

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД ДОНЕЦЬКИЙ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ВИЩА ШКОЛА ЕКОНОМІКИ
ТА МЕНЕДЖМЕНТУ

ФАКУЛЬТЕТ ЕКОНОМІКИ

КАФЕДРА «ЕКОНОМІКА І МАРКЕТИНГ»

КУРС ЛЕКЦІЙ

з нормативної навчальної дисципліни циклу природничо-наукової та
загальноєкономічної підготовки

СТАТИСТИКА

для студентів денної (заочної; денно-заочної) форми навчання

Напрямок(и) підготовки: 6.030504 Економіка підприємства – ЕГП, ЕПЕК, ЕПМ,
6.030507 Маркетинг - МПР

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД ДОНЕЦЬКИЙ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ВИЩА ШКОЛА ЕКОНОМІКИ
ТА МЕНЕДЖМЕНТУ

ФАКУЛЬТЕТ ЕКОНОМІКИ

КАФЕДРА «ЕКОНОМІКА І МАРКЕТИНГ»

КУРС ЛЕКЦІЙ

з нормативної навчальної дисципліни циклу природничо-наукової та
загальноекономічної підготовки

СТАТИСТИКА

для студентів денної (заочної; денно-заочної) форми навчання

Напрямок(и) підготовки: 6.030504 Економіка підприємства – ЕГП, ЕПЕК, ЕПМ,
6.030507 Маркетинг - МПР

Розглянуто
на засіданні кафедри
«Економіка і маркетинг»
Протокол №____
від «_29» «____04_____» 2009р.

Затверджено на засіданні
Навчально-видавничої
Ради ДонНТУ
Протокол №_3_
від «_24» «____06_____»
2009р.

Донецьк, 2010

Курс лекцій з нормативної навчальної дисципліни циклу природничо-наукової та загальноекономічної підготовки «Статистика» для студентів денної (заочної, денно-заочної) форми навчання галузі знань: 0305 «Економіка та підприємництво», напрями підготовки: 6.030504 «Економіка підприємства», 6.030507 «Маркетинг» / Укл.: О.В. Мізіна. - Донецьк: ДоНТУ, 2010. – 134 с.

У посібнику розглянуто базові питання й положення організації та проведення статистичного дослідження соціально-економічних явищ і процесів, висвітлено теоретичні аспекти статистичного дослідження ринкових перетворень. Викладено основні теми курсу «Статистика». Розглянуті методи розрахунку й аналізу абсолютних, відносних і середніх величин, питання статистичного вивчення варіації і форми розподілу, вибіркового метод і статистична оцінка параметрів розподілу, методи перевірки статистичних гіпотез про форму розподілу, статистичні методи вимірювання взаємозв'язків, статистичні індекси, ряди динаміки.

Зміст узгоджено з типовою нормативною програмою дисципліни «Статистика». Навчальний посібник призначений для студентів напрямів підготовки 6.030504 „Економіка підприємства” та 6.030507 „Маркетинг” галузі знань 0305 „Економіка та підприємництво” денної і заочної форм навчання. Призначення спеціальності: підготовка фахівців-маркетологів і економістів для роботи на підприємствах різних галузей промисловості, в комерційних банках, державних і комерційних організаціях.

Укладач

О.В. Мізіна, доц.

Відповідальний
за випуск

В.І. Тутова

ПЕРЕДМОВА

Вдосконалення господарського механізму на сучасному етапі, розвиток ринкових відносин в економіці, перехід економіки до інтенсивного шляху розвитку, необхідність підвищення ефективності суспільного виробництва поставили задачі підвищення рівня професійних, економічних і спеціальних знань керівників всіх галузей народного господарства. Необхідний системний, комплексний підхід до аналізу виробничої діяльності об'єднань, підприємств і їх структурних підрозділів. Збільшується потреба економічної практики у взаємозв'язаному аналізі результатів виробничої і фінансової діяльності на всіх рівнях управління. Це вимагає не тільки оцінки формування зведених показників, але і кількісної оцінки участі структурних підрозділів в загальній динаміці.

Особливе значення при цьому набуває вивчення методології економіко-статистичного аналізу. Статистика виявляє кількісні зміни масових суспільних явищ. Головним прийомом в статистичному дослідженні є формування і вивчення узагальнюючих статистичних показників. Такі показники, з одного боку, виступають предметом дослідження, а з іншою – служать спеціальним прийомом дослідження. Статистичні показники відображають суспільно-економічні явища, які знаходяться в складних взаємозв'язках. Взаємозв'язок явищ вимагає сумісного застосування різних статистичних методів і показників. Абсолютні, відносні, середні показники, показники структури і динаміки, доповнюючі їх взаємозв'язані системи показників в дослідженні повинні застосовуватися комплексно. До основних задач статистичного вивчення відносяться вимірювання об'ємів і рівня, структурних зрушень, динаміки і зв'язку суспільних явищ, чинників, що впливають на загальну динаміку і величини їх впливу. Подальше вдосконалення методології розрахунку показників, розширення застосування статистичних методів при соціально-економічній оцінці організаційно-технічних і господарських рішень та в цілому в економічній практиці є важливою передумовою підвищення рівня управління суспільним виробництвом.

Тема 1. Методологічні засади статистики.

1. Предмет статистики.
2. Основні категорії статистики
3. Статистична методологія

1. Предмет статистики.

Слово «*статистика*» (від лат. status — стан речей) означає кількісний облік масових, насамперед соціально-економічних, явищ і процесів. Статистикою називають також науку, яка об'єднує принципи та методи роботи з масовими числовими даними — кількісними характеристиками зазначених явищ і процесів.

Предметом статистики є величини і кількісні співвідношення між масовими суспільними явищами, закономірності їх формування, розвитку та взаємозв'язку, які розглядаються в безпосередньому зв'язку з їх якісним змістом.

У наведеному визначенні предмета статистики підкреслюються три принципові його особливості. По-перше, статистика вивчає кількісний бік суспільних явищ, по-друге, вона розглядає кількість у зв'язку з якістю явищ та процесів в конкретних умовах місця й часу, по-третє, вона вивчає не поодинокі, а масові явища,

Кількісний бік суспільних явищ — це насамперед їх розміри. Вивчаючи кількісні сторони явищ, статистика відображає їх в числах-показниках, тобто характеризує міру явищ, встановлює схожість й відмінність окремих рис, об'єднує елементи в групи.

Вивчення кількісних аспектів суспільних явищ нерозривно пов'язане з їх якісними характеристиками. Так, групуючи населення за віком, статистика виокремлює якісно відмінні сукупності: осіб дошкільного та шкільного віку, працездатного та пенсійного. Тобто, перш ніж виконувати розрахунки, потрібно визначити якісні властивості та межі кожної такої сукупності. Явища суспільного життя динамічні, вони безперервно змінюються й розвиваються, відповідно кількісну сторону статистика може вивчати тільки в конкретних умовах місця й часу.

Третя особливість предмета статистики зумовлюється масовістю суспільних явищ, їх повторюваністю у просторі або з плином часу.

Для масового явища характерна участь у ньому багатьох елементів, істотні властивості яких однакові або схожі між собою. Так, акціонерів певної фірми об'єднує спільна власність, менеджерів — управлінські функції, а робітників — відповідні виробничі завдання. Наявність будь-яких властивостей у окремого, поодинокого елемента — випадковість. Проте тільки-но численні елементи об'єднуються в одне ціле, сукупна дія випадковостей дає результат, практично незалежний від випадку.

Розглядаючи суспільні явища як масові й спираючись на облік усієї сукупності фактів, що їх стосуються, статистика мовою чисел характеризує

ступінь розвитку таких явищ, напрям і швидкість їх змін, щільність взаємозв'язків і взаємозалежностей.

Статистика — багатогалузева наука. Вона складається з окремих самостійних розділів, які водночас тісно пов'язані між собою. Виокремлюють чотири складові цієї науки.

1. Теорія статистики розглядає категорії статистичної науки, а також спільні для будь-яких масових явищ методи й засоби аналізу.

2. Економічна статистика вивчає явища і процеси, що відбуваються в економіці, розробляє систему економічних показників та методи вивчення економіки країни чи регіону як єдиного цілого.

3. Галузеві статистики (промислова, фінансова, соціальної інфраструктури і т. ін.) розробляють зміст і методи обчислення показників, які відбивають особливості кожної окремої галузі.

4. Соціальна статистика вивчає соціальні умови та характер праці, рівень життя, прибутків, споживання матеріальних благ і послуг населенням.

2. Основні категорії статистики.

Основними поняттями (категоріями) в статистиці є поняття статистичної закономірності й статистичної сукупності.

Закономірність — це повторюваність, послідовність і порядок у масових процесах.

Статистичні закономірності притаманні лише сукупностям, так як об'єктивною основою їх існування є складне переплетіння причин, які формують масовий процес, — основних, спільних для всіх подій масового процесу, та індивідуальних для кожної з них окремо, але випадкових для маси.

Статистичні закономірності виявляються по-різному. Їх можна об'єднати в чотири групи:

1. Закономірності розвитку (динаміки) явищ. *Так, статистика свідчить про збільшення кількості населення Земної кулі, зростання тривалості життя, зменшення середнього віку одруження тощо.*

2. Закономірності розподілу елементів сукупності. *Це може бути розподіл населення за віком, сімей — за кількістю дітей, комерційних банків — за статутним фондом.*

3. Закономірності структурних зрушень. *Прикладом може бути збільшення частки міського населення в загальній його кількості, збільшення частки населення похилого віку в сільській місцевості.*

4. Закономірності зв'язку між явищами. *Наприклад, залежність продуктивності праці від фондоозброєності, собівартості продукції — від продуктивності праці, урожайності — від родючості ґрунту, попиту — від ціни на товар.*

Статистична сукупність — це певна множина елементів, поєднаних умовами існування й розвитку.

Так, статистичною є сукупність комерційних банків країни. Їх об'єднує характер банківських послуг, хоча капітал, кредитно-інвестиційний портфель, прибуток та інші ознаки в них різні.

При вивченні кваліфікації працівників елементом сукупності виступає конкретний працівник, а межі сукупності обмежені рамками підприємства. З другого боку,

промисловість можна розглядати як сукупність підприємств та в даному випадку елементом є підприємство.

Специфічна риса статистики — узагальнення даних. Передумовою такого узагальнення має бути вимірювання, тобто приписування явищу числових значень. Статистичним еквівалентом властивостей, притаманних елементам сукупності, є **ознака**.

Ознаки - властивості, характерні риси або особливості об'єктів (явищ), які можуть бути охарактеризовані низкою статистичних величин.

Кожний елемент сукупності характеризується багатьма ознаками, значення яких змінюються від елемента до елемента або від одного періоду до іншого. Ознака, яка набуває в межах сукупності різних значень, називається такою, що варіює, а відмінність, коливання значень ознаки — **варіацією**. Наприклад, ознаки людини: вік, стать, сімейний стан, освіта тощо; ознаки підприємства: спеціалізація, форма власності, рентабельність виробництва і т. ін.

Одні ознаки виражаються числами, інші — словесно. Їх називають відповідно **кількісними** і **атрибутивними** (описовими). Серед атрибутивних ознак одні чітко окреслені (стать, професія, галузь), інші невизначені (суб'єктивні оцінки, твердження, думки). Для атрибутивних ознак вимірювання означає реєстрацію наявності або відсутності властивості, яка вивчається.

Ознаки мають різний рівень вимірювання, що відображується у відповідних типах шкал. Вирізняють три їх типи: номінальну, порядкову, метричну.

Номінальна шкала — шкала найменувань. «Оцифрування» ознак цієї шкали виконується так, щоб подібним елементам відповідало одне й те саме число, а неподібним — різні числа. Для ідентифікації найменувань шкали використовуються натуральні числа 1, 2, 3, ... або певні числові коди.

Номінальні ознаки, які мають лише два протилежні значення (наприклад, задоволений / незадоволений), називають альтернативними. Їх ідентифікують числами «1» або «0» залежно від наявності чи відсутності властивості.

Порядкова (рангова) шкала встановлює не лише відношення подібності елементів, а й відношення послідовності — порядку. Це відношення типу «більше ніж», «краще ніж» і т. ін. Кожній позначці шкали приписується число — ранг. Такими числами можуть бути: 1, 2, 3 ... n ; 0, 25, 50, 75, 100; -2, -1, 0, 1, 2, тобто значення будь-якої монотонно зростаючої функції, що відповідають послідовності значень ознаки, не враховуючи відстань між ними.

Метрична шкала — це звичайна шкала дійсних чисел. За допомогою метричної шкали вимірюються натурально-речові явища, ресурси та результати господарсько-фінансової діяльності. За характером варіації ознаки метричної шкали поділяються на дискретні та неперервні.

Дискретні ознаки мають лише окремі цілочислові значення: кількість укладених на біржі угод, кількість дітей у сім'ї тощо. **Неперервні ознаки** мають будь-які значення в певних межах варіації. Наприклад, вік людини в межах від 0 до 100 і більше років. До неперервних належать також розрахункові ознаки, а саме: народжуваність, урожайність, балансова ліквідність тощо.

Окремо взяті елементи будь-якої сукупності характеризуються практично необмеженим числом різних ознак.

Оскільки статистика вивчає масові процеси, індивідуальні значення ознак систематизуються, зводяться в єдине ціле. Узагальнюючою характеристикою явищ є **статистичний показник**. На відміну від ознак, які реєструються, статистичні показники розраховуються. Це може бути простий підсумок елементів сукупності або значень ознаки, результат порівняння двох величин або складніших розрахунків.

Величина показника може мінятися залежно від методологічних особливостей його побудови.

3. Статистична методологія

Статистична методологія — це комплекс спеціальних, притаманних лише статистиці методів і прийомів дослідження. Вона ґрунтується на загальнофілософських (діалектична логіка) і загальнонаукових (порівняння, аналіз, синтез) принципах.

