sum t}; \qquad a_1 = \frac{n\sum yt - \sum y \sum t}{n\sum t^2 - \sum t \sum t}.$$

Отже, для визначення параметрів a_0 і a_1 необхідно мати чотири суми: y; yt; t; t^2.

Якщо ж $t = 0$, то тоді формули для обчислення параметрів a_0 і a_1, спрощуються, набираючи такого вигляду:

$$a_0 = \frac{\sum y \sum t^2 - \sum yt \sum t}{n\sum t^2 - \sum t \sum t} = \frac{\sum y \sum t^2}{n\sum t^2} = \frac{\sum y}{n} = \bar{y};$$

$$a_1 = \frac{n\sum yt - \sum y \sum t}{n\sum t^2 - \sum t \sum t} = \frac{n\sum yt}{n\sum t^2} = \frac{\sum yt}{\sum t^2}.$$

Значення t можна обчислити за формулою:

$$\sum t = \frac{n(n+1)}{2}.$$

За умов, що $t = 0$, значення t^2 можна відшукати за формулами:

при парному числі рівнів $\sum t^2 = \frac{(n-1)n(n+1)}{3} = \frac{n(n-1)}{3}$;

при непарному числі рівнів $\sum t^2 = \frac{(n-1)n(n+1)}{12} = \frac{n(n-1)}{12}$;

Для обчислення параметрів a_0 і a_1 можна користуватися формулами:

$$a_0 = \bar{y} - a_1\bar{t}; \qquad a_1 = \frac{\sum (y - \bar{y})(t - \bar{t})}{\sum (t - \bar{t})^2}.$$

Порядок вирівнювання за рівнянням прямої лінії проілюструємо на прикладі ряду динаміки урожайності соняшнику (табл. 10.10).

Обгрунтуємо вибір математичного рівняння для вирівнювання динамічного ряду. З даних таблиці видно, що зростання урожайності відбувається рівномірно. Побудова лінійної діаграми (див. рис. 10.1) показує, що ламана крива за своєю формою близька до прямої лінії. Виходячи з цього, доцільніше цей ряд динаміки вирівнювати за рівнянням прямої лінії (лінійним трендом):

$$\bar{y}_t = a_0 + a_1 t.$$

Параметри a_0 і a_1 шуканої прямої, яка задовольняє способу найменших квадратів, знайдемо, розв'язавши таку систему нормальних рівнянь:

$$\begin{cases} \sum y = a_0 n + a_1 \sum t; \\ \sum yt = a_0 \sum t + a_1 \sum t^2. \end{cases}$$

Отже, щоб визначити параметри рівняння, **необхідно знайти такі чотири суми:** y; yt; t; t^2.

Усі розрахунки зведемо в табл. 10.10.

Таблиця 10.10

Розрахункові дані для аналітичного вирівнювання динамічного ряду урожайності соняшнику способом найменших квадратів

Рік	Урожайність, ц/га y	1 спосіб				2 спосіб				Різниця між фактичною і вирівняною урожайністю $y-\bar{y}_t$	Квадрат різниці $(y-\bar{y}_t)^2$
		t	yt	t^2	Вирівняна урожайність, ц/га $\bar{y}_t$	t	yt	t^2	Вирівняна урожайність, ц/га $\bar{y}_t$		
1988	10,3	1	10,3	1	11,0	– 7	– 72,1	49	11,0	– 0,7	0,49
1989	12,1	2	24,2	4	11,5	– 6	– 72,6	36	11,5	0,6	0,36
1990	11,6	3	34,8	9	12,0	– 5	– 58,0	25	12,0	– 0,4	0,16
1991	13,1	4	52,4	16	12,6	– 4	– 52,4	16	12,6	0,5	0,25
1992	13,9	5	69,5	25	13,1	– 3	– 41,7	9	13,1	0,8	0,64
1993	12,9	6	77,4	36	13,6	– 2	– 25,8	4	13,6	– 0,7	0,49
1994	14,1	7	98,7	49	14,2	– 1	– 14,4	1	14,2	– 0,1	0,01
1995	15,0	8	120,0	64	14,7	0	0	0	14,7	0,3	0,09
1996	16,4	9	147,6	81	15,2	1	16,4	1	15,2	1,2	1,44
1997	15,9	10	159,0	100	15,8	2	31,8	4	15,8	0,1	0,01
1998	16,9	11	185,9	121	16,3	3	50,7	9	16,3	0,6	0,36
1999	13,8	12	165,6	144	16,8	4	55,2	16	16,8	– 3,0	9,00
2000	16,6	13	215,8	169	17,3	5	83,0	25	17,3	– 0,7	0,49
2001	17,8	14	249,2	196	17,9	6	106,8	36	17,9	– 0,1	0,01
2002	20,0	15	300,0	225	18,4	7	140,0	49	18,4	1,6	2,56
Разом	220,4	120	1910,4	1240	220,4	0	147,2	280	220,4	0	16,36

Вирівнювання динамічного ряду проведемо двома способами – звичайним і спрощеним (способом відліку від умовного початку).

І спосіб. Використовуючи отримані величини, розв'яжемо систему рівнянь:

$$\begin{cases} 220{,}4 = 15a_0 + 120a_1 \\ 1910{,}4 = 120a_0 + 1240a_1 \end{cases} \left| \begin{array}{l} :15 \\ :20 \end{array} \right.$$

Поділимо обидва рівняння на коефіцієнти при a_0: перше рівняння на 15, а друге – на 120, а потім віднімемо від другого рівняння перше:

$$\begin{cases} 14{,}69 = a_0 + 8{,}00a_1; \\ 15{,}92 = a_0 + 10{,}33a_1; \end{cases}$$

$$1{,}23 = 2{,}33a_1.$$

Звідси $a_1 = 0{,}5279 = 0{,}53$ ц/га. $1{,}23 = 2{,}33a_1$.

Визначимо a_0, підставивши в одне з рівнянь значення a_1:

$$220{,}4 = 15a_0 + 120 \cdot 0{,}53;$$

$$a_0 = \frac{220{,}4 - 63{,}2}{15} = \frac{157{,}2}{15} = 10{,}48 \text{ ц/га}$$

Параметри рівняння можна визначити і за іншими, зручнішими формулами:

$$a_0 = \frac{\sum y \sum t^2 - \sum yt \sum t}{n\sum t^2 - \sum t \sum t} = \frac{220{,}4 \cdot 1240 - 1910{,}4 \cdot 120}{15 \cdot 1240 - 120 \cdot 120} = 10{,}48 \text{ ц/га};$$

$$a_1 = \frac{n\sum yt - \sum y \sum t}{n\sum t^2 - \sum t \sum t} = \frac{15 \cdot 1910{,}4 - 220{,}4 \cdot 120}{15 \cdot 1240 - 120 \cdot 120} = 0{,}53 \text{ ц/га},$$

або $a_0 = \bar{y} - a_1\bar{t}$,

де $\bar{y} = \dfrac{\sum y}{n} = \dfrac{220{,}4}{15} = 14{,}69$; $\bar{t} = \dfrac{\sum t}{n} = \dfrac{120}{15} = 8$;

$$a_1 = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{\bar{t^2} - (\bar{t})^2}; \quad \overline{yt} = \frac{\sum yt}{n} = \frac{1910{,}4}{15} = 127{,}36;$$

$$\bar{t^2} = \frac{\sum t^2}{n} = \frac{1240}{15} = 82{,}67; \quad (\bar{t})^2 = 8^2 = 64.$$

$$a_1 = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{\bar{t^2} - (\bar{t})^2} = \frac{127{,}36 - 14{,}69 \cdot 8}{82{,}67 - 64{,}00} = \frac{9{,}84}{18{,}67} = 0{,}53 \text{ ц/га};$$

$$a_0 = \bar{y} - a_1\bar{t} = 14{,}69 - 0{,}5287 \cdot 8 = 10{,}48 \text{ ц/га}.$$

Перевіримо правильність розв'язання системи рівнянь, виходячи з рівності:

$$\tilde{y} = a_0 + a_1\bar{t}; \quad 14{,}69 = 10{,}48 + 0{,}5287 \cdot 8 = 14{,}69.$$

Таким чином, рівняння прямої лінії, що вирівнює ряд динаміки, має вигляд:

$$\tilde{y}_t = a_0 + a_1 t = 10{,}48 + 0{,}53t.$$

Коефіцієнт регресії $a_1 = 0{,}53$ ц/га показує, що в середньому за досліджуваний період урожайність соняшнику щорічно підвищувалася на 0,53 ц/га. Коефіцієнт $a_0 = 10{,}48$ ц/га – значення вирівняної урожайності для року в динамічному ряду, взятого за початок відліку (1995 р., коли $t = 0$).

Підставляючи в отримане рівняння значення ($t = 1, 2, ..., 15$), дістанемо вирівняні (розрахункові) значення урожайності. Наприклад:

для 1988 р. $\tilde{y}_{t=1} = 10{,}48 + 0{,}53 \cdot 1 = 11{,}0$ ц/га;

для 1989 р. $\tilde{y}_{t=2} = 10{,}48 + 0{,}53 \cdot 2 = 11{,}5$ ц/га і т.д.

Перевіримо правильність всіх розрахунків, порівнюючи суми фактичної і вирівняної урожайності:

$$\sum y = \sum \tilde{y}_t \quad ; \quad 220{,}4 = 220{,}4.$$

2 спосіб. Для спрощення розрахунків використаємо спосіб відліку від умовного початку. Виразимо значення дат (t) у відхиленнях від дати, взятої за умовний початок (вона знаходиться в центрі ряду динаміки: $t = 0$ в 1995 р. (табл. 10.10). Система рівнянь спрощується, оскільки $t = 0$:

$$\begin{cases} \sum y = na_0; \\ \sum yt = a_1 \sum t^2. \end{cases}$$

Звідси

$$a_0 = \frac{\sum y}{n} = \frac{220{,}4}{15} = 14{,}69 \quad \text{ц/га};$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{147{,}2}{280} = 0{,}53 \quad \text{ц/га}.$$

Параметри рівняння можна знайти і за іншими, наведеними у першому способі формулами розв'язання прикладу. Їхні значення будуть такими самими.

Рівняння лінійного тренду має вигляд:

$$\tilde{y}_t = a_0 + a_1 t = 14{,}69 + 0{,}53t.$$

Параметр $a_0 = 14{,}69$ ц/га – значення вирівняної урожайності для центрального в динамічному ряду року, взятого за початок відліку. Для 1995 р. $t = 0$, тоді $\tilde{y}_t = 14{,}69 + 0{,}53 \cdot 0 = 14{,}69$ ц/га. Параметр

a_0 дорівнює середній урожайності в динамічному ряду $\bar{y} = \sum y : n = 220{,}4 : 15 = 14{,}69$ ц/га. Коефіцієнт регресії $a_1 = 0{,}53$ ц/га характеризує середнє щорічне збільшення урожайності. Він має таке саме значення і зміст, що й при вирівнюванні першим способом.

Вирівняні (теоретичні) рівні урожайності ($\tilde{y}_t$) обчислюють аналогічно тому, як це було зроблено в першому способі розрахунків.

Щоб оцінити ступінь наближення лінійного тренду до фактичних даних динамічного ряду, розрахуємо залишкове середнє квадратичне відхилення і коефіцієнт варіації. Для цього обчислимо відхилення фактичної урожайності від вирівняної $(y - \tilde{y}_t)$, їх квадрати $(y - \tilde{y}_t)^2$ і їх суму $\sum (y - \tilde{y}_t)^2$ (гр. 11 і 12, табл. 10.10).

Залишкове середнє квадратичне відхилення становитиме:

$$\sigma_{\text{зал}} = \sqrt{\frac{\sum (y - \tilde{y}_t)^2}{n}} = \sqrt{\frac{16{,}36}{15}} = \sqrt{1{,}09} = 1{,}04 \quad \text{ц/га}.$$

Коефіцієнт варіації дорівнює:

$$V = \frac{\sigma}{\bar{y}} 100\% = \frac{1{,}04}{14{,}69} \cdot 100\% = 7{,}1\%.$$

Отже, коливання фактичної урожайності навколо прямої лінії в середньому становить 1,04 ц/га, або 7,1%. Невеликий коефіцієнт варіації вказує на те, що рівняння прямої лінії досить точно відображає тенденцію зміни урожайності в часі.

Водночас аналіз динамічного ряду урожайності свідчить про те, що не зважаючи на значне коливання урожайності по роках, чітко простежується тенденція її підвищення і прискорення приростів в останні роки. Тому логічно припустити, що досліджуваний ряд динаміки можна вирівнювати за рівнянням параболи другого порядку.

Вирівнювання рядів динаміки **за параболою другого порядку** здійснюється в тих випадках, коли зміна рівнів ряду відбувається приблизно рівномірним прискоренням або уповільненням ланцюгових абсолютних приростів.

Рівняння параболи другого порядку характеризується трьома параметрами:

$$\tilde{y}_t = a_0 + a_1 t + a_2 t^2,$$

де a_2 – показник щорічного прискорення (або уповільнення, якщо a_2 зі знаком мінус) абсолютного приросту рівнів ряду.

Параметри параболи другого порядку a_0, a_1 і a_2 знаходять з такої системи нормальних рівнянь:

$$\begin{cases}\sum y = a_0 n + a_1\sum t + a_2\sum t^2;\\ \sum yt = a_0\sum t + a_1\sum t^2 + a_2\sum t^3;\\ \sum yt^2 = a_0\sum t^2 + a_1\sum t^3 + a_2\sum t^4.\end{cases}$$

При $t = 0$ і $t^3 = 0$ система рівнянь значно спрощується, набуваючи такого вигляду:

$$\begin{cases}\sum y = a_0 n + a_2\sum t^2;\\ \sum yt = a_1\sum t^2;\\ \sum yt^2 = a_0\sum t^2 + a_2\sum t^4.\end{cases}$$

З цієї системи a_1 визначають елементарно з другого рівняння

$$a_1 = \frac{\sum yt}{\sum t^2},$$

а a_0 і a_1 з системи двох рівнянь з двома невідомими:

$$\begin{cases}\sum y = a_0 n + a_2\sum t^2,\\ \sum yt^2 = a_0\sum t^2 + a_2\sum t^4.\end{cases}$$

Використовуючи дані попереднього прикладу про динаміку урожайності соняшнику за 1988–2002 рр. (табл. 10.10), проведемо вирівнювання ряду за рівнянням параболи другого порядку способом найменших квадратів (табл. 10.11). Для спрощення розрахунків візьмемо $\sum t = 0$ і $\sum t^3 = 0$.

Щоб визначити параметри a_0, a_1 і a_2, рівняння параболи другого порядку, розв'яжемо таку систему рівнянь:

$$\begin{cases}\sum y = a_0 n + a_2\sum t^2;\\ \sum yt = a_1\sum t^2;\\ \sum yt^2 = a_0\sum t^2 + a_2\sum t^4.\end{cases}$$

Підставимо знайдені величини в систему рівнянь:

$$\begin{cases}220{,}4 = 15a_0 + 280a_2;\\ 147{,}2 = 280a_1;\\ 4119{,}4 = 280a_0 + 9352a_2.\end{cases}$$

Таблиця 10.11

Розрахунок даних для вирівнювання динамічного ряду урожайності соняшнику за параболою другого порядку способом найменших квадратів

Рік	Урожайність, ц/га	Умовне позначення часу	Розрахункові величини				Вирівняна урожайність, ц/га	Різниця між фактичною і вирівняною урожайністю, ц/га	Квадрат різниці
	y	t	t^2	t^4	yt	yt^2	$\bar{y}_t$	$y-\bar{y}_t$	$y-\bar{y}_t$
1988	10,3	– 7	49	2401	– 72,1	504,7	10,8	0,5	0,25
1989	12,1	– 6	36	1296	– 72,6	435,6	11,5	0,6	0,36
1990	11,6	– 5	25	625	– 58,0	290,0	12,0	– 0,4	0,16
1991	13,1	– 4	16	256	– 52,4	209,6	12,6	0,5	0,25
1992	13,9	– 3	9	81	– 41,7	125,1	13,1	0,8	0,64
1993	12,9	– 2	4	16	– 25,8	51,6	13,6	– 0,7	0,49
1994	14,1	– 1	1	1	– 14,1	14,1	14,1	0	0
1995	15,0	0	0	0	0	0	14,7	0,3	0,09
1996	16,4	1	1	1	16,4	16,4	15,2	1,2	1,44
1997	15,9	2	4	16	31,8	63,6	15,7	0,2	0,04
1998	16,9	3	9	81	50,7	152,1	16,3	0,6	0,36
1999	13,8	4	16	256	55,2	220,8	16,8	– 3,0	9,00
2000	16,6	5	25	625	83,0	415,0	17,4	– 0,8	0,64
2001	17,8	6	36	1296	106,8	640,8	17,9	– 0,1	0,01
2002	20,0	7	49	2401	140,0	980,0	18,7	1,3	1,69
Разом	220,4	0	280	9352	147,2	4119,4	220,4	0	15,42

З другого рівняння визначимо значення a_1 :

$$a_1 = \frac{147,2}{280} = 0,5257 = 0,53 \quad \text{ц/га}.$$

Розв'язавши перше і третє рівняння, матимемо значення параметрів a_0 і a_2:

$$\begin{cases} 220,4 = 15a_0 + 280a_2; \\ 4119,4 = 280a_0 + 9352a_2. \end{cases}$$

Для вирівнювання коефіцієнтів при a_0 розділимо перше рівняння на 15, а друге – на 280. Дістанемо:

$$\begin{cases} 14,69 = a_0 + 18,67a_2; \\ 14,71 = a_0 + 33,40a_2. \end{cases}$$

Віднявши з другого рівняння перше, отримаємо:

$$0,02 = 14,73a_2.$$

Звідси

$$a_2 = \frac{0{,}02}{14{,}73} = 0{,}0014 \quad ц/га.$$

Визначимо **значення** параметру a_0, підставивши **значення a_2 в перше** рівняння:

$$220{,}4 = 15a_0 + 280 \cdot 0{,}0014;$$

$$a_0 = \frac{220{,}4 - 0{,}39}{15} = \frac{220{,}01}{15} = 14{,}67 \quad ц/га.$$

Розв'язання системи рівнянь дає таке значення шуканих параметрів рівняння (в ц/га):

$$a_0 = 14{,}67;\ a_1 = 0{,}53;\ a_2 = 0{,}0014.$$

Параметри рівняння параболи другого порядку можна визначити не тільки способом підстановки, а й безпосередньо користуючись формулами, що спрощують розрахунки:

$$a_0 = \frac{\sum y \sum t^4 - \sum yt^2 \sum t^2}{n\sum t^2 - \sum t^2 \sum t^2} = \frac{220{,}4 \cdot 9352 - 4119{,}4 \cdot 280}{15 \cdot 9352 - 280 \cdot 280} = 14{,}67 \text{ ц/га};$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{147{,}2}{280} = 0{,}5257 = 0{,}53 \quad ц/га;$$

$$a_2 = \frac{n\sum yt^2 - \sum y \sum t^2}{n\sum t^4 - \sum t^2 \sum t^2} = \frac{15 \cdot 4119{,}4 - 220{,}4 \cdot 280}{15 \cdot 9352 - 280 \cdot 280} = 0{,}0014 \text{ ц/га},$$

Перевіримо правильність розрахунку параметрів, підставляючи в нормальне рівняння їхні числові значення:

$$\begin{cases} \sum y = a_0 n + a_2 \sum t^2; \ \ 220{,}4 = 15 \cdot 14{,}67 + 0{,}014 \cdot 280 = 220{,}4; \\ \sum yt = a_1 \sum t^2; \ \ 147{,}2 = 0{,}5257 \cdot 280 = 147{,}2; \\ \sum yt^2 = a_0 \sum t^2 + a_2 \sum t^4; \ \ 4119{,}4 = 14{,}67 \cdot 280 + 0{,}0014 \cdot 9352 = 4119{,}4. \end{cases}$$

Отже, параболічний тренд має такий вигляд:

$$\tilde{y}_t = a_0 + a_1 t + a_2 t^2 = 14{,}67 + 0{,}53t + 0{,}0014t^2.$$

Пояснимо значення знайдених коефіцієнтів: $a_0 = 14{,}67$ ц/га – це початок відліку або вирівняне значення урожайності для центрального в ряду динаміки року, взятого за початок умовного відліку (1995 р., коли $t = 0$); $a_1 = 0{,}53$ ц/га – середній щорічний приріст урожайності: $a_2 = 0{,}0014$ ц/га – середнє щорічне прискорення приросту урожайності.

Обчислимо за рівнянням параболи згладжене значення урожайності, підставляючи до рівняння замість t його числові значення (від – 7 до + 7; гр. 8, табл. 10.11):

в 1988 р. при $t = -7$; $\tilde{y}_t = 14,67 + 0,53(-7) + 0,0014(-7)^2 = 10,8$ ц/га,

в 1989 р. при $t = -6$; $\tilde{y}_t = 14,67 + 0,53(-6) + 0,0014(-6)^2 = 11,5$ ц/га і т.д.

Перевіримо правильність розрахунків:

$$\sum y = \sum \tilde{y}_t \text{ ; } 220,4 = 220,4.$$

Як видно з розрахунків, вирівняні рівні урожайності дуже близькі до фактичних рівнів. Отже, парабола другого порядку досить точно відображає тенденцію зміни урожайності на досліджуваному відрізку часу.

Щоб оцінити ступінь наближення параболічного тренду до фактичних даних динамічного ряду, обчислимо залишкове середнє квадратичне відхилення і коефіцієнт варіації. Для цього визначимо відхилення ($y - \tilde{y}_t$), квадрат відхилення $(y - \tilde{y}_t)^2$ та їхню суму $\sum (y - \tilde{y}_t)^2$ (гр. 9 і 10, табл. 10.11).

Залишкове середнє квадратичне відхилення становитиме:

$$\sigma_{зал} = \sqrt{\frac{\sum (y - \tilde{y}_t)^2}{n}} = \sqrt{\frac{15,42}{15}} = \sqrt{1,028} = 1,01 \quad \text{ц/га.}$$

Коефіцієнт варіації дорівнює:

$$V = \frac{\sigma}{\bar{y}} \cdot 100\% = \frac{1,01}{14,69} \cdot 100\% = 6,9\%.$$

Отже, коливання фактичної урожайності навколо параболи другого порядку в середньому становить 1,01 ц/га, або 6,9%.

Невеликий коефіцієнт варіації вказує на те, що параболічний тренд досить точно відображає тенденцію зміни урожайності за досліджуваний відрізок часу.

Порівняємо результати вирівнювання динамічного ряду урожайності за лінійним і параболічним трендом. Залишкове середнє квадратичне відхилення, добуте за рівнянням параболи другого порядку, дещо менше, ніж залишкове середнє квадратичне відхилення, отримане за рівнянням прямої лінії ($1,01 < 1,04$). Отже, парабола другого порядку точніше відтворює тенденцію зміни урожайності у часі, ніж пряма лінія.

Однак несуттєві відмінності в σ і V допускають можливість вирівнювання даного ряду динаміки і за лінійним трендом.

Для обґрунтування вибору вирівнювання ряду динаміки за лінійним або параболічним трендом можна оцінити істотність відмінностей між залишковими дисперсіями по F-критерію Фішера за правилами, викладеними в курсі математичної статистики.

Фактичне дисперсійне відношення становитиме:

$$F_{\text{факт}} = \frac{\sigma^2_{\text{по прямій}}}{\sigma^2_{\text{по параболі}}} = \frac{1{,}09}{1{,}028} = 1{,}06.$$

Табличне значення F-критерію при кількості ступенів свободи варіації $k = n - 1 = 15 - 1 = 14$ і рівні значущості = 0,05 становитиме 3,74.

Оскільки $F_{\text{табл}} > F_{\text{факт}}$ (3,74 > 1,06), то відмінності в залишкових дисперсіях є випадковими, відтак не можна віддати перевагу якомусь-небудь способу вирівнювання.

Якщо рівні ряду динаміки виявляють тенденцію до постійності ланцюгових темпів зростання, тобто рівні ряду змінюються в геометричній прогресії, то вирівнювати цей ряд слід за **рівнянням показникової кривої**:

$$\tilde{y}_t = a_0 a_1^t,$$

де a_0 і t мають попередній зміст, a_1 - середній темп зростання вирівняного рівня за одиницю часу.

Техніка вирівнювання за рівнянням показникової кривої аналогічна техніці вирівнювання за рівнянням прямої лінії, відмінність лише у тому, що вирівнюються тут не рівні ряду, а їх логарифми.

Прологарифмувавши рівняння показникової кривої, дістанемо рівняння прямої лінії, в якому рівні ряду динаміки замінені на їхні логарифми:

$$\lg \tilde{y}_t = \lg a_0 + t \lg a_1.$$

Параметри рівняння a_0 і a_1 дістанемо з такої системи нормальних рівнянь:

$$\begin{cases} \sum \lg y = n \lg a_0 + \lg a_1 \sum t; \\ \sum t \lg y = \lg a_0 \sum t + \lg a_1 \sum t^2. \end{cases}$$

Прирівнявши $t = 0$, дістанемо:

$$\begin{cases} \sum \lg y = n \lg a_0; \\ \sum t \lg y = \lg a_1 \sum t^2. \end{cases}$$

Звідси $\lg a_0 = \dfrac{\sum \lg y}{n}$; $\lg a_1 = \dfrac{\sum t \lg y}{\sum t^2}$.

Отже, для обчислення параметрів рівняння a_0 і a_1 потрібно знайти три суми: $\sum \lg y$; $\sum t \lg y$; $\sum t^2$.

Порядок вирівнювання ряду динаміки за рівнянням показникової кривої способом найменших квадратів проілюструємо на такому прикладі (табл. 10.12).

Таблиця 10.12

Розрахунок даних для вирівнювання динамічного ряду урожайності помідорів за показниковою кривою способом найменших квадратів

Рік	Урожайність, ц/га	Умовне позначення часу	Розрахункові величини				Вирівняні значення урожайності, ц/га
	y	t	t^2	$\lg y$	$t\lg y$	$\lg \tilde{y}_t$	$\tilde{y}_t$
1996	85	– 3	9	1,9294	– 5,7882	1,9309	85,3
1997	91	– 2	4	1,9590	– 3,9180	1,9562	90,4
1998	95	– 1	1	1,9777	– 1,9777	1,9815	95,8
1999	102	2	0	2,0086	0	2,0068	101,6
2000	108	1	1	2,0334	2,0334	2,0321	107,7
2001	115	2	4	2,0607	4,1214	2,0574	114,2
2002	120	3	9	2,0792	6,2376	2,0827	121,0
Разом	**716**	**0**	**28**	**14,0480**	**0,7085**	**14,0476**	**716,0**

Використовуючи знайдені величини, визначимо параметри рівнянь:

$$\lg a_0 = \frac{\sum \lg y}{n} = \frac{14{,}0480}{7} = 2{,}0068;$$

$$\lg a_1 = \frac{\sum t \lg y}{\sum t^2} = \frac{0{,}7085}{28} = 0{,}0253.$$

За таблицями антилогарифмів $a_0 = 101{,}58$; $a_1 = 1{,}06$.