Згідно з принципами діалектичної логіки статистика будь-яке суспільне явище розглядає не ізольовано, а у взаємозв'язку з іншими, виявляє фактори, які спричиняють варіацію значень ознак, оцінює ефекти впливу факторів і щільність причинно-наслідкових зв'язків.

Суспільні явища динамічні, тому статистика вивчає їх у розвитку, оцінюючи тенденції та циклічні коливання, інтенсивність динаміки та структурних зрушень.

Розглядаючи сукупності елементів, статистика, з одного боку, вивчає окремі частини явищ, визначає в них схожі риси і відмінності, об'єднує елементи в групи, вирізняючи окремі типи й форми явищ, а з іншого — узагальнює інформацію як за окремими групами (типами), так і за сукупністю в цілому, тобто відображає явища в цілому.

Особливості статистичної методології пов'язані, по-перше, з точним вимірюванням і кількісним описуванням масових суспільних явищ; по-друге, з використанням узагальнюючих показників для характеристики об'єктивних статистичних закономірностей.

Будь-яке статистичне дослідження послідовно проходить три етапи. Перший етап — збирання первинного статистичного матеріалу реєстрацією фактів чи опитуванням респондентів. На другому етапі зібрані дані підлягають систематизації та групуванню — від характеристики окремих елементів переходять до узагальнюючих показників у формі абсолютних, відносних чи середніх величин. Третій етап передбачає аналіз варіації, динаміки, взаємозв'язків.

Етапи об'єднуються метою дослідження.

Так, **масове спостереження** дає інформаційну базу для статистичних узагальнень і характеристики об'єктивних закономірностей. Статистичні дані мають безперечну доказову силу саме тому, що вони спираються не на окремі факти, а на їх сукупність.

На другому етапі — етапі узагальнення даних масового спостереження — елементи сукупності класифікують за певними ознаками, наприклад, народжених можна класифікувати за статтю та місцем народження, видобуток вугілля — за шахтами або за періодами. Впорядковану таким чином статистичну сукупність називають **статистичним рядом**. Залежно від способу класифікації розрізняють **ряди розподілу** та **ряди динаміки**. **Ряд розподілу** — це результат **групування** елементів сукупності у статистиці (станом на певний момент чи за певний інтервал часу). **Ряд динаміки** класифікує значення статистичних показників у часі (за періодами чи моментами часу), описує динаміку розвитку масового процесу.

Передумовою використання статистичних методів у конкретному дослідженні має бути визначення суті явища, що вивчається, його істотних властивостей. **Теоретичний аналіз** дає всебічне уявлення про природу й логіку предмета пізнання. Це — об'єктивна основа методологічних рішень.

Тема 2. Статистичне спостереження.

1. Суть й організаційні форми статистичного спостереження.
2. Види та способи спостереження.
3. Програмно-методологічні та організаційні питання статистичного спостереження
4. Помилки спостереження й контроль достовірності даних.

1. Суть й організаційні форми статистичного спостереження.

Проблема інформаційного забезпечення є першочерговою для будь-якої сфери діяльності. **Інформаційне забезпечення** — це сукупність відомостей про явища та процеси суспільного життя, які орієнтовані на певних користувачів. Отже, **інформація** — це продукт збирання та обробки даних, що має відповідне аналітичне призначення.

Збір статистичної інформації або отримання достовірних статистичних даних основна задача статистичного спостереження (далі спостереження).

Статистичне спостереження — це спланована, науково організована реєстрація масових даних про будь-які соціально-економічні явища та процеси.

Від інших методів збирання даних статистичне спостереження відрізняється характером і масовістю даних та способами їх отримання.

З огляду на різноманітність сфер спостереження застосовують різні організаційні форми збирання даних.

Нині поширені дві організаційні **форми спостереження**: звітність та спеціально організовані спостереження.

Звітність — це форма спостереження, згідно з якою кожний суб'єкт діяльності регулярно подає відповідну інформацію до державних органів статистики та певних відомств у вигляді документів (звітів) спеціально затвердженої форми.

Звітність формується на основі первинного обліку, що забезпечує зіставлення та контроль даних та характеризується такими властивостями, як обов'язковість, систематичність, вірогідність.

Обов'язковість означає, що звіти неодмінно подають усі зареєстровані суб'єкти діяльності, додержуючи уніфікованої форми та затвердженого переліку показників.

Систематичність передбачає регулярне, своєчасне складання та подання звітності у затверджені терміни.

Вірогідність — дані, наведені у звітності, мають відповідати дійсності й виключати будь-які викривлення (приховування та приписки). Звітність складається на підставі первинних даних оперативного та бухгалтерського обліку.

Залежно від рівня затвердження та призначення звітність поділяється на зовнішню та внутрішню. **Зовнішню** затверджують та збирають органи держстатистики, міністерства та відомства, **внутрішню** — розробляють самі суб'єкти діяльності для власних оперативних, управлінських інших потреб. Наприклад, аналіз ринкової ситуації, визначення власних ресурсів, прогнозування діяльності ґрунтується на даних внутрішньої звітності.

За частотою подання звітність поділяється на періодичну та річну. Періодична звітність (місячна, квартальна, піврічна) охоплює показники поточної діяльності суб'єктів, річна — підбиває головні підсумки фінансово-виробничої діяльності суб'єктів за рік. Залежно від терміновості подання звіти можуть передаватися телетайпом, поштою, електронною поштою.

Спеціально організовані спостереження охоплюють сфери життя та діяльності, що не уловлюються звітністю. До них відносять: перепис, спеціальне обстеження, опитування та інше.

Перепис — суцільне або вибіркове спостереження масових явищ з метою визначення їх розміру та складу на певну дату. Перепис здійснюється періодично або одноразово. Так, переписи населення в більшості країн світу відбуваються раз на 10 років.

Особливістю переписів є те, що вони проводяться одночасно по всій території за єдиною для всіх одиниць програмою.

Спеціальні обстеження — несуцільне спостереження окремих масових явищ згідно з певною тематикою, що виходить за межі звітності. Наприклад, бюджетні обстеження домогосподарств, маркетингові обстеження.

Опитування — це, як правило, несуцільне спостереження думок, мотивів, оцінок, що реєструються зі слів респондентів.

Винятком є суцільне опитування всього населення — **референдум**: масове волевиявлення щодо принципових соціально-політичних та економічних питань.

2. Види та способи спостережень.

Класифікуючи статистичні спостереження, визначають їх вид і спосіб реєстрації даних.

Види спостереження розрізняють за двома критеріями: ступенем охоплення одиниць і часом реєстрації даних.

За ступенем охоплення спостереження бувають суцільними та несуцільними.

Суцільні спостереження — це обстеження, під час яких реєструються всі без винятку одиниці сукупності. До цього виду належать обстеження у формі звітності, а також більшість переписів.

Несуцільні спостереження — це обстеження, що мають на меті реєструвати не всі одиниці сукупності, а лише їх певну частину. До таких спостережень належать вибіркове, основного масиву, монографічне, анкетне, моніторинг.

Вибіркове спостереження — це обстеження, під час якого реєструється деяка частина одиниць сукупності, відібрана у випадковому порядку. *Прикладом можуть бути вибірккові обстеження суб'єктів малого бізнесу, обстеження рівня знань студентів державних і недержавних вищих закладів освіти, бюджетів домашніх господарств, а також обстеження якості товарів і продукції тощо.*

Обстеження основного масиву — це обстеження переважної частини одиниць сукупності, що відіграють визначальну роль у характеристиці об'єкта спостереження. *Прикладом може бути обстеження міст із найвищим рівнем забруднення атмосферного повітря або обстеження діяльності групи найвпливовіших комерційних банків. Його також використовують при вивченні рівня цін на ринках. Реєструють рівень цін на продукти, реалізовані на ринках великих міст, частка торгівлі на яких в загальному обсязі складає майже 70%.*

Монографічне обстеження — це детальне обстеження окремих типових одиниць сукупності з метою їх досконалого вивчення. *Прикладом може бути обстеження діяльності фондової біржі або обстеження стану справ збанкрутілої компанії.*

Анкетне спостереження — це обстеження певної частини одиниць сукупності внаслідок неповного повернення від респондентів заповнених реєстраційних формулярів (анкет). *Прикладом є анкетне обстеження мешканців регіону щодо їх ставлення до процесу приватизації землі або обстеження студентів щодо якості викладання фахових дисциплін.*

Моніторинг — це спеціально організоване систематичне спостереження за станом певного середовища. *Наприклад, моніторинг рівня радіаційного забруднення на територіях, що постраждали внаслідок аварії на ЧАЕС. Проводиться також моніторинг бюджетів окремих соціальних груп населення (фермерів, пенсіонерів, студентських сімей).*

За часом реєстрації фактів спостереження поділяються на:

- поточне,
- періодичне.
- одноразове.

Поточне спостереження — це систематична реєстрація фактів щодо явищ, у міру їх виникнення. *Наприклад, реєстрація новонароджених, реалізаційні процеси (збут і реалізація продукції)*

Періодичне спостереження проводиться через певні (як правило, рівні) проміжки часу. Прикладом можуть бути переписи населення, виробничих площ, технологій тощо, а також обстеження суб'єктів бізнесу щодо перспектив інвестування.

Одноразове спостереження проводиться в міру виникнення потреби в дослідженні явища чи процесу. Наприклад, маркетингове дослідження щодо адаптації товару до місцевого ринку або обстеження думки населення щодо впровадження страхової медицини.

Статистичні спостереження здійснюються **трьома способами**:

- безпосередній облік фактів,
- документальний облік,
- опитування.

Безпосередній облік — це обстеження, під час якого обліковець особисто реєструє факти підрахунком, вимірюванням, оцінюванням, оглядом. Наприклад, реєстрація товарних потоків, що перетинають митні кордони, облік готівкової грошової маси в банках, інвентаризація майна.

Документальний облік — це обстеження, коли факти реєструють за даними, наведеними в документах первинного обліку. Наприклад, у такий спосіб складають статистичну звітність і визначають усі економічні показники макро- та мікрорівнів: обсяги матеріальних, трудових і фінансових ресурсів, розмір доходів, капітальних вкладень, обсяги експорту та імпорту товарів тощо.

Опитування може здійснюватись по-різному: експедиційним способом, самореєстрацією, кореспондентським та анкетним шляхом.

Експедиційний спосіб — це реєстрація фактів спеціально підготовленими обліковцями (реєстраторами) з одночасною перевіркою точності реєстрації.

Самореєстрація означає, що факти фіксують самі респонденти після попереднього інструктажу з боку реєстраторів-обліковців.

Кореспондентський спосіб — це реєстрація фактів про явища та процеси на місцях їх виникнення спеціально підготовленими особами та надсилання результатів до відповідних інстанцій. Цей спосіб широко використовується в дослідженнях ринку товарів і послуг на рівні окремих регіонів, а також під час відстежування руху товарів у специфічних умовах ринку.

При анкетному опиті анкети респондентам вручають особисто або висилають поштою.

3. Програмно-методологічні та організаційні питання статистичного спостереження

Статистичне спостереження передбачає: підготовку спостереження; реєстрацію даних та формування бази даних. Підготовка статистичного спостереження — найвідповідальніший етап, оскільки тут вирішуються основні **методологічні питання** (що і як вивчатиметься, на які запитання потрібно дістати відповіді) та **організаційні питання** (хто, де, коли проводить спостереження і що для цього необхідно). Тобто на цьому етапі складається докладний **план статистичного спостереження**, який охоплює методологічні та організаційні питання (див. рис.2.1).



Рисунок 2.1 - План статистичного спостереження.

Мета спостереження — здобути статистичні дані, які є підставою для узагальненої характеристики стану та розвитку явища або процесу з визначенням відповідної закономірності.

Кінцевою метою спостереження є підготовка управлінських рішень та вжиття відповідних заходів. Згідно мети визначають об'єкт та одиницю спостереження.

Об'єкт спостереження - це сукупність явищ, що підлягають обстеженню. Необхідно чітко визначити його межі та істотні ознаки. Це дозволяє запобігти розбіжностям у тлумаченні результатів обстеження. Для цього застосовуються **цензи**. **Ценз** — це набір кількісних обмежувальних ознак. Наприклад, при дослідженні суб'єктів малого бізнесу враховують, що в Україні малим вважається підприємство з чисельністю зайнятих не більш як 50 осіб та обсягом виручки, що не перевищує 1 млн грн. Тому цензом буде максимально припустима кількість зайнятих та обсяг виручки.

Об'єкт спостереження як сукупність складається з окремих елементів — одиниць спостереження. *Одиниця спостереження* — це первинний елемент об'єкта, що є джерелом інформації: підприємство, організація, сім'я.

Елемент (одиниця) сукупності — це первинна одиниця, від якої дістають інформацію. Вони є носіями ознак, які підлягають реєстрації та саме вони піддаються безпосередньому обстеженню.

Елемент сукупності та одиниця спостереження можуть збігатися (як, зокрема, під час перепису). І відповідних прикладів існує чимало: у бюджетних обстеженнях господарств такими одиницями є домашні господарства, під час вивчення громадської думки населення — кожний респондент — людина, яка висловлює свою думку.

Програма спостереження — це перелік запитань, на які потрібно дістати відповіді в результаті спостереження, тобто вона містить перелік ознак, що підлягають реєстрації

Зміст та кількість запитань формують згідно з метою спостереження і реальними можливостями його проведення (грошовими, трудовими витратами та терміном реєстрації). Складають такий перелік запитань, який на підставі щонайменшої кількості залучених даних дає якнайбільше інформації.

Організаційна частина плану визначає місце, час й органи спостереження, графік підготовки та інструктажа кадрів, матеріально-технічну базу.

Насамперед з'ясовується, на який *орган* покладено відповідальність за проведення обстежень, їх підготовку. Залежно від масштабності об'єкта спостереження діють такі групи зазначених органів.

1. Центральні органи державної статистики, а саме Держкомітет статистики України та його регіональні відділення.

2. Статистичні відділи міністерств і відомств.

3. Спеціальні інститути, агентства, міжнародні організації. До таких інститутів належать, зокрема, Інститут соціології, філіал Інституту суспільної думки Геллапа (Соціо-Геллап), Міжнародний інститут соціології та інші.

4. Аналітичні відділи окремих економічних структур (підприємств, організацій, фірм, банків, бірж, страхових товариств тощо).

Місце обстеження — це пункт, в якому перебуває одиниця спостереження і реєструються дані.

Час спостереження розділяють на об'єктивний та суб'єктивний.

Об'єктивним називають час, до якого відносяться дані спостереження. Коли об'єктом спостереження є **процес**, то вибирають інтервал часу, протягом якого нагромаджуються дані. Якщо об'єктом спостереження є **певний стан**, то вибирають певний момент. Момент часу, станом на який реєструються дані, називається **критичним**.

Наприклад, виробництво видів продукції враховується за певний період, а наявність житлових фондів - на певну дату. Число угод, укладених на універсальній біржі, може реєструватися як за деякий період (протягом місяця), так і за станом на певні моменти — дні біржових торгів, що відбуваються, як правило, один раз на тиждень.

Суб'єктивним часом називають період, протягом якого реєструються дані.

Якщо термін представлення місячного звіту до 5 лютого, то суб'єктивний час (час складання звіту) буде з 1 по 5 лютого, а об'єктивний - один місяць.

4. Помилки спостереження й контроль достовірності даних.

Точність та достовірність даних є найважливішою вимогою статистики. Контроль даних спостереження водночас є методологічним і організаційним питанням. Якщо його розглядати як засіб *попередження помилок*, то питання наближається до методологічного. Якщо контроль — це *виявлення та виправлення помилок*, то йдеться швидше про організаційний захід.

Точність - міра відповідності даних спостереження дійсній їх величині.

Достовірність даних - міра об'єктивного відображення ними сутності явищ та процесів.

Помилки спостереження - розбіжності між даними спостереження й дійсними значеннями показників.

Розрізняють помилки реєстрації та репрезентативності.

Помилки репрезентативності виникають тільки при несуттєвому спостереженні в тих випадках, коли відібрана частина сукупності не повністю відображає склад сукупності в цілому.

Помилками реєстрації називаються такі, які виникають в результаті неправильного встановлення фактів або їх неправильного запису.

Помилки бувають випадковими або систематичними.

Випадкові помилки виникають унаслідок дії випадкових причин й спотворюють дані в ту або іншу сторону. Вплив їх на узагальнюючі показники зрівноважується.

Систематичні помилки приводять до значних відхилень загальних результатів спостережень. Вони можуть бути навмисні та ненавмисні (наприклад, ненавмисною є помилка, пов'язана з тенденцією округляти свій вік людьми старшого покоління).

Помилки спостереження виявляються шляхом перевірки та контролю достовірності даних. По-перше здійснюють зовнішній контроль формулярів спостережень, перевіряють правильність та повноту їх заповнення. Потім здійснюють логічний та арифметичний контроль.

Логічний контроль — це перевірка сумісності даних, яка полягає в порівнянні взаємозалежних ознак. Логічний контроль звичайно полягає в зіставленні відповідей на взаємозв'язані питання, що дозволяє виявити несумісність відповідей. Арифметичний контроль полягає в перевірці всіх узагальнюючих показників і в узгодженні тих показників, які виводяться один з іншого.

Тема 3. Зведення і групування статистичних даних.

1. Суть статистичного зведення.
2. Класифікації та групування.
3. Основні питання методології групувань.

1. Суть статистичного зведення.

Зареєстровані у процесі статистичного спостереження значення ознак відбивають увесь діапазон варіації, яка об'єктивно існує в сукупності. При цьому у поодиноких відомостях губиться загальне, закономірне. Для встановлення закономірностей, закладених в даних, необхідна їх систематизація. Така систематизація відбувається завдяки зведенню.

Статистичне зведення - це впорядкування, систематизація і наукова обробка даних.

Суть статистичного зведення полягає в тому, що матеріали спостереження класифікують та агрегують. Елементи сукупності за певними ознаками об'єднують у групи, класи, типи, а інформацію про них агрегують як у межах груп, так і в цілому по сукупності. Основне завдання зведення — виявити типові риси та закономірності масових явищ чи процесів.

Зведення є основою подальшого аналізу статистичної інформації. Складові статистичного зведення такі:

- 1) розробка програми систематизації та групування даних;
- 2) обґрунтування системи показників для характеристики груп і сукупності в цілому;
- 3) проектування макетів таблиць, в яких подаються результати зведення;
- 4) визначення технологічних схем обробки інформації, програмного забезпечення;
- 5) підготовка даних до обробки на комп'ютері, формування автоматизованих банків даних;
- 6) безпосереднє зведення, узагальнення, розрахунок показників.

Тобто, програма систематизації та групування даних передбачає вибір групувальних ознак і правил формування груп. Результати статистичного зведення подаються у формі статистичних таблиць, макети яких розроблюються разом з програмою обробки даних.

За формою обробки даних зведення бувають **централізованими** та **децентралізованими**.

У статистичній практиці інформація обробляється переважно **децентралізовано**. Так, у разі обробки статистичної звітності зведення здійснюється від нижчої до вищої ланки управління: звіти підприємств зводяться регіональними статистичними органами, підсумки по регіонах передаються в Держкомстат, де узагальнюються в цілому по країні. Матеріали переписів, одноразових статистичних обстежень, соціологічних опитувань надходять до єдиного центру, де й обробляються. Така форма зведення називається **централізованою**.

2. Класифікації та групування.

Групування полягає в розділенні сукупності на групи по істотним для них ознаках.

Традиційно цей поділ виконують за такою схемою: із множини ознак, які описують явище, добирають розмежувальні, а потім сукупність поділяють на групи та підгрупи відповідно до значень цих ознак.

Головний принцип будь-якого поділу ґрунтується на двох положеннях:

- 1) в один клас, групу об'єднуються елементи певною мірою подібні між собою;
- 2) ступінь подібності між елементами, які належать до одного класу, значно вищий, ніж між елементами, що належать до різних класів.

Основою групування може бути будь-яка атрибутивна чи кількісна ознака, що має градації. Таку ознаку називають **групувальною**. Залежно від складності масового явища (процесу) та мети дослідження групувальних ознак може бути одна, дві й більше.

У статистичній практиці часто вдаються до розбиття сукупностей за атрибутивними ознаками — **класифікації** та **номенклатури**. Їх розробляють міжнародні й національні статистичні органи.

У міжнародній статистиці відома галузева класифікація видів економічної діяльності, стандартна класифікація занять, стандартна торговельна класифікація. Різновидом класифікації є товарні номенклатури, наприклад Брюссельська митна номенклатура. Кожній класифікаційній позиції надається **код** (шифр), який замінює її назву і є постійним засобом ідентифікації під час передавання інформації по каналах зв'язку, комп'ютерної обробки тощо.

Поряд з класифікацією використовують **групування**, на які покладаються такі аналітичні функції:

- 1) вивчення структури та структурних зрушень;
- 2) визначення типів соціально-економічних явищ, виокремлення однорідних груп і підгруп;
- 3) виявлення взаємозв'язків між ознаками.

Згідно з цими функціями розрізняють три види групувань: структурне, типологічне, аналітичне.

Структурне групування характеризує склад якісно однорідної сукупності за певними ознаками, оскільки у межах такої сукупності значення ознак варіюються. Різновидом структурних групувань є **ряди розподілу**. Ряд розподілу складається з двох елементів: варіант та частот. Варіантами є окремі значення груповочної ознаки, а частотами - числа, які показують, скільки разів повторюються окремі значення варіант. Замість частоти може бути використана **частість** (частка або відсоток від загальної чисельності одиниць).

За допомогою таких групувань вивчають, наприклад, склад населення по вікових групах, склад підприємств по числу працюючих (див. табл. 3.1).

Типологічне групування — це поділ якісно неоднорідної сукупності на класи, соціально-економічні типи, однорідні групи. Основне завдання такого групування — ідентифікація типів. Вибір групувальної ознаки та кількісних міжгрупових меж ґрунтується на всебічному теоретичному аналізі суті явища,

його характерних рис та особливостей формування в конкретних умовах часу та простору.

Прикладом такого угруповання є розподіл населення України по сферах виробничої діяльності (табл.3.2).

Таблиця 3.1 - Угрупування підприємств по числу працюючих

Чисельність працюючих, люд.	Кількість підприємств	
	одиниць	% до підсумку
До 50	431	51,4
51 - 200	291	34,7
201 - 600	90	10,7
601 - 1000	15	1,8
Понад 1000	12	1,4
Разом	839	100,0

Таблиця 3.2 - Розподіл населення України, зайнятого в народному господарстві

Сфери діяльності	200... р.
Всього зайнято в народному господарстві у тому числі	100,0
в галузях матеріального виробництва	73,8
в невиробничій сфері	26,2

Структурні й типологічні групування описові: вони характеризують структуру сукупності, виокремлюють характерні риси та особливості. Типологічні групування відрізняються від структурних тільки метою дослідження, форми їх повністю співпадають (табл.3.3).

Таблиця 3.3 - Схема структурного й типологічного угруповання

Межі угруповання по істотній ознаці	Кількість одиниць сукупності	Система показників		
::	::	::	::	::
::	::	::	::	::
Разом				

Скориставшись групуванням, можна також виявити наявність та напрям зв'язку між ознаками, з яких одна розглядається як результат, інша — як фактор, що впливає на результат. Тобто розглядається як мінімум дві ознаки: факторна (причина) й результативна (слідство). Прикладом є таблиця 3.4

Таблиця 3.4- Залежність урожайності пшениці від терміну збирання

Термін збирання	Збиральна посівна площа, га	Урожайність, ц/га
Своєчасно	330	42
З незначним запізненням	550	36
Зі значним запізненням	120	22
У цілому по сукупності	1000	35

Аналітичні групування проводяться по факторній ознаці та в кожній групі визначається середня величина результативної ознаки (табл.3.5). За наявності зв'язку між ознаками середні групові систематично зростають (прямий зв'язок) або зменшуються (зворотний зв'язок).

Таблиця 3.5 - Схема аналітичного угруповання

Межі угруповання по факторній ознаці, X_i	Кількість одиниць сукупності f_i	Середнє значення результативної ознаки, Y_i
	f_1	Y_1
	f_2	Y_2
	$::$	$::$
Разом	$\sum f_i$	x

Поділ групувань на три види певною мірою відносний. Адже часто групування універсальні: одночасно виділяються типи, визначається склад сукупності й виявляється взаємозв'язок між ознаками.

3. Основні питання методології групувань.

Як відомо, при проведенні групування ознаки можуть бути атрибутивними або кількісними. Для атрибутивної ознаки число груп відповідає числу їх різновидів.

Наприклад, при розподілі населення по рівню освіти виділяють 5 груп: з вищим, незакінченим вищим, середнім спеціальним, середнім, неповним середнім. Якщо ознака альтернативна, можливі 2 групи: працівники задоволені або не задоволені професією.

Якщо групувальна ознака кількісна, постає питання про кількість груп та межі кожної з них. Кількість груп залежить від ступеня варіації групувальної ознаки та обсягу сукупності. Так, для дискретної ознаки, діапазон варіації якої обмежений (кількість дітей у сім'ї, тарифний розряд тощо), груп, як правило, стільки, скільки варіант ознаки. Якщо розглядається неперервна ознака (стаж роботи працівника, собівартість продукції) або у разі значної варіації дискретної ознаки (кількість працюючих на підприємстві, кількість укладених на біржі угод) діапазон варіації розбивається на m інтервалів.

Орієнтовно оптимальна кількість груп визначається за стандартними процедурами, зокрема за формулою Стерджеса:

$$m = 1 + 2,30259 \lg n,$$

де n — обсяг сукупності;

m — число інтервалів.

Інтервали на практиці утворюють за трьома формальними принципами:

- рівності інтервалів;
- кратності інтервалів;
- рівності частот.

У структурних і аналітичних групуваннях та якщо значення ознаки змінюються рівномірно найчастіше застосовують принцип рівності інтервалів. Ширину кожного інтервалу визначають за формулою

$$h = \frac{X_{\max} - X_{\min}}{m} \quad (2.1)$$

де $X_{\max}$, $X_{\min}$ - найбільше та якнайменше значення ознаки,
 m - кількість груп.

Визначаючи межі інтервалів, ширину h доцільно округлювати.

Якщо діапазон варіації ознаки надто широкий і поділ значень нерівномірний, беруть нерівні інтервали, зокрема сформовані за принципом кратності, коли ширина кожного наступного інтервалу в k раз більша (менша), ніж попереднього.

Наприклад, припустимо, що прибутковість активів комерційних банків коливається від 1 до 42%, а прибутковість капіталу — від 11 до 165%. В табл. 3.6 при групуванні за прибутковістю активів скористались принципом рівних інтервалів, а за прибутковістю капіталу — принципом кратності інтервалів ($k = 2$).

Таблиця 3.6 – Варіанти формування інтервалів групувань за рівнем прибутковості, %

Прибутковість активів	Прибутковість капіталу
До 10	11 — 20
10 — 20	21 — 40
20 — 30	41 — 80
30 і більше	81 і більше

Перший та останній інтервали (або один із них) відкриті, тобто мають лише одну межу (верхню чи нижню). За допомогою відкритих інтервалів усі крайні значення ознаки, що варіює, зводяться в одну групу, завдяки чому групування стає компактним.