Отже, рівняння показникової кривої, що відображує **тенденцію** зміни урожайності помідорів буде таким:

$$\lg \tilde{y}_t = \lg a_0 + t \lg a_1 = 2{,}0068 + 0{,}0253t$$

або, потенціюючи наведене рівняння, дістанемо:

$$\tilde{y}_t = a_0 a_1^{\,t} = 101{,}58 \cdot 1{,}06^t.$$

Знайдене рівняння має такий зміст: параметр a_0 є середньою геометричною фактичного ряду динаміки, а параметр a_1 – середнім коефіцієнтом зростання цього ж ряду динаміки. Віднявши від нього одиницю і перемноживши на 100%, матимемо темп зростання. Отже, за 1996–2002 рр. середньорічний темп зростання урожайності помідорів становив:

$$(1{,}06 - 1) \cdot 100\% = 6{,}0\%.$$

Вирівнювання за показниковою кривою здійснюється з врахуванням усіх значень рівнів вихідного ряду динаміки на відміну від вирівнювання по середньому коефіцієнту зростання, яке проводиться лише на основі співвідношення двох крайніх рівнів ряду.

Розрахуємо за рівнянням показникової кривої вирівняні значення урожайності помідорів, підставляючи в рівняння замість t його числові значення (від – 3 до + 3).

Наприклад, для 1996 р. це вирівняне значення буде таким:

$$\lg \tilde{y}_t = 2{,}0068 + 0{,}0253\ (-\ 3) = 1{,}9309.$$

Потенціюючи, отримуємо $\lg \tilde{y}_{t=-3} = 85{,}3$ ц/га.

Аналогічно визначають вирівняні рівні динаміки і для решти періодів часу.

Вирівняні рівні можна також розрахувати за звичайним рівнянням показникової кривої:

$$\tilde{y}_t = a_0 a_1{}^t = 101{,}58 \cdot 1{,}06^t,$$

підставляючи замість його числові значення (від – 3 до + 3).

Наприклад, для 1996 р. вирівняне значення буде таким:

$$\lg \tilde{y}_{t=-3} = 101{,}58 \cdot 1{,}06^{-3} = \frac{101{,}58}{1{,}06^3} = \frac{101{,}58}{1{,}191} = 85{,}3 \quad \text{ц/га і т.д.}$$

Логарифми вирівняних рівнів і самі рівні наведені в двох останніх графах табл. 10.12. Результати обчислення двома способами збігаються.

Як видно з розрахунків, вирівняні рівні урожайності дуже близькі до фактичних рівнів ряду динаміки. Отже, показникова крива досить точно відображає тенденцію зміни урожайності помідорів за досліджуваний відрізок часу.

При вирівнюванні за **гіперболою**, рівняння якої має вигляд:

$$\tilde{y}_t = a_0 + \frac{a_1}{t},$$

параметри a_0 і a_1 знаходять способом найменших квадратів з такої системи нормальних рівнянь:

$$\begin{cases} \sum y = a_0 n + a_1 \sum \frac{1}{t}; \\ \sum y \frac{1}{t} = a_0 \sum \frac{1}{t} + a_1 \sum \frac{1}{t^2}. \end{cases}$$

10.4. Факторний аналіз рядів динаміки

Важливе місце у вивченні динаміки соціально-економічних явищ належить факторному аналізу, метою якого є дослідження впливу окремих факторів на кількісні і якісні зміни явища в часі. В аналізі динаміки сільськогосподарського виробництва важливо насамперед оцінити залежність результативних показників від комплексу економічних і природних факторів.

Для здійснення факторного аналізу рядів динаміки статистика використовує ряд методів і прийомів, серед яких найбільш поширені такі: приведення рядів динаміки до однієї основи, порівняння кількох паралельних рядів результативних і факторних показників (одного результативного і одного факторного показників, одного результативного і кількох факторних показників, одного факторного і кількох результативних показників), укрупнення періодів, розчленування досліджуваної сукупності на якісно однорідні групи і підгрупи, тобто побудова простих і комбінаційних групувань, застосування дисперсійного і кореляційного методів аналізу та ін.

Застосування деяких з перелічених прийомів факторного аналізу рядів динаміки розглянемо на таких прикладах.

Найпростішим і розповсюдженим прийомом вивчення залежності результативних показників від факторів, що визначають тенденцію, є прийом **приведення рядів динаміки до однієї** основи. В табл. 10.13 подана динаміка виходу валової продукції сільського господарства і основних факторів в ТОВ за 1997–2002 рр.

Таблиця 10.13

Динаміка виходу валової продукції сільського господарства і основних факторів виробництва в ТОВ за 1997–2002 рр.

Рік	В розрахунку на 100 га сільськогосподарських угідь			
	Вартість валової продукції сільського господарства, тис. грн.	Вартість основних виробничих фондів сільськогосподарського призначення (фондозабезпеченість), тис. грн.	Енергетичних потужностей (енергозабезпеченість), кінських сил	Кількість внесених мінеральних добрив, ц діючої речовини
1997	66,7	103,4	246	87,0
1998	71,4	112,5	257	85,1
1999	75,0	120,4	275	94,4
2000	88,9	135,8	291	113,5
2001	84,3	146,7	327	110,8
2002	90,9	159,0	362	134,6

Дані таблиці свідчать про те, що збільшення виходу валової продукції на 100 га сільськогосподарських угідь (результативний показник) у динаміці в ТОВ супроводжувалося закономірним зростанням факторів інтенсивності сільськогосподарського виробництва (фондозабезпеченості, енергозабезпеченості і кількості внесених мінеральних добрив – факторні показники).

Для проведення порівняльного аналізу чотирьох рядів динаміки і виявлення більш чіткої залежності результативного показника від факторів інтенсивності виробництва здійснено перетворення вихідних рядів динаміки.

Потреба перетворення рядів динаміки зумовлена тим, що різнойменні показники наведених чотирьох паралельних рядів безпосередньо непорівнянні між собою. Крім того, вони виражені в різних одиницях вимірювання.

Щоб привести подібні ряди динаміки до порівнянного вигляду, використаємо прийом приведення їх до однієї основи. З цією метою розрахуємо базисні темпи зростання, взявши за постійну базу порівняння рівні 1997 р. Добуті дані виразимо в процентах (табл. 10.14).

Таблиця 10.14

Динаміка виходу валової продукції сільського господарства і факторів інтенсивності виробництва в ТОВ за 1997–2002 рр. (в % до 1997 р.)

Рік	В розрахунку на 100 га сільськогосподарських угідь			
	Вартість валової продукції сільського господарства, тис. грн.	Вартість основних виробничих фондів сільськогосподарського призначення (фондозабезпеченість), тис. грн.	Енергетичних потужностей (енергозабезпеченість), кінських сил	Кількість внесених мінеральних добрив, ц діючої речовини
1997	100,0	100,0	100,0	100,0
1998	107,0	108,8	104,5	97,8
1999	112,4	116,4	111,8	108,5
2000	133,3	131,3	118,3	130,4
2001	126,4	141,9	132,9	127,4
2002	136,3	153,8	147,2	154,7

Порівняння темпів зростання виходу валової продукції і факторів виробництва свідчить про випереджаючі темпи зростання факторів інтенсивності виробництва (в 1,5 рази) порівняно з темпами зростання виходу валової продукції (в 1,363, або на 36,3%). Це означає, що в даному господарстві в динаміці вихід валової продукції на одиницю факторів мав тенденцію до зниження.

Зроблені за факторним аналізом рядів динаміки висновки доповнимо розрахунком коефіцієнтів випередження. Коефіцієнти випередження визначимо як відношення базисних темпів зростання за однакові відрізки часу по двох динамічних рядах.

Так, коефіцієнт випередження зростання фондозабезпеченості порівняно зі зростанням валової продукції становить 1,13 (1,538 : 1,363), зростання енергозабезпеченості порівняно зі зростанням валової продукції – 1,08 (1,472 : 1,363), зростання кількості внесених мінеральних добрив порівняно зі зростанням валової продукції – 1,13 (1,547 : 1,363). Отже, темпи зростання факторів інтенсивності виробництва – фондозабезпеченості, енергозабезпеченості, кількості внесених добрив – випереджали темпи зростання валової продукції сільського господарства відповідно в 1,13; 1,08 і 1,13 рази. Важливу роль у проведенні факторного аналізу рядів динаміки має прийом **зіставлення кількох паралельних рядів.** Найбільш широко цей прийом використовується для вивчення залежності результативної ознаки від комплексу факторів. Покажемо це на прикладі табл. 10.15.

Таблиця 10.15

Динаміка продуктивності корів і факторів молочного скотарства в ТОВ

Показник	*У середньому за триріччя*		
	1994–1996 рр.	1997–1999 рр.	2000–2002 рр.
Надій молока на корову, кг	3452	3817	4264
Витрати кормів з розрахунку на корову, ц кормових одиниць	40,5	44,7	48,1
у тому числі концентрованих	9,4	10,2	12,3
Частка чистопородних корів, %	91	96	100
Вихід телят на 100 корів, гол.	92	95	98

Дані таблиці свідчать про те, що підвищення продуктивності корів у динаміці зумовлено зростанням рівня і якості годівлі, удосконаленням породної структури стада, а також кращим використанням корів для одержання приплоду.

Факторний аналіз рядів динаміки може бути здійснений на основі ряду прийомів виявлення основної тенденції, викладених у попередньому параграфі, зокрема з використанням прийому укрупнення періодів (табл. 10.16).

Таблиця 10.16

Динаміка урожайності зернових культур і її факторів в господарствах району

Періоди	Урожайність, ц/га	Метеоумови		Економічні умови	
		Сума опадів за травень-червень, мм	Сума опадів за серпень-вересень, мм	Внесено мінеральних добрив на 1 га зернових культур, ц діючої речовини	Вартість силових і робочих машин на 100 га ріллі, тис. грн.
1985-1989	21,7	129,5	104,3	0,85	46,5
1990-1994	25,9	106,2	90,8	1,05	68,7
1995-1999	30,8	117,4	101,5	1,57	76,8
2000-2002	38,5	133,7	109,9	2,15	81,5

Для виявлення залежності урожайності зернових культур від факторів інтенсифікації динамічний ряд урожайності був розділений на специфічні для району періоди. Це дало змогу по періодах нівелювати вплив метеорологічних умов (періоди за метеоумовами в цілому суттєво не відрізняються один від одного) і проявити вплив факторів інтенсифікації (кількості внесених мінеральних добрив і забезпеченості силовими і робочими машинами).

При факторному аналізі рядів динаміки важливе значення має розгляд показників, що безпосередньо визначають рівень досліджуваного явища. Якщо досліджуваний показник є складеним, то його необхідно розкласти на відповідні прості з наступним аналізом кожного з них. Наприклад, надій на корову можна розглядати як добуток двох показників: витрат кормів на одну корову і окупності кормів (вихід молока на 1 ц кормових одиниць). Собівартість 1 ц молока є, нарешті, відношення виробничих витрат на одну корову до продуктивності і т.д.

10.5. Інтерполяція і екстраполяція. Прогнозування суспільних явищ

Під час аналізу рядів динаміки доводиться стикатися з такими випадками, коли в рядах відсутні дані про їхні рівні за той або інший період. Такі дані можуть бути відсутні або всередині ряду, або спочатку чи в кінці його.

Приблизне визначення відсутніх рівнів усередині однoякісного періоду, коли відомі рівні, що лежать по обидві сторони невідомого, називають **інтерполяцією ряду динаміки.** Приблизне визначення невідомих рівнів, що лежать за його межами, тобто в майбутньому (або в минулому), називають **екстраполяцією ряду динаміки.** Відповідно екстраполювання може здійснюватися як у бік майбутнього (перспективна екстраполяція), так і у бік минулого (ретроспективна екстраполяція). По суті, екстраполяція являє собою продовження ряду динаміки на основі виявленої закономірності зміни рівнів за досліджуваний відрізок часу.

Інтерполяцію (як і екстраполяцію) здійснюють виходячи з припущення, що зміни в межах періоду, що виражають закономірність розвитку, відносно стійкі, тобто що ні виявлена тенденція, ні її характер не зазнали і не зазнають суттєвих змін у тому проміжку часу, рівні якого нам невідомі.

Щоб мати досить надійні результати обчислення відсутніх рівнів, інтерполяцію та екстраполяцію слід проводити в межах однорідних періодів, яким властива одна закономірність розвитку.

Інтерполяцію і екстраполяцію ряду динаміки можна проводити різними способами. Найпростішим способом є використання середніх характеристик досліджуваного ряду динаміки: середнього абсолютного приросту (при стабільних ланцюгових абсолютних приростах) і середнього коефіцієнту зростання (при стабільних темпах зростаннях). Однак визначення відсутніх рівнів ряду динаміки, і особливо при екстраполяції, найчастіше пов'язують з аналітичним вирівнюванням рядів способом найменших квадратів, який дає точніші результати. При цьому для виходу за межі періоду, для якого знайдена залежність від часу, досить продовжити значення незалежної змінної – часу.

Дослідження динаміки суспільних явищ і виявлення основної тенденції їх розвитку в минулому дають основу для визначення їхніх майбутніх розмірів.

Велику роль в плануванні має екстраполяція, яка дає змогу прогнозувати соціально-економічні явища. Прогнозування є важливим етапом планової роботи.

Під **прогнозуванням** розуміють процес наукового виявлення можливих шляхів і результатів майбутнього розвитку соціально-економічних явищ, оцінку показників, що характеризують ці явища для більш або менш віддаленого майбутнього.

Розрізняють **короткострокові** прогнози (від кількох днів до одного року), **середньострокові** (від одного року до 5 років) і довгострокові прогнози (понад 5 років).

Застосування екстраполяції для прогнозування базується на припущенні, що характер динаміки, тобто певна закономірність (тенденція) зміни досліджуваного явища, яка мала місце для певного періоду часу в минулому, збережеться на обмеженому відрізку в майбутньому. Така екстраполяція справедлива, якщо система розвивається еволюційно в досить стабільних умовах. Чим крупніша система, тим більш імовірно збереження параметрів її зміни, звісно, на невеликий строк.

Користуючись цим методом, слід пам'ятати, що можливості використання отриманих кривих для прогнозування надто обмежені, тому що зміна величини ознаки не є власне функцією часу. Крім того, закономірності і тенденції теперішнього часу не можна механічно переносити на майбутнє.

У зв'язку з цим прогнозуванню має передувати ретельний аналіз комплексу взаємопов'язаних факторів, які в майбутньому будуть визначати тенденцію розвитку досліджуваного соціально-економічного явища.

Принципове значення у встановленні прогнозного рівня мають два питання. Перше стосується проблеми встановлення періоду завчасності (на яку віддаленість), на який можна визначати майбутній рівень ряду. На практиці виходять з такого положення. Якщо досліджуване явище зазнає суттєвих змін, то віддаленість слід брати невеликою (не більше двох-трьох років), якщо ж явище в часі змінюється незначно, то віддаленість прогнозованого рівня можна брати до п'яти років.

Друге питання стосується визначення минулого періоду, за яким повинна встановлюватися основна тенденція розвитку явища. За базу для прогнозування не можна, очевидно, брати короткий період, бо він для даного явища може виявитися не досить типовим через дію випадкових факторів. Недоцільно брати за основу і дуже тривалий період, оскільки умови розвитку явища в часі можуть істотно змінюватися. Отже, потрібно брати оптимальний (якісно однорідний), не дуже довгий і не дуже короткий ряд, для рівнів якого характерні однакові умови розвитку.

Розглянемо методику прогнозування по лінійному тренду на прикладі ряду динаміки урожайності соняшнику (табл. 10.10).

Нагадаємо, що в результаті розв'язання рівняння прямої лінії знайдена така залежність урожайності соняшнику в часі:

$$\tilde{y}_t = a_0 + a_1 t = 10{,}48 + 0{,}53t,$$

де коефіцієнт регресії $a_1 = 0{,}53$ ц/га характеризує середній щорічний приріст урожайності за досліджуваний період. Використаємо знайдене рівняння лінійного тренду для прогнозування урожайності соняшнику на перспективу (2003–2005 рр.). Ряд динаміки обмежений 2002 роком, для якого $t = 15$. Для 2002-2005 рр. t відповідно дорівнює 16, 17 і 18.

Підставимо значення t в рівняння лінійного тренду і одержимо такий прогноз урожайності по роках:

на 2003 р. ($t = 16$) = 10,48 + 0,53 · 16 = 19,0 ц/га;

на 2004 р. ($t = 17$) = 10,48 + 0,53 · 17 = 19,5 ц/га;

на 2005 р. ($t = 18$) = 10,48 + 0,53 · 18 = 20,0 ц/га.

Зобразимо фактичний і вирівняний ряд динаміки урожайності соняшнику графічно (рис. 10.1).

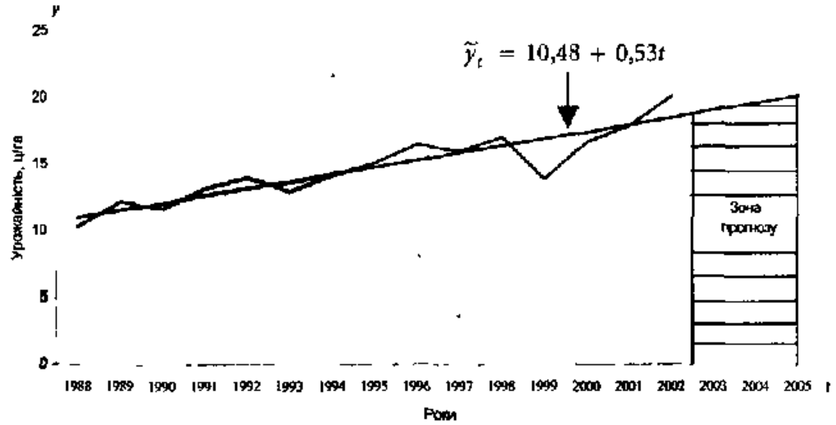


Рис. 10.1. Динаміка і прогнозування урожайності соняшнику в ТОВ району на 2003–2005 рр.

Відмітимо, що за графіком і рівнянням лінійного тренду прогноз урожайності має збігатися.

10.6. Аналіз сезонних коливань

У практиці дослідження динамічних рядів часто доводиться мати справу з аналізом сезонних коливань рівнів рядів.

Сезонними коливаннями називають періодичні внутрішньорічні коливання, зумовлені зміною пори року. Такі коливання спостерігаються в багатьох галузях народного господарства. Особливо вони характерні для сільського господарства, де виробництво продукції значною мірою залежить від природних умов. Продуктивність тварин, використання трудових ресурсів і техніки, переробка сільськогосподарської продукції тощо має явно виражений сезонний характер.

При вивченні сезонних коливань перед статистикою ставляться такі завдання: по-перше, встановити загальну тенденцію зміни досліджуваного явища у часі, по-друге, охарактеризувати ступінь сезонності, по-третє, виявити фактори, що викликають сезонні коливання.

Аналіз сезонних коливань дає змогу дати кількісну оцінку інтенсивності сезонних змін і розробити заходи щодо їх послаблення.

Щоб виявити сезонні коливання, аналізують місячні рівні ряду за один рік або кілька років.

Сезонні коливання в статистиці вимірюють за допомогою розрахунку спеціальних показників – **індексів сезонності**. Показники сезонності у вигляді сезонної хвилі можуть бути розраховані різними способами. Способи розрахунку показників сезонності залежать від характеру основної тенденції ряду динаміки.

При стабільній тенденції в ряді динаміки, в якому внутрішньорічні коливання ознаки відбуваються навколо деякого постійного рівня, показники сезонності визначають як процентне відношення рівнів за кожний місяць до середньомісячного рівня за рік.

Однак місячні рівні за один рік можуть бути нетиповими через вплив випадкових причин. Тому на практиці індекси сезонності визначають за місячними даними за кілька років (три роки і більше). В цьому разі для кожного місяця встановлюють середню величину рівня за кілька років (наприклад, три роки), далі з них розраховують середньомісячний рівень для всього ряду. Після цього кожен середньомісячний рівень порівнюють з середньомісячним річним рівнем за кілька років, а знайдений результат перемножують на сто процентів.

Проведемо аналіз сезонності яйценосності курей в ТОВ (табл. 10.17).

Сезонність яйценосності курей охарактеризуємо за допомогою індексів сезонності – процентне відношення окремих рівнів до середнього рівня даного ряду динаміки.

У зв'язку з тим, що місячні дані одного року внаслідок впливу випадкових факторів можуть бути нетиповими для виявлення тенденції розвитку, обчислимо індекси сезонності не за один рік, а в середньому за три роки.

Для розрахунку індексів сезонності спочатку для кожного місяця обчислимо середню величину яйценосності курей за три роки, а потім з них визначимо середньорічний рівень для трирічча і знайдемо процентне відношення середніх для кожного місяця до середньорічного рівня (індекси сезонності).

Таблиця 10.17

Динаміка яйценосності курей в ТОВ по місяцях

Місяць	Рік			Разом за три роки	У середньому за три роки $\bar{y}_i$	Показники сезонності $I_s = \frac{\bar{y}_i}{\bar{y}_{заг}} \cdot 100\%$
	2000	2001	2002			
1	11	12	12	35	11,7	60,0
2	17	16	17	50	16,7	85,6
3	18	19	18	55	18,3	93,8
4	19	21	18	58	19,3	99,0
5	24	26	29	79	26,3	134,9
6	26	28	27	81	27,0	138,5
7	28	27	25	80	26,7	136,9
8	26	24	24	74	24,7	126,7
9	21	21	19	61	20,3	104,1
10	15	16	15	46	15,3	78,5
11	10	14	14	38	12,7	65,1
12	13	16	16	45	15,0	76,9
Разом	228	240	234	702	19,5	1200,0
У середньому	19,0	20,0	19,5	58,5	19,5	100,0

Виконаємо вказані розрахунки. Визначимо рівні яйценосності курей для кожного місяця за формулою середньої арифметичної простої:

$$\bar{y} = \frac{y_1 + y_2 + ... + y_n}{n} = \frac{\sum y}{n},$$

де y - місячні рівні, n - число місяців.

У січні $\bar{y} = \frac{y_1 + y_2 + y_3}{n} = \frac{11 + 12 + 13}{3} = \frac{35}{3} = 11,7$ шт.

У лютому $\bar{y} = \frac{y_1 + y_2 + y_3}{n} = \frac{17 + 16 + 17}{3} = \frac{50}{3} = 16,7$ шт. і т.д.

За обчисленими середньомісячними рівнями визначимо загальний середній рівень для трьох років (гр. 6, табл. 10.17):

$$\bar{y}_{заг} = \frac{\sum y}{n} = \frac{11,7 + 16,7 + ... + 15,0}{12} = \frac{234}{12} = 19,5 \text{ шт.}$$

Значення загального середнього рівня можна визначити також і за загальними даними за окремі роки:

$$\bar{y}_{заг} = \frac{\sum \bar{y}}{n} = \frac{228 + 240 + 234}{36} = \frac{702}{36} = 19{,}5 \quad шт.$$

або за даними про середні рівні за кожен рік:

$$\bar{y}_{заг} = \frac{\sum \bar{y}}{n} = \frac{19{,}0 + 20{,}0 + 19{,}5}{3} = \frac{58{,}5}{3} = 19{,}5 \quad шт.$$

Обчислимо індекси сезонності яйценосності курей в ТОВ (**гр. 6, табл. 10.17**):

$$у\ січні\ \ I_s = \frac{\bar{y}_{січня}}{\bar{y}_{заг}} \cdot 100\% = \frac{11{,}7}{19{,}5} \cdot 100\% = 60{,}0\%;$$

$$у\ лютому\ \ I_s = \frac{\bar{y}_{лютого}}{\bar{y}_{заг}} \cdot 100\% = \frac{16{,}7}{19{,}5} \cdot 100\% = 85{,}6\% \quad і\ т.д.$$

Оскільки середній індекс сезонності для всіх 12 місяців має дорівнювати 100%, то сума індексів повинна становити 1200.

Зобразимо сезонну хвилю яйценосності курей графічно за допомогою побудови лінійної діаграми (рис. 10.2).

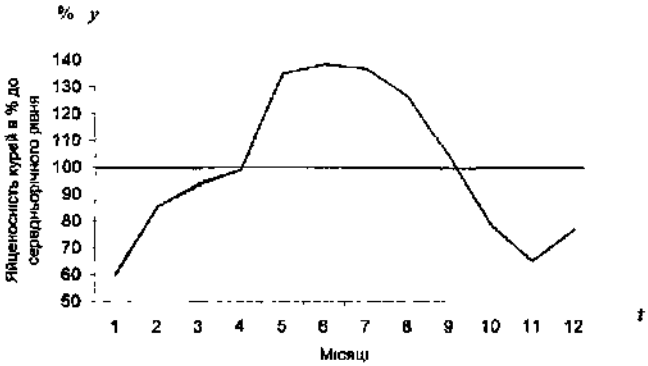


Рис. 10.2. Сезонна хвиля яйценосності курей в ТОВ

З даних таблиці і графіка видно, що сезонність яйценосності курей в ТОВ має чітко виражений характер: найвища яйценосність спостерігається у червні–липні, а найменша – у грудні–січні.

Як характеристики сезонності можуть бути використані показники варіації: середнє квадратичне відхилення і коефіцієнт варіації. Чим менша величина цих показників, тим меншою є сезонність досліджуваного явища.

Завдання для самоконтролю до розділу 10

1. Що таке ряди динаміки і яка їх роль в статистичному аналізі?

2. З яких елементів складається ряд динаміки?

3. Які Ви знаєте види рядів динаміки? Наведіть приклади.

4. Яких умов треба дотримуватися при побудові рядів динаміки?

5. Який вид середніх величин використовується при розрахунку середнього рівня моментного і інтервального ряду динаміки?

6. Що таке базисні і ланцюгові показники динаміки?

7. Назвіть показники динаміки і розкажіть як вони розраховуються.

8. Як розраховується середній темп зростання в рядах динаміки?