Принцип рівних частот використовують нечасто і переважно в аналітичних групуваннях, щоб уникнути зважування групових середніх. При утворенні інтервалів за цим принципом не буває нечисленних груп, що дозволяє отримати достовірну характеристику кожної групи.

Групування проводять за однією або за декількома ознаками. Групування по одній ознаці є *простим*, по декількох - *складним*. Останнє може бути *комбінаційним*, якщо в його основі послідовно скомбіновано дві або більше ознаки, або *багатовимірним*, якщо воно проводиться по декількох ознаках одночасно.

Наприклад, угрупування населення за віком та статтю буде комбінаційним, якщо в кожній віковій групі виділені підгрупи за статтю. А угрупування сімей по рівню споживання продовольчих й непродовольчих товарів одночасно багатовимірне.

Тема 4. Узагальнюючі статистичні показники.

1. Суть і види статистичних показників.
2. Абсолютні величини.
3. Відносні величини.
4. Середні величини.
5. Система статистичних показників.

1. Суть і види статистичних показників.

Статистичний показник — це міра, тобто єдність якісного і кількісного відображення певної властивості соціально-економічного явища чи процесу.

Якісний зміст показника визначається суттю явища і відбивається в його назві: народжуваність, продуктивність праці, прибутковість тощо. Кількісна сторона подається числом та його вимірником. Оскільки статистика вивчає суспільні явища в конкретних умовах простору і часу, значення будь-якого показника визначається щодо цих атрибутів.

Сполучною ланкою між якісним змістом і числовим вираженням є — **модель показника** (правило побудови), яка розкриває його структуру, установлюючи, що, де, коли і як має бути виміряне. У моделі обґрунтовуються одиниці вимірювання, технологія збирання даних, обчислювальні операції.

Модель показника має надзвичайно важливе значення для забезпечення **вірогідності** (достовірності) статистичної інформації. Достовірність включає два аспекти: адекватність і точність.

Точність - міра відповідності даних спостереження дійсній їх величині.

Адекватність - здатність показника відобразити саме ту властивість, яка передбачена програмою дослідження.

Для однієї й тієї самої якісної категорії можна побудувати кілька показників. Наприклад, рентабельність виробництва можна оцінити на основі балансового або чистого прибутку, стосовно капіталу чи активів.

Адекватність досягається логічним й статистичним обґрунтуванням показника на стадії його проектування. Точність вимірювання залежить від структури показника, організації спостережень та обробки даних.

Показники розділяються по різних ознаках.

За **способом обчислення** розрізняють первинні й вторинні (похідні) показники. *Первинні* визначаються зведенням даних спостереження й представляються у формі абсолютних величин (кількість та сума внесків в банк). На базі первинних визначаються похідні показники *першого порядку* (середній розмір внеску), а при порівнянні останніх маємо *похідні показники другого порядку* (індекс середнього розміру внеску).

За **ознакою часу** показники розділяються на інтервальні та моментні. *Інтервальні* характеризують явище за певний час (день, декаду, місяць, рік) й залежать від тривалості періоду, за який розраховуються. Наприклад, місячний обсяг товарної продукції. *Моментні* показники дають кількісну характеристику явища на певний момент часу (наприклад, протяжність трубопроводів на кінець року). Інтервальні й моментні показники можуть бути первинними й похідними.

Особливістю первинних інтервальних показників є їх *адитивність*, тобто можливість підсумовування. Похідні показники здебільшого неаддитивні.

Серед статистичних показників існують пари взаємообернених показників, які паралельно характеризують одне й те саме явище. Наприклад, продуктивність праці за одиницю часу — прямий показник, трудомісткість одиниці продукції — обернений.

2. Абсолютні величини.

Абсолютні величини характеризують розміри соціально-економічних явищ. Ідеться про обсяги сукупності чи окремих її частин (кількість елементів) та відповідні їм обсяги значень ознаки. Так, на товарній біржі це кількість заявок і обсяги попиту на купівлю товару.

Абсолютні величини - це іменовані числа, тобто кожна з них має свою одиницю вимірювання: штуки, тонни, кіловати, гривні тощо.

Натуральні вимірники відбивають притаманні явищам фізичні властивості. Так, видобуток вугілля вимірюється в тоннах, газу — у кубічних, виробництво тканин — у квадратних метрах і т. д. Іноді використовують *комбіновані натуральні вимірники*: вантажооборот транспорту в тонно-кілометрах, споживання електроенергії — у кіловат-годинах і т. д. Якщо постає потреба звести воєдино кілька різновидів одного явища, то беруть *умовно-натуральні* вимірники. При цьому роль спільної міри, еталона для розрахунків і порівнянь відіграє один різновид. Перерахунок в умовні одиниці виконується за допомогою спеціальних коефіцієнтів-сумірників. Так, у тоннах умовного палива складається паливний баланс. Перерахунок натурального палива в умовне здійснюється згідно з калорійним еквівалентом. Еталоном є кам'яне вугілля, теплотворна здатність якого становить 7000 калорій на 1 кг. Для донецького вугілля калорійний еквівалент — 0,87 ... 0,90, для торфу — 0,37 ... 0,40, для газу — 1,2 і т. д.

Для визначення обсягу трудових ресурсів чи затрат праці на виробництво продукції, оцінювання трудомісткості продукції використовуються *трудові вимірники* (людино-година, людино-день).

Узагальнюючи облікові дані навіть на рівні окремого суб'єкта господарювання, а тим паче на рівні галузей чи економіки в цілому, використовують *вартісні (грошові)* вимірники.

Не завжди абсолютна величина є результатом зведення даних по сукупності в цілому чи за окремим її складовим. Часто вона визначається за певним правилом, певною методикою на основі інших показників. Так обчислюють валовий внутрішній продукт, національний дохід, прогнози тощо.

Існує певна множина абсолютних величин, які обліковуються у формі балансу. Така форма передбачає розрахунок показника за джерелами формування та напрямками використання, а це дає змогу визначити не лише сумарний показник, а й усі його компоненти. Складаються баланси підприємств, матеріальні баланси найважливіших продуктів, палива, трудових ресурсів і т. ін. Широко використовуються також динамічні баланси за схемою:

$$\begin{aligned} & \text{Залишок на початок періоду} + \text{Надходження} - \text{Витрати} = \\ & = \text{Залишок на кінець періоду.} \end{aligned}$$

3. Відносні величини.

Поглиблений соціально-економічний аналіз фактів потребує різного роду порівнянь. Результатом порівняння є **відносна статистична величина**, яка характеризує міру кількісного співвідношення різнойменних чи однойменних показників.

Кожна відносна величина являє собою дріб, чисельником якого є порівнювана величина, а знаменником — **база порівняння**.

Різноманітність співвідношень реального життя для свого відображення потребує різних за змістом і статистичною природою відносних величин. Відповідно до аналітичних функцій відносні величини можна класифікувати так:

I. Відношення різнойменних показників

Відносні величини інтенсивності

II. Відношення однойменних показників

1) відносні величини структури;

2) відносні величини координації.

3) відносні величини порівнянь;

4) відносні величини динаміки.

Відносні величини інтенсивності є результатом порівняння різнойменних абсолютних величин: у чисельнику — обсяги певного явища (кількість подій, фактів), у знаменнику — обсяг середовища, якому це явище (подія) властиве. Таке співвідношення характеризує інтенсивність (ступінь) поширення явища в середовищі. Це іменована величина, в якій об'єднуються одиниці вимірювання чисельника й знаменника. Наприклад, виробництво продукції на душу населення.

У формі відносних величин інтенсивності обчислюється низка показників технічного рівня виробництва (електроозброєність праці), ефективності використання ресурсів (фондовіддача), економічного розвитку країни (валовий внутрішній продукт на душу населення), життєвого рівня населення (забезпеченість сімей товарами культурно-побутового призначення), інших аспектів суспільного життя.

Якщо обсяги явища незначні відносно обсягів середовища, то результат їх співвідношення збільшується в 100, 1000 і більше разів. Так, демографічні явища (народжуваність, смертність, шлюбність) розраховуються на 1000 осіб, забезпеченість лікарями, підприємствами громадського харчування — на 10 000 осіб, захворюваність, злочинність — на 100 000 осіб. Такі показники називаються відповідно **промille** (‰), **продецимілле** (‰₀₀), **просантимілле** (‰₀₀₀).

Відносні величини структури - це відношення між розмірами частини й цілого. Вони характеризують склад, структуру сукупності. Форма представлення - питома вага або відсотки. Сума відносних величин структури

дорівнює 1 або 100%. Різницю між відповідними частками двох сукупностей називають **процентним пунктом**.

Відносні величини координації - це співвідношення між окремими частинами цілого або відношення окремих частин сукупності до однієї з них, прийнятої за базу порівняння. Наприклад, число міських жителів, що доводяться на 100 сільських; число жінок, що доводяться на 100 чоловіків. Ці величини виражаються у відсотках, проміллі або кратних відношеннях (наприклад, на 100 чоловіків доводиться 114 жінок).

Відносні величини просторових порівнянь обчислюються як відношення однойменних показників, що характеризують різні об'єкти або території й мають однакову тимчасову визначеність.

Важливу роль у статистичному аналізі відіграє **порівняння зі стандартом**. При цьому фактичні значення показника порівнюють з певним еталоном — нормативом, стандартом, оптимальним рівнем.

Такими величинами порівняння є виконання договірних зобов'язань, використання виробничих потужностей, дотримання норм витрат електроенергії тощо.

Для показників, які не мають визначеного еталона (захворюваність, злочинність тощо), базою порівняння може бути максимальне чи мінімальне значення або середня по сукупності в цілому.

Динамікою називається зміна соціально-економічного явища в часі, а **відносна величина динаміки** характеризує напрям та інтенсивність зміни. Ці величини визначаються співвідношенням значень показника за два періоди чи моменти часу. При цьому базою порівняння може бути або попередній або більш віддалений у часі рівень.

Наприклад, інвестиції в підприємства становили, тис. грн.: у 2008 р. — 420, у 2009 р. — 588. Порівнюючи значення показника, дістаємо темпи зростання інвестицій: $588 : 420 = 1,4$.

Якщо значення показника зменшується, відносна величина динаміки буде меншою за одиницю. Передумовою обчислення відносних величин динаміки є порівнянність даних за одиницею вимірювання, методикою розрахунку показника, масштабом об'єкта.

Для деяких показників встановлюють планові завдання. Шляхом порівняння планових й фактичних значень показників обчислюють відносні величини: **планового завдання** та **виконання плану**. Якщо позначити фактичний рівень поточного періоду y_1 , базового y_0 та плановий рівень $y_{пл}$, то:

- відносну величину динаміки можна розрахувати по формулі $K_d = y_1/y_0$,
- планового завдання — $K_{пз} = y_{пл}/y_0$,
- виконання плану $K_{вп} = y_1/y_{пл}$.

Очевидно, що ці величини взаємозв'язані, тобто

$$\frac{y_1}{y_0} = \frac{y_1}{y_{пл}} * \frac{y_{пл}}{y_0} \quad \text{або} \quad K_d = K_{пз} \times K_{вп}$$

4. Середні величини.

Середня величина є узагальнюючою мірою ознаки, що варіює, у статистичній сукупності. Показник у формі середньої характеризує рівень ознаки в розрахунку на одиницю сукупності. В середній взаємно компенсуються індивідуальні відмінності елементів та узагальнюються типові риси. Вона відображає характерний рівень ознаки, властивої всім елементам сукупності. Середня характеризуватиме типовий рівень лише за умови, що сукупність якісно однорідна. Крім того, типовий рівень ознаки проявляє себе лише у разі узагальнення *масових фактів*.

Середня дозволяє здійснювати порівняльний аналіз декількох сукупностей. Зокрема, можливо порівняння однотипних підприємств по рівню середньомісячного (середньорічного) обсягу виробництва, або аналіз динаміки середнього рівня обсягу виробництва одного підприємства за ряд періодів.

Не слід змішувати середні з відносними величинами інтенсивності. Середня завжди узагальнює кількісну варіацію ознаки, яка в тій чи іншій мірі властива всім без виключення елементам сукупності. Відносна величина інтенсивності має зовсім інший характер. Наприклад, з розрахунку на одну сім'ю за матеріалами спостережень за рік виготовлено 4,6 л плодово-ягідних вин. Це, природно, не означає, що кожна з обстежених сімей обов'язково виготовляла якусь кількість вина. В даному випадку виробництво вина як ознака не є властивим всім сім'ям. Це відносна величина інтенсивності.

У кожному конкретному випадку використовується певний вид середньої, зокрема: середня арифметична; середня гармонічна; середня геометрична; середня квадратична і т. д. Середня вимірюється в тих самих одиницях, що й ознака. Залежно від характеру первинної інформації середня будь-якого виду може бути простою чи зваженою.

Середня арифметична.

Оскільки для більшості соціально-економічних явищ характерна адитивність обсягів (виробництво цукру, витрати палива тощо), то найпоширенішою є арифметична середня, яка обчислюється діленням загального обсягу значень ознаки на обсяг сукупності.

За первинними, незгрупованими даними обчислюється **середня арифметична проста**:

$$\bar{x} = \frac{\text{Обсяг значень ознаки}}{\text{Обсяг сукупності}} = \frac{\sum_{i=1}^n x_i}{n}.$$

де x_i – індивідуальні значення ознаки (варіанти,)

n – число варіант (обсяг сукупності).

За цією ж формулою обчислюються середні у динамічному ряду.

Якщо в січні фірма продала 315 одиниць продукції, у лютому — 305 од., а в березні — 340 од., то середньомісячний продаж продукції: $(315 + 305 + 340) : 3 = 320$ одиниць.