9. Назвіть прийоми вирівнювання рядів динаміки.

10. У чому суть прийому приведення рядів динаміки до однієї основи і змикання рядів динаміки?

11. У чому суть прийому і які умови застосування прийому укрупнення періодів?

12. Охарактеризуйте прийоми вирівнювання рядів динаміки способом ковзної середньої, по середньому абсолютному приросту і середньому коефіцієнту зростання

13. Як здійснюється виявлення тенденції способом найменших квадратів?

14. Які правила підбору кривих для вирівнювання рядів динаміки?

15. Що таке інтерполяція і екстраполяція рядів динаміки, їх значення і застосування?

16. Як виконується прогноз на майбутнє за допомогою рівняння тренду?

17. Що таке автокореляція в рядах динаміки? Як вона вимірюється?

18. Як виміряти сезонні коливання в рядах динаміки?

19. За наведеними даними розрахуйте базисні і ланцюгові показники динаміки: абсолютний приріст, темп зростання, процент приросту, абсолютне значення одного процента приросту. Обчисліть середній рівень ряду динаміки, середній абсолютний приріст. Зробіть висновки.

Рік	1998	1999	2000	2001	2002
Вартість основних фондів підприємства, млн грн.	2,5	2,6	2,8	2,6	3,1

20. Тенденція витрат фірми на капітальні вкладення (тис. грн.) за останні 10 років описується трендовим рівнянням $\tilde{y}_t = 47{,}6 + 25{,}4t$. Поясніть економічний зміст параметрів рівняння.

Розділ 11

Індекси

11.1. Поняття про індекси і їх роль в статистико-економічному аналізі

У практиці статистичного аналізу сільськогосподарського виробництва часто доводиться мати справу з явищами, для оцінки зміни яких розглянутих вище середніх і відносних величин недостатньо. Наприклад, агрофірма виконала план виробництва зерна на 107%, овочів – на 102%, молока – на 91%, м'яса – на 101%. Виникає питання, чи виконаний план виробництва в цілому і на скільки? Складність розв'язання цього питання полягає в тому, що наведені вище елементи сукупності (різні види продукції) різнорідні і безпосереднє додавання ні їх обсягів, ні процентів виконання плану не має економічного змісту. Така ж ситуація має місце при оцінці зміни цін, собівартості продукції, продуктивності праці, рентабельності та інших ознак по всіх продуктах рослинництва і тваринництва. Для оцінки загальної зміни подібних явищ у статистиці використовують індекси.

Слово «індекс» (index) в перекладі з латинської означає покажчик, показник. Індекс – відносна величина, що характеризує зміну суспільних явищ у часі, просторі або порівняно з планом. Однак, індексом не слід вважати будь-яку відносну величину порівняння. За допомогою індексів характеризується зміна складних суспільних явищ.

Індексами у статистиці називають складні відносні показники, що характеризують середню зміну сукупності, яка складається з несумірних елементів.

За допомогою індексів вирішуються такі основні завдання:

1) характеристика зміни складного масового соціально-економічного явища в динаміці, просторі та в порівнянні з планом;

2) визначення ступеня впливу окремих факторів на ті чи інші результативні показники;

3) вивчення взаємозв'язку між соціально-економічними явищами;

4) оцінка впливу структурних зрушень на результативні показники.

Як бачимо, за допомогою індексів вирішуються завдання, подібні тим, що розв'язувались з використанням відносних величин, які характеризують зміну окремих сумірних елементів; індекси характеризують зміну складних масових явищ, що складаються з безпосередньо несумірних елементів.

Перераховані завдання доводиться вирішувати на всіх рівнях управління: від окремого підприємства до народного господарства країни в цілому. В аналізі сільськогосподарського виробництва індекси широко використовуються для характеристики змін всіх основних економічних та техніко-економічних показників: обсягів виробництва та реалізації продукції, продуктивності праці, собівартості продукції, урожайності сільськогосподарських культур, продуктивності тварин, використання машин і обладнання тощо.

В індексному методі застосовується певна символіка, тобто система умовних позначень, за допомогою яких будують і записують індекси.

Кожна індексована величина має своє символічне позначення (звичайно у вигляді латинської літери). Кількість продукції позначається через q, ціна за одиницю продукції – через p, собівартість одиниці продукції – через z, затрати праці на одиницю продукції – через t, урожайність – через y, посівна площа – через s і т.д.

Індивідуальний індекс позначається літерою "i" і наділяється підрядковим знаком індексованого показника. Так, i_p означає індивідуальний індекс цін певного виду продукції. Зведений індекс позначається літерою I і також супроводжується підрядковим знаком індексованого показника. Наприклад, I_q – зведений індекс фізичного обсягу продукції, I_z – зведений індекс собівартості продукції і т.д.

Для відображення періодів часу використовують спеціальні позначення, які пишуть знизу символу індексованої величини або величини, що використовується як вага (сумірник). Період, з яким порівнюють, називають **базисним**, а період, який порівнюють – **поточним**, або **звітним**. Дані базисного періоду, позначають підрядковим знаком «0», а звітного – «1». Наприклад, кількість продукції за базисний і звітний період відповідно позначається q_0 і q_1. Для планового рівня застосовується знак «пл.». Так, кількість продукції за планом позначається $q_{пл}$ і т.д.

Індекси розраховуються з точністю до 0,0001. Така точність зумовлена взаємозв'язком індексів.

Індекси показують, у скільки разів (на скільки процентів) рівень звітного періоду вищий (нижчий) за рівень базисного періоду. Якщо індекс більший за одиницю, або більший за 100%, то це свідчить про те, що рівень у звітному періоді підвищився, а якщо індекс менший за одиницю, або менший за 100%, то це свідчить про зменшення рівня у звітному періоді порівняно з базисним періодом.

В статистичному аналізі часто постає завдання – дати узагальнену характеристику зміни сукупності, елементи якої безпосередньо несумірні. Наприклад, необхідно встановити зміну динаміки фізичного обсягу виробленої або реалізованої продукції за кількома різнорідними видами продукції.

Зміну обсягу виробництва або реалізації кожного виду продукції окремо можна визначити за допомогою індивідуального індексу (коефіцієнта зростання), як відношення обсягів виробництва або реалізації продукції в звітному і базисному періодах:

$$i_q = \frac{q_1}{q_0}.$$

Зміну ж усіх видів (або їх більшості) простим співставленням мас продукції охарактеризувати неможливо, бо різнорідні види продукції безпосередньо не підсумовуються. Немає економічного смислу і додавання індивідуальних індексів. Тому для визначення загальної зміни *обсягу продукції і цін одних лише індивідуальних індексів не досить.*

Щоб привести продукцію до порівняного вигляду і розв'язати проблему підсумовування, потрібно всі види продукції звести до єдиного змісту. Оскільки всі види продукції втілюють в собі певні витрати виробництва (затрати живої і минулої праці) і мають вартість, яка виступає у вигляді ціни, то економічно обґрунтованим сумірником різнорідних видів продукції стає ціна одиниці продукції. Іншими сумірниками можуть бути затрати праці на одиницю продукції, собівартість одиниці продукції, вміст поживних речовин тощо.

Найчастіше сумірником різнорідної продукції є ціна. Перемноживши ціну за одиницю продукції на її кількість, дістанемо вартісний (ціновий) вираз кожного виду продукції. Підсумувавши вартість всіх видів продукції, матимемо загальну вартість їх за певний період.

Порівняння вартості продукції звітного і базисного періодів дає загальний індекс вартісного обсягу продукції:

$$I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0},$$

де $\sum q_1p_1$ - вартість всієї продукції за базисний період;

$\sum q_0p_0$ – вартість всієї продукції за звітний період.

Обчислення цього індексу покаже зміну вартісного обсягу продукції внаслідок впливу двох факторів: зміни фізичного обсягу продукції і цін за одиницю продукції.

Щоб визначити зміну вартісного обсягу продукції за рахунок зміни фізичного обсягу продукції, потрібно усунути вплив цін. Цього можна досягти, якщо продукцію звітного і базисного періодів обчислити в однакових (фіксованих, порівнянних) цінах.

Тоді індекс фізичного обсягу продукції буде мати вигляд:

$$I_q = \frac{\sum q_1p_0}{\sum q_0p_0},$$

де $\sum q_1p_0$ – вартість всієї продукції звітного періоду в порівнянних (незмінних) цінах (умовна вартість продукції).

Щоб визначити середню зміну вартісного обсягу продукції внаслідок зміни цін, слід взяти незмінним обсяг продукції. Тоді індекс цін буде мати вигляд:

$$I_p = \frac{\sum q_1p_1}{\sum q_1p_0}.$$

У наведених індексах є дві величини: одну, зміна якої вивчається, називають **індексованою**, а другу постійну, що приводить різнорідні елементи сукупності до порівняного вигляду – **коефіцієнтом сумірності**, або **вагою**. Так, в індексі фізичного обсягу індексованою величиною виступає обсяг продукції (q_0 і q_1), а вагою – незмінна ціна (p_0). В індексі цін індексуються ціни (p_0 і p_1), а незмінним залишається фізичний обсяг продукції (q_1).

Деякі види індексів дістали назву за індексованою величиною. Наприклад, в індексі собівартості продукції індексованою величиною є собівартість, в індексі урожайності – урожайність і т.д. Сумірники (ваги) індексів виражають у вартісних, трудових, умовно-натуральних одиницях вимірювання, а також у вигляді відносних величин структури.

Побудова індексів полягає в зведенні різнорідних елементів складних соціально-економічних явищ до співставного вигляду і порівняння рівнів показників, що відносяться або до різних періодів часу, або до різних територій, або до планового завдання та фактичного його виконання.

За допомогою індексів можна порівнювати не тільки у часі, але і в просторі. Наприклад, можна побудувати індекс фізичного обсягу продукції, в якому буде зіставлено обсяг продукції двох господарств або районів чи груп господарств.

Отримані на основі індексного методу показники використовуються для характеристики розвитку явищ у часі, по території, вивчення структури і взаємозв'язків, виявлення ролі факторів у зміні складних явищ.

Значення індексних показників полягає не тільки в тому, що вони дають відносну характеристику досліджуваних явищ, але і в тому, що на їх основі можна обчислити абсолютні різниці між чисельником і знаменником, які дають досить важливу інформацію про абсолютні прирости обсягів виробництва і реалізації продукції, економії затрат праці і матеріальних засобів та ін.

Методику обчислення та економічний зміст зведених індексів покажемо на прикладі аналізу реалізації продукції в господарстві за два періоди (табл. 11.1).

Таблиця 11.1

Дані для розрахунку індексів вартісного та фізичного обсягів продукції і цін

Продукція	Вихідні дані				Розрахункові дані				
	кількість продукції, тис. ц		ціна реалізації 1 ц, грн.		індивідуальні індекси		виручка від реалізації, тис. грн.		
	базисний період	звітний період	базисний період	звітний період	обсягу продукції	цін	базисний період	звітний період	умовна
	q_0	q_1	p_0	p_1	$i_q = \frac{q_1}{q_0}$	$i_p = \frac{p_1}{p_0}$	q_0p_0	q_1p_1	q_1p_0
Зерно	25,6	30,7	16,83	17,66	1,1992	1,0493	430,8	542,2	516,7
Картопля	18,7	15,0	23,29	22,41	0,8021	0,9622	435,5	336,2	394,4
Молоко	22,5	26,4	21,24	23,08	1,1733	1,0866	477,9	609,3	560,7
М'ясо свиней	1,9	1,7	159,37	163,12	0,8947	1,0235	302,8	277,3	270,9
Разом	–	–	–	–	–	–	1647,0	1765,0	1697,7

Використовуючи індексний метод аналізу і вихідні дані табл. 11.1, встановимо загальну зміну вартісного обсягу продукції за два періоди і вплив на його зміну фізичного обсягу продукції і цін.

Індивідуальні індекси (коефіцієнти зростання) показують, що у звітному періоді порівняно з базисним фізичний обсяг реалізації зерна та молока збільшився відповідно на 19,92 і 17,33%, а картоплі і м'яса свиней скоротився на 19,79 і 10,53%, ціни реалізації по зерну, молоку і м'ясу свиней збільшились відповідно на 4,93%; 8,66 та 2,35%, а по картоплі зменшились на 3,78%.

Визначимо відносну зміну вартісного обсягу реалізованої продукції по всіх видах, для чого розрахуємо індекс вартісного обсягу продукції:

$$I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{1765}{1647} = 1{,}0716, \text{ або } 107{,}16\%.$$

Цей індекс показує, що в динаміці внаслідок дії двох факторів (обсягу продукції і цін) вартісний обсяг збільшився на 7,16%.

Різниця між чисельником і знаменником індексу характеризує *абсолютний приріст вартісного обсягу продукції за два періоди:*

$$\Delta qp = \sum q_1 p_1 - \sum q_0 p_0 = 1765 - 1647 = 118 \text{ тис. грн.}$$

Щоб визначити зміну вартісного обсягу продукції за рахунок зміни тільки фізичного обсягу продукції, обчислимо індекс фізичного обсягу продукції, в якому продукція звітного і базисного періодів оцінена в єдиних цінах:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{1697{,}7}{1647{,}0} = 1{,}0308, \text{ або } 103{,}08\%.$$

Отже, вартісний обсяг реалізованої продукції у звітному періоді порівняно з базисним внаслідок зростання фізичного обсягу продукції збільшився на 3,08%.

Абсолютний приріст вартісного обсягу реалізованої продукції за рахунок зростання фізичного обсягу становить:

$$\Delta q = \sum q_1 p_0 - \sum q_0 p_0 = 1697{,}7 - 1647{,}0 = 50{,}7 \text{ тис. грн.}$$

Вплив зміни цін реалізації на зміну вартісного обсягу продукції встановимо за допомогою індексу цін:

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{1765{,}0}{1697{,}7} = 1{,}0396, \text{ або } 103{,}96\%.$$

Отже, вартісний обсяг реалізованої продукції внаслідок зміни цін збільшився на 3,96%.

Абсолютний приріст вартісного обсягу продукції внаслідок зростання цін реалізації становить:

$$\Delta p = \sum q_1 p_1 - \sum q_1 p_0 = 1765{,}0 - 1697{,}7 = 67{,}3 \text{ тис. грн.}$$

Обчислені індекси і абсолютні прирости пов'язані між собою такою рівністю:

$$I_{qp} = I_q \cdot I_p; \qquad 1{,}0716 = 1{,}0308 \cdot 1{,}0396;$$

$$\Delta_{qp} = \Delta_q + \Delta_p; \qquad 118{,}0 = 50{,}7 + 67{,}3.$$

Отже, вартісний обсяг реалізованої продукції по господарству за два періоди збільшився на 7,16%, в тому числі внаслідок зростання фізичного обсягу продукції на 3,08% і збільшення цін реалізації на 3,96%.

Обчислені агрегатні індекси показують середню зміну вартісного обсягу продукції, а тому аналіз одержаних результатів доповнимо аналізом по окремих видах продукції.

Загальне збільшення фізичного обсягу продукції в середньому на 3,08% пояснюється більш високим зростанням обсягів реалізації зерна на 19,92% і молока на 17,33% порівняно із скороченням обсягів реалізації картоплі на 19,79% і м'яса свиней на 10,53%.

Аналіз індивідуальних індексів цін по окремих видах продукції показує, що, не дивлячись на зниження цін реалізації по картоплі на 3,78%, в середньому ціни підвищились на 3,96% внаслідок їх зростання по зерну на 4,93%, молоку 8,66% і м'ясу свиней на 2,35%.

11.2. Класифікація індексів

У статистичному аналізі для всебічної характеристики розвитку складних соціально-економічних явищ і визначення ролі факторів у формуванні результативних показників використовуються різні форми і види індексів, що викликає необхідність відповідної їх класифікації. Економічні індекси класифікуються за такими ознаками: ступенем охоплення елементів сукупності, формою побудови, базою порівняння, характером ваг, складом явищ і змістом індексованих величин.

За **ступенем охоплення** сукупності індекси поділяють на індивідуальні, групові і загальні.

Індивідуальними називають індекси, що характеризують зміну окремих елементів складного соціально-економічного явища. Їх прикладом можуть бути зміна обсягу виробництва окремих видів продукції (зерна, молока, м'яса і т.д.), цін, собівартості виробництва окремих видів продукції тощо. Якщо індекси охоплюють не всі елементи складного явища, а лише частину, то їх називають **груповими**, або **субіндексами**, наприклад, якщо визначаються зміни фізичного обсягу продукції або цін по рослинництву або тваринництву. Загальні, або **тота-**

льні, індекси характеризують зведені (узагальнюючі) результати спільної зміни всіх одиниць досліджуваної сукупності. Ці індекси охоплюють усі явища, наприклад, підприємство, сільське господарство, народне господарство.

Залежно від **форми побудови** розрізняють агрегатні і середні індекси. Останні поділяють на середні арифметичні і середні гармонічні. Середні індекси – похідні, їх дістають в результаті перетворення агрегатних індексів. Перетворення агрегатного індексу в середній з індивідуальних полягає в підстановці або в чисельник, або в знаменник замість індексованого показника його вираз через відповідний індивідуальний індекс. Якщо така заміна виконана в чисельнику, то агрегатний індекс перетворюється в середній арифметичний, якщо ж в знаменнику – то в середній гармонічний з індивідуальних індексів. Як правило, середній арифметичний індекс застосовується при індексуванні об'ємних показників, а середній гармонічний – при індексуванні якісних показників.

Способи побудови індексів залежать від змісту і методології розрахунку досліджуваних статистичних показників, наявної вихідної інформації, цілей і завдань дослідження.

Індекси, подібні розглянутим вище, називають **агрегатними**, оскільки і чисельник і знаменник являють собою агрегати, з'єднання різнорідних елементів. Свою назву вони дістали від латинського слова «aggrega», що означає «приєдную». Агрегатна форма індексів є основною формою індексів. Вона дістала найбільш широке застосування в статистичній практиці. В агрегатних індексах чисельник і знаменник подані підсумком добутків двох показників, один з яких змінюється, тобто виступає в ролі індексованої величині, а другий залишається незмінним і виступає в ролі сумірника.

За допомогою агрегатних індексів можна обчислити не тільки відносну зміну явища, але й абсолютні розміри цієї зміни. Різниця між чисельником і знаменником індексу характеризує абсолютну зміну складного явища за рахунок індексованої величини.

До розрахунку **середнього арифметичного індексу** удаються в тих випадках, коли за вихідними даними відомі індивідуальні індекси фізичного обсягу продукції ($i_q = q_1 : q_0$) і вартість продукції кожного виду за базисний період ($q_0 p_0$). Тоді загальний індекс фізичного обсягу можна обчислити як середню зважену із індивідуальних індексів. Для цього замінено невідому кількість продукції звітного періоду (q_1) добутком $i_q q_0$. Ця можливість випливає з формули індивідуального індексу фізичного обсягу продукції. Тоді чисельник агрегатної форми індексу фізичного обсягу продукції одержить вираз $\sum q_1 p_0 = \sum i_q q_0 p_0$, а сам індекс фізичного обсягу прийме вигляд:

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}.$$

Одержана формула являє собою середню арифметичну із індивідуальних індексів фізичного обсягу продукції, зважену вартістю продукції базисного періоду.

Наведемо приклад обчислення середнього арифметичного індексу за такими даними (табл. 11.2).

Таблиця 11.2

Дані для розрахунку середнього арифметичного індексу фізичного обсягу продукції

Продукція	Вихідні дані		Розрахункові дані	
	вартість продукції за базисний період, тис. грн	зміна фізичного обсягу у звітному періоді порівняно з базисним	індивідуальні індекси фізичного обсягу продукції	умовна вартість реалізованої продукції, тис. грн.
	$q_0 p_0$	%	i_q	$i_q q_0 p_0$
Зерно	350	+8	1,08	378,0
Картопля	126	– 2	0,98	123,5
Овочі	94	без змін	1,00	94,0
Разом	570	–	–	595,5

Оскільки за вихідними даними відомі індивідуальні індекси фізичного обсягу продукції (i_q) і вартість кожного виду продукції ($q_0 p_0$) за базисний період, середню зміну фізичного обсягу по всіх видах продукції визначимо за формулою середнього арифметичного індексу фізичного обсягу продукції

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{595{,}5}{570{,}0} = 1{,}0447, \text{ або } 104{,}47\%.$$

Отже, в середньому фізичний обсяг продукції у звітному періоді порівняно з базисним збільшився на 4,47%, або на 25,5 тис. грн. ($\sum i_q q_0 p_0 - \sum q_0 p_0 = 595{,}5 - 570{,}0$).

Середній гармонічний індекс являє собою середню гармонічну із індивідуальних індексів і розраховується в тих випадках, коли відомі індивідуальні індекси цін (i_p) і вартість кожного виду продукції за звітний період (q_1p_1), але невідомі дані щодо ціни за одиницю продукції за базисний період (p_0). Для отримання середнього гармонічного індексу цін в знаменнику агрегатного індексу цін $I_p = \frac{\sum q_1p_1}{\sum q_1p_0}$ ціну базисного періоду (p_0) замінено рівним їй відношенням $p_0 = \frac{p_1}{i_p}$. В результаті знаменник агрегатної форми індексу цін одержить вираз $\sum q_1p_0 = \sum \frac{1}{i_p} q_1p_1$, а сам індекс цін буде мати такий вигляд:

$$I_p = \frac{\sum q_1p_1}{\sum \frac{q_1p_1}{i_p}}.$$

Порядок обчислення середнього гармонічного індексу покажемо на прикладі даних табл. 11.3.

Таблиця 11.3

Дані для розрахунку середнього арифметичного індексу фізичного обсягу продукції

Продукція	Вихідні дані		Розрахункові дані	
	вартість продукції за звітний період, тис. грн.	зміна цін у звітному періоді порівняно з базисним	індивідуальні індекси цін	умовна вартість реалізованої продукції, тис. грн.
	q_1p_1	%	i_p	$\frac{q_1p_1}{i_p}$
Зерно	426	+7	1,07	398,1
Картопля	147	– 3	0,97	151,5
Овочі	105	+4	1,04	101,0
Разом	**678**	–	–	**650,6**

Оскільки за вихідними даними відомі індивідуальні індекси цін (i_p) і вартість кожного виду продукції за звітний період (q_1p_1), середню зміну цін по всіх видах продукції визначимо за формулою середнього гармонічного індексу цін

$$I_p = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_p}} = \frac{426 + 147 + 105}{\frac{426}{1{,}07} + \frac{147}{0{,}97} + \frac{105}{1{,}04}} = \frac{678{,}0}{650{,}6} = 1{,}0421,\ \text{або}\ 104{,}21\%.$$

Отже, в середньому ціни реалізації продукції у звітному році порівняно з базисним збільшились на 4,21%, а абсолютний приріст вартісного обсягу продукції внаслідок зростання цін становить 27,4 тис. грн.

$$(\sum q_1 p_1 - \sum \frac{q_1 p_1}{i_p} = =678{,}0 - 650{,}6).$$

Залежно від **бази порівняння** індекси поділяють на: **динамічні** – характеризують відносну зміну складних соціально-економічних явищ у часі; **планові** – використовуються для визначення відносної величини планового завдання і узагальнюючої характеристики ступеня виконання плану; **територіальні** – виражають співвідношення складних масових явищ у просторі (між підприємствами, районами, областями, республіками, країнами і т.п.).

Більш докладно розглянемо динамічні індекси, оскільки вони найчастіше застосовуються в індексному аналізі.

При вивченні розвитку соціально-економічних явищ у динаміці виникає потреба порівнювати дані не за два, а за три періоди. У таких випадках необхідно вибрати базу порівняння. Залежно від бази порівняння розрізняють індекси з постійною (базисні) і змінною базою порівняння (ланцюгові).

В **базисних індексах** всі періоди порівнюються з одним, взятим за базу, а в **ланцюгових** – кожен наступний період порівнюється з безпосередньо попереднім.

Базисні і ланцюгові індекси можуть бути індивідуальними і загальними.

Індивідуальні базисні і ланцюгові індекси являють собою різновидність базисних і ланцюгових відносних величин динаміки. Тому способи обчислення тих й інших показників тотожні.

Обчислення ж загальних базисних і ланцюгових індексів має свої особливості.

За **характером ваг** базисні і ланцюгові індексі класифікують на індекси з постійними і змінними вагами. При обчисленні індексів з постійними вагами за ваги для всього ряду приймаються сумірники будь-якого одного періоду. При обчисленні індексів із змінними вагами за ваги кожен раз беруть сумірники іншого періоду.

Наведемо приклади базисних і ланцюгових індексів з постійними і змінними вагами і покажемо їх взаємозв'язок. Для прикладу використаємо індекс фізичного обсягу продукції.

Базисні індекси з постійними вагами:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \; I_{q_{2/0}} = \frac{\sum q_2 p_0}{\sum q_0 p_0}; \; I_{q_{3/0}} = \frac{\sum q_3 p_0}{\sum q_0 p_0} \text{ і т.д.}$$

Базисні індекси зі змінними вагами:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \; I_{q_{2/0}} = \frac{\sum q_2 p_1}{\sum q_0 p_1}; \; I_{q_{3/0}} = \frac{\sum q_3 p_2}{\sum q_0 p_2} \text{ і т.д.}$$

Ланцюгові індекси з постійними вагами:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \; I_{g_{2/0}} = \frac{\sum q_2 p_0}{\sum q_1 p_0}; \; I_{q_{3/0}} = \frac{\sum q_3 p_0}{\sum q_2 p_0} \text{ і т.д.}$$

Ланцюгові індекси із змінними вагами:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \; I_{q_{2/1}} = \frac{\sum q_2 p_1}{\sum q_1 p_1}; \; I_{q_{3/2}} = \frac{\sum q_3 p_2}{\sum q_2 p_2} \text{ і т.д.}$$

Між ланцюговими і базисними індексами існує такий взаємозв'язок: для індексів з постійними вагами добуток ланцюгових індексів дорівнює базисному індексу крайніх періодів. Наприклад, для індексів фізичного обсягу продукції:

$$I_{q_{1/0}} \cdot I_{q_{2/1}} \cdot I_{q_{3/2}} = I_{q_{3/0}};$$

$$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$$

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_2 p_0}{\sum q_1 p_0} \cdot \frac{\sum q_3 p_0}{\sum q_2 p_0} = \frac{\sum q_3 p_0}{\sum q_0 p_0}.$$

Частка від ділення наступного базисного індексу з постійними вагами на попередній дорівнює відповідному ланцюговому індексу:

$$I_{q_{2/0}} : I_{q_{1/0}} = I_{q_{2/1}};$$

$$\downarrow \quad \downarrow \quad \downarrow$$

$$\frac{\sum q_2 p_0}{\sum q_0 p_0} : \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum q_2 p_0}{\sum q_1 p_0}.$$

Обчислення базисних і ланцюгових індексів розглянемо на такому прикладі (табл. 11.4).

Таблиця 11.4

Динаміка обсягів продукції і цін

Вид продукції	Обсяг реалізації, тис. ц			Ціна реалізації 1 ц продукції, грн.		
	2000 р.	2001 р	2002 р.	2000 р.	2001 р	2002 р.
	q_0	q_1	q_2	p_0	p_1	p_2
A	22	25	26	17	18	20
B	5	6	5	10	9	11
C	17	19	18	5	6	7

Визначимо базисні індекси фізичного обсягу реалізованої продукції з постійними вагами (в цінах 2000 р.):

$$I_1 = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{25 \cdot 17 + 6 \cdot 10 + 19 \cdot 5}{22 \cdot 17 + 5 \cdot 10 + 17 \cdot 5} = \frac{580}{509} = 1{,}1395;$$

$$I_2 = \frac{\sum q_2 p_0}{\sum q_0 p_0} = \frac{26 \cdot 17 + 5 \cdot 10 + 18 \cdot 5}{22 \cdot 17 + 5 \cdot 10 + 17 \cdot 5} = \frac{582}{509} = 1{,}1434.$$

Ланцюгові індекси фізичного обсягу з постійними вагами:

$$I_1 = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{25 \cdot 17 + 6 \cdot 10 + 19 \cdot 5}{22 \cdot 17 + 5 \cdot 10 + 17 \cdot 5} = \frac{580}{509} = 1{,}1395.$$

Перший ланцюговий і перший базисний індекси з постійними вагами завжди дорівнюють один одному:

$$I_2 = \frac{\sum q_2 p_0}{\sum q_1 p_0} = \frac{26 \cdot 17 + 5 \cdot 10 + 18 \cdot 5}{25 \cdot 17 + 6 \cdot 10 + 19 \cdot 5} = \frac{582}{580} = 1{,}0034.$$

Обчислимо базисні і ланцюгові індекси цін із змінними вагами. За змінні ваги в цих індексах візьмемо обсяг реалізованої продукції у звітному періоді.

Базисні індекси із змінними вагами:

$$I_1 = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{25 \cdot 18 + 6 \cdot 9 + 19 \cdot 6}{25 \cdot 17 + 6 \cdot 10 + 19 \cdot 5} = \frac{618}{580} = 1{,}0655;$$

$$I_2 = \frac{\sum q_2 p_2}{\sum q_2 p_0} = \frac{26 \cdot 20 + 5 \cdot 11 + 18 \cdot 7}{26 \cdot 17 + 5 \cdot 10 + 18 \cdot 5} = \frac{701}{582} = 1{,}2045.$$

Ланцюгові індекси із змінними вагами:

$$I_1 = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{25 \cdot 18 + 6 \cdot 9 + 19 \cdot 6}{25 \cdot 17 + 6 \cdot 10 + 19 \cdot 5} = \frac{618}{580} = 1{,}0655;$$

$$I_2 = \frac{\sum q_2 p_2}{\sum q_2 p_1} = \frac{26 \cdot 20 + 5 \cdot 11 + 18 \cdot 7}{26 \cdot 18 + 5 \cdot 9 + 18 \cdot 6} = \frac{701}{621} = 1{,}1288 .$$

За **складом явища** розрізняють індекси постійного і змінного складу. Індекси, в яких змінюється одна величина, називають **індексами постійного складу** (індекси фізичного обсягу продукції, цін, собівартості, продуктивності праці та ін.), а дві і більше величин – **змінного складу** (індекси вартісного обсягу продукції, валового збору, загальних витрат та ін.).

Відношення індексу змінного складу до індексу постійного складу дає **індекс структурних зрушень.**

Індекси змінного складу можна подати у вигляді добутку двох і більше взаємопов'язаних індексів постійного складу. Таке розкладання можна використовувати для аналізу зміни складного явища під впливом певних факторів. Так, наприклад, індекс вартісного обсягу продукції можна подати як добуток індексів фізичного обсягу продукції і цін:

$$\begin{array}{ccccc} I_{qp} & = & I_q & \cdot & I_p; \\ \downarrow & & \downarrow & & \downarrow \\ \frac{\sum q_1 p_1}{\sum q_0 p_0} & = & \frac{\sum q_1 p_0}{\sum q_0 p_0} & \cdot & \frac{\sum q_1 p_1}{\sum q_1 p_0}. \end{array}$$

Індекс валового збору можна подати як добуток двох взаємопов'язаних індексів: урожайності, розміру і структури посівних площ:

$$\begin{array}{ccccc} I_{ys} & = & I_y & \cdot & I_s; \\ \downarrow & & \downarrow & & \downarrow \\ \frac{\sum y_1 s_1}{\sum y_0 s_0} & = & \frac{\sum y_1 s_1}{\sum y_0 s_1} & \cdot & \frac{\sum y_0 s_1}{\sum y_0 s_0}. \end{array}$$

За **змістом індексованих величин** розрізняють індекси об'ємних і якісних показників. **Індекси об'ємних показників** характеризують зміну обсягу явища, числа одиниць сукупності. До цієї групи індексів відносять індекси фізичного обсягу продукції, розміру і структури посівних площ та ін. **Індекси якісних показників** відображають зміну ознак, властивостей одиниць сукупності. До індексів якісних показників відносять індекси цін, собівартості продукції, продуктивності праці, урожайності, продуктивності тварин та ін.

Індекси об'ємних показників (базисні і ланцюгові), як правило, будуються з постійними вагами. Так, при визначенні індексів фізичного обсягу продукції для всіх періодів використовують порівнянні ціни одного з попередніх періодів.

Індекси якісних показників (базисні і ланцюгові) обчислюються переважно із змінними вагами, оскільки в цих індексах застосовуються ваги поточних періодів, а поточний період від індексу до індексу змінюється. Щоб запобігти впливу на величину індексу відмінностей в структурі об'ємного показника, який відіграє роль ваги (наприклад, в структурі посівних площ, стада, затрат праці і т.д.), застосовують індекси, обчислені за тією самою стандартною структурою. У цьому разі ланцюгові і базисні індекси якісних показників визначаються з постійними вагами.

При побудові індексів важливе значення має вибір сумірників (ваг), оскільки індекси, обчислені з різними вагами, дають різні економічні, а інколи й такі, що суперечать дійсності, результати. Тому при виборі ваг потрібно керуватися тим, що індекси мають бути актуальними і являти собою систему взаємопов'язаних показників.

У практиці статистики, як правило, при обчисленні індексів об'ємних показників беруть ваги базисного періоду, а при обчисленні індексів якісних показників – ваги звітного періоду. Крім того, для того щоб добуток індексів об'ємного і якісного показників (наприклад, I_q і I_p) був рівним індексу їх загальної взаємодії (I_{qp}), необхідно, щоб один із індексів був побудований з вагами базисного періоду (I_q), а другий – з вагами звітного періоду (I_p):

11.3. Найважливіші економічні індекси та їх взаємозв'язок

Крім індексів цін, вартісного і фізичного обсягу продукції, в статистиці широке застосування знаходять індекси продуктивності праці, собівартості продукції, урожайності, продуктивності тварин та ін.

Для характеристики зміни продуктивності праці використовуються два індекси продуктивності праці (трудовий і вартісний).

Якщо продуктивність праці виражається показником затрат робочого часу на одиницю продукції, то індивідуальний індекс продуктивності праці будують так:

$$i_t = \frac{t_0}{t_1} = \frac{T_0}{q_0} : \frac{T_1}{q_1}.$$

де t_0 і t_1 – затрати робочого часу на одиницю продукції відповідно в базисному і звітному періодах; $T_0 = q_0\ t_0$ і $T_1 = q_1\ t_1$ – затрати робочого часу на всю продукцію відповідно в базисному і звітному періодах; q_0 і q_1 – кількість виробленої продукції в базисному і звітному періодах.

Агрегатний **індекс продуктивності праці (трудовий)** виражається формулою:

$$I_t = \frac{\sum t_0 q_1}{\sum t_1 q_1},$$

де $\sum t_0 q_1$ – **умовні затрати** робочого часу на всю продукцію звітного періоду;

$\sum t_1 q_1$ – фактичні затрати робочого часу на всю продукцію звітного періоду.

На відміну від наведених вище формул інших агрегатних індексів в цьому індексі базисна величина індексованого показника (t_0) знаходиться в чисельнику, а звітна величина (t_1) – в знаменнику. Це зумовлено тим, що затрати праці на одиницю продукції і продуктивність праці пов'язані оберненою залежністю.

Тому індивідуальний і агрегатний індекс затрат робочого часу (трудомісткості) мають відповідно вигляд:

$$i_{\substack{\text{затрат} \\ \text{роб часу}}} = \frac{t_1}{t_0} \quad \text{і} \quad I_{\substack{\text{затрат} \\ \text{роб часу}}} = \frac{\sum t_1 q_1}{\sum t_0 q_1}.$$

Чим менше затрати робочого часу на виробництво одиниці продукції, тим вищою при інших рівних умовах є продуктивність праці.

Перетворивши агрегатну формулу індексу продуктивності праці заміною $t_0 = i_t q_1$, дістанемо середній арифметичний індекс продуктивності праці:

$$I_t = \frac{\sum i_t q_1 t_1}{\sum q_1 t_1}.$$

Якщо затрати робочого часу на всю продукцію звітного періоду ($t_1\ q_1$) позначити через T_1, то наведена вище формула середнього арифметичного індексу продуктивності праці буде мати вигляд:

$$I_t = \frac{\sum i_t T_1}{\sum T_1}.$$

Трудовий індекс продуктивності праці є індексом постійного складу. Він характеризує зміну продуктивності праці за рахунок одного фактора – зміни затрат робочого часу. Якщо індекс продуктивності праці більше одиниці, то різниця між чисельником і знаменником характеризує економію затрат робочого часу за рахунок підвищення продуктивності праці.

Вартісний індекс продуктивності праці обчислюється за формулою

$$I_{\omega} = \frac{\sum q_1 p_0}{\sum T_1} : \frac{\sum q_0 p_0}{\sum T_0},$$

де q_1 і q_0 - кількість продукції звітного і базисного періодів;

$\sum T_1 = \sum t_1 q_1$ і $\sum T_0 = \sum t_0 q_0$ – затрати робочого часу на всю продукцію у звітному і базисному періодах;

p_0 - порівнянна (незмінна) ціна за одиницю продукції.

В цьому індексі порівнюються середній виробіток (в гривнях) за одиницю робочого часу за звітний і базисний періоди. Вартісний індекс продуктивності праці є індексом змінного складу. Він характеризує зміну продуктивності праці внаслідок зміни двох факторів – затрат робочого часу на виробництво окремих видів продукції і структури (асортименту) продукції.

Вартісний індекс продуктивності праці можна подати як індекс, що характеризує зміну фізичного обсягу продукції, та індекс затрат робочого часу:

$$I_{\omega} = \frac{\sum q_1 p_0}{\sum T_1} : \frac{\sum q_0 p_0}{\sum T_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} : \frac{\sum T_1}{\sum T_0} = I_q : I_t .$$

Розрахунок і аналіз індексів продуктивності праці розглянемо на такому прикладі (*табл.* 11.5).

Співставляючи умовні і фактичні затрати праці на всю продукцію звітного періоду, розрахуємо трудовий індекс продуктивності праці:

$$I_t = \frac{\sum t_0 q_1}{\sum t_1 q_1} = \frac{602{,}5}{572{,}3} = 1{,}0528, \text{ або } 105{,}28\%.$$

Отже, продуктивність праці у звітному періоді порівняно з базисним у середньому підвищилася на 5,28%.

Різниця між чисельником та знаменником індексу $\sum t_0 q_1 - \sum t_1 q_1 =$ $= 602{,}5 - 572{,}3 = 30{,}2$ тис. людино-годин показує економію робочого часу, отриману в результаті зростання продуктивності праці.

Трудовий індекс продуктивності праці може бути розрахований і за формулою середнього арифметичного індексу:

$$I_t = \frac{\sum i_t T_1}{\sum T_1} = \frac{1{,}0417 \cdot 42{,}5 + 0{,}9692 \cdot 84{,}5 + 0{,}9322 \cdot 72{,}6 + 1{,}1125 \cdot 118{,}4 + 1{,}1089 \cdot 143{,}9 + 1{,}0627 \cdot 110{,}4}{572{,}3} =$$

$$= \frac{602{,}5}{572{,}3} = 1{,}05228, \text{ або } 105{,}28\%.$$

Таблиця 11.5

Дані для розрахунку індексів продуктивності праці

Продукція	Вихідні дані					Розрахункові дані					
	кількість продукції, тис. ц		затрати праці на 1 ц, людино-годин		порівнянна ціна 1 ц, грн.	індивідуальні індекси продуктивності праці	затрати праці на всю продукцію, тис. людино-годин			вартість валової продукції, тис. грн	
	базисний період	звітний період	базисний період	звітний період			базисний період	звітний період	умовні	базисний період	звітний період
	q_0	q_1	t_0	t_1	p_0	$i_t = \frac{t_0}{t_1}$	$t_0q_0 = T_0$	$t_1q_1 = T_1$	$t_0q_1 = T_{умов}$	q_0p_0	q_1p_0
Зерно	54	59	0,75	0,72	22,19	1,0417	40,5	42,5	44,2	1198,3	1309,2
Цукрові буряки	125	130	0,63	0,65	5,98	0,9692	78,8	84,5	81,9	747,5	777,4
Картопля	44	41	1,65	1,77	25,27	0,9322	72,6	72,6	67,6	1111,9	1036,1
Овочі	30	36	3,66	3,29	30,29	1,1125	109,8	118,4	131,8	908,7	1090,4
Молоко	26	28	5,70	5,14	30,44	1,1089	148,2	143,9	159,6	791,4	852,3
Приріст великої рогатої худоби	3	4	29,34	27,61	186,31	1,0627	88,0	110,4	117,4	588,9	745,2
Разом	–	–	–	–	–	–	537,9	572,3	602,5	5316,7	5810,6

Таким чином, одержано такий самий результат, що й при визначенні агрегатного індексу продуктивності праці.

Співставляючи між собою середні рівні продуктивності праці (виробництво валової продукції у порівнянних цінах на одну людино-годину) за два періоди (ϖ_1 і ϖ_0), розрахуємо вартісний індекс продуктивності праці:

$$I_\varpi = \frac{\varpi_1}{\varpi_0} = \frac{\sum q_1 p_0}{\sum T_1} : \frac{\sum q_0 p_0}{\sum T_0} = \frac{5810{,}6}{572{,}3} : \frac{5316{,}7}{537{,}9} = \frac{10{,}153}{9{,}884} = 1{,}0272,$$

або 102,72%.

Отже, продуктивність праці у звітному періоді порівняно з базисним підвищилася на 0,269 грн. (10,153 – 9,884), або на 2,72%.

Перетворивши формулу вартісного індексу продуктивності праці, його можна подати як відношення індексів фізичного обсягу продукції і загальних затрат праці:

$$I_\varpi = \frac{\sum q_1 p_0}{\sum T_1} : \frac{\sum q_0 p_0}{\sum T_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} : \frac{\sum T_1}{\sum T_0} = I_q : I_T =$$

$$= \frac{5810{,}6}{5316{,}7} : \frac{572{,}3}{537{,}9} = \frac{1{,}0929}{1{,}0640} = 1{,}0272, \quad \text{або} \quad 102{,}72\%.$$

Отже, збільшення фізичного обсягу продукції на 9,29% було забезпечено приростом загальних затрат праці на 6,40%.

Зміна собівартості виробництва окремого виду продукції за два окремих періоди може бути визначена за допомогою індивідуального індексу:

$$i_z = \frac{z_1}{z_0},$$

де z_1 і z_0 – собівартість одиниці продукції у базисному і звітному періодах, грн.

Для характеристики зміни середньої собівартості у динаміці будують **індекс собівартості змінного складу:**

$$I_{\bar{z}} = \frac{\bar{z}_1}{\bar{z}_0} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0},$$

де $\bar{z}_1$ і $\bar{z}_0$ – середня собівартість одиниці продукції в базисному і звітному періодах; z_1 і z_0 – собівартість одиниці продукції в базисному і звітному періодах; q_1 і q_0 – кількість продукції в базисному і звітному періодах.

Цей індекс показує зміну собівартості виробництва продукції внаслідок впливу двох факторів (зміни собівартості і структури продукції) і є індексом змінного складу.

Щоб усунути вплив зміни структури продукції на динаміку середньої собівартості, необхідно розрахувати для двох періодів середню собівартість при тій самій структурі продукції. Для собівартості це фіксування тієї самої структури дістає відображення в такій формулі:

$$I_z = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_1}{\sum q_1}.$$

Після скорочення на $\sum q_1$ цей індекс набуває вигляду формули агрегатного **індексу собівартості постійного складу:**

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1},$$

де $\sum z_1 q_1$ – фактичні витрати на виробництво всієї продукції у звітному періоді; $\sum z_0 q_1$ – умовні витрати на виробництво всієї продукції у звітному періоді.

Різниця між чисельником і знаменником індексу показує суму економії від зниження собівартості (якщо різниця матиме знак мінус) або суму перевитрат (якщо різниця матиме знак плюс).

Індекс собівартості постійного складу характеризує середню зміну собівартості за рахунок одного фактора – зміни самих витрат на виробництво одиниці продукції і не відображує її зміни внаслідок зміни структури продукції.

Для встановлення ступеня впливу структурних зрушень на зміну середньої собівартості обчислюють **індекс структурних зрушень** – відношення індексу змінного складу до індексу постійного складу:

$$I_{\text{стр}} = I_{\bar{z}} : I_z = \left(\frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} \right) : \frac{\sum z_1 q_1}{\sum z_0 q_1} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0}.$$

Після перетворення дістанемо:

$$I_{\text{стр}} = \frac{\sum z_0 q_1}{\sum z_0 q_0} : \frac{\sum q_1}{\sum q_0}.$$

З цієї формули видно, що при випуску однорідної продукції індекс структурних зрушень можна обчислити як відношення індексу фізичного обсягу продукції, сумірною та базисною собівартістю, до індексу фізичного обсягу тієї самої продукції в натуральних одиницях вимірювання.

Між індексами загальних витрат, собівартості та фізичного обсягу продукції існує такий взаємозв'язок, який випливає з такого положення: взаємозв'язки індексів конкретних економічних явищ зумовлені взаємозв'язками відтворюваних ними явищ. Так, загальні витрати на виробництво продукції (zq) є добутком собівартості (z) на обсяг виробленої продукції (q). Звідси

$$i_{zq} = i_z \cdot i_q \text{ і } I_{zq} = I_z \cdot I_q,$$

що підтверджується такою рівністю:

$$\frac{\sum z_1 q_1}{\sum z_0 q_0} = \frac{\sum z_1 q_1}{\sum z_0 q_1} \cdot \frac{\sum z_0 q_1}{\sum z_0 q_0}.$$

Звідси індекс фізичного обсягу продукції можна дістати як частку від ділення індексу загальних витрат на індекс собівартості виробництва продукції.

Для характеристики ефективності або окупності витрат використовується **індекс середніх витрат на одну гривню** або тисячу гривень продукції:

$$I_{\text{середніх витрат}} = \frac{\sum z_1 q_1}{\sum p_0 q_1} : \frac{\sum z_0 q_0}{\sum p_0 q_0},$$

де p_0 – порівнянна (незмінна) ціна одиниці продукції.

Індекс середніх витрат є індексом змінного складу. Він характеризує зміну середніх витрат за рахунок двох факторів: зміни собівартості і структурних зрушень (асортименту продукції).

Індекс середніх витрат можна розкласти на два взаємопов'язаних індекси: індекс загальних витрат та індекс фізичного обсягу і подати так:

$$I_{\text{середніх витрат}} = \frac{\sum z_1 q_1}{\sum z_0 q_0} : \frac{\sum p_0 q_1}{\sum p_0 q_0} = I_{zq} : I_q.$$

Якщо необхідно визначити ступінь виконання плану по собівартості продукції, розраховують індекс виконання плану по собівартості:

$$I_{\text{викон. плану}} = \frac{\sum z_1 q_1}{\sum z_{пл} q_1}.$$

На величину цього індексу впливає не тільки виконання плану собівартості по кожному виду продукції, але й зсув фактичного асортименту продукції (q_1) порівняно із запланованим, оскільки план по зниженню собівартості звичайно складено за плановим асортиментом:

$$I_{\text{план. завдан.}} = \frac{\sum z_{\text{пл}} q_{\text{пл}}}{\sum z_0 q_{\text{пл}}}.$$

Щоб зсув у асортименті продукції не впливав на показник виконання плану по собівартості, індекс виконання плану слід було б обчислювати з плановими вагами:

$$I_{\text{викл. плану}} = \frac{\sum z_1 q_{\text{пл}}}{\sum z_{\text{пл}} q_{\text{пл}}}.$$

Слід зазначити, що величини індексів продуктивності праці вартісного і середніх витрат залежать не тільки від зміни рівня собівартості і затрат праці на одиницю продукції, але й від зміни складу продукції, тобто співвідношення продуктів з різними витратами на одну гривню продукції або різним її виходом на одиницю часу. Тому індекс середніх витрат на одну гривню і вартісний індекс продуктивності праці слід застосовувати при відсутності істотних відмінностей в складі продукції за два періоди.

Порядок розрахунку індексів собівартості і середніх витрат на одну гривню продукції розглянемо на такому прикладі (табл. 11.6).

Таблиця 11. 6

Дані для розрахунку індексів собівартості і середніх витрат на одну гривню продукції

Продукція	Вихідні дані					Розрахункові дані					
	кількість продукції, тис. ц		затрати праці на 1 ц, грн.		порівнянна ціна 1 ц, грн	індивідуальні індекси собівартості	затрати на продукцію, тис. грн.			вартість валової продукції, тис. грн.	
	базисний період	звітний період	базисний період	звітний період			базисний період	звітний період	умовні	базисний період	звітний період
	q_0	q_1	z_0	z_1	p_0	$i_z = \frac{z_1}{z_0}$	$z_0 q_0$	$z_1 q_1$	$z_0 q_1$	$q_0 p_0$	$q_1 p_0$
Зерно	54	59	7,60	6,77	22,19	0,8908	410,4	399,4	448,4	1198,3	1309,2
Цукрові буряки	125	130	4,35	4,50	5,98	1,0345	543,8	585,0	565,5	747,5	777,4
Картопля	44	41	13,13	14,09	25,27	1,0731	577,7	577,7	538,3	1111,9	1036,1
Овочі	30	36	15,18	12,13	30,29	0,7991	455,4	436,7	546,5	908,7	1090,4
Молоко	26	28	27,44	25,42	30,44	0,9264	713,4	711,8	768,3	791,4	852,3
Приріст великої рогатої худоби	3	4	272,02	247,611	186,31	0,9103	816,1	990,4	1088,1	588,9	745,2
Разом	–	–	–	–	–	–	3516,8	3701,0	3955,1	5316,7	5810,6

Співставляючи фактичні та умовні витрати на виробництво всієї продукції, визначимо агрегатний індекс собівартості:

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1} = \frac{3701{,}0}{3955{,}1} = 0{,}9358\text{, або } 93{,}58\%.$$

Отже, собівартість виробництва продукції у звітному періоді порівняно з базисним в середньому знизилася на 6,42% (93,58 – 100).

Різниця між чисельником і знаменником індексу $\sum z_1 q_1 - \sum z_0 q_1 =$ 3701,0 – 3955,1 = – 254,1 тис. грн. показує економію у витратах за рахунок зниження собівартості продукції.

Знаючи індивідуальні індекси собівартості і фактичні витрати на всю продукцію звітного періоду, загальний індекс собівартості продукції можна розрахувати за формулою середнього гармонічного індексу:

$$I_z = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}} =$$

$$= \frac{3701{,}0}{\frac{399{,}4}{0{,}8908} + \frac{585{,}0}{1{,}0345} + \frac{577{,}7}{1{,}0731} + \frac{436{,}7}{0{,}7991} + \frac{711{,}8}{0{,}9264} + \frac{990{,}4}{0{,}9103}} =$$

$$= \frac{3701{,}0}{3955{,}1} = 0{,}9358,$$

або 93,58%.

Отже, отримано такий самий результат, що й при розрахунку агрегатного індексу собівартості.

Співставляючи між собою середні рівні витрат на одну гривню продукції у звітному і базисному періодах, розрахуємо індекс середніх витрат на одну гривню сукупної продукції:

$$I_{\text{середніх витрат}} = \frac{\sum z_1 q_1}{\sum p_0 q_1} : \frac{\sum z_0 q_0}{\sum p_0 q_0} = \frac{3701{,}0}{5810{,}6} : \frac{3516{,}8}{5316{,}7} = \frac{0{,}6369}{0{,}6614} = 0{,}9629,$$

або 96,29%.

Отже, середні витрати на одну гривню продукції у звітному періоді порівняно з базисним знизилися на 0,0245 грн. (0,6369 – 0,6614), або на 3,71%.

Індекс середніх витрат на одну гривню продукції можна подати як відношення індексів загальних витрат і фізичного обсягу продукції:

$$I_{\text{середніх витрат}} = \frac{\sum z_1q_1}{\sum p_0q_1} : \frac{\sum z_0q_0}{\sum p_0q_0} = \frac{\sum z_1q_1}{\sum z_0q_0} : \frac{\sum z_0q_1}{\sum z_0q_0} =$$

$$= I_{zq} : I_q = \frac{3701{,}0}{3516{,}8} : \frac{5810{,}6}{5316{,}7} = \frac{1{,}0524}{1{,}0929} = 0{,}9629, \quad \text{або} \quad 96{,}29\%.$$

Знайдений індекс показує, що загальні витрати на всю продукцію звітного періоду підвищилися на 5,24% при зростанні фізичного обсягу продукції на 9,29%. Більш високі темпи зростання фізичного обсягу продукції (9,29%) порівняно із зростанням витрат на її виробництво (5,24%) призвели до зниження витрат на одну гривню на 3,71%.