Моментні показники замінюються середніми як півсума значень на початок і кінець періоду. Якщо моментів більш ніж два, а інтервали часу між ними рівні, то в чисельнику до півсуми крайніх значень додають усі проміжні, а знаменником є число інтервалів, яке на одиницю менше від числа значень ознаки. Таку формулу називають **середньою хронологічною**:

$$\bar{x} = \frac{\frac{x_1 + x_n}{2} + x_2 + x_3 + \dots + x_{n-1}}{n-1}.$$

Наприклад, на фірмі залишки обігових коштів на початок кожного місяця I кварталу становили, млн грн.: січень — 70, лютий — 82, березень — 77, квітень — 80. Середньомісячний залишок обігових коштів, млн грн.:

$$\bar{x} = \frac{\frac{70+80}{2} + 82 + 77}{4-1} = \frac{234}{3} = 78.$$

У великих за обсягом сукупностях окремі значення ознаки (варіанти) можуть повторюватись. У такому разі обсяг значень ознаки можна визначити

як $\sum_{i=1}^k x_i f_i$ (де f_i — частота варіант, причому $f_1 + f_2 + \dots + f_k = n$). Такий процес множення у статистиці називають **зважуванням**. Вагами можуть бути частоти або частки (відносні величини структури), іноді інші величини.

Значення ознаки осереднюються за формулою **середньої арифметичної зваженої**:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i}.$$

При розрахунку середньої за інтервальним рядом розподілу в якості варіант використовують середнє значення інтервалу як напівсуму його граничних значень. Ширину відкритого інтервалу умовно приймають такою ж, як і в сусідньому закритому інтервалі.

Функціонально середня зважена більш навантажена ніж проста середня, оскільки враховує повторюваність кожної варіанти і певною мірою відображує склад сукупності. Значення середньої зваженої залежить не лише від значень варіант, а й від структури сукупності. Чим більшу вагу мають малі значення ознаки, тим менша середня, і навпаки.

У структурованій сукупності при розрахунку середньої зваженої варіантами можуть бути як окремі значення ознаки, так і **групові середні** $\bar{x}_j$, кожна з яких має відповідну вагу у вигляді групових частот f_j . Таку середню називають **загальною середньою**.

$$\bar{x} = \frac{\sum_{j=1}^m \bar{x}_j f_j}{\sum_{j=1}^m f_j}.$$

де m - кількість груп

Для альтернативної ознаки, яка приймає взаємо виключені значення 1 або 0, середня величина є часткою елементів сукупності d_1 з ознакою, що цікавить дослідника:

$$\bar{x} = \frac{\sum x f}{\sum f} = \frac{1 f_1 + 0 f_0}{f_1 + f_0} = \frac{f_1}{\sum f} = d_1$$

Наприклад, з 230 працівників 46 мають намір змінити професію. В цьому випадку $d_1 = 46:230 = 0,20$, тобто середній рівень потенційної професійної мобільності працівників складає 20%.

Для ознак порядкової (рангової) шкали використовуються різні варіанти оцифрування. Так, окремим пунктам 3-бальної шкали можна надати значень 1, 2, 3 або -1, 0, 1. Очевидно, що розраховані для цих варіантів оцифрування середні бали будуть відрізнятися. Тому для ознак такої шкали звичайно розраховується *нормований середній бал*, який відображує рівень підтримки по питанню, що досліджується. Спочатку ранжують значення ознаки в порядку зростання якості (починаючи з 1), а далі розраховують нормований бал за формулою

$$\bar{K} = \frac{\bar{x} - R/2}{x'}$$

де $\bar{x}$ - середньозважений ранг (x – ранг ознаки),

R - розмах шкали рангів (різниця між найбільшим та найменшим значеннями рангової шкали),

x' - середина шкали рангів (яка дорівнює напівсумі граничних рангів).

За даними таблиці 4.1 оцінимо ставлення населення до смертної кари.

Таблиця 4.1 - Ставлення населення до смертної кари

Варіанти відповідей	Число респондентів	Ранги	xf
Категорично проти	21	1	21
Не визначився	32	2	64
Повністю підтримую	47	3	141
Разом	100		226

$$\bar{K} = \frac{2,26 - 2/2}{2} = 0,63$$

Тобто, рівень підтримки населенням смертної кари складає у середньому 63%.

Середня арифметична має певні **властивості**:

1. Алгебраїчна сума відхилень всіх варіант від середньої дорівнює нулю. (тобто в середній взаємно компенсуються додатні та від'ємні відхилення окремих варіант).

2. Сума квадратів відхилень варіант від середньої арифметичної є меншою, ніж від будь-якої іншої величини, тобто

$$\sum (x - \bar{x})^2 f = \min$$

3. Якщо всі варіанти збільшити (зменшити) на одну й ту саму величину A або в A раз, то й середня зміниться аналогічно.

4. Значення середньої залежить не від абсолютних значень ваг, а від пропорцій між ними. Тобто якщо частоту кожної з груп зменшити або збільшити в одну й ту ж кількість раз, то середня при цьому не зміниться.

Згідно з цією властивістю замість абсолютних ваг — частот f_j — можна використати відносні ваги у вигляді часток або процентів.

Наприклад, на акції трьох різних компаній очікується щорічний прибуток, %: 15, 22, 18. За умови, що інвестор розподілив свої внески між акціями цих компаній у пропорції 30, 20 та 50%, очікуваний прибуток від такого портфеля акцій

$$\bar{x} = \frac{15 \cdot 30 + 22 \cdot 20 + 18 \cdot 50}{100} = 17,9\%.$$

Третю й четверту властивості використовують для спрощення обчислення середньої за рядом розподілу методом, який називається відліком від умовного нуля.

Формула розрахунку середньої методом відліку від умовного нуля має вигляд:

$$\bar{x} = \frac{\sum_{i=1}^m \frac{x_i - x_0}{k} f_i}{\sum f_i} * k + x_0$$

де x_0 — умовний нуль (за який приймають варіанту з найбільшою частотою);

k — загальний множник, що міститься у варіантах або у відхиленнях варіант від початку відліку,

Таким чином $x' = \frac{x_i - x_0}{k}$ — умовна варіанта

Розрахунок середньої з x' проводиться за формулою:

$$\bar{x}' = \frac{\sum x' f}{\sum f};$$

Тобто

$$\bar{x} = \bar{x}' k + x_0$$

Приклад. Обчислити середню арифметичну розподілу підприємств за обсягом випуску продукції за даними таблиці 4.2

Таблиця 4.2 - Розподіл підприємств за обсягом випуску продукції

Обсяг випуску продукції, тис. од.	Середина інтервалу, x_i	Кількість підприємств, f_i	$x_i - x_0$ ($x_0 = 16,55$)	$\frac{x_i - x_0}{k}$ ($k = 0,3$)	$\frac{x_i - x_0}{k} * f_i$
1	2	3	4	5	6
14,3-14,6	14,45	1	-2,1	-7	-7
14,6-14,9	14,75	2	-1,8	-6	-12
14,9-15,2	15,05	8	-1,5	-5	-40
15,2-15,5	15,35	26	-1,2	-4	-104
15,5-15,8	15,65	65	-0,9	-3	-195
15,8-16,1	15,95	120	-0,6	-2	-240
16,1-16,4	16,25	181	-0,3	-1	-181
16,4-16,7	16,55	201	0	0	0
16,7-17,0	16,85	170	0,3	1	170
17,0-17,3	17,15	120	0,6	2	240
17,3-17,6	17,45	64	0,9	3	192
17,6-17,9	17,75	28	1,2	4	112
17,9-18,2	18,05	10	1,5	5	50
18,2-18,5	18,35	3	1,8	6	18
18,5-18,8	18,65	1	2,1	7	7
Разом		1000			10

$$x = (10/1000) * 0,3 + 16,55 = 16,553 \text{ тис. од.}$$

Середня гармонічна та деякі особливості обчислення середніх величин.

При розрахунку середньої з обернених показників використовують середню гармонічну.

Припустимо, що кожний з трьох робітників за годину виготовив 2, 3 та 4 деталі (прямий показник). Тоді витрати робочого часу на виготовлення однієї деталі трьома працівниками складають 1/2, 1/3 та 1/4 години відповідно (обернений показник). Середні витрати часу на виготовлення однієї деталі можна розрахувати за формулою середньої гармонічної простої:

$$\bar{x} = \frac{n}{\sum 1/x} = \frac{3}{1:(1/2) + 1:(1/3) + 1:(1/4)} = \frac{3}{2 + 3 + 4} = \frac{1}{3} \text{ години.}$$

Більш широко використовується середня гармонічна зважена

$$\bar{x} = \frac{\sum z}{\sum (z/x)}, \quad z = x f \quad (4.1)$$

По суті це перетворена середня арифметична. Її використовують тоді, коли показник, виступаючий вагами f , відсутній та його слід додатково визначати на підставі відомих варіант x і множення варіант на частоту xf .

Приклад. Визначити середню продуктивність праці, виходячи з даних таблиці 4.3

Таблиця 4.3 - Дані про продуктивність праці та обсяг виробництва однієї з ділянок підприємства

Бригада	Середня продуктивність праці, нат.од./люд.	Обсяг виробництва нат. од.
Перша	40,0	16000
Друга	45,5	27300
Разом	x	43300

Середню продуктивність праці по ділянці в цілому визначають розподілом сумарного обсягу виробництва на чисельність працюючих. Чисельність грає в даному випадку роль ваг. Цей показник відсутній, проте його легко визначити, розділивши обсяг виробництва по кожній бригаді на середню продуктивність праці по бригаді. Підставивши необхідні дані у формулу (4.1), отримаємо наступний формальний запис:

$$\bar{X} = \frac{16000 + 27300}{\frac{16000}{40,0} + \frac{27300}{45,5}} = \frac{43300}{400 + 600} = 43,3 \text{ нат.од./люд.}$$

В практиці аналізу достатньо часто вимагається визначити середню не абсолютної величини, а відносної. Наприклад, маючи відомості про ступінь виконання державного замовлення на одну й ту ж продукцію по двох підприємствах, необхідно визначити середній відсоток виконання держзамовлення. Вибір форми середньої, залежить від наявної інформації. Основою вибору є логічна формула показника або вихідне співвідношення середньої. Так, рентабельність реалізації обчислюється відношенням:

$$BCC = \frac{\text{Прибуток від реалізації}}{\text{Обсяг реалізації}}.$$

Якщо в розпорядженні є безпосередні дані, що характеризують знаменник вихідного співвідношення середньої, то середня обчислюється за формулою арифметичної зваженої.

Якщо знаменник вимагає додаткових обчислень, то середня визначається по формулі середньої гармонійної.

Приклад. Відомий загальний обсяг продукції по декількох її видах та частка експортної продукції кожного вигляду. Вимагається знайти частку експортної продукції в загальному обсязі продукції всіх видів (таблиця 4.4).

Таблиця 4.4 - Виробництво продукції підприємствами регіону

Вид продукції	Загальний обсяг виробництва, млн.шт.	Частка експортної продукції %
А	591,6	25,0
Б	221,4	19,3
Разом	813,0	х

Запишемо вихідне співвідношення середньої:

$$\text{Частка експортної продукції, \%} = \frac{\text{Обсяг експортної продукції}}{\text{Загальний обсяг виробництва}} * 100$$

Числове значення чисельника співвідношення відсутнє та вимагає додаткових розрахунків. Загальний обсяг експортної продукції вигляду А можна визначити як $0,25 * 591,6 = 147,9$ млн.шт., експортної продукції вигляду Б - $0,193 * 221,4 = 42,7$ млн.шт. Обсяг експортної продукції по обох видах щодо загального обсягу виробництва й складає середню частку. Ці обчислення можна формалізувати та записати таким чином:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{0,25 * 591,6 + 0,193 * 221,4}{591,6 + 221,4} = \frac{190,6}{813,0} = 0,234$$

Тобто, з логічної суті задачі витікає необхідність використання середньоарифметичної зваженої.

Приклад. Відомі частки забракованої продукції в загальному обсязі перевіреної продукції по видах таї абсолютна кількість забракованої продукції по видах. Вимагається визначити частку забракованої продукції в загальному обсязі перевіреної продукції (таблиця 4.5).

Таблиця 4.5 - Результати вибіркової перевірки продукції

Вид продукції	Забраковано, шт.	Частка бракованої продукції в загальному обсязі перевіреної, %
А	473	6,1
В	107	11,1
З	153	13,5
Разом	733	х

Запишемо вихідне співвідношення середньої:

$$\text{Частка бракованої продукції} = \frac{\text{Обсяг бракованої продукції}}{\text{Загальний обсяг перевіреної продукції}}$$

Знаменник співвідношення відсутній і його слід визначити додатково, виходячи з наявної інформації. Це повинна бути середня гармонійна зважена

$$\begin{aligned}\bar{x} &= \frac{\sum z}{\sum \left(\frac{z}{x}\right)} = \frac{473 + 107 + 153}{473/0,061 + 107/0,111 + 153/0,135} = \\ &= \frac{733}{9851} = 0,074 \cdot \text{або} \cdot 7,4\%\end{aligned}$$

Середня геометрична

Якщо визначальна властивість сукупності формується як добуток індивідуальних значень ознаки, використовується середня геометрична:

$$\bar{x} = \sqrt[n]{x_1 x_2 x_3 \dots x_n} = \sqrt[n]{\prod_1^n x_i},$$

де Π — символ добутку;

x_i — відносні величини динаміки, виражені кратним відношенням j -го значення показника до попереднього ($j - 1$)-го.

Наприклад, внаслідок інфляції споживчі ціни за три роки зросли в 2,7 раза, в тому числі за перший рік у 1,8 раза, за другий — в 1,2, за третій — в 1,25 раза. Як визначити середньорічний темп зростання цін? Середня арифметична $(1,8 + 1,2 + 1,25) : 3 = 1,416$ не забезпечує визначальної властивості: за три роки за цією середньою ціни зросли б у $1,416 \cdot$

$1,416 \cdot 1,416 = 2,84$, а не в 2,7 раза. Визначальна властивість $\prod_1^n x_i = 2,7$ забезпечується лише геометричною середньою:

$$\bar{x} = \sqrt[3]{1,8 \cdot 1,2 \cdot 1,25} = \sqrt[3]{2,7} = 1,394$$

Коли часові інтервали не однакові, розрахунок виконують за формулою *середньої геометричної зваженої*:

$$\bar{x} = \sqrt[n]{\prod_1^m x_j^{n_j}},$$

де n_j — часовий інтервал,

$$\sum n_j = n,$$

m — кількість інтервалів.

5. Система економічних показників.

Будь-який показник відтворює лише одну грань процесу або явища. Комплексна характеристика соціально-економічних явищ передбачає використання системи показників. Кожний показник системи має самостійне значення і водночас є складовою узагальнюючої властивості. Систему показників визначають як ієрархічну структуру, на верхньому рівні якої знаходиться узагальнюючий інтегральний показник, на нижньому -

приватні показники, які безпосередньо вимірюються та об'єднуються в блоки (див. рис.4.1). Наприклад, показники якості можна розглядати як блок системи показників конкурентноспроможності технології.

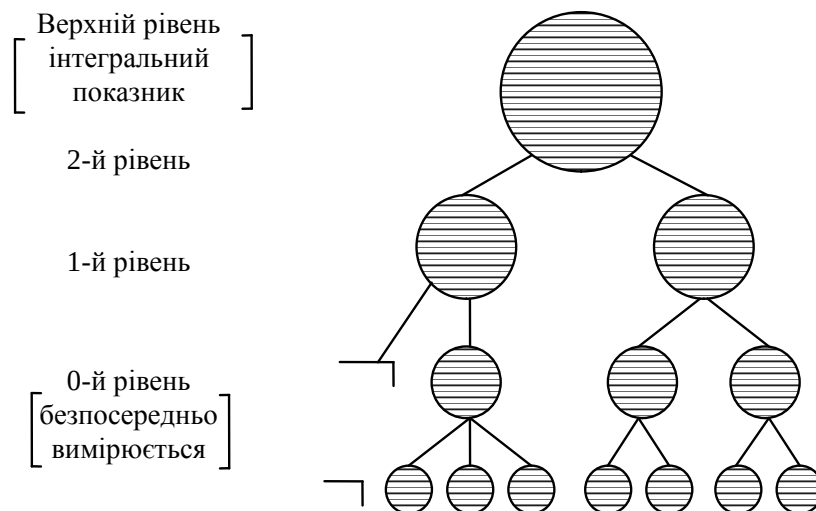


Рисунок 4.1 -. Схема ієрархії системи показників

Визначення інтегральних оцінок ґрунтується на взаємозв'язках показників системи. По характеру взаємозв'язок показників може бути адитивним $a = b + c$, мультиплікативним $a = b * c$, кореляційним – $a = f(b; c)$ або змішаним, наприклад адитивно-мультиплікативним $a = bc + d$, адитивно-кореляційним $a = f(b) + c$.

Будь-яка система показників дає інформацію якісно відмінну від тієї, яку несуть окремі показники.

Складність окремих суспільних явищ зумовила появу інтегральних **комплексних оцінок**, які обчислюються комбінуванням показників нижніх рівнів. Конструювання **інтегральних оцінок** ґрунтується на стандартизації показників, зведенні їх до одного виду. З-поміж інтегральних оцінок широко використовуються оцінки у вигляді **багатовимірних середніх**. Суть багатовимірної середньої полягає в заміні індивідуальних значень множини показників j -го елемента сукупності x_{ij} відносними величинами P_{ij} . Базою порівняння можуть бути середні значення показників по сукупності в цілому $\bar{x}_i$ або еталонні значення $x_{i,st}$ (норма, стандарт):

$$P_{ij} = \frac{x_{ij}}{\bar{x}_i} \quad \text{або} \quad P_{ij} = \frac{x_{ij}}{x_{i,st}}.$$

Середню арифметичну з відношень P_{ij} називають **багатовимірною**. Вона визначається для кожного j -го елемента i є інтегральною оцінкою певного явища саме для цього елемента:

$$\overline{P}_j = \frac{\sum_{i=1}^m P_{ij}}{m},$$

де m — число показників.

Серед показників системи вирізняють стимулятори і дестимулятори. Показники-стимулятори свідчать про високий рівень i -го параметра при $P_{ij} > 1$, дестимулятори — при $P_{ij} < 1$. Щоб звести їх до однозначної характеристики, для дестимуляторів відношення P_{ij} обчислюють як обернену величину.

Якщо показники вважаються різновагомими, кожному з них надається певна вага і розрахунок виконується за формулою середньої арифметичної зваженої:

$$\overline{P}_j = \sum_{i=1}^m P_{ij} d_i,$$

де d_i — вага i -го показника; визначається вона, як правило, експертами

так, щоб $\sum_{i=1}^m d_i = 1$.

Розглянемо розрахунок багатовимірної середньої на прикладі інвестиційної привабливості цінних паперів. Система показників містить:

- 1) рентабельність активів (x_1 , норм $> 20\%$);*
- 2) коефіцієнт капіталізації (x_2 , норм $< 10\%$);*
- 3) оборотність активів (x_3 , норм $= 0,67$ обороту);*
- 4) коефіцієнт заборгованості (x_4 , норм $< 0,7$).*

Значення показників для п'яти емітентів наведено в таблиці 4.6. Усі вони мають певні нормативи, а тому при стандартизації за базу порівняння доцільно взяти нормативи. Серед показників x_1 і x_3 — стимулятори (норматив має нижню межу), x_2 та x_4 — дестимулятори (норматив має верхню межу). Припустимо, що вплив цих показників на інвестиційну привабливість акції однаковий. Тоді інтегральна оцінка визначається як середня арифметична проста. За розрахунками найбільш привабливим для інвестора є перше підприємство, для якого багатовимірна середня становить

$$\overline{P}_j = \frac{3,80 + 2,17 + 9,55 + 3,18}{4} = 4,67$$

Таблиця 4.6 – Багатовимірні середні інвестиційної привабливості цінних паперів.

N_n	Значення показників				Відносні величини P_{ij}				$\bar{P}_j$
	x_1	x_2	x_3	x_4	$\frac{x_{1j}}{x_{1,st}}$	$\frac{x_{2,st}}{x_{2j}}$	$\frac{x_{3j}}{x_{3,st}}$	$\frac{x_{4,st}}{x_{4j}}$	
1	76,0	4,6	6,4	0,22	3,80	2,17	9,55	3,18	4,67
2	47,2	3,9	0,5	0,34	2,36	2,56	0,74	2,06	2,88
3	23,5	12,8	1,1	0,75	1,17	0,78	1,64	0,89	1,12
4	36,3	10,1	1,4	0,58	1,82	0,99	2,09	1,21	1,53
5	16,5	2,8	1,2	0,43	0,83	3,57	1,79	1,63	1,96

Таким чином, статистичний аналіз виконують у двох напрямках:

1) замість ізольованих характеристик окремих сторін предмета розглядають зв'язки і відношення, виявляють фактори, які впливають на рівень і варіацію показників, оцінюють ефекти їх впливу;

2) вивчають динаміку показників, напрям і швидкість змін, визначають характер і рушійні сили розвитку.

Тема 5. Аналіз рядів розподілу

1. Закономірність розподілу.
2. Характеристики центра розподілу.
3. Характеристики варіації.
4. Характеристики форми розподілу
5. Види та взаємозв'язок дисперсій
6. Криві розподілу. Перевірка статистичних гіпотез про відповідність емпіричного та теоретичного розподілів.

1. Закономірність розподілу

Статистична сукупність формується під впливом різних причин та умов: з одного боку — типових, спільних для всіх елементів сукупності, а з іншого — випадкових, індивідуальних. Ці чинники взаємозв'язані, а їх спільна взаємодія визначає як індивідуальні значення ознак, так і розподіл останніх у межах сукупності. Характерні властивості структури статистичної сукупності відбиваються в рядах розподілу.

Ряд розподілу складається з двох елементів: **варіант** — значень групової ознаки x_j та **частот (часток)** f_j . Саме у співвідношенні варіант і частот виявляється **закономірність розподілу**.

Залежно від статистичної природи варіант **ряди розподілу** поділяються на **атрибутивні** та **варіаційні**. Частотними характеристиками будь-якого ряду є абсолютна чисельність j -ї групи — **частота** f_j та відносна частота j -ї групи —

частка d_j . Очевидно, що $\sum_{j=1}^m f_j = n$, а $\sum_{j=1}^m d_j = 1$, або 100%.

Додатковою характеристикою варіаційних рядів є кумулятивна частота (частка), що являє собою результат послідовного об'єднання груп і підсумовування відповідних їм частот (часток). **Кумулятивна частота** S_{f_j} (частка S_{d_j}) характеризує обсяг сукупності зі значеннями варіант, які не перевищують x_j (табл. 5.1).

Таблиця 5.1- Частотні характеристики рядів розподілу

Значення варіант x_j	Частоти f_j	Частки d_j	Кумулятивні	
			частоти s_{f_j}	частки s_{d_j}
x_1	f_1	d_1	f_1	d_1
x_2	f_2	d_2	$f_1 + f_2$	$d_1 + d_2$
x_3	f_3	d_3	$f_1 + f_2 + f_3$	$d_1 + d_2 + d_3$
...	...	...	...	...
x_m	f_m	d_m	Σf_j	1
Разом	Σf_j	1	×	×

Варіаційний ряд може бути дискретним або інтервальним. Якщо варіаційний ряд інтервальний з нерівними інтервалами, то його частотні характеристики непорівнянні. Тоді, аналізуючи розподіл, використовують **щільність** частоти (частки) на одиницю інтервалу, тобто $g_j = f_j : h_j$ або $g_j = d_j : h_j$.

Поглиблений аналіз закономірностей розподілу передбачає характеристику особливостей сукупності, зокрема:

- а) визначення типового рівня ознаки, який є центром тяжіння;
- б) вимірювання варіації ознаки, ступеня згрупованості індивідуальних значень ознаки навколо центра розподілу;
- в) оцінювання особливостей варіації, ступеня її відхилення від симетрії;
- г) оцінювання нерівномірності розподілу значень ознаки між окремими елементами сукупності, тобто ступінь їх концентрації.

Базою аналізу закономірностей розподілу є варіаційний ряд — дискретний або інтервальний — з рівними інтервалами.

2. Характеристики центра розподілу

Центром тяжіння будь-якої статистичної сукупності є типовий рівень ознаки, узагальнююча характеристика всього розмаїття її індивідуальних значень. Такою характеристикою є **середня величина** $\bar{X}$.

За даними ряду розподілу середня обчислюється як арифметична зважена (вагами є частоти f_j або частки d_j):

$$\bar{x} = \frac{\sum_{j=1}^m x_j f_j}{\sum_{j=1}^m f_j}, \quad \bar{x} = \sum_{j=1}^m x_j d_j$$

де j — номер групи;
 m — число груп.

В інтервальних рядах як варіанту x_j використовують середину інтервалу.

Окрім типового рівня важливе значення має домінанта, тобто найбільш поширене значення ознаки. Таке значення називають **модой** (M_o).

Мода (M_o) - це та варіанта, яка частіше за все повторюється у ряді розподілу.

В дискретному ряді - це варіанта, якій відповідає найбільша частота.

Наприклад, при реєстрації жіночого взуття, проданого впродовж дня в магазині, встановлено, що самим ходовим був розмір 23,5, тобто $M_o = 23,5$ (див. табл. 5.2)

Таблиця 5.2 - Розподіл проданого взуття за розмірами

Розмір взуття	22,0	22,5	23,0	23,5	24,0	24,5	25 і більш	Разом
Кількість пар	15	36	70	102	93	76	58	450
Кумулятивні частоти	15	51	121	223	316	392	450	x

В інтервальному ряді легко знаходиться лише модальний інтервал, а сама мода визначається приблизно за формулою

$$M_o = x_0 + h \frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \quad (5.1)$$

де x_0 - нижня межа модального інтервалу,

h, f_m - ширина й частота модального інтервалу,

f_{m-1}, f_{m+1} - частота інтервалів, що передують та подальший щодо модального.

Хай в результаті вибіркового обстеження працівників одного з цехів машинобудівного заводу отримані такі дані (див. табл. 5.3)

Модальний інтервал складає від 6 до 9 років, оскільки відповідна йому частота максимальна $f_m = 230$. Підставимо дані у формулу (5.1) і обчислимо моду

$$M_o = 6 + 3 * \frac{230 - 125}{(230 - 125) + (230 - 225)} = 6 + 2,9 = 8,9$$

Це означає, що такий стаж є найтипівішим для працівників цеху.

Таблиця 5.3- Розподіл працівників цеху за стажом роботи

Стаж, років	Чисельність працівників	Кумулятивні частоти
До 3-х	40	40
3 - 6	125	165
6 - 9	230	395
9 - 12	225	620
12 - 15	144	764
15 і більше	36	800
Разом	800	x

Характеристикою центра розподілу вважається також медіана.

Медіана (Me) - це варіанта, яка ділить ранжируваний ряд на дві рівні за чисельністю частини.

Якщо непарне число варіант записати в порядку збільшення або зменшення, то центральна з них буде медіаною. При парній кількості варіант, медіана розраховується як середня арифметична двох центральних варіант. При визначенні медіани за даними ряду розподілу використовують кумулятивні частоти, які полегшують пошук центральної варіанти. Перш за все, слід встановити порядковий номер центральної варіанти, для чого загальну кількість варіант, збільшену на 1, ділять на 2. Потім по ряду кумулятивних частот визначають, в якій групі знаходиться даний номер.