За наявними даними можна також обчислити і перевірити взаємозв'язок між індексами загальних витрат, собівартості і фізичного обсягу продукції за собівартістю:

$$
\begin{array}{ccccc}
I_{zq} & = & I_z & \cdot & I_q; \\
\downarrow & & \downarrow & & \downarrow \\
\dfrac{\sum z_1q_1}{\sum z_0q_0} & = & \dfrac{\sum z_1q_1}{\sum z_0q_1} & \cdot & \dfrac{\sum z_0q_1}{\sum z_0q_0}; \\
\downarrow & & \downarrow & & \downarrow \\
\dfrac{3701{,}0}{3516{,}8} & = & \dfrac{3701{,}0}{3955{,}1} & \cdot & \dfrac{3955{,}1}{3516{,}8}; \\
\downarrow & & \downarrow & & \downarrow \\
1{,}0524 & = & 0{,}9358 & \cdot & 1{,}1246.
\end{array}
$$

Отже, загальні витрати на виробництво продукції у звітному періоді порівняно з базисним зросли в середньому на 5,24%. Це зростання відбулося внаслідок зростання фізичного обсягу продукції на 12,46% при зниженні собівартості продукції на 6,42%.

Аналогічно побудові індексів собівартості будують індекси урожайності і продуктивності тварин. Смисл і зміст індексів урожайності розглядається в параграфі 11.5 даного розділу.

11.4. Територіальні індекси

У статистичному аналізі часто виникає потреба в зіставленні рівнів складних соціально-економічних явищ у просторі: по підприємствах, районах, містах, областях, регіонах, країнах, тобто в обчисленні територіальних індексів. Узагальнюючі показники, які характеризують співвідношення рівнів складних соціально-економічних явищ у просторі, тобто в розрізі територій і об'єктів, називають **територіальними** індексами.

Загальні принципи побудови індексів при територіальних порівняннях багато в чому подібні до вивчення складних статистичних сукупностей у часі. Разом з тим побудова територіальних індексів має свою специфіку і певні труднощі порівняно з динамічними індексами. Якщо, наприклад, потрібно порівняти між собою два райони (район *A* і район *B*), щоб визначити в якому з них вищий рівень цін на ринках, то відразу ж виникає дві проблеми: що взяти за базу порівняння і ваги-сумірники індексу.

Вибір бази порівняння залежить від цілей і завдань дослідження. При порівнянні двох сукупностей будь-яка з них може бути взята за базу порівняння. В разі, якщо один з районів є передовим (низька собівартість, висока продуктивність праці і рентабельність виробництва і т.п.), порівняння з ним може мати рацію. В решті випадків для одержання об'єктивних висновків щодо міри відмінностей територіальні індекси мають бути обчислені як з вагами сукупності, взятої за базу порівняння, так і з вагами порівнюваної сукупності.

При динамічних порівняннях, як уже зазначалося, за ваги в агрегатному індексі цін беруть кількість виробленої продукції у звітному періоді. Але при територіальних порівняннях поняття «звітний період» і «базисний період» мають умовне значення. Якщо порівнювати район *A* з районом *B*, то базою буде рівень цін в районі *B*, а за ваги треба брати кількість продукції в районі *A*.

У цьому випадку індекс цін матиме вигляд:

$$I_p = \frac{\sum p_A q_A}{\sum p_B q_A}.$$

Проте зовсім не обов'язково порівнювати район *A* з районом *B*. На тій самій підставі можна порівнювати район *B* з районом *A*. При такому зіставленні базою буде рівень цін району *A*, а вагою «звітного періоду» – кількість продукції району *B*. Отже, індекс цін повинен мати такий вигляд:

$$I_p = \frac{\sum p_B q_B}{\sum p_A q_B}.$$

Наприклад, за даними табл. 11.7 потрібно визначити, в якому з двох населених пунктів і наскільки вищий рівень цін на ринках.

Таблиця 11.7

Ціни і кількість продукції, проданої на ринках двох населених пунктів

Продукція	Пункт *A*		Пункт *B*	
	продано, т	ціна 1 кг, грн.	продано, т	ціна 1 кг, грн.
	q_A	p_A	q_B	p_B
Капуста	250	0,26	300	0,22
Помідори	60	0,80	50	0,85

Якщо в індексі цін фіксувати ваги того пункту, який порівнюється з іншим, то можна побудувати і обчислити два індекси:

$$I_p = \frac{\sum p_A q_A}{\sum p_B q_A} = \frac{0{,}26 \cdot 250 + 0{,}80 \cdot 60}{0{,}22 \cdot 250 + 0{,}85 \cdot 60} = \frac{113}{106} = 1{,}0660\text{, або } 106{,}60\%,$$

тобто стосовно складу продукції, проданої в пункті *A*, рівень цін в пункті *A* порівняно з пунктом *B* вищий на 6,60%.

Однак, порівнюючи ціни пункту *B* з цінами пункту *A*, дістанемо:

$$I_p = \frac{\sum p_B q_B}{\sum p_A q_B} = \frac{0{,}22 \cdot 300 + 0{,}85 \cdot 50}{0{,}26 \cdot 300 + 0{,}80 \cdot 50} = \frac{108{,}5}{118{,}0} = 0{,}9195\text{, або } 91{,}95\%,$$

тобто стосовно складу продукції, проданої в пункті B, ціни в цьому пункті нижчі, чим в пункті *A* на 8,05%.

Таким чином, в кожному пункті рівень цін виявляється різним, якщо відношення рівнів цін вимірювати стосовно кола продукції порівнюваного пункту.

Такі відмінності в територіальних індексах виникають внаслідок того, що порівнювані сукупності відрізняються за структурою виробництва, посівів, стада, складу працівників і т.д. В цих випадках для одержання об'єктивних висновків і однозначної відповіді потрібно здійснити вирівнювання сукупностей за структурою. В теорії і практиці статистики пропонуються різні способи розв'язання цієї проблеми, в тому числі і **спосіб стандартних ваг**. Цей спосіб полягає в тому, що значення індексованої величини зважується не за вагами якого-небудь одного територіального підрозділу, а за вагами області, економічного району, республіки, країни, в яких знаходяться порівнювані райони.

В нашому прикладі за ваги можна використати кількість проданої продукції в населених пунктах *A* і *B* у цілому ($q = q_A + q_B$):

$$I_p = \frac{\sum p_A q}{\sum p_B q}.$$

Тоді, який би район не був взятий за базу порівняння, результати не будуть заперечувати один одному. Так, в нашому прикладі дістанемо:

$$I_p = \frac{\sum p_A q}{\sum p_B q} = \frac{0{,}22(250 + 300) + 0{,}80(60 + 50)}{0{,}22(250 + 300) + 0{,}85(60 + 50)} = \frac{231{,}0}{214{,}5} = 1{,}0769,$$

або 107,69%;

$$I_p = \frac{\sum p_B q}{\sum p_A q} = \frac{214{,}5}{231{,}0} = 0{,}9286, \text{ або } 92{,}86\%.$$

Отже, стосовно кола продукції, проданої в обох пунктах в цілому, рівень цін в пункті A на 7,69% вищий, чим в пункті B.

В територіальних індексах об'ємних показників за ваги можуть бути взяті середні рівні відповідних якісних показників (цін, собівартості, трудомісткості, урожайності і т.д.), обчислені в середньому по порівнюваних територіальних підрозділах:

$$I_q = \frac{\sum q_A \overline{p}}{\sum q_B \overline{p}},$$

де $\overline{p} = \dfrac{p_A q_A + p_B q_B}{q_A + q_B}$.

Так, в нашому прикладі середня ціна капусти становить: 0,24 грн. $(0{,}26 \cdot 250 + 0{,}22 \cdot 300) : (250 + 300)$, а середня ціна помідорів – 0,82 грн. $(0{,}80 \cdot 60 + 0{,}85 \cdot 50) : (60 + 50)$. Використовуючи ці середні ціни за ваги індексу фізичного обсягу продукції, дістанемо:

$$I_q = \frac{\sum q_A \overline{p}}{\sum q_B \overline{p}} = \frac{250 \cdot 0{,}24 + 60 \cdot 0{,}82}{300 \cdot 0{,}24 + 50 \cdot 0{,}82} = \frac{109{,}2}{113{,}0} = 0{,}9664,$$

або = 96,64%,

тобто обсяг продажу продукції (в порівнянних середніх цінах) в пункті A був на 3,36% меншим, чим в пункті B.

11.5. Індексний аналіз

У статистичному аналізі важлива роль належить індексному методу, який дозволяє у відносному та абсолютному виразі оцінити вплив окремих факторів на результативний показник. В основі індексного методу аналізу лежить прийом розкладання індексів змінного складу, які характеризують зміну загального обсягу явища, на індекси постійного (фіксованого) складу, що його складають.

Добуток індексів об'ємного (кількості продукції, площі посіву та ін.) та якісного (ціни, урожайності та ін.) показників дає індекс змінного складу. Зіставляючи між собою індекси змінного і постійного складу, можна визначити вплив структурних зрушень, оцінка впливу яких має велике значення в індексному аналізі складних явищ.

Можливі дві схеми розкладання індексів змінного складу.

I схема. Спочатку індекс змінного складу розкладається на два індекси постійного складу:

$$I_{\text{загального обсягу}} = I_{\text{обсягу і структури}} \cdot I_{\text{якісного показника}};$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\frac{\sum q_1 x_1}{\sum q_0 x_0} = \frac{\sum q_1 x_0}{\sum q_0 x_0} \cdot \frac{\sum q_1 x_1}{\sum q_1 x_0},$$

де x – якісний показник; q – об'ємний показник.

Потім індекс обсягу і структури розкладають на:

$$I_{\text{обсягу і структури}} = I_{\text{обсягу}} \cdot I_{\text{структури}};$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\frac{\sum q_1 x_0}{\sum q_0 x_0} = \frac{\sum q_1}{\sum q_0} \cdot \left(\frac{\sum q_1 x_0}{\sum q_0 x_0} : \frac{\sum q_1}{\sum q_0}\right)$$

Внаслідок двох етапів розкладання матимемо:

$$I_{\text{загально обсягу}} = I_{\text{обсягу}} \cdot I_{\text{структури}} \cdot I_{\text{якісного показника}}.$$

II схема. Спочатку індекс змінного складу розглядається як добуток індексів обсягу і середнього рівня якісного показника:

$$I_{\text{загального обсягу}} = I_{\text{обсягу}} \cdot I_{\substack{\text{середнього рівня} \\ \text{якісного показника}}};$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\frac{\sum q_1 x_1}{\sum q_1 x_0} = \frac{\sum q_1}{\sum q_0} \cdot \frac{\bar{x}_1}{x_0}.$$

Потім індекс середнього рівня якісного показника знаходять як добуток індексів якісного показника і структури:

$$\underset{\text{якісного показника}}{I_{\text{середнього рівня}}} = \underset{\text{показника}}{I_{\text{якісного}}} \cdot I_{\text{структури}};$$

$$\frac{\bar{x}_1}{\bar{x}_0} = \frac{\bar{x}_1}{\bar{x}_{\text{умов}}} \cdot \frac{\bar{x}_{\text{умов}}}{\bar{x}_0},$$

де $\bar{x}_{\text{умов}} = \dfrac{\sum x_0 q_1}{\sum q_1}$.

Внаслідок двох етапів розкладання матимемо ті самі індекси, що й за I схемою.

$$I_{\text{загального обсягу}} = I_{\text{обсягу}} \cdot \underset{\text{показника}}{I_{\text{якісного}}} \cdot I_{\text{структури}}.$$

Покажемо практичне застосування зазначених схем аналізу на прикладі індексного методу аналізу валового збору і середньої урожайності зернових культур (табл. 11.8).

Таблиця 11.8

Площі посіву, урожайність і валовий збір зернових культур в ТОВ за 2001–2002 рр.

Культури	Посівна площа, га		Урожайність, ц/га		Валовий збір, тис. ц		
	2001 р	2002 р	2001 р	2002 р.	2001 р.	2002 р	умовний
	S_0	S_1	y_0	y_1	y_0S_0	y_1S_1	y_0S_1
Озима пшениця	1028	1100	38,7	44,6	39,8	49,1	42,6
Ячмінь ярий	260	234	29,4	32,5	7,6	7,6	6,9
Кукурудза	219	273	50,3	55,6	11,0	15,2	13,7
Горох	112	95	21,7	20,9	2,4	2,0	2,1
Овес	84	68	24,3	22,7	2,0	1,5	1,6
Інші культури	47	50	21,0	17,5	1,0	0,9	1,0
Разом	**1750**	**1820**	**36,46**	**41,92**	**63,8**	**76,3**	**67,9**

За вихідними даними табл. 11.8 потрібно, використовуючи індексний метод аналізу, визначити загальну зміну валового збору зернових культур за два роки і встановити залежність валового збору від розміру і структури посівних площ, середньої урожайності і урожайності окремих культур. Перевірити взаємозв'язок між індексами.

Обчислимо індекс валового збору:

$$I_{yS} = \frac{\sum y_1 S_1}{\sum y_0 S_0} = \frac{76,3}{63,8} = 1,1959, \text{ або } 119,59\%.$$

Індекс показує, що валовий збір зерна в 2002 р. порівняно з 2001 р. збільшився на 19,59%.

Для визначення ступеня впливу на валовий збір окремих факторів спочатку розкладемо індекс змінного складу (індекс валового збору) на два індекси постійного складу (урожайності та розміру і структури посівних площ), а потім індекс розміру і структури посівних площ на індекс розміру і індекс структури посівних площ.

Отже, індекс валового збору можна подати як добуток таких трьох індексів:

$$I_{yS} = I_y \cdot I_S \cdot I_{\substack{\text{структури} \\ \text{посівних площ}}}.$$

Застосуємо два прийоми розкладу цього індексу.

І прийом аналізу. Подамо індекс валового збору **як добуток двох** індексів: **урожайності** окремих культур і розміру і **структури посівних площ:**

$$\begin{array}{ccccc} I_{yS} & = & I_y & \cdot & I_{\text{розміру і структури}}; \\ \downarrow & & \downarrow & & \downarrow \\ \dfrac{\sum y_1S_1}{\sum y_0S_0} & = & \dfrac{\sum y_1S_1}{\sum y_0S_1} & \cdot & \dfrac{\sum y_0S_1}{\sum y_0S_0}. \end{array}$$

Розкладемо індекс розміру і структури посівних **площ на індекс** розміру та індекс структури посівних площ:

$$I_{\text{розміри і структури}} = I_{\text{розміру}} \cdot I_{\text{структури}};$$

$$I_{\text{структури}} = \frac{I_{\text{розміру і структури}}}{I_{\text{розміру}}}.$$

Визначимо намічені індекси.

Індекс урожайності окремих культур (постійного складу):

$$I_y = \frac{\sum y_1S_1}{\sum y_0S_1} = \frac{76{,}3}{67{,}9} = 1{,}1237, \text{ або } 112{,}37\%.$$

Індекс розміру і структури посівних площ:

$$I_{\substack{\text{розміру і} \\ \text{структури}}} = \frac{\sum y_0S_1}{\sum y_0S_0} = \frac{67{,}9}{63{,}8} = 1{,}0643, \text{ або } 106{,}43\%.$$

Індекс розміру посівних площ:

$$I_S = \frac{\sum S_1}{\sum S_0} = \frac{1820}{1750} = 1{,}0400, \text{ або } 104{,}00\%.$$

Індекс структури посівних площ:

$$I_{\text{структури}} = \frac{I_{\text{розміру і структури}}}{I_{\text{розміру}}} = \frac{\sum y_0 S_1}{\sum y_0 S_0} : \frac{\sum S_1}{\sum S_0} =$$

$$= \frac{67{,}9}{63{,}8} : \frac{1820}{1750} = \frac{1{,}0643}{1{,}0400} = 1{,}0234, \quad \textit{або} \quad 102{,}34\%.$$

Перевіримо взаємозв'язок індексів і правильність розрахунків:

$$I_{yS} = I_y \cdot I_S \cdot I_{\text{структури}};$$

$$1{,}1959 = 1{,}1237 \cdot 1{,}0400 \cdot 1{,}0234.$$

Індекси показують, що збільшення валового збору зерна на 19,59% головним чином зумовлено підвищенням урожайності окремих культур на 12,37%, розширенням посівних площ на 4,00% і внаслідок удосконалення структури посівних площ на 2,34%. Удосконалення структури посівних площ означає збільшення в 2002 р. порівняно з 2001 р. в загальній посівній площі зернових культур площі культур з вищою урожайністю.

II прийом аналізу. У зв'язку з тим, що посівні площі однорідних культур порівнянні і їх можна підсумовувати, валовий збір зернових культур можна подати як добуток середньої урожайності на посівну площу:

2001 р. $\sum y_0 S_0 = \bar{y}_0 \sum S_0$;

2002 р. $\sum y_1 S_1 = \bar{y}_1 \sum S_1$;

умовна $\sum y_0 S_1 = \bar{y}_{\text{умов}} \sum S_1$.

Виходячи з цього, індекс урожайності і індекс структури посівних площ можуть бути розраховані за формулами не агрегатних індексів, а індексів середніх рівнів.

Поділивши індекс урожайності постійного складу на $\dfrac{\sum S_1}{\sum S_1}$, отримаємо індекс середніх рівнів урожайності в 2002 р. і умовну

$$I_y = \frac{\sum y_1 S_1}{\sum y_0 S_1} : \frac{\sum S_1}{\sum S_1} = \frac{\sum y_1 S_1}{\sum S_1} : \frac{\sum y_0 S_1}{\sum S_1} = \frac{\bar{y}_1}{\bar{y}_{\text{умов}}},$$

де $\bar{y}_1$ і $\bar{y}_{\text{умов}}$ – відповідно середня урожайність в 2002 р. і умовна.

Індекс структури посівних площ через середні рівні може бути поданий в такому вигляді:

$$I_{\text{структура}} = I_{\text{розміру і структури}} : I_{\text{розміру}} =$$

$$= \frac{\sum y_0 S_1}{\sum y_0 S_0} : \frac{\sum S_1}{\sum S_0} = \frac{\sum y_0 S_1}{\sum S_1} : \frac{\sum y_0 S_0}{\sum S_0} = \frac{\bar{y}_{\text{умов}}}{\bar{y}_0},$$

де $\bar{y}_0$ середня урожайність в 2001 р.

Відмінність урожайності $\bar{y}_{\text{умов}}$ і $\bar{y}_0$ зумовлена тільки структурою посівних площ, оскільки урожайність по окремих культурах є фіксованою (за базисним роком). Тоді індекс валового збору можна розкласти таким чином:

$$\begin{array}{ccccccc} I_{yS} & = & I_y & \cdot & I_{\text{стр}} & \cdot & I_S; \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \dfrac{\sum y_1 S_1}{\sum y_0 S_0} & = & \dfrac{\bar{y}_1}{\bar{y}_{\text{умов}}} & \cdot & \dfrac{\bar{y}_{\text{умов}}}{\bar{y}_0} & \cdot & \dfrac{\sum S_1}{\sum S_0}. \end{array}$$

Визначимо середню урожайність за 2001 р., 2002 р. і умовну:

$$\bar{y}_0 = \frac{\sum y_0 S_0}{\sum S_0} = \frac{63{,}8}{1{,}750} = 36{,}46 \text{ ц/га};$$

$$\bar{y}_1 = \frac{\sum y_1 S_1}{\sum S_1} = \frac{76{,}3}{1{,}820} = 41{,}92 \text{ ц/га};$$

$$\bar{y}_{\text{умов}} = \frac{\sum y_0 S_1}{\sum S_1} = \frac{67{,}9}{1{,}820} = 37{,}31 \text{ ц/га}.$$

Обчислимо необхідні індекси:

$$\begin{array}{ccccccc} I_{yS} & = & I_y & \cdot & I_{\text{умов}} & \cdot & I_S; \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \dfrac{\sum y_1 S_1}{\sum y_0 S_0} & = & \dfrac{\bar{y}_1}{\bar{y}_{\text{умов}}} & \cdot & \dfrac{\bar{y}_{\text{умов}}}{\bar{y}_0} & \cdot & \dfrac{\sum S_1}{\sum S_0}; \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \dfrac{76{,}3}{63{,}8} & = & \dfrac{41{,}92}{37{,}31} & \cdot & \dfrac{37{,}31}{36{,}46} & \cdot & \dfrac{1820}{1750}; \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ 1{,}1959 & = & 1{,}1237 & \cdot & 1{,}0234 & \cdot & 1{,}0400. \end{array}$$

Таким чином, маємо ті самі індекси, що й у першому прийомі аналізу.

Наявність обчислених індексів дає змогу провести аналіз зміни середньої урожайності, яка залежить безпосередньо від двох факторів: урожайності окремих культур і структури посівних площ:

$$I_{\bar{y}} = I_y \cdot I_{\text{структури}};$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$\frac{\bar{y}_1}{\bar{y}_0} = \frac{\bar{y}_1}{\bar{y}_{\text{умов}}} \cdot \frac{\bar{y}_{\text{умов}}}{\bar{y}_0};$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$\frac{41,92}{36,46} = \frac{41,92}{37,31} \cdot \frac{37,31}{36,46};$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$1,1498 = 1,1237 \cdot 1,0234.$$

Індекси показують, що середня урожайність в 2002 р. порівняно з 2001 р. підвищилась на 14,98% здебільшого внаслідок зростання урожайності окремих культур на 12,37%, а також внаслідок удосконалення структури посівних площ на 2,34%.

Завдання для самоконтролю до розділу 11

1. Що в статистиці називається індексом?
2. Яка роль індексного методу аналізу в економічних дослідженнях?
3. Які завдання вирішуються за допомогою індексів?
4. Що таке індексована величина?
5. Які розрізняють види індексів?
6. Які індекси називають індивідуальними, а які загальними? Що вони характеризують?
7. У чому суть побудови агрегатних індексів?
8. Як будуються середні індекси з індивідуальних?
9. Назвіть види індексів: за ступенем охоплення елементів сукупності, базою порівняння, способом розрахунку, характером індексованих величин, ваг і коефіцієнтів сумірництва.
10. Обґрунтуйте вибір ваг і коефіцієнтів сумірництва в індексах.
11. Розкрийте зміст, види і способи розрахунку індексів фізичного обсягу, цін, продуктивності праці і собівартості продукції.

12. Який зв'язок між ланцюговими і базисними індексами?

13. Наведіть приклади взаємопов'язаних індексів.

14. Що характеризують індекси структурних зрушень і як вони розраховуються?

15. Що характеризує різниця між чисельником і знаменником в агрегатних індексах?

16. Дайте визначення і наведіть формули територіальних індексів.

17. За наведеними даними розрахуйте індекси цін, вартісного і фізичного обсягу продукції. Обчисліть абсолютні прирости вартісного обсягу продукції за рахунок зміни цін і фізичного обсягу продукції. Використовуючи взаємозв'язок індексів і абсолютних приростів, перевірте правильність визначених показників. Зробіть висновки щодо знайдених результатів.

Вид продукції	Кількість продукції, тис. шт.		Ціна одиниці продукції, грн.	
	базисний рік	звітний рік	базисний рік	звітний рік
А	20	25	40	42
Б	6	5	315	292
В	48	53	73	84

18. Як у середньому змінилися ціни на промислові товари у звітному періоді порівняно з базисним, якщо фізичний обсяг продукції збільшився на 6,8%, а товарооборот – на 9,5%.

Розділ 12

Статистичні графіки

12.1. Роль і значення графічного методу

Поряд з таблицями для характеристики результатів статистичного зведення і обробки масових даних широко застосовують статистичні графіки.

Статистичним графіком називають наочне масштабне зображення статистичних даних за допомогою геометричних ліній, точок, знаків, фігур, географічних картосхем та інших графічних засобів.

Графічний метод настільки міцно увійшов в арсенал засобів наукового узагальнення і в методику наукових досліджень, що сучасну науку неможливо собі уявити без його застосування. Особливо велика роль цього методу в статистичних дослідженнях, де вивчаються складні взаємозв'язки, тенденції, закономірності соціально-економічних явищ і процесів в динаміці і просторі.

Застосування графічного методу у вивченні масових соціально-економічних явищ досить різнопланове. Так, графіки застосовують для характеристики змін загальних явищ і процесів у часі, вивчення структури явищ, порівняння, контролю виконання плану, дослідження взаємозв'язків між результативними і факторними ознаками, зображення розміщення явищ у просторі, ступеня розповсюдженості по території тих чи інших явищ, міжнародних порівнянь і зіставлень та в інших випадках.

Особливо велика роль графіків у пропаганді передового досвіду, прогресивних тенденцій, закономірностей, нових технологій, наукових досягнень та ін.

В ряді випадків графіки є незамінним засобом аналізу, дослідження і виявлення взаємозв'язків, закономірностей і тенденцій суспільних явищ (наприклад, в кореляційному аналізі і динамічних рядах).

Статистичні графіки застосовуються для того, щоб зробити статистичні матеріали наочними, доступними, зрозумілими і такими, які б сприяли кращому їх аналізу. Завдання полягає в тому, щоб у кожному випадку вибрати найкраще графічне зображення, яке б відповідало характеру величин і більш повно розкривало їх зміст.

Графічне зображення статистичних даних здійснюється шляхом використання геометричних фігур, точок, ліній та інших символічних образів. Числові значення статистичних величин переводяться в графічні образи за допомогою масштабу. Вміле розміщення графічних зображень створює діаграму, на якій статистичні дані представлені в наочній формі, яка дає уявлення про загальні закономірності і тенденції розвитку досліджуваних явищ. При правильній побудові графіки стають виразнішими, доступнішими, сприяють кращому аналізу статистичних показників, їх узагальненню і вивченню.

Графічне зображення допомагає глибше і наочніше охарактеризувати багато статистичних показників, полегшує сприйняття і запам'ятовування певних фактів.

Графіки є найефективнішою формою відображення даних з точки зору їх сприйняття. Вони незамінні у випадку необхідності одночасного огляду декількох величин у часі або у просторі. Графічне зображення дає змогу одним поглядом охопити як всю сукупність явищ в цілому, так і окремі її частини, скласти цілісне уявлення про досліджувані явища.

Внаслідок цього певна інформація за допомогою графіків може бути засвоєна незрівнянно швидше, ніж будь-якими іншими способами, а самі графіки виявляються більш наочними, ніж статистичні таблиці, продовженням і доповненням яких вони є.

Статистичні графіки можуть дати нові знання про предмет вивчення, які у вихідному цифровому матеріалі безпосередньо не виявляються. Виявлення закономірностей, які притаманні тим чи іншим явищам, факторів, які їх визначають, диференціації явищ у часі та просторі – завдання, які ефективно вирішуються з використанням графічного методу.

Графіки у виразній, доступній, лаконічній і компактній формі дозволяють наочно зображати різні статистичні показники, допомагаючи досліднику виявляти їх рівні, співвідношення, зміну тенденцій, склад і структуру складних сукупностей.