Визначимо медіанний розмір проданого взуття за даними таблиці 5.2. Порядковий номер центральної варіанти рівний $(450+1)/2=225,5$. З ряду кумулятивних частот видно, що варіанта складає 24,0, оскільки її номер знаходиться саме в цій групі розмірів. Таким чином, $Me = 24,0$.

В інтервальному ряді розподілу аналогічно визначається медіанний інтервал. Конкретне значення медіани визначається за формулою

$$Me = x_0 + h \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} \quad (5.2)$$

де x_0 - нижня межа медіанного інтервалу,

h - ширина медіанного інтервалу,

S_{m-1} - кумулятивна частота інтервалу, що передує медіанному,

f_m - частота медіанного інтервалу.

Обчислимо медіану, скориставшись даними табл. 5.3:

$$Me = 9 + 3 \frac{\frac{800}{2} - 395}{225} = 9,1$$

Медіана, як і мода, не залежить від крайніх значень ознаки; сума модулів відхилень варіант від медіани мінімальна, тобто вона має властивість лінійного мінімуму:

$$\sum_1^m |x_j - \text{Me}| f_j = \min$$

Цю властивість медіани можна використати при проектуванні розміщення зупинок міського транспорту, заготівельних пунктів тощо.

Мода і медіана - це особливий вид середніх величин. На відміну від абстрактної середньої арифметичної, ці характеристики завжди співпадають з конкретними варіантами.

Окрім моди і медіани, в аналізі закономірностей розподілу використовуються також квартилі та децилі. **Квартилі** — це варіанти, які поділяють обсяги сукупності на чотири рівні частини, **децилі** — на десять рівних частин. Ці характеристики визначаються на основі кумулятивних частот (часток) за аналогією з медіаною, яка є другим квартилем або п'ятим децилем.

3. Характеристики варіації

В одних сукупностях індивідуальні значення ознаки щільно групуються навколо центра розподілу, в інших — значно відхиляються. Чим менші відхилення, тим однорідніша сукупність, а отже, тим більш надійні й типові характеристики центра розподілу, передусім середня величина. Вимірювання ступеня коливання ознаки, її варіації — невід'ємна складова аналізу закономірностей розподілу.

Для вимірювання та оцінювання варіації використовуються абсолютні та відносні характеристики. До абсолютних належать: розмах варіації, середнє лінійне та середнє квадратичне відхилення, дисперсії; відносні характеристики подаються низкою коефіцієнтів.

Розмах варіації - це різниця між найбільшим та якнайменшим значеннями ознаки $R = x_{\max} - x_{\min}$.

В інтервальному ряді розподілу розмах варіації визначають як різницю між верхньою межею останнього інтервалу й нижньою межею першого або як різницю між середніми значеннями цих інтервалів.

Перевагою варіаційного розмаху є простота його обчислення й тлумачення. Проте, коли частоти крайніх варіант надто малі, варіаційний розмах неадекватно характеризує варіацію. У таких випадках використовують квартильні або децильні розмахи. Квартильний розмах охоплює 50% обсягу сукупності

$$R_Q = Q_3 - Q_1.$$

Інші абсолютні характеристики варіації базуються на відхиленнях індивідуальних значень ознаки від середньої величини. Оскільки $\sum_1^m (x_j - \bar{x}) f_j = 0$ то при розрахунку такого роду показників використовують або модулі, або квадрати відхилень. В результаті маємо наступні характеристики варіації: середнє лінійне l та середнє квадратичне σ відхилення та дисперсію σ^2 (див. таблицю 5.4)

Таблиця 5.4 - Розрахунок узагальнюючих характеристик варіації

Показник	Середнє відхилення	
	лінійне	квадратичне
За даними: - незгрупованими	$\bar{l} = \frac{\sum_1^n x - \bar{x} }{n}$	$\sigma = \sqrt{\frac{\sum_1^n (x - \bar{x})^2}{n}}$
- згрупованими	$\bar{l} = \frac{\sum_1^m x_j - \bar{x} f_j}{\sum_1^m f_j}$	$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}}$

Очевдно, що дисперсія – це середній квадрат відхилень. Середнє лінійне $\bar{l}$ та середнє квадратичне σ відхилення є безпосередніми мірами варіації. Це іменовані числа (в одиницях вимірювання ознаки), за змістом вони ідентичні, проте завдяки математичним властивостям $\sigma > \bar{l}$. Коли обсяг сукупності досить великий і розподіл ознаки, що варіює, наближається до нормального, то $\sigma = 1,25\bar{l}$, а $R = 6\sigma$. Значення ознаки в межах $\bar{x} \pm \sigma$ мають 68,3% обсягу сукупності, у межах $\bar{x} \pm 2\sigma$ — 95,4%, у межах $\bar{x} \pm 3\sigma$ — 99,7%. Це відоме «правило трьох сигм».

Розглянуті абсолютні характеристики варіації мають одиниці вимірювання ознаки. При порівнянні варіації різних ознак або варіації однієї ознаки в різних сукупностях використовують відносні характеристики - коефіцієнти варіації.

Коефіцієнти варіації розраховують за формулами:

- лінійний

$$V_{\bar{l}} = \frac{\bar{l}}{\bar{x}}; \quad (5.3)$$

- квадратичний

$$V_{\sigma} = \frac{\sigma}{\bar{x}}; \quad (5.4)$$

- осциляції

$$V_R = \frac{R}{\bar{x}}; \quad (5.5)$$

- квартильный

$$V_Q = \frac{R_Q}{2Me} \quad (5.6)$$

Для оцінки однорідності сукупності та порівняння варіацій найбільш часто використовують квадратичний коефіцієнт варіації. В економічних розрахунках вважають, що сукупність є однорідною, а середня - типовою, якщо коефіцієнт варіації не перевищує 10-15% (в математиці ця величина складає 33%). Якщо при розрахунках використовується тільки один вид коефіцієнта варіації, зустрічається його позначення як K_V .

4. Характеристики форми розподілу

Аналіз закономірностей розподілу передбачає оцінювання ступеня однорідності сукупності, асиметрії та ексцесу розподілу.

Однорідність сукупності — передумова використання інших статистичних методів (середніх величин, регресійного аналізу тощо).

В однорідних сукупностях розподіли одновершинні (одномодальні). Багатoverшинність свідчить про неоднорідний склад сукупності, про різнотиповість окремих складових. *Критерієм однорідності* сукупності вважається квадратичний коефіцієнт варіації,

З-поміж одновершинних розподілів є симетричні та асиметричні (скошені), гостро- та плосковершинні. У **симетричному розподілі** рівновіддалені від центра значення ознаки мають однакові частоти, в **асиметричному** — вершина розподілу зміщена. Напрямок асиметрії протилежен напрямку зсуву вершини. Якщо вершина зміщена вліво, то це правостороння асиметрія та навпаки.

В симетричному розподілі характеристики центру мають однакові значення $\bar{x} = Me = Mo$; при правосторонній асиметрії $\bar{x} > Me > Mo$, при лівобічній $\bar{x} < Me < Mo$.

Чим більша асиметрія, тим більше відхилення $(\bar{x} - Mo)$. Очевидно, найпростішою **мірою асиметрії** є відносне відхилення $A = \frac{\bar{x} - Mo}{\sigma}$, яке характеризує напрям і міру скошеності в середині розподілу; при правосторонній асиметрії $A > 0$, при лівосторонній — $A < 0$.

Іншою властивістю одновершинних розподілів є ступінь зосередженості елементів сукупності навколо центра розподілу. Цю властивість називають **ексцесом** розподілу.

Асиметрія та **ексцес** — дві пов'язані з варіацією властивості форми розподілу. Комплексне їх оцінювання виконується на базі **моментів розподілу**.

Момент розподілу - це середня арифметична k -го ступеню відхилення $(x-a)$. В загальному вигляді момент розподілу розраховується за формулою:

$$m_k = \frac{\sum (x-a)^k}{n} \quad (5.7 \text{ а})$$

або

$$m_k = \frac{\sum (x-a)^k * f}{\sum f} \quad (5.7 \text{ б})$$

де m_k - момент k -го порядку,

x - варіанти ряду,

f - частоти ряду,

n - обсяг вибірки,

k та a - постійні числа.

Залежно від величини a моменти розділяють на *початкові* $a=0$, *центральні* $a=x$ та *умовні* $a=x_0$, де x_0 - деяка варіанта ряду, звичайно близька до його середини. Ступінь k визначає порядок моменту.

Початковий момент k -го порядку виражається формулою:

$$M_k = \frac{\sum x^k f}{\sum f} \quad (5.8)$$

Центральний момент k -го порядку виражається формулою:

$$m_k = \frac{\sum (x - \bar{x})^k f}{\sum f} \quad (5.9)$$

Умовний момент k -го порядку виражається формулою:

$$m_k^* = \frac{\sum (x - x_0)^k f}{\sum f} \quad (5.10)$$

Очевидно, що початковий момент 1-го порядку є середня арифметична, 2-го - середній квадрат значень ознаки. Центральний момент 2-го порядку характеризує дисперсію.

Центральні моменти 3-го і 4-го порядків характеризують відповідно асиметрію та ексцес. У симетричному розподілі $m_3=0$. Чим більша скошеність ряду, тим більше значення величини. Для того щоб характеристика скошеності не залежала від масштабу вимірювання ознаки, для порівняння ступеня асиметрії різних розподілів використовується стандартизований момент $As=m_3/\sigma^3$, званий також **коефіцієнтом асиметрії**, який на відміну від коефіцієнта скошеності залежить від крайніх значень ознаки. При *правосторонній* асиметрії коефіцієнт $As>0$, при *лівобічній* $As<0$. Тому правосторонню асиметрію називають позитивною (додатною), а лівобічну – негативною (від'ємною). Вважається, що при $As<0,25$ асиметрія *низька*, якщо As не перевищує $0,5$ - *середня* та при $As>0,5$ - *висока*.

Для вимірювання ексцесу використовують стандартизований момент 4-го порядку $E=m_4/\sigma^4$. При *симетричному*, близькому до нормального розподілі $E=3$, при *гостровершинному* розподілі $E>3$, при *плосковершинному* $E<3$.

Розрахунок центральних моментів m_3 і m_4 за даними інтервального ряду розподілу доцільно проводити за формулами:

$$m_3 = \frac{\sum \left(\frac{x - \bar{x}}{h} \right)^3 * f}{\sum f} * h^3; \quad (5.11)$$

$$m_4 = \frac{\sum \left(\frac{x - \bar{x}}{h} \right)^4 * f}{\sum f} * h^4$$

де h - ширина інтервалу або будь-яке число,
 f - частота або частість інтервалу.

Аналіз закономірностей розподілу можна поглибити, якщо описати його певною функцією.

5. Види та взаємозв'язок дисперсій

Дисперсія, або середній квадрат відхилення σ^2 посідає особливе місце у статистичному аналізі соціально-економічних явищ. Вона є невіддільним і важливим елементом інших статистичних методів, зокрема дисперсійного аналізу.

Для ознак метричної шкали дисперсія — це середній квадрат відхилень індивідуальних значень ознаки від середньої. Залежно від даних вона може бути простою та зваженою:

- для незгрупованих даних

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i} \quad (5.12 \text{ а})$$

- для згрупованих

$$\sigma^2 = \frac{\sum_{j=1}^m (x_j - \bar{x})^2 f_j}{\sum_{j=1}^m f_j} \quad (5.12 \text{ б})$$

Більш зручно проводити розрахунки дисперсії за формулою різниці квадратів:

$$D = \sigma^2 = \overline{x^2} - (\bar{x})^2 \quad (5.13)$$

де $\overline{x^2}$ - середній квадрат значень варіюючої ознаки;
 $(\bar{x})^2$ - квадрат середньої величини.

Перетворена формула дисперсії (5.13) має наступний вигляд:

- для незгрупованих даних

$$\sigma^2 = \frac{\sum x^2}{n} + \left(\frac{\sum x}{n} \right)^2;$$

- для згрупованих даних

$$\sigma^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum x f}{\sum f} \right)^2;$$

Дисперсія альтернативної ознаки обчислюється як добуток часток:

$$\sigma^2 = d_1 d_0, \quad (5.14)$$

де d_1 — частка елементів сукупності, яким властива ознака,

d_0 — частка решти елементів ($d_0 = 1 - d_1$).

Застосуємо основну формулу дисперсії до цих характеристик структури:

$$\begin{aligned} \sigma^2 &= \frac{\sum (x - \bar{x})^2 f}{\sum f} = \frac{(1 - d_1)^2 d_1 + (0 - d_1)^2 d_0}{d_1 + d_0} = \\ &= d_0^2 d_1 + d_1^2 d_0 = d_1 d_0 (d_1 + d_0) = d_1 d_0. \end{aligned}$$

Якщо, скажімо, у збиральному цеху частка висококваліфікованих робітників становить $d_1 = 0,2$, то дисперсія частки $\sigma^2 = 0,2(1 - 0,2) = 0,16$.

Очевидно, що за відсутності варіації $\sigma^2 = 0$. Максимальне значення дисперсії складає 0,25, коли $d_1 = d_0 = 0,5$, а саме, коли одиниці сукупності порівну розподілені між двома значеннями.

Дисперсія альтернативної ознаки широко використовується при проектуванні вибірових обстежень, обробці даних соціологічних опитувань, статистичному контролі якості продукції тощо. За відсутності первинних даних про розподіл сукупності припускають, що $d_1 = d_0 = 0,5$ і використовують максимальне значення дисперсії $\sigma^2 = 0,5 \cdot 0,5 = 0,25$ (див. підрозд. 5.4).