Пізнавальна цінність статистичних графіків пояснюється їх здатністю відображати реальну дійсність у простому, ясному і наочному вигляді. Візуальна інтерпретація об'єктивних статистичних показників дає змогу полегшити пізнання предмету дослідження, робить його більш дохідливим.

Наочно зображуючи статистичні дані, полегшуючи їх сприйняття, графіки допомагають виявити найбільш характерні співвідношення і зв'язки явищ, виявити основні тенденції, закономірності розвитку досліджуваних явищ. Цим пояснюється широке використання графіків для популяризації статистичних даних.

Специфічною особливістю графіків є їх лаконічність, простота кодування інформації та однозначність тлумачення записів у символічній формі. До окремих особливостей статистичних графічних зображень належать також їх виразність, дохідливість, універсальність, доступність для огляду та ін.

Слід нагадати, що побудова графіка виправдана, якщо він дає будь-які переваги порівняно з цифрами, які зведені в ряди або таблиці. Побудований графік повинен бути цікавим за змістом, доцільним за формою, економним за використанням графічних засобів, простим і зрозумілим для читача, добре виконаним технічно. Вдало побудовані графіки здатні викликати інтерес читача і привернути увагу до зображуваних статистичних даних; вони полегшують запам'ятовування і більш швидке розуміння суті у співвідношенні зображуваних цифр.

12.2. Основні елементи графіка. Правила побудови статистичних графіків

Основні елементи графіка такі: поле графіка, геометричні знаки, просторові орієнтири, масштаб, експлікація графіка.

Поле графіка – простір, в якому розміщуються геометричні знаки, що утворюють графік. Він характеризується форматом і співвідношенням сторін.

Розмір графіка повинен відповідати його призначенню. Для демонстрації на лекції або доповіді застосовуються графіки великих форматів, для ілюстрації наукового звіту або для розміщення в книзі, статті, курсовому проекті (роботі) – невеликі графіки.

Суттєвим фактором забезпечення найкращого зорового сприйняття відображуваних статистичних даних є вибір пропорцій співвідношення сторін графіка. Співвідношення сторін графіка визначається законами геометричної гармонії і вимогами забезпечення неспотвореного зорового сприйняття графічного образу. Як показує практика побудови і аналізу графіків, найбільш зручні формати із співвідношенням сторін (ординат і абсцис) від 1 : 1,3 до 1 : 1,5. Водночас це не заперечує можливості застосування квадратної форми графіків, яка в окремих випадках є дуже зручною формою відображення статистичних даних.

Геометричні знаки – це сукупність геометричних чи інших графічних знаків, за допомогою яких відображаються статистичні дані і створюється графічний образ. Це точки, прямі і криві лінії та їх відрізки, площини (кола, квадрати та ін.), об'ємні фігури (куби, кулі та ін.), геометричні фігури (знаки – символи, зображення предметів). Геометричні знаки становлять основу графіка, його мову. Залежно від типу геометричних знаків графіки поділяють на лінійні, точкові, стовпчикові, стрічкові, квадратні, крутові, секторні, фігурні та ін.

Важливим моментом побудови графіків є вибір графічного знаку. Вибір графічного знаку визначається характером вихідної інформації, а також основною метою, яка закладена в даний графік. Вдалий його вибір сприяє максимальному досягненню мети графіка і найбільш виразному зображенню статистичних даних. Так, наприклад, якщо метою дослідження є вивчення обсягу будь-якого виду продукції в динаміці, то вихідні дані можуть бути зображені як у вигляді стовпчикової, крутової, квадратної та інших діаграм, так і за допомогою лінійної діаграми. Для відображення обсягу виробництва продукції доцільно використати площинну (стовпчикову, крутову, квадратну та ін.) діаграму, а для відображення динаміки – лінійну діаграму.

Просторові орієнтири визначають розташування геометричних знаків у полі графіка. Вони задаються у вигляді координатних сіток (діаграми) або контурних ліній (картограми). Координатна сітка створюється перетином ліній, які проходять через поділки горизонтальної та вертикальної шкали. Для побудови графіка, як правило, використовується система прямокутних (декартових) координат, зокрема права верхня частина координатного поля, але нерідко зустрічаються графіки, які побудовані за принципом полярних координат (крутові, секторні, радіальні та інші діаграми). Криволінійні контурні лінії застосовують в статистичних картах (картограмах, картодіаграмах) як засіб просторової орієнтації.

На горизонтальній шкалі (вісь абсцис) прямокутних діаграм, як правило, відкладають незалежні змінні (часові відрізки, періоди, об'єкти та ін.), на вертикальній (вісь ординат) – залежні змінні (наприклад, значення результативних показників).

На координатній сітці графіка обов'язково повинна бути вказана основна горизонтальна (нульова) лінія (вісь абсцис). Для наочності її виділяють потовщеною лінією. Якщо рівні відображуваних явищ такі, що основна частина координатної сітки залишається невикористаною, то на шкалі робиться розрив ($\wr\wr$, $\approx$), який виключає непотрібну частину сітки, але з обов'язковою вказівкою нульової лінії. Це дасть змогу рівномірніше заповнити поле графіка. Невключення нуля у вертикальну шкалу є поширеною помилкою, яка спотворює зображення. Це може призвести до неправильного висновку.

Масштабні орієнтири статистичних графіків включають масштаб і масштабні шкали. **Масштабом графіка** називають умовну міру переведення числової величини в графічну. Його звичайно виражають довжиною відрізка, прийнятого за одиницю зображуваної статистичної величини. Наприклад, 1 см на графіку становить 10 га посівної площі. Масштаб може бути показаний або масштабним відрізком або масштабною шкалою. Числове значення масштабу краще вказувати тільки на відмітках, що відповідають круглим числам. Усі проміжні відмітки читають шляхом відліку від найближчого числа, позначеного на масштабній шкалі.

Отримання оптимальної пропорції досягається підбором (пробна побудова кількох варіантів) або досвідченістю упорядника діаграми.

Вибираючи масштаб, слід виходити з того, щоб усі статистичні дані, які потрібно нанести на графік, розмістилися на полі графіка. На вертикальній шкалі графіка обов'язково має бути нульова відмітка.

Вертикальну і горизонтальну шкали слід будувати так, щоб нульове значення було обов'язково на графіку. Якщо ж такі шкали побудувати неможливо або недоцільно, слід дати розрив цих шкал. Такий розрив припускається при збереженні змісту графіка.

При виборі масштабу довжину шкали ділять на різницю крайніх величин явищ, що зображаються. Припустимо, довжина шкали дорівнює 10 см, мінімальне значення явища, що зображується, дорівнює 20 га, а максимальне – 120 га, тоді масштаб становить 10 : (120 – 20) = 0,1 см, тобто 0,1 см на масштабній шкалі буде відповідати одиниці даного явища.

Одним з основних елементів графіка є **масштабна шкала** графіка, тобто лінія, окремі крапки чи риски якої можуть бути прочитані як певні числа.

Масштабні шкали, як правило, розміщуються зліва і знизу графіка. Для побудови шкал рекомендується користуватися міліметровим папером з готовою сіткою. На шкалах повинен розміщуватися весь діапазон зображуваних цифрових даних, звичайно з деякими заокругленнями. Якщо, наприклад, максимальна величина урожайності в досліджуваній сукупності становить 48,5 ц/га, то очевидно, що на шкалі мають бути поділки, що містять 49 або 50 ц/га. Тому останнє число на шкалі має дещо перевищувати максимальний рівень ознаки.

Масштабна шкала складається з трьох елементів: 1) ліній, які є носієм чи опорою шкали; 2) поділок або позначок шкали (точки або риски, які розміщені в певному порядку на носії шкали); 3) цифрових позначень чисел, що відповідають певним точкам або рискам.

Носіями шкали можуть бути пряма лінія (осі координат) або крива лінія (коло, дуга).

Довжину відрізка між двома сусідніми поділками називають **графічним інтервалом**, а різницю між числовими значеннями цих поділок – **числовим інтервалом**.

Масштабні шкали можуть бути прямолінійними, криволінійними, неперервними, перервними, рівномірними і нерівномірними.

Прямолінійними називають шкали, в яких пряма лінія поділена на сантиметри і міліметри, **криволінійними**, в яких крива лінія (коло) поділена на 360^0.

Неперервна шкала застосовується для величин, що безперервно змінюються (всім точкам відповідає певне число, а усі проміжні значення можуть бути інтерпольовані). **Перервна** шкала – шкала з величинами, проміжне значення яких не інтерполюється (наприклад, якщо поділки шкали представлені річними даними, то точка між двома роками нічого не означає, оскільки масштаб не передбачав місячних даних).

Рівномірною (арифметичною) називається шкала, в якій рівним відрізкам (поділкам) на шкалі відповідають рівні числові значення. На рівномірній шкалі графічні матеріали пропорційні абсолютним розмірам статистичних показників. Так, якщо значення показника зростає у два рази, то відрізок, що її відображає, повинен відповідно збільшуватись у два рази. Такі шкали мають переважне застосування в статистичних графіках.

Шкала, на якій рівним графічним відрізкам відповідають нерівні числові значення, називають **нерівномірною**. Прикладом нерівномірної шкали може бути логарифмічна шкала, в якій рівним графічним відрізкам відповідають не рівні абсолютні числа, а рівні їх відношення (логарифми).

При побудові логарифмічної шкали слід виконати такі операції: 1) визначити довжину шкали (для спрощення і зручності за основу побудови логарифмічної шкали приймають відрізок завдовжки 10 см);

2) визначити логарифми чисел від 1 до 10; 3) логарифми чисел помножити на довжину шкали, знайдені числа послідовно відкласти на носії шкали і відповідно нанести позначки в натуральних числах.

Слід зазначити, що на логарифмічній шкалі відлік починається не від нуля, як в рівномірній, а від одиниці, оскільки $\lg 1 = 0$.

У випадку, якщо логарифмічна шкала нанесена на обидві осі координат, координатну сітку називають **логарифмічною**, а якщо тільки на одну з осей – **напівлогарифмічною.** При побудові статистичних графіків використовують напівлогарифмічну сітку, на яку на осі ординат наносять логарифмічну шкалу.

Експлікація графіка – це словесне тлумачення його змісту. Вона включає назву графіка, написи вздовж масштабних шкал і змістовних значень застосовуваних геометричних знаків.

Графіки можуть супроводжуватися умовними позначеннями, що розкривають зміст застосованих геометричних знаків. Пояснення до вертикальних і горизонтальних шкал повинні розкривати зміст показників, що відображаються, одиниці їх вимірювання. Статистичний графік – це знакова модель, без експлікації його неможливо прочитати і зрозуміти, тобто перенести значення з формалізованої системи характеристики дійсності на саму дійсність.

Одним з найважчих і найважливіших завдань побудови графіка є відшукування правильної його **композиції**, під якою розуміють поєд*нання всіх його елементів. Правильна композиція графіка означає:* ретельний відбір з наявного цифрового статистичного матеріалу даних, що підлягають графічному зображенню; вибір виду графіка; вибір формату (розміру і співвідношення сторін) графіка; підбір масштабу та геометричних знаків і їх розміщення в полі графіка; правильне розміщення і поєднання всіх елементів графіка тощо.

Раціональне розміщення матеріалу на полі графіка створює цілісне уявлення про досліджувані явища.

Створення правильної композиції графіка повинно переслідувати головну ціль – отримати компактне, просте і логічне зображення досліджуваного явища і водночас підкреслювати ті чи інші особливості цього явища (динаміку, тенденції, зв'язки, закономірності, склад, структуру і т.п.).

Не менш важливим завданням композиції графіка є його художнє та естетичне оформлення. Графік повинен притягувати увагу, забезпечуючи водночас легкість його читання та засвоєння.

Щоб композиція графіка відповідала зазначеним вище вимогам, необхідно при побудові графіків виконувати певні правила.

До побудови графіків відносять ті самі вимоги, що й до побудови таблиць. Кожен графік повинен мати чітку і повну назву, що відображає зміст досліджуваного явища, час і місто показників, що наводяться.

У графіку, крім заголовка, обов'язково необхідно наводити і другий текст, в який входять назва і цифри масштабу, назва ліній, цифри, що характеризують окремі частини графіка, умовні позначення, посилання на джерела даних, одиниці вимірювання та ін. Усі пояснювальні написи і заголовки графіка, як і в статистичній таблиці, повинні чітко, коротко і точно розкривати його зміст. Назву графіка, як правило, розміщують в нижній його частині. Написи на графіку повинні бути виконані чітко і охайно. Їх пропонується, як правило, робити горизонтально, тому що вертикальний текст менш зручний для читання. Пояснювальні написи можуть бути розміщені як на самому графіку, так і за його межами. Останній спосіб застосовують у випадках, коли не вистачає місця на полі графіка.

Масштаб на горизонтальній і вертикальній шкалах має бути оптимальним, таким, що не перекручує реальне співвідношення явищ, які аналізуються. Горизонтальну шкалу (на осі абсцис) слід будувати зліва направо, а вертикальну (на осі ординат) – знизу вверх. Цифри шкали слід наносити ліворуч та знизу або вздовж осей. Якщо числові дані не включені у графік, бажано їх дати окремо у формі таблиці. Нульові лінії (вертикальну та горизонтальну) рекомендується відокремлювати на графіку відмінно від усіх ліній координатної сітки. Густота координатної сітки має бути оптимальною і не ускладнювати читання графіка. У зв'язку з цим не слід перевантажувати графіки великою кількістю графічних знаків. Наприклад, на лінійних діаграмах рекомендується наносити не більше як 5-6 ліній; секторну діаграму не слід поділяти більше як на 4-5 секторів і т.п. Особливо не слід завантажувати графік цифрами. Він потрібний для того, щоб замінювати цифри, тому їх слід вписувати у графік лише у крайніх випадках (наприклад, у секторній діаграмі, де важко поглядом вловити співвідношення секторів).

Графік має бути наочним, зрозумілим, легко читатися та по можливості художньо оформленим. З цією метою лінії на графіку можуть бути зображені різним кольором або рисунком (суцільною, пунктирною, точковою, точково-пунктирною лінією).

12.3. Види статистичних графіків і способи їх побудови

Статистичні графіки відрізняються великою різноманітністю. Залежно від способу побудови їх можна поділити на дві великі групи: діаграми і статистичні карти.

Діаграми – це умовне зображення числових величин та їх співвідношень за допомогою геометричних знаків. Термін «діаграма» тотожний терміну «статистичний графік». Діаграми є найбільш розповсюдженим видом графіків. Виділяють такі основні види діаграм: лінійні, стовпчикові, стрічкові, квадратні, секторні, радіальні, трикутні, фігурні, знак Варзара та ін.

Залежно від кола розв'язуваних завдань усі діаграми можна поділити на діаграми порівняння, структури та динаміки.

Розглянемо методику і техніку побудови статистичних графіків, що найчастіше застосовуються на практиці.

Найрозповсюдженішим видом показових діаграм є **лінійні діаграми**, які використовуються здебільшого для характеристики динамічних рядів та рядів розподілу. Поряд з цим лінійні діаграми широко використовуються для вивчення взаємозв'язків між явищами, порівняння кількох показників, ходу виконання планів тощо.

Лінійні діаграми дають можливість зображати явища у вигляді ліній, які з'єднують точки, розташовані у координатному полі. Ламані лінії, що утворюються, показують характер розвитку явища у часі або особливості його розподілу за величиною якої-небудь ознаки або зв'язку явищ.

За способом побудови – це графіки з рівномірною (арифметичною) шкалою. При їх побудові використовують прямокутну систему координат. Розташування будь-якої точки в цій системі визначається двома параметрами – абсцисою та ординатою. Іноді поле в межах осей координат для зручності нанесення геометричних знаків та читання графіка покривається горизонтальними і вертикальними лініями, проведеними за прийнятим масштабом. Ці лінії утворюють **координатну числову сітку**.

На горизонтальній осі (вісь абсцис) відкладають однакові за довжиною відрізки, що відображають періоди (роки, місяці, декади, дні і т.д.). На вертикальній осі (вісь ординат) у певному масштабі наносять значення досліджуваної величини. На перетині перпендикулярів відповідних значень досліджуваної ознаки і часових дат до осей координат отримують точки. Ламана лінія, яка з'єднує ці точки, характеризує зміну досліджуваного явища у часі.

Побудову простої лінійної діаграми розглянемо на такому прикладі (табл. 12.1).

Таблиця 12.1

Динаміка виробництва молока в агрофірмі за 1993–2002 рр.

Роки	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Валове виробництво молока тис. ц	27,0	27,9	26,5	28,0	29,7	32,0	31,0	33,5	34,2	37,2

В прямокутній системі координат на ось абсцис нанесемо показники часу (роки з 1993 по 2002), беручи масштаб річного періоду таким, що дорівнює 1 см. Тоді довжина горизонтальної шкали буде дорівнювати 10 см (1 см $\times$ 10 років). На осі ординат, виходячи із оптимального співвідношення осей ординат і абсцис як 1 : 1,5, тобто завдовжки 6,7 см (10 : 1,5), нанесемо в певному масштабі значення валового виробництво молока від 26 до 38 тис. ц. Візьмемо 1 см за 2 тис. ц. При цьому для більшої наочності на осі ординат зробимо розрив, оскільки мінімальне значення виробництва молока значно відрізняється від нуля. На полі графіка точками відкладемо відповідні значення валового виробництва молока по роках. Отримані точки з'єднаємо відрізками прямої лінії (рис. 12.1).

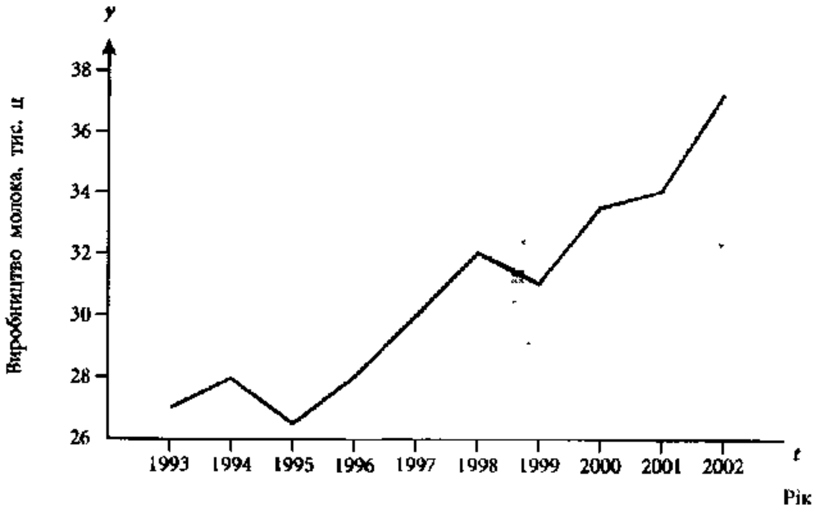


Рис. 12.1. Динаміка валового виробництва молока в агрофірмі за 1993–2002 рр.

Побудований графік показує постійне зростання виробництва молока. Ламана крива, яка має невеликі зламування, безперервно прямує вгору.

Лінійні діаграми можуть бути побудовані з метою вивчення взаємозв'язків між двома ознаками: результативним і факторним (наприклад, між урожайністю і якістю ґрунтів). При цьому на осі абсцис відкладають значення факторної ознаки (якості ґрунтів), а на осі ординат – значення результативної ознаки (урожайності).

Лінійні діаграми зручні для зображення кількох паралельних рядів з метою їх порівняння (наприклад, динаміки продуктивності корів і рівня годівлі або інших якісно відмінних ознак). У цьому разі будують дві (при двох ознаках) або кілька шкал. Другу шкалу будують праворуч.

Особливе місце мають лінійні діаграми зі спеціальними базовими лініями. Найбільш типовими є два випадки. В першому випадку значення вертикальної шкали на початку координат приймають за 100%, тобто лінія, що виходить з цієї точки, відображає рівень базисної величини, яка дорівнює 100%. Всі значення величин, які перевищують базисну, розташовують вище цієї лінії, а значення, які менше рівня базисної величини, розташовують нижче.

У другому випадку при зображенні відхилень від середнього значення рівня (частіше у процентах) базова лінія, що характеризує середній рівень, є нульовою. Додатні відхилення (перевищення) від середнього рівня відкладають вище цієї лінії, від'ємні – нижче.

Діаграми у вигляді вертикальних стовпчиків і стрічок є найбільш простим і досить ефективним видом графічного зображення для аналізу соціально-економічних явищ.

Стовпчикові та стрічкові діаграми застосовуються переважно для порівняння різних показників у просторі і у часі, а також аналізу структури явищ.

Стовпчикові діаграми – це графіки, в яких різні величини представлені у вигляді стовпчиків однакової ширини, які розташовані один від одного на однаковій відстані або щільно. Якщо стовпчики розташовані не по вертикалі, а по горизонталі, то такі діаграми називаються **стрічковими**.

Основа порівняння в стовпчикових і стрічкових діаграмах – лінійна (одномірна). Висота стовпчиків і довжина стрічок відповідно з прийнятим масштабом пропорційна величині зображуваних явищ.

При побудові стовпчикових (стрічкових) діаграм потрібно додержуватись таких основних правил. Основи стовпчиків (стрічок) повинні бути рівними. Стовпчики (стрічки) можуть бути розміщені на однаковій відстані один від одного або щільно. Звичайно додержуються правила, щоб ширина проміжків була вдвоє меншою за ширину самих стовпчиків (стрічок). Висота стовпчиків і довжина стрічок повинні строго відповідати зображуваним цифрам.

Рекомендується включення в діаграму масштабної шкали, яка дає змогу визначити висоту стовпчика і довжину стрічки. Шкала може співпадати з гранню першого стовпчика або стрічки або розташовуватися на окремій лінії зліва (в стовпчиковій діаграмі) або у верхній частині (в стрічковій діаграмі). Шкала, на якій встановлюється висота стовпчиків або довжина стрічок, повинна бути безперервною і починатися з нуля. Написи і вказівки цифр в кінці стовпчиків (стрічок) робити не рекомендується, бо це може створити зорове подовження стовпчиків (стрічок). Цифри показників краще всього писати всередині стовпчиків (стрічок) або розташувати в один ряд над ними на рівні закінчення шкали по осі ординат.

Стовпчики (стрічки) для кращої наочності можуть бути зафарбовані суцільною фарбою, якщо стовпчик (стрічка) відображає ціле явище, або кількома фарбами, якщо зображуються порівняння різних структур явищ, кожному з яких відведена частина стовпчика (стрічки).

Стрічковою діаграмою можна зображувати те саме, що й стовпчиковою. Однак вертикальні стовпчики краще стрічок, якщо числа виражають ідею висоти (рівень зростання) і якщо пояснювальні написи до кожного стовпчика невеликі. Горизонтальні стрічки більш наочні, якщо зображувані величини виражають ідею подовженості (автомобільних доріг, річок і т.п.) і якщо пояснювальний текст до них невеликий.

Стовпчикові і стрічкові діаграми краще за лінійні передусім у тих випадках, коли порівнюваних величин не так багато, порушується безперервність у часі (порівнюють не суміжні періоди) і потрібно звернути увагу не на відносну зміну, а на абсолютну величину порівнюваних рівнів.

Таблиця 12.2

Посівні площі сільськогосподарських культур в ТОВ району за 2000 і 2002 рр., тис. га

Рік	Культури				Разом
	зернові	технічні	картопля і овоче-баштанні	кормові	
2000	26,4	2,9	4,5	21,2	55,0
2002	33,1	4,0	3,3	19,6	60,0

Порядок побудови стовпчикової діаграми покажемо на такому прикладі (табл. 12.2).

Порівняємо за допомогою стовпчикової діаграми загальну посівну площу в ТОВ району за 2000 і 2002 рр., яка відповідно становила 55,0 і 60,0 тис. га.

Для цього на осі абсцис побудуємо два стовпчика з основами по 3 см на відстані 1 см один від одного. Масштаб на осі ординат візьмемо рівним такому співвідношенню: 10 тис. га на 1 см. Цифри, що характеризують загальний розмір посівної площі, напишемо всередині стовпчиків. *Для наочності стовпчики рекомендується зафарбувати або* заштрихувати.

Побудуємо стовпчикову діаграму (рис. 12.2).

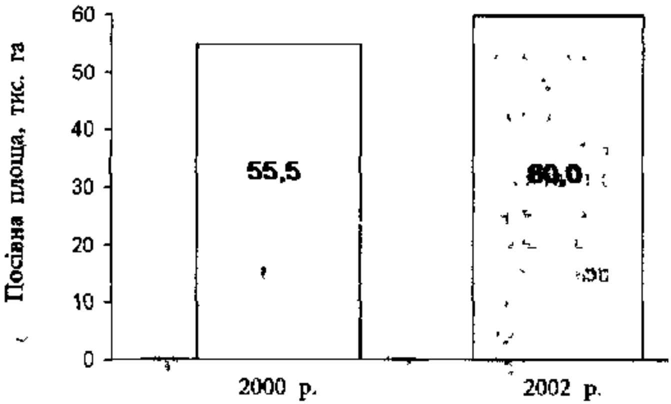


Рис. 12.2. Посівні площі в ТОВ району за 2000 і 2002 рр.

Таблиця 12.3

Структура посівних площ в ТОВ району за 2000 і 2002 рр.

Культури	2000 р.			2002 р.		
	тис. га	в % до підсумку	в градусах	тис. га	в % до підсумку	в градусах
Зернові	26,4	48,0	173	33,1	55,2	199
Технічні	2,9	5,3	19	4,0	6,7	24
Картопля, овоче-баштанні	4,5	8,2	30	3,3	5,5	20
Кормові	21,2	38,5	138	19,6	32,6	117
Разом	**55,0**	**100,0**	**360**	**60,0**	**100,0**	**360**

Стовпчикова діаграма може бути використана не тільки для характеристики загального розміру, а й структури того або іншого явища. При побудові стовпчикової структурної діаграми висоту стовпчика беруть за 100% і поділяють на частини пропорційно структурі явищ (табл. 12.3).

Щоб полегшити читання і аналіз таких діаграм, окремі складові частини розфарбовують різним кольором або штрихуванням.

Використовуючи дані табл. 12.3, побудуємо стовпчикову структурну діаграму (рис. 12.3).

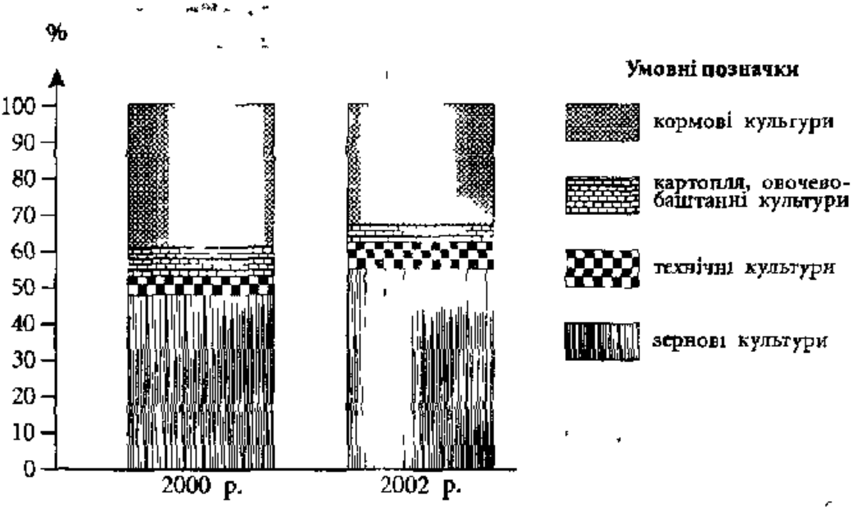


Рис. 12.3. Структура посівних площ в ТОВ району в 2000 і 2002 рр.

Під круговими (квадратними) діаграмами розуміють графіки, що виражають *однорідні величини* через *площі* кругів (квадратів).

Щоб побудувати кругову і квадратну діаграми з порівнюваних статистичних величин, потрібно добути квадратні корені, а потім зобразити круги і квадрати із сторонами, пропорційними одержаним результатам.

Порівняємо між собою за 2002 р. розмір посівних площ технічних (4,0 тис. га), картоплі і овоче-баштанних (3,3 тис. га) культур. Корені квадратні з посівних площ відповідно становлять $\sqrt{4,0}=2,0$ і $\sqrt{3,3}=1,8$.

Якщо прийняти 1 см за 1 одиницю, то посівні площі можна зобразити або кругами з радіусами 2,0 і 1,8 см, або квадратами із сторонами 2,0 см для технічних культур і 1,8 см для картоплі і овочебаштанних культур.

Побудуємо в прийнятому масштабі кругову і квадратну діаграми (рис. 12.4). На квадратних і кругових діаграмах на відміну від стовпчикових масштаб можна не наводити, але в кожній геометричній фігурі слід показати ті числові значення, які вони зображують.

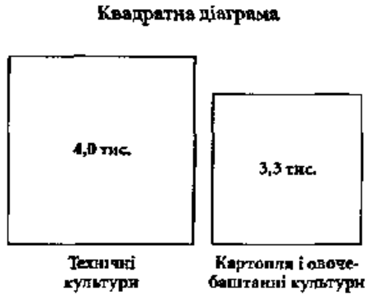


Рис. 12.4. Посівні площі технічних, картоплі і овоче-баштанних культур в ТОВ району в 2002 р.

Як видно з рис. 12.4, користуватися круговими і квадратними діаграмами для порівняння рівнів показників (розміру, обсягу) потрібно з обережністю, оскільки вони не дають точного уявлення про дійсні співвідношення порівнюваних величин і можуть навіть перекручувати уявлення про них. Це зумовлено тим, що візуально складно визначити, на скільки і у скільки разів одна фігура більша або менша за іншу. Крім того, той, хто читає діаграму, по-різному її сприймає, беручи за порівняння або висоту фігури, або її площу.

У зв'язку з цим перевагу слід віддавати здебільшого одномірним (лінійним) порівнянням, використовуючи для цих цілей стовпчикові або стрічкові діаграми.

Секторна діаграма являє собою круг, розділений радіусами на окремі сектори, кожний з яких характеризує питому вагу відповідної частини в загальному обсязі зображуваної величини. Секторні діаграми використовуються переважно для характеристики структури явищ. При порівнянні різних структур загальні площі кіл беруть однаковими. Кожний сектор виділяють за кольором або штрихом; крім того, в кожному секторі нерідко дають і цифрове позначення його питомої ваги. При малому куті сектора експлікація до нього вказується стрілкою.

При побудові секторної діаграми коло поділяють на сектори, площі яких пропорційні часткам частин досліджуваного явища. Площа круга зображує загальний розмір явища і беруть її такою, що дорівнює 100%, або 360^0. Перед побудовою діаграми абсолютні величини переводять у проценти, а проценти в градуси. Кожен процент дорівнює $3{,}6^0$ (360 : 100).

Послідовність розміщення секторів визначається їхньою величиною: самий великий розміщується зверху, а решта – за рухом годинникової стрілки в порядку зменшення.

В секторній діаграмі можна в основний круг вписати малий круг, вказавши в ньому базу, рівну 100%.

Інколи замість круга використовують півкола, поділені на сектори, де 1% дорівнює $1{,}8^0$.

Секторні діаграми наочні тільки тоді, коли досліджувана сукупність поділяється не більш як на 4-5 секторів і спостерігається значна структурна диференціація. Якщо сукупність поділяється на більшу кількість секторів і структурна диференціація незначна, то для зображення структури явищ доцільно застосовувати стовпчикову або стрічкову діаграму.

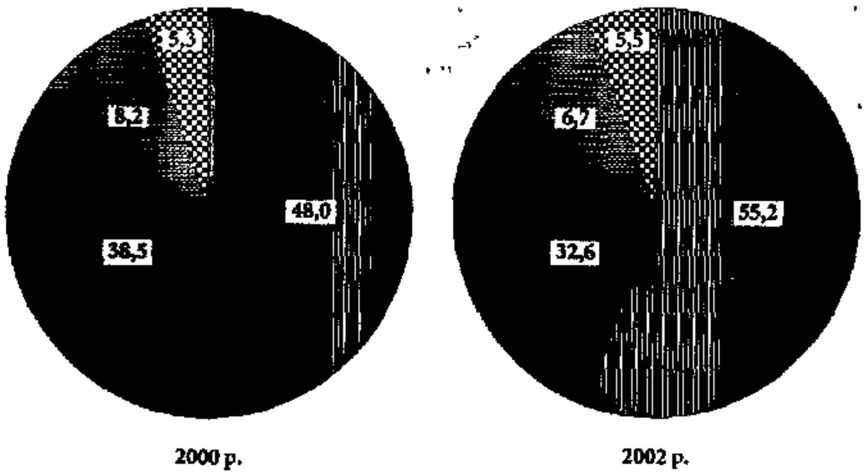


Рис. 12.5. Структура посівних площ в ТОВ району в 2000 і 2002 рр.

Побудову секторної діаграми покажемо на прикладі даних табл. 12.3. Для побудови секторної діаграми використаємо дані цієї таблиці і умовні позначки рис. 12.3. За допомогою транспортира знайдемо потрібні кути і поділимо однакові круги на сектори. Для кращої наочності графіка сектори круга зобразимо різним штрихуванням (рис. 12.5).

Радіальні діаграми використовуються для зображення явищ, які періодично змінюються у часі (переважно сезонних коливань). Для побудови їх використовують полярну систему координат. Круг поділяється на 12 рівних частин, кожна з яких означає певний місяць. Величину радіуса беруть за середньомісячний рівень (100%) і відповідно до цього масштабу на променях, починаючи від центра круга, відкладають відрізки, що зображують місячні рівні. Кінці цих відрізків з'єднують між собою, внаслідок чого утворюється замкнена фігура дванадцятикутник, який характеризує сезонні коливання досліджуваного явища.

В радіальній діаграмі за ось абсцис беруть коло, а за ось ординат – його радіуси, які є носіями масштабної шкали з точкою відліку від центра кола.

Залежно від того, який зображується цикл досліджуваного явища – замкнений або продовжуваний (з періоду в період) – розрізняють замкнені і спіральні радіальні діаграми. Наприклад, якщо зображуються дані по місяцях за кілька років, то при з'єднанні рівня грудня даного року з рівнем січня цього ж року діаграма буде замкненою; при з'єднанні рівня грудня даного року з рівнем січня наступного року утвориться спіральна діаграма. Спіральна діаграма застосовується в тому разі, якщо поряд з сезонними коливаннями відбувається систематичне зростання досліджуваного явища.

Таблиця 12.4

Виробництво молока в агрофірмі по місяцях 2002 р.

Місяць	Виробництво молока, тис. т	Показники сезонності, %	Довжина радіуса, см
I	1526	76,3	2,3
II	1616	80,8	2,4
III	1826	91,3	2,7
IV	1930	96,5	2,9
V	2236	111,8	3,3
VI	2516	125,8	3,8
VII	2568	128,4	3,9
VIII	2370	118,5	3,6
IX	2122	106,1	3,2
X	1908	95,4	2,9
XI	1802	90,1	2,7
XII	1580	79,0	2,4
Разом	**24000**	–	–

Проілюструємо побудову замкненої радіальної діаграми на такому прикладі (табл. 12.4).

Визначимо середній рівень ряду динаміки – середньомісячне виробництво молока:

$$\bar{x} = \frac{\sum x}{n} = \frac{24000}{12} = 2000 \quad ц.$$

Розрахуємо показники сезонності як відношення рівня кожного місяця в процентах до середньомісячного. Наприклад, для січня (1526 : 2000) ·100 = 76,3%, для лютого (1616 : 2000) · 100 = 80,8% і т.д.

Обчислимо довжину радіуса для кожного місяця, прийнявши середньомісячний рівень рівним 3 см. Тоді величина радіуса для січня становитиме (3 ·76,3) : 100 = 2,3 см, для лютого – (3 ·80,8) : 100 = 2,4 см і т.д.

Графік показує, що найбільше виробництво молока спостерігається в травні–серпні, а найменше – в грудні–березні.

Побудуємо радіальну діаграму (рис. 12.6).

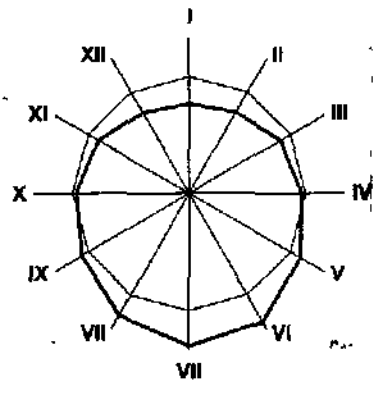


Рис. 12.6. Сезонність виробництва молока в агрофірмі

Особливим різновидом площинних діаграм є **графічні статистичні знаки**, запропоновані російським статистиком В.Є.Варзаром (1851–1940). **Знак Варзара** являє собою площинну діаграму у вигляді прямокутника. Цей вид діаграм використовують для зображення показників, які є результатом перемноження двох інших пов'язаних між собою показників – факторів.

Застосування знаків Варзара ефективне для аналізу показників валового виробництва за складовими (в рослинництві – через урожайність і площу, в тваринництві – продуктивність однієї голови і поголів'я) та іншими складними показниками з виявленням ролі окремих факторів. Так, наприклад, за допомогою знаків Варзара можна графічно зобразити динамічні і територіальні зміни таких показників, як валовий збір сільськогосподарських культур (добуток урожайності на посівну площу), валова продукція (добуток продуктивності праці – виробництво продукції на одного працівника – на чисельність працівників), обсяг вантажоперевезень (добуток виробітку на одну автомашину на середньоспискову чисельність автомашин) тощо.

Знак Варзара будують у вигляді прямокутника, основа і висота якого визначаються за масштабом двома факторами – співмножниками, а площа – величиною результативного показника – добутку. Побудову знака Варзара (рис. 12.7) розглянемо на такому прикладі (табл. 12.5).

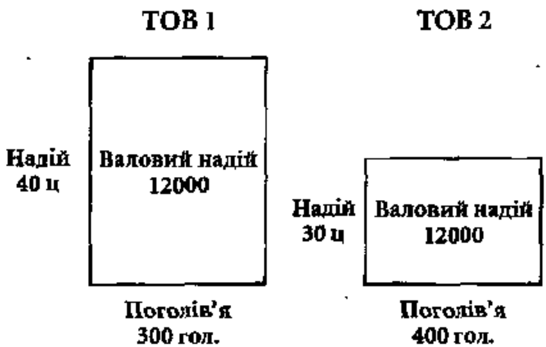


Рис. 12.7. Поголів'я, продуктивність корів і валовий надій молока в ТОВ

Таблиця 12.5

Поголів'я, продуктивність корів і валовий надій молока в ТОВ

ТОВ	Середньорічне поголів'я корів, гол.	Надій на корову, ц	Валовий надій, ц
1	300	40	12000
2	400	30	12000

Для побудови діаграми приймемо такий масштаб: на осі ординат (надій на корову) 10 ц на 1 см, по осі абсцис (середньорічне поголів'я) – 100 голів на 1 см.

Як видно з рис. 12.7, валовий надій молока в двох ТОВ однаковий, але в ТОВ 1 його одержали від меншого поголів'я за рахунок більш високої продуктивності корів.

Лінійні напівлогарифмічні діаграми (графіки відношень) будують таким чином, щоб одна із шкал позначалась як логарифмічна, а друга – як арифметична. В даному випадку логарифмічний масштаб наносять на осі ординат, а на осі абсцис розташовують рівномірну (арифметичну) шкалу для відліку часу за прийнятими інтервалами (роки, місяці, дні і т.д.).

Логарифмічний масштаб характерний тим, що в ньому відрізки шкали пропорційні не зображуваним числовим величинам, а їх логарифмам.

Напівлогарифмічні діаграми доцільно і ефективно застосовувати для порівняння відносних змін в динамічних рядах з істотно різними абсолютними рівнями. *Цей вид діаграм має особливу цінність для зображення пропорційних і процентних відношень, оскільки на цьому графіку кут нахилу кривої виражає відносні зміни, наприклад, темпів зростання.*

Різниці ординат точок кривої (їх приріст) пропорційні темпам зростання, так само, як і на звичайній шкалі ці ординати пропорційні рівням ряду. Отже, за логарифмічною шкалою можна визначити процентне відношення між будь-якими її точками.

Перевага напівлогарифмічних діаграм в аналізі рядів динаміки полягає в тому, що вони дають більш правильне уявлення про темпи динаміки. Тому лінійні напівлогарифмічні діаграми ще називають **діаграмами темпів**. Діаграма з рівномірною арифметичною шкалою правильно передає абсолютні прирости обсягів того або іншого явища, а відносні прирости (темпи) спотворює. Так, якщо виробництво продукції зростає рівномірно, збільшуючись з року в рік, наприклад, на 3%, то це означає, що абсолютні річні прирости будуть весь час збільшуватися. На рівномірній координатній сітці лінія динаміки матиме вигляд зростаючої кривої, а на напівлогарифмічній – вид прямої. Напівлогарифмічний графік правильніше покаже темпи зміни досліджуваного явища, що має велике значення для аналізу динаміки, особливо за тривалий період.

Проілюструємо це на такому прикладі. На рис. 12.8 зображено зростання урожайності кукурудзи на зерно в 2002 р. порівняно з 1994 р. двох підприємств: підприємства І з 20 до 30 ц/га і підприємства ІІ – з 40 до 60 ц/га.

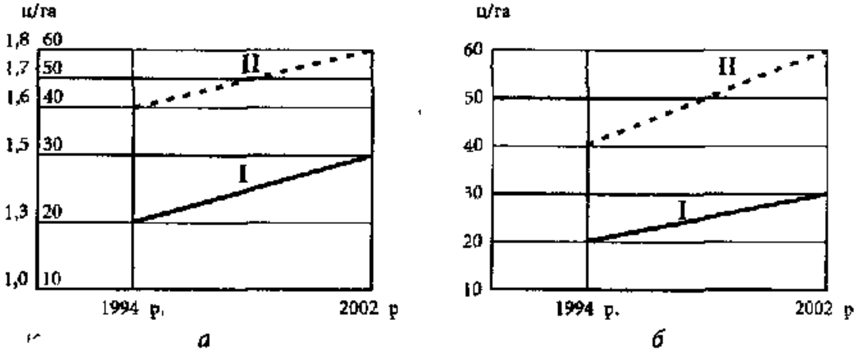


Рис. 12.8. Напівлогарифмічна (а) і арифметична (б) лінійні діаграми динаміки урожайності кукурудзи на зерно

В обох підприємствах однаковий темп зростання урожайності, тобто збільшення в 1,5 рази, що визначається напівлогарифмічною діаграмою (рис. 12.8, *а*) без розрахунку відносних величин, а на арифметичній діаграмі (рис. 12.8, *б*) на напрям ліній впливає величина абсолютного приросту (у підприємства I зростання урожайності становить 10 ц/га, а у підприємства II – 20 ц/га).

Техніка побудови логарифмічної шкали така (рис. 12.9).

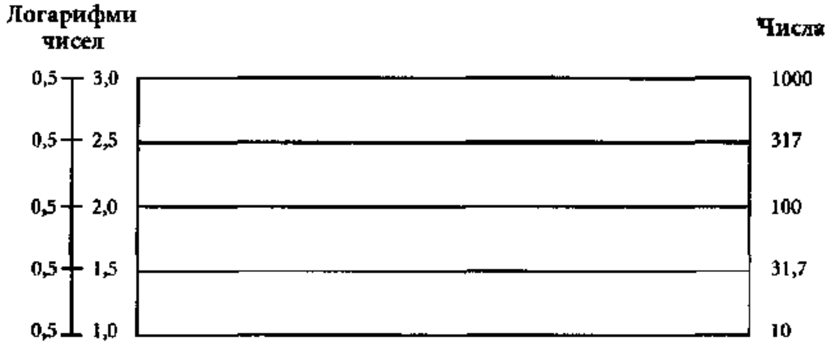


Рис. 12.9. Логарифмічна шкала

Спочатку потрібно знайти логарифми вихідних чисел, накреслити ординату і поділити її на кілька рівних частин. Потім нанести на ординату (або рівну їй паралельну лінію) відрізки, пропорційні абсолютним приростам цих логарифмів.

Далі записати відповідні логарифми чисел і їх антилогарифми, *наприклад 0,0000; 0,3010; 0,4771; 0,6021; ...; 1,0000,* що дає *1, 2, 3, 4,* ..., 10. Одержані антилогарифми дають кінцевий вид шуканої шкали на ординаті. На логарифмічній шкалі немає нульової базової лінії, оскільки $\lg 0$ є $-\infty$.

Наведемо приклад побудови напівлогарифмічної діаграми. Припустимо, що потрібно зобразити на графіку рівень енергозабезпеченості підприємства за 1970–2000 рр. За ці роки вона зросла в 3,3 рази. Знайдемо логарифми для кожного рівня динамічного ряду (табл. 12.6).

Таблиця 12.6

Динаміка енергозабезпеченості підприємства за 1970–2000 рр.

Роки	1970	1975	1980	1985	1990	1995	2000
Енергетичні потужності з розрахунку на 100 га посівної площі, кВт-год.	215	316	398	501	562	631	708
Логарифми чисел	2,33	2,50	2,60	2,70	2,75	2,80	2,85

Обчисливши мінімальне і максимальне значення логарифмів енергозабезпеченості, побудуємо масштаб з таким розрахунком, щоб усі дані розташувались на графіку.

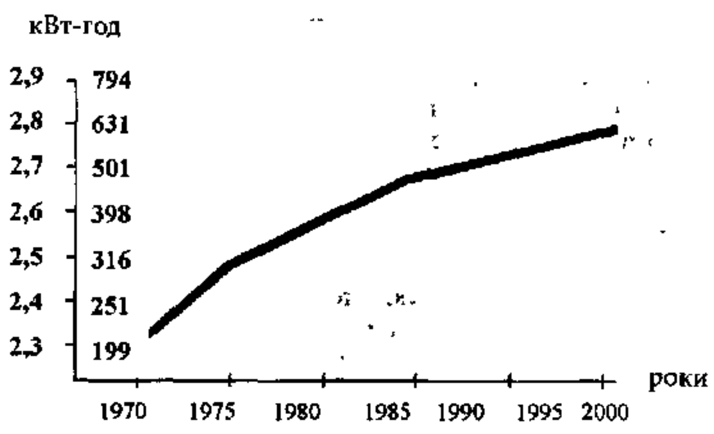


Рис 12.10. Динаміка енергозабезпеченості підприємства за 1970–2000 рр.

Враховуючи масштаб, знайдемо відповідні точки, які з'єднаємо прямими лініями, внаслідок одержимо графік (рис. 12.10) з використанням логарифмічного масштабу на осі ординат. Він називається діаграмою на напівлогарифмічній сітці.

Повною логарифмічною діаграмою він стане в тому випадку, якщо на осі абсцис буде побудований логарифмічний масштаб. В рядах динаміки це ніколи не застосовується, бо логарифмування часу втрачає усяке значення.

Застосування логарифмічного масштабу дає змогу без будь-яких обчислень характеризувати динаміку явища. Якщо крива на логарифмічному масштабі дещо відхилена від прямої і стає увігнутою до осі абсцис, значить має місце зниження темпів; коли крива наближається до прямої - стабільність темпів; якщо вона віддиляється від прямої в сторону, опуклу до осі абсцис (як в нашому прикладі), досліджуване явище (енергозабезпеченість) має тенденцію до зростання із зростаючими темпами.

Напівлогарифмічні діаграми широко використовуються при зображенні відносних змін величин, виражених в різних одиницях вимірювання. Це дає змогу на одній діаграмі порівнювати темпи зростання середньої заробітної плати, продуктивності праці, випуску продукції та інші показники.

У цьому разі логарифмічна шкала поділяється на декілька циклів. Цифри кожного циклу в 10 разів більше цифр нижчого циклу. Наприклад, в межах нижчого циклу значення показника змінюються від 1 до 10, другого від 10 до 100 і т.д. Кожен цикл відповідає зміні характеристики логарифма на одиницю. Для побудови логарифмічної шкали потрібно масштаб циклу (наприклад, один цикл дорівнює 5 см) помножити на логарифми чисел від 1 до 10 і одержані добутки нанести на вертикальну шкалу в межах кожного циклу. Число самих циклів визначається амплітудою коливання рівнів (різницею між максимальним і мінімальним значенням).

Найвиразнішими і такими, що легко сприймаються, є **зображувальні (картинні, фігурні діаграми)**, на яких дають художнє зображення будь-якого явища. Геометричні знаки (точки, лінії і т.п.) в зображувальних діаграмах замінюються фігурами, що якоюсь мірою символізують зовнішній образ зображуваного на графіку явища. Наприклад, поголів'я корів, свиней, овець в господарстві або господарствах району (області, країни) може бути зображено рядом фігур цих тварин.

Достоїнство таких графіків полягає у високому ступені наочності, в одержанні найкращого відображення порівнюваних явищ. Фігурні діаграми можуть бути побудовані за двома основними принципами: кількісним і пропорційним. Для кількісного діаграмування характерне використання рівних за розміром фігур – знаків, число яких показує величину зображуваних явищ.

Фігури, що зображують ту чи іншу величину, розташовують зліва направо на однаковій відстані. В цьому відношенні фігурні діаграми є різновидом лінійних діаграм, в яких поєднуються позитивні сторони стовпчикових і стрічкових діаграм і переваги символічного зображення порівняно з геометричним.

При побудові фігурних діаграм кожній фігурі надається конкретне числове значення і певні стандартні розміри. Сама ж досліджувана статистична величина зображується певною кількістю однакових за розміром фігур. Однак здебільшого не вдається зобразити статистичний показник цілою кількістю фігур. Останню з них доводиться ділити на частини, тому що за масштабом один знак є занадто великою одиницею вимірювання. Складність її визначення є недоліком фігурних діаграм. Проте велика точність зображення статистичних даних у фігурних діаграмах не вимагається, і результати одержуються цілком задовільними.

Принцип пропорційної побудови фігурних діаграм ґрунтується на відображенні величини порівнюваних показників розміром фігур у відповідній пропорції із зображуваними явищами. Недоліки цього способу, по суті, повторюють недоліки розглянутих вище кругових і квадратних діаграм внаслідок можливого досить значного перекручування при сприйнятті у виняткових випадках.

Покажемо приклад побудови фігурної діаграми за даними про чисельність фермерських господарств у двох областях (умовно *А* і *Б*) в 2002 р.

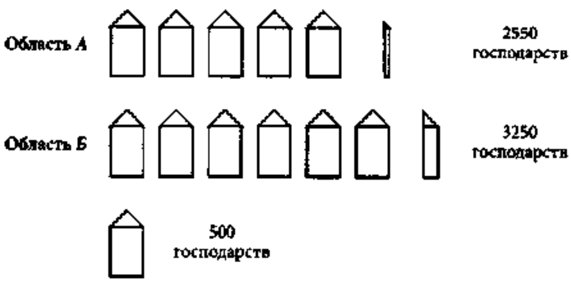


Рис. 12.11. Чисельність фермерських господарств у двох областях в 2002 р.

Приймемо умовно за один фігуро-знак 500 господарств. Тоді число фермерських господарств в області *А* в кількості 2550 буде зображено 5,1 фігурами, а в області *Б* в кількості 3250 – 6,5 фігурами.

Особливе місце в системі графічних зображень звітних і планових даних займають **контрольно-планові графіки.** Основне завдання цих графіків – оперативна характеристика виконання тих або інших виробничих процесів і їх відповідність плановим завданням. Великою перевагою контрольно-планових графіків є те, що вони дають змогу наочного порівняння виконання плану по великому колу взаємопов'язаних об'єктів (бригад, ланок, видів робіт і т.д.).

Контрольно-планові графіки бувають різних видів і ступеня складності залежно від числа об'єктів і кількості ознак, що підлягають графічному зображенню.

Серед великого різноманіття контрольно-планових графіків для вивчення ходу виконання плану найчастіше використовують графік Ганта. *Цей вид графіка зображує рівень виконання плану по кількох* об'єктах як за окремі періоди, так і за звітний період в цілому.

Порядок побудови контрольно-планового графіка (графіка Ганта) розглянемо на такому прикладі (табл. 12.7).

Таблиця 12.7

Дані про хід сівби озимої пшениці по двох відділеннях агрофірми

Показник	Відділення № 1				Відділення № 2			
	Дні роботи							
	25.08	26.08	27.08	28.08	25.08	26.08	27.08	28.08
Денне завдання, га	200	200	200	200	150	150	150	150
Фактично виконано, га	180	200	220	180	180	120	140	160
В % до завдання	90	100	110	90	120	80	93	107
Фактично посіяно з початку робіт до кінця відповідного дня (наростаючий підсумок), га	180	380	600	780	180	300	440	600

Цей графік побудуємо на спеціально розграфленій сітці, в якій по горизонталі в певному масштабі відкладемо періоди часу (дні), а по вертикалі вкажемо об'єкти спостереження.

Кожний відрізок по горизонталі, який означає відрізок часу (день), приймемо рівним 100% і поділимо його на 5 рівних частин, кожна з яких дорівнює 20% планового завдання. На графіку цю шкалу дамо зростаючим підсумком. При цьому позначки 0 і 100%, щоб не ускладнювати графік – не даватимемо.

Ступінь виконання плану на графіку зобразимо двома лініями: тонкою і жирною. Тонка лінія покаже ступінь виконання плану за весь минулий період. Жирна лінія є своєрідною сумарною лінією, що характеризує ступінь виконання плану за станом на кожен день.

Побудуємо контрольно-плановий графік (рис. 12.12).

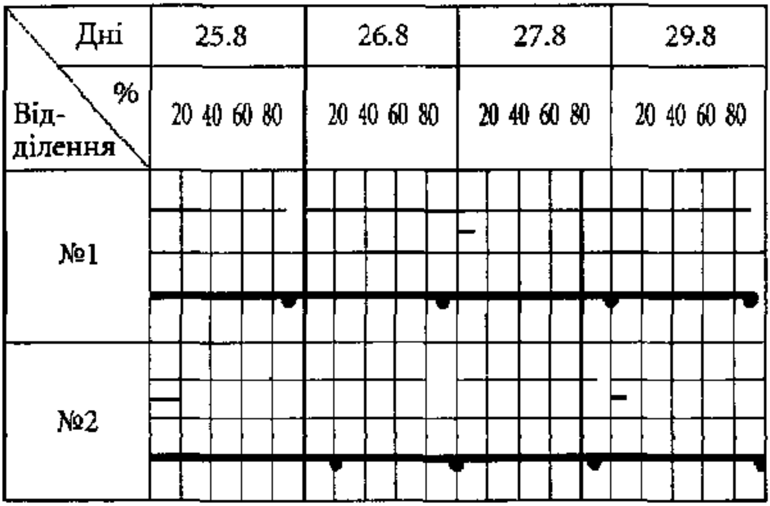


Рис. 12.12. Графік виконання плану сівби озимої пшениці по відділеннях агрофірми

Побудову графіка почнемо з даних по першому відділенню. Планове завдання першого дня (25.08) виконано на 90%. Відповідно до цього показника на діаграмі проведемо тонку лінію, яка займає 90% всього денного проміжку, або 4,5 поділки. За другий день (26.08) план виконано на 100%. Відповідно проведемо тонку лінію, яка займає весь денний відрізок (100%). Планове завдання третього дня (27.08) виконано на 110%. Тому тонка лінія займе всі п'ять частин денного відрізку (100%), під лінією в першій частині проведемо додатковий відрізок, який дорівнює половині першої частини (10%). Ця додаткова лінія показує перевиконання плану за третій день на 10%. За четвертий день (28.08) план сівби виконано на 90%. На діаграмі проведемо тонку лінію, яка займає чотири частини денного відрізку (80%) і половину останньої п'ятої частини, що становить 10% планового завдання.

Нанесемо на графік жирну лінію. На ній одночасно з її побудовою зробимо позначки про щодобове виконання плану. За перший день її довжина буде дорівнювати довжині тонкої лінії (90%). За другий день план виконано на 100%, а за перший і другий день сумарний процент становитиме 190% (90 + 100), тобто жирну лінію за перший день доведемо до 100%, а за другий день – до поділки, що відповідає 90%. Жирна лінія показує, що за перші два дні план не виконано на 10%, або на 20 га (400 – 380). За третій день план виконано на 110%, а за три дні разом сумарний процент становитиме 300% (90 + 100 + 110). Жирна лінія займе три повних денних відрізки. За четвертий день план сівби виконано на 90%. Відповідно сумарний процент за 4 дні становитиме 390% (90 + 100 + 110 + 90). Жирну лінію за четвертий день продовжимо до відмітки 90%.

Аналогічно нанесемо тонку і жирну лінії відповідно до даних відділення № 2.

Аналіз контрольно-планового графіку показує, що за перші чотири дні посівних робіт відділення № 1 не виконало план на 10%, або на 20 га (800 – 780), а відділення № 2 виконало планове завдання на 100%.

До статистичних карт відносять картограми і картодіаграми. Для характеристики територіального розміщення яких-небудь соціально-економічних явищ (щільність населення по регіонах країни, розподіл районів за рівнем урожайності, продуктивності тварин і т.д.) застосовують картограми.

Картограма являє собою схематичну географічну карту, на якій різною фарбою або штрихуванням зображено розподіл будь-якого явища в межах зображуваної на карті території.

Картограми можуть бути виконані по матеріалах окремого господарства (внутрішньогосподарський розріз по бригадах, відділках, полях сівозміни), району (в розрізі господарств), області (в розрізі районів) і т.д.

Картограми поділяють на фонові і точкові. **Фонова картограма** – вид картограми, на якій штрихами різної густоти або фарбою різного ступеня насиченості зображають інтенсивність якого-небудь показника в межах територіальної одиниці. Цей вид картограм, як правило, використовується для зображення середніх і відносних показників.

При побудові фонових картограм попередньо здійснюється групування даних за досліджуваною ознакою. При цьому звичайно виділяють невелику кількість груп (не більше 5-6), а для наочності інтервали заокруглюють. Для кожної групи встановлюється своє штрихування (інтенсивність її повинна зростати з мірою зростання показника). Група господарств (район), що попадають в той або інший інтервал, позначається на карті відповідним штрихуванням.

Точкова діаграма – вид картограми, на якій рівень досліджуваного показника зображується за допомогою точок. Точки зображують одну одиницю сукупності або деяку їх кількість, показуючи на географічній карті щільність або частоту появи певної ознаки. Цей вид картограм використовується здебільшого для відображення розміщення і концентрації абсолютних показників – площі угідь, посівів, чисельності худоби і т.д. При цьому розмір того або іншого показника по територіальних одиницях характеризується певною кількістю точок.

При побудові точкової діаграми всі точки, що наносяться на карту, повинні мати однаковий розмір, оскільки кожна з них характеризує певну величину. Точки легко підрахувати на карті. Необхідно продумати, яку величину буде означати кожна точка; якщо це дуже мале значення, то потрібна буде дуже велика кількість точок, і, навпаки, мала кількість точок не дасть враження густоти.

Побудуємо точкову картограму розміщення посівів цукрових буряків в південній підзоні області (рис. 12.13).

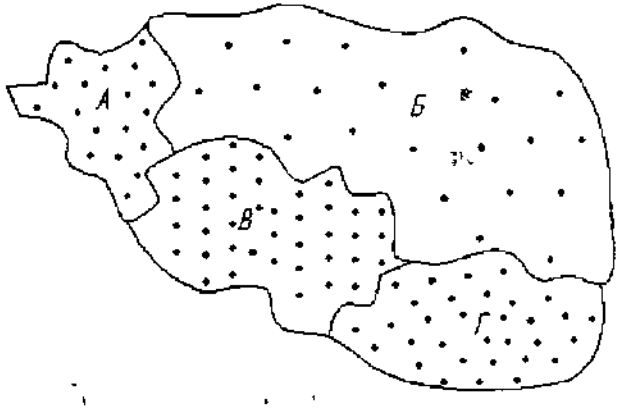


Рис. 12.13. Точкова картограма розміщення посівів цукрових буряків по районах південної підзони області

Точкова діаграма побудована за такими даними щодо площі посіву цукрових буряків: в районі А – 3500 га, Б – 5000, В – 6000, Г – 8000 га. Масштаб виражено таким співвідношенням: 1 точка дорівнює

500 га. Відповідно кількість точок по районах становитиме. А – 7 точок, Б – 10, В – 12, Г – 16 точок.

Загальним недоліком картограм є те, що в межах виділеної територіальної дільниці коливання статистичних показників чітко не виявляється. Крім того, недоцільно зображувати абсолютні величини, оскільки вони відносяться до різних за величиною територій

Картограма – це поєднання схематичної географічної карти з однією із розглянутих вище діаграм (круговою, квадратною, стовпчиковою та ін.). Прикладом картограм є географічна карта, на якій чисельність великих міст зображена у вигляді кіл різної величини.

Завдання для самоконтролю до розділу 12

1. Що таке статистичні графіки, їх роль і значення в аналізі масових суспільних явищ?

2. Які основні елементи графіка?

3. Як класифікуються статистичні графіки?

4. З якою метою використовуються лінійні, стовпчикові і секторні діаграми ?

5. Як будують радіальну діаграму? Що зображують за допомогою радіальної діаграми?

6. Яке призначення мають контрольно-планові графіки і як їх будують?

7. Як будують квадратні і кругові діаграми?

8. Що таке картограми і картодіаграми, у чому їх відмінність?

9. Назвіть основні вимоги до побудови статистичних графіків?

10. Розкрийте загальні правила аналізу статистичних графіків.

11. За наведеними даними зобразіть структуру основних виробничих фондів, для чого побудуйте стовпчикову і секторну діаграми.

Вартість основних фондів, тис. грн.	в тому числі			
	будівлі і споруди	машини і обладнання	транспортні засоби	інші основні фонди
3200	2080	480	256	384

ДОДАТКИ

Додаток 1

Таблиця значень функції $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

t	0	1	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9	10	11
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127

Продовження додатка I

1	2	3	4	5	6	7	8	9	10	11
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0664	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0092	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,3	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
4,0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

***Усі значення помножені на 10000**

Додаток 2

Таблиця значень інтеграла ймовірностей $\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{\frac{t^2}{2}} dt$ *

t	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1113	1993	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4312	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4981	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6680	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8415	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8690	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8926	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9127	9164	9181	9199	9216	9223	9279	9265
1,8	9281	9297	9297	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9439	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9556	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9651	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9729	9743	9749	9756	9762	9768	9744	9780
2,3	9786	9791	9791	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9840	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9879	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9909	9915	9917	9920	9922	9924	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	9973	9974	9975	9976	9976	9977	9978	9979	9979	9980

*Усі значення помножені на 10000

Додаток 3

Критичні точки розподілу *t*-Стьюдента

Число ступенів свободи k	Рівень значущості (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,32	12,70	31,82	63,70	318,30	637,00
2	2,92	4,30	6,97	9,92	22,33	31,60
3	2,35	3,18	4,54	5,84	10,22	12,90
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,5
60	1,68	2,02	2,42	2,70	3,31	3,55
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Рівень значущості α (одностороння критична область)					

Додаток 4

Таблиця значень *F* для рівня значущості $\alpha = 0{,}05$

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	604	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,91
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,79
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,72	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,47	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,53
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Додаток 5

Таблиця значень *F* для рівня значущості α = 0,01

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	4052,10	4999,00	5403,50	5625,10	5764,10	5889,40	5981,30	6105,80	6234,20	6366,50
2	98,49	99,01	99,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,01	2,57
19	8,18	5,93	5 01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,94	5,72	4,87	4,31	3,99	3,75	3,45	3,12	2,75	2,30
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,48	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,07	2,74	2,37	1,90
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,82
45	7,23	5,11	4,25	3,77	3,45	3,23	2,94	2,61	2,23	1,75
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,56	2,18	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,00

Додаток 6

Критичні точки розподілу χ^2

Число ступенів свободи k	Рівень значущості α					
	0,99	0,95	0,10	0,05	0,01	0,001
1	0,000016	0,0039	2,7	3,8	6,6	10,8
2	0,020	0,103	4,6	6	9,2	13,8
3	0,115	0,352	6,3	7,8	11,3	16,3
4	0,30	0,711	7,8	9,5	13,3	18,5
5	0,55	1,15	9,2	11,1	15,1	20,5
6	0,87	1,64	10,6	12,6	16,8	22,5
7	1,24	2,17	12,0	14,1	18,5	24,3
8	1,65	2,73	13,4	15,5	20,1	21,6
9	2,09	3,33	14,7	16,9	21,7	27,9
10	2,56	3,94	16,0	18,3	23,2	29,6
11	3,05	4,57	17,3	19,7	24,7	31,3
12	3,57	5,23	18,5	21,0	26,2	32,9
13	4,11	5,89	19,8	22,4	27,7	34,5
14	4,66	6,57	21,1	23,7	29,1	36,1
15	5,23	7,26	22,3	25,0	30,6	37,7
16	5,81	7,96	23,5	26,,3	32,0	39,3
17	6,41	8,67	24,8	27,6	33,4	40,8
18	7,01	9,39	26,0	28,9	34,8	42,3
19	7,63	10,1	27,2	30,1	36,2	43,8
20	8,26	10,9	28,4	31,4	37,6	45,3
21	8,9	11,6	29,6	32,7	38,9	46,8
22	9,54	12,3	30,8	33,9	40,3	48,3
23	10,2	15,1	32,0	35,2	41,6	49,7
24	10,9	13,8	33,2	36,4	43,0	51,2
25	11,5	14,6	34,4	37,7	44,3	52,6
26	12,2	15,4	35,6	38,9	45,6	51,1
27	12,9	16,2	36,7	40,1	47,0	55,5
28	13,6	16,9	37,9	41,3	48,3	56,9
29	14,3	17,7	39,1	42,6	49,6	58,3
30	15,0	18,5	40,3	43,8	50,9	59,7

Додаток 7

Таблиця значень критерію Q-Тьюкі (α = 0,05)

k	l								
	2	3	4	5	6	7	8	9	10
1	17,97	26,98	32,82	37,08	40,41	43,12	45,40	47,36	49,07
2	6,085	8,331	9,798	10,88	11,74	12,44	13,03	13,54	13,99
3	4,501	5,910	6,825	7,502	8,037	8,478	8,853	9,177	9,469
4	3,927	5,040	5,757	6,287	6,707	7,053	7,347	7,602	7,826
5	3,635	4,602	5,218	5,673	6,033	6,330	6,582	6,802	6,995
6	3,461	4,339	4,896	5,305	5,628	5,985	6,122	6,319	6,493
7	3,344	4,165	4,681	5,060	5,359	5,606	5,815	5,998	6,158
8	3,261	4,041	4,529	4,886	5,167	5,399	5,597	5,767	5,918
9	3,199	3,949	4,415	4,756	5,024	5,244	5,432	5,595	5,739
10	3,151	3,877	4,327	4,654	4,912	5,124	5,305	5,461	5,599
11	3,113	3,820	4,256	4,574	4,823	5,028	5,202	5,353	5,487
12	3,082	3,733	4,199	4,508	4,751	4,950	5,119	5,265	5,395
13	3,055	3,735	4,151	4,453	4,690	4,885	5,049	5,192	5,318
14	3,033	3,702	4,111	4,407	4,639	4,829	4,990	5,131	5,254
15	3,014	3,674	4,076	4,367	4,595	4,782	4,940	5,077	5,198
16	2,998	3,649	4,046	4,333	4,557	4,741	4,897	5,031	5,150
17	2,984	3,628	4,020	4,303	4,524	4,705	4,858	4,991	5,108
18	2,971	3,609	3,997	4,277	4,495	4,673	4,824	4,956	5,071
19	2,960	3,593	3,977	4,253	4,469	4,645	4,794	4,924	5,038
20	2,950	3,578	3,958	4,232	4,445	4,620	4,768	4,896	5,008
30	2,888	3,486	3,845	4,102	4,302	4,464	4,602	4,720	4,824
120	2,800	3,356	3,685	3,917	4,096	4,241	4,363	4,468	4,560
∞	2,772	3,314	3,633	3,858	4,030	4,170	4,286	4,387	4,474

k	l									
	11	12	13	14	15	16	17	18	19	20
1	50,59	51,96	53,20	54,33	55,36	56,32	57,22	58,04	58,83	59,96
2	14,39	14,75	15,08	15,38	15,65	15,91	16,14	16,37	16,57	16,77
3	9,717	9,946	10,15	10,35	10,53	10,69	10,84	10,98	11,11	11,24
4	8,027	8,208	8,373	8,525	8,664	8,794	8,914	9,028	9,134	9,233
5	7,168	7,324	7,466	7,596	7,717	7,828	7,932	8,030	8,122	8,208
6	6,649	6,789	6,917	7,034	7,143	7,244	7,338	7,426	7,508	7,587
7	6,302	6,431	6,550	6,658	6,759	6,852	6,939	7,020	7,097	7,170
8	6,054	6,175	6,287	6,389	6,483	6,571	6,653	6,729	6,802	6,870
9	5,867	5,983	6,089	6,086	6,276	6,359	6,437	6,510	6,519	6,644
10	5,722	5,833	5,935	6,028	6,114	6,194	6,269	6,339	6,405	6,437
11	5,605	5,713	5,811	5,901	5,984	6,062	6,134	6,202	6,265	6,326
12	5,511	5,615	5,710	5,798	5,878	5,953	6,023	6,089	6,151	6,209
13	5,431	5,533	5,625	5,711	5,789	5,862	5,931	5,995	6,155	6,112
14	5,364	5,463	5,254	5,637	5,714	5,786	5,852	5,915	5,974	6,029
15	5,306	5,404	5,493	5,574	5,649	5,720	5,785	5,846	5,904	5,958
16	5,256	5,352	5,439	5,520	5,596	5,662	5,727	5,786	5,843	5,897
17	5,212	5,307	5,392	5,471	5,544	5,612	5,675	5,734	5,790	5,842
18	5,174	5,267	5,352	5,429	5,501	5,568	5,630	5,688	5,743	5,794
19	5,140	5,231	5,315	5,391	5,462	5,528	5,589	5,647	5,701	5,752
20	5,108	5,199	5,282	5,357	5,427	5,493	5,553	5,610	5,663	5,714
30	4,917	5,001	5,077	5,147	5,211	5,271	5,327	5,379	5,429	5,475
120	4,641	4,714	4,781	4,842	4,898	4,950	4,998	5,044	5,086	5,126
∞	4,552	4,622	4,685	4,743	4,796	4,845	4,891	4,934	4,974	5,012

Додаток 8

Значення критерію *z*, які відповідають різним величинам коефіцієнта кореляції *r* від 0 до 0,99

r	Сотi частки *r*									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,497	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,662	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,775	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,950	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647
0,99	2,647	2,700	2,759	2,826	2,903	2,995	3,106	3,250	3,453	3,800

Додаток 9

Критичні значення вибіркового коефіцієнта кореляції при різних числах ступенів волі і рівнях значущості

Число ступенів свободи $k = n - 2$	Рівень значущості		Число ступенів свободи $k = n - 2$	Рівень значущості	
	$\alpha = 0{,}05$	$\alpha = 0{,}01$		$\alpha = 0{,}05$	$\alpha = 0{,}01$
1	0,997	0,999	24	0,388	0,496
2	0,950	0,990	25	0,381	0,487
3	0,878	0,959	26	0,374	0,478
4	0,811	0,917	27	0,367	0,470
5	0,754	0,874	28	0,361	0,463
6	0,707	0,834	29	0,355	0,456
7	0,666	0,798	30	0,349	0,449
8	0,632	0,765	35	0,325	0,418
9	0,602	0,735	40	0,304	0,393
10	0,576	0,708	45	0,288	0,372
11	0,553	0,684	50	0,273	0,354
12	0,532	0,661	60	0,250	0,325
13	0,514	0,641	70	0,232	0,302
14	0,497	0,623	80	0,217	0,283
15	0,482	0,606	90	0,205	0,267
16	0,468	0,590	100	0,195	0,254
17	0,456	0,575	125	0,174	0,228
18	0,444	0,561	150	0,159	0,208
19	0,433	0,549	200	0,138	0,181
20	0,423	0,537	300	0,113	0,148
21	0,413	0,526	400	0,098	0,128
22	0,404	0,515	500	0,088	0,115
23	0,396	0,505	1000	0,062	0,081

Додаток 10

5%-й і 1%-й рівень імовірності коефіцієнтів автокореляції

Розмір вибірки	Додатні значення r_a		Від'ємні значення r_a	
	Рівень			
	5%-й	1%-й	5%-й	1%-й
5	0,253	0,297	– 0,753	– 0,798
6	0,345	0,447	– 0,708	– 0,863
7	0,370	0,510	– 0,674	– 0,799
8	0,371	0,531	– 0,625	– 0,764
9	0,366	0,533	– 0,593	– 0,737
10	0,360	0,525	– 0,564	– 0,705
11	0,353	0,515	– 0,539	– 0,679
12	0,348	0,505	– 0,516	– 0,655
13	0,341	0,495	– 0,497	– 0,634
14	0,335	0,485	– 0,479	– 0,615
15	0,328	0,475	– 0,462	– 0,597
20	0,299	0,432	– 0,399	– 0,524
25	0,276	0,398	– 0,356	– 0,473
30	0,257	0,370	– 0,324	– 0,433
35	0,242	0,347	– 0,300	– 0,401
40	0,229	0,329	– 0,279	– 0,376
45	0,218	0,313	– 0,262	– 0,356
50	0,208	0,301	– 0,248	– 0,339
55	0,199	0,289	– 0,236	– 0,324
60	0,191	0,278	– 0,225	– 0,310
65	0,184	0,268	– 0,216	– 0,298
70	0,178	0,259	– 0,207	– 0,287
75	0,174	0,250	– 0,201	– 0,276
80	0,170	0,246	– 0,195	– 0,271
85	0,165	0,239	– 0,189	– 0,263
90	0,161	0,233	– 0,184	– 0,255
95	0,157	0,227	– 0,179	– 0,248
100	0,154	0,221	– 0,174	– 0,242

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Альбом наглядных пособий по общей теории статистики. - М.: Финансы и статистика, 1991. - 80 с.
2. Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. Теорія статистики: Навч. посібник. - К.: Либідь, 2001 - 320 с.
3. Горкавий В.К. Статистика: Підручник. - К.: Вища школа, 1995. - 415 с.
4. Громыко Г.Л. Общая теория статистики: Практикум. - М.: ИНФРА-М, 2000. - 139 с.
5. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под ред. И.И. Елисеевой - М.: Финансы и статистика, 2001. - 368 с.
6. Ефимова М.Р. Общая теория статистики: Учебник. - М.: ИНФРА-М, 2000. - 416 с.
7. Ефимова М.Р., Ганченко О.И., Петрова Е.В. Практикум по общей теории статистики. - М.: Финансы и статистика, 2001. - 368 с.
8. Ефимова М.Р., Рябцев В.М. Общая теория статистики: Учебник. - М.: Финансы и статистика, 1991. - 304 с.
9. Кожухарь Л.И. Основы общей теории статистики: Учеб. пособие. - М.: Финансы и статистика, 1999. - 144 с.
10. Мармоза А.Т. Практикум по математической статистике: Учеб. пособие. - К.: Вища школа, 1990. - 190 с.
11. Мармоза А.Т. Практикум з теорії статистики: Навч. посібник. - К.: Ельга, Ніка-Центр, 2003. - 344 с.
12. Общая теория статистики: Учебник / А.И. Харламов, О.Э. Башина, В.Т. Бабурин и др.; Под ред. А.А. Спирина, О.Э. Башиной. - М.: Финансы и статистика, 2000. - 440 с.
13. Практикум по статистике: Учеб. пособие / А.П. Зинченко и др. - М.: Колос, 2001. - 392 с.
14. Практикум по теории статистики: Учеб. пособие / Под ред. Р.А. Шмойловой. - М.: Финансы и статистика, 2002. - 416 с.
15. Статистика. Курс лекций / Харченко Л.П., Долженкова В.Г., Ионин В.Г. и др.; Под ред. В.Г. Ионина. - Новосибирск: Издательство НГАЭиУ; М.: ИНФРА-М, 2000. - 310 с.
16. Статистика. Підручник / С.С. Герасименко та ін. - К.: КНЕУ, 2000. - 467 с.
17. Степаненко М.В. Статистика: Навч. посібник. - К.: Вища школа, 1991. - 205 с.
18. Теория статистики: Учебник / Под ред. Р.А. Шмойловой. - М.: Финансы и статистика, 2001. - 560 с.
19. Филимонов В.С., Гуртовник Е.А. Практикум по статистике: Учеб. пособие. - М.: Статистика, 1987. - 128 с.

ЗМІСТ

РОЗДІЛ 6. Вибіркове спостереження

РОЗДІЛ 7. Перевірка статистичних гіпотез

РОЗДІЛ 8. Дисперсійний аналіз

РОЗДІЛ 9. Кореляційний аналіз

Відомості про автора

Мармоза Анатолій Тимофійович – доцент кафедри аналізу і статистики Державного агроекологічного університету, кандидат економічних наук, народився 2 грудня 1939 року.

Після служби в армії (1958–1961) закінчив Луганський сільськогосподарський інститут за спеціальністю «Економіка і організація сільськогосподарського виробництва» (1966). Потім працював економістом з оплати праці в радгоспі, асистентом кафедри статистики Луганського сільськогосподарського інституту (1967–1968). Закінчив очну аспірантуру при кафедрі статистики Московської сільськогосподарської академії, де здобув вчений ступінь кандидата економічних наук за спеціальністю «Статистика» (1972).

З березня 1972 року працює в Державному агроекологічному університеті на посадах асистента, старшого викладача, доцента, завідуючого кафедрою, заступника декана, декана економічного факультету.

Все науково-педагогічне життя (понад 35 років) присвятив одній дисципліні – статистиці.

Має понад 30 наукових праць, серед яких два навчальних посібника з грифами міністерств:

1. Практикум по математической статистике. – К.: Выща школа, 1990. – 190 с.

2. Практикум з теорії статистики. – К.: Ельга, Ніка-Центр, 2003. – 344 с.

Нагороджений знаком «Відмінник аграрної освіти та науки» II ступеня Міністерства аграрної політики України.

Домашня адреса:
10002, м. Житомир-2,
вул. Гагаріна, буд. 7, кв. 27.
Телефон: 8-0412-34-03-57.

НАВЧАЛЬНЕ ВИДАННЯ

\4армоза Лнатолій Тимофійович

ТЕОРІІ СТАТИСТИКИ

Навчальний поабник

Коректор І. Бугрєєва

ГПдписано до друку 28.07.2003 Формат 60x84/16 ,
Пашр офсетний Друк офсетний. Умовн друк. арк 22,79.
Тираж 2000 прим. Зам №3-188.

Ффма «Ельга» 01042, Кшв, вул.Глазунова, 4/47 ;
e-mail.c]ga@svitonlme com
Свщоцтво №23495978 вщ 27 04.95

Видавництво «НіКа-Центр» 01135, Кшв-135, аУс 192;
т /ф. (044) 242-61-56; e-mail-psyhea@uprotel.net ua, servic57@i.conMW
Свщоцтво про внесения до Державного реєстру суб'ектш пидавничо! справи ДК №1399 вщ 18 06 2003

В1ддруковано з дшпозитивш
у АТ "КНИГА". 04655 ГСП, Кшв, вул Артема, 25