У тому випадку, коли номінальна ознака приймає не два, а більше значень дисперсія розраховується за формулою

$$\sigma^2 = d_1 d_2 \dots d_j \dots d_m = \prod d_j \quad (5.15)$$

де d_j - частка j -ї групи;

m - кількість груп.

Дисперсія має певні математичні властивості. Сформулюємо найважливіші з них:

1. Дисперсія постійної величини (A) рівна нулю: $D(A) = 0$.
2. Дисперсія не міняється, якщо всі варіанти збільшити або зменшити на одне й те ж число: $D(x + A) = D(x)$.
3. Якщо всі варіанти помножити (розділити) на число A , дисперсія збільшиться (зменшиться) в A^2 раз.

4. Якщо вагу збільшити або зменшити в одне й те ж число раз, то дисперсія не зміниться.

5. Дисперсія щодо середньої арифметичної менше ніж середній квадрат відхилення від будь-якого числа x_0 на величину $(x_0 - \bar{x})^2$. Ця властивість носить назву властивості мінімальності дисперсії від середньої.

$$\frac{\sum (x - x_0)^2 * f}{\sum f} = D + (x_0 - \bar{x})^2$$

Практично дисперсія розраховується по спрощуючій обчислення формулі, званій «відліком від умовного нуля»:

$$\sigma^2 = \frac{\sum_{i=1}^m \left(\frac{x_i - x_0}{k} \right)^2 f_i}{\sum f_i} * k^2 - (\bar{x} - x_0)^2 \quad (5.16)$$

де k й x_0 - довільні постійні.

Приклад. Обчислити дисперсію і середнє квадратичне відхилення розподілу підприємств за обсягом випуску продукції за даними таблиці 5.5

Проміжні обчислення при знаходженні середньої арифметичної та дисперсії по спрощуючих формулах мають багато загального. Тому розрахунки можна вести паралельно.

$$\bar{X} = (10/1000) * 0,3 + 16,55 = 16,553 \text{ тис. од.}$$

$$\sigma^2 = \frac{4064}{1000} * 0,32 - (16,553 - 16,55)^2 = 0,3657$$

Тоді середнє квадратичне відхилення

$$\sigma = \sqrt{0,3657} \approx 0,605 \text{ тис. од.}$$

Таблиця 5.5 - Розподіл підприємств за обсягом випуску продукції

Обсяг випуску продукції, тис. од.	Середині інтервалу, x_i	Кількість підприємств, f_i	$x_i - x_0$ ($x_0 = 16,55$)	$\frac{x_i - x_0}{\kappa}$ ($\kappa = 0,3$)	$\frac{x_i - x_0}{\kappa} * f_i$	$(\frac{x_i - x_0}{\kappa})^2 * f_i$
1	2	3	4	5	6	7
14,3-14,6	14,45	1	-2,1	-7	-7	49
14,6-14,9	14,75	2	-1,8	-6	-12	72
14,9-15,2	15,05	8	-1,5	-5	-40	200
15,2-15,5	15,35	26	-1,2	-4	-104	416
15,5-15,8	15,65	65	-0,9	-3	-195	585
15,8-16,1	15,95	120	-0,6	-2	-240	480
16,1-16,4	16,25	181	-0,3	-1	-181	181
16,4-16,7	16,55	201	0	0	0	0
16,7-17,0	16,85	170	0,3	1	170	170
17,0-17,3	17,15	120	0,6	2	240	480
17,3-17,6	17,45	64	0,9	3	192	576
17,6-17,9	17,75	28	1,2	4	112	448
17,9-18,2	18,05	10	1,5	5	50	250
18,2-18,5	18,35	3	1,8	6	18	108
18,5-18,8	18,65	1	2,1	7	7	49
Разом		1000			10	4064

Якщо сукупність розбито на групи за певною ознакою x , то для будь-якої іншої ознаки y можна обчислити дисперсію як у цілому по сукупності, так і в кожній групі. Центром розподілу сукупності в цілому є загальна середня

$$\bar{y} = \frac{\sum_{i=1}^n y}{n} \quad \text{або} \quad \bar{y} = \frac{\sum y_i f_i}{\sum f_i}$$

центром розподілу в j -й групі — групова середня

$$\bar{y}_j = \frac{\sum y}{f_j} \quad \text{або} \quad \bar{y}_j = \frac{\sum y_i f_i}{f_j}$$

де $f_j = \sum f_i$ - обсяг групи j .

Відхилення індивідуальних значень ознаки y від загальної середньої $\bar{y}$ можна подати як дві складові: $(y - \bar{y}) = (y - \bar{y}_j) + (\bar{y}_j - \bar{y})$. Узагальнюючими характеристиками цих відхилень є дисперсії: загальна, групова та міжгрупова.

Загальна дисперсія характеризує варіацію ознаки y навколо загальної середньої:

$$\sigma^2 = \frac{\sum_1^n (y - \bar{y})^2}{n} \quad \text{або} \quad \sigma^2 = \frac{\sum f_i (y_i - \bar{y})^2}{n}$$

Групова дисперсія характеризує варіацію відносно групової середньої:

$$\sigma_j^2 = \frac{\sum_{f_j} (y - \bar{y}_j)^2}{f_j} \quad \text{або} \quad \sigma_j^2 = \frac{\sum f_i (y_i - \bar{y}_j)^2}{f_j}$$

Оскільки в групи об'єднуються певною мірою схожі елементи сукупності, то варіація в групах, як правило, менша, ніж у цілому по сукупності. Якщо причинні комплекси, що формують варіацію в різних групах, неоднакові, то й групові дисперсії різняться між собою.

Узагальнюючою мірою внутрішньогрупової варіації є **середня з групових дисперсій**:

$$\overline{\sigma^2} = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j}$$

Різними є й групові середні $\bar{y}_j$. Мірою варіації їх навколо загальної середньої є **міжгрупова дисперсія**

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^m f_j}$$

Отже, загальна дисперсія складається з двох частин. Перша характеризує внутрішньогрупову, друга — міжгрупову варіацію.

Взаємозв'язок дисперсій називається **правилом (теоремою) розкладання дисперсії (варіації)**:

$$\sigma^2 = \overline{\sigma^2} + \delta^2$$

6. Криві розподілу. Перевірка статистичних гіпотез про відповідність емпіричного та теоретичного розподілів.

Аналіз закономірностей розподілу можна поглибити, якщо описати його певною функцією, яка називається **теоретичною кривою**. Теоретична крива описує закономірність співвідношення варіант та частот. Найпоширенішою є нормальна крива. Вона використовується як стандарт, з яким порівнюють інші

розподіли. Деякі розподіли, які не є нормальними, приводять до такого вигляду перетворенням змінної величини x (наприклад, заміною значень x їх логарифмами $\lg x$).

Нормальний закон розподілу має місце у разі, коли діє велике число дрібних незалежних або слабо залежних випадкових величин (причин), які грають в загальній сумі приблизно однакову роль та серед них не можна виділити головних.

Найважливішою задачею наукового дослідження є встановлення зв'язків між двома величинами або економічними явищами. В економічних явищах часто доводиться розглядати зв'язки як між випадковими, так і зв'язки між випадковими й невипадковими величинами (оскільки не всі чинники, що впливають на економічні процеси – випадкові величини). Вивчення таких зв'язків здійснюється за допомогою кореляційно-регресійного аналізу. Зокрема, даний розділ статистики розглядає стохастичні зв'язки між величинами, коли одна з них y , будучи випадковою величиною, реагує на зміну іншої величини x (як випадкової, так і невипадкової) зміною свого закону розподілу.

Кореляційний аналіз (вивчаючий зв'язки між випадковими величинами) пред'являє жорсткі вимоги до початкової інформації. Зокрема він припускає, що розподіл випадкової величини повинен підпорядковувати нормальному закону.

Регресійний аналіз (вивчаючий зв'язки між випадковими і невипадковими величинами) пред'являє менш жорсткі вимоги. Його проведення можливо навіть у разі деякої відмінності розподілу випадкової величини від нормальної, що особливо важливо в економічних дослідженнях, де розподіл показників часто асиметрично.

Проте, у будь-якому випадку, як правило, вимагається оцінити ступінь близькості закону розподілу випадкової величини до нормального.

Якщо безперервна випадкова величина підкоряється закону нормального розподілу, то вона має наступну щільність розподілу

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} * e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$

У цьому випадку для неї справедлива функція

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{u^2}{2}} du \quad (5.17)$$

Ця функція називається нормованою функцією Лапласа, значення якої табульовані. Вона характеризує площу під кривою в проміжку від 0 до t .

Щоб оцінити вірогідність попадання в інтервал від $-\infty$ до x розраховують $F(x)$.

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (5.18)$$

де $F(x)$ – функція нормального розподілу.

Очевидно, що $F_{(x)} = 1/2 + \Phi(t)$. Для визначення вірогідності попадання нормально розподіленої випадкової величини x в заданий інтервал $(x_1; x_2)$ знаходять різницю $F_{(x_2)} - F_{(x_1)}$, тобто

$$P(x_1 < x < x_2) = F_{(x_2)} - F_{(x_1)} = (1/2 + \Phi(t_2)) - (1/2 + \Phi(t_1)) = \Phi(t_2) - \Phi(t_1) \quad (5.19)$$

Функції $F_{(x)}$ й $\Phi(t)$ базуються на стандартизованих відхиленнях

$$t_1 = \frac{x_1 - \bar{x}}{\sigma}; \quad (5.20)$$

$$t_2 = \frac{x_2 - \bar{x}}{\sigma}$$

де x_1, x_2 – відповідно нижня та верхня межа інтервалу,
 $\bar{x}$ - середньоарифметичне значення x ,
 σ - середнє квадратичне відхилення.

Тоді

$$P(x_1 < x < x_2) = F\left(\frac{x_2 - \bar{x}}{\sigma}\right) - F\left(\frac{x_1 - \bar{x}}{\sigma}\right) = \Phi\left(\frac{x_2 - \bar{x}}{\sigma}\right) - \Phi\left(\frac{x_1 - \bar{x}}{\sigma}\right)$$

Значення функцій табульовані (додатки 5 і 12), наприклад:

t	0	0,33	0,72	1,00	1,38	1,78	2,83	∞
$F_{(x)}$	0,500	0,628	0,764	0,841	0,916	0,962	0,997	1

При негативних значеннях t функція складає $(1 - F_{(x)})$. Наприклад, $F(-1,38) = 1 - 0,916 = 0,084$.

Частоти, відповідні теоретичній кривій, називають **теоретичними**. Для нормального розподілу їх визначають за формулою

$$f' = np_j = n (F_{(x_i)} - F_{(x_{i-1})}) \quad (5.21)$$

де n - обсяг сукупності,

p_j - оцінка вірогідності попадання в j -й інтервал (p можна знайти використовуючи також значення функції $\Phi(t)$ (див. формулу 5.17)).

Приклад. Для визначення теоретичних частот використовуємо дані таблиці 5.6

$$\bar{x} = 3940/80 = 49,25; \quad \sigma = [(195200/80) - 49,252]^{1/2} = 3,8$$

$$t_1 = (44 - 49,25)/3,8 = -1,38 \quad F(-1,38) = 0,084$$

$$t_4 = (56 - 49,25)/3,8 = 1,78 \quad F(1,78) = 0,962.$$

Теоретичні частоти складають, наприклад, для першої групи $f'1 = 80 \cdot 0,076 = 6,08$, для другої $f'2 = 80 \cdot 0,288 = 23,04$.

Таблиця 5.6 - Обчислення теоретичних частот нормальної кривої.

Інтервал	Частота	x	xf	x^2f	t_{i-1}	t_i	$F_{(X_{i-1})}$	$F_{(X_i)}$	$p=F_{(X_i)} - F_{(X_{i-1})}$	f'
40-44	5	42	210	8820	-2,43	-1,38	0,008	0,084	0,076	6,08
44-48	26	46	1196	55016	-1,38	-0,33	0,084	0,372	0,288	23,04
48-52	32	50	1600	80000	-0,33	0,72	0,372	0,764	0,392	31,36
52-56	13	54	702	37908	0,72	1,78	0,764	0,962	0,198	15,84
56-60	4	58	232	13456	1,78	2,83	0,962	0,998	0,036	2,88
Разом	80	$\bar{x}$	3940	195200	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	79,2

Між теоретичними та емпіричними частотами є відхилення, які можуть мати випадковий характер або бути результатом невідповідності теоретичної кривої реальному характеру розподілу. Для об'єктивної оцінки істотності відхилень $f-f'$ використовують критерії згоди, частіше всього критерій Пірсона χ^2 та Колмогорова λ .

Статистичну характеристику критерію χ^2 визначають за формулою

$$\chi^2 = \sum \frac{(f - f')^2}{f'} \quad (5.22)$$

Розрахунок критерію для вищезгаданого прикладу наведено у таблиці 5.7. Тобто χ^2 складає 1,53

Таблиця 5.7- Розрахунок критерію

Номер групи	Частота		Відхилення $f - f'$	$(f - f')^2$	$\frac{(f - f')^2}{f'}$
	f	f'			
1	5	6,08	-1,08	1,166	0,19
2	26	23,04	2,96	8,762	0,38
3	32	31,36	0,64	0,410	0,01
4	13	15,84	-2,84	8,066	0,51
5	4	2,88	1,12	1,254	0,44
Разом	80	79,2	0,8	$\bar{x}$	1,53

Існує два способи оцінки близькості емпіричного розподілу до теоретичного:

1. За першим визначається вірогідність $P(\chi^2)$ досягнення критерієм даної величини. Якщо ця вірогідність перевищує 0,05, то відхилення фактичних частот від теоретичних вважаються випадковими. Якщо ж $P(\chi^2) < 0,05$, то відхилення вважаються істотними, а емпіричний розподіл - відмінним від теоретичного.

2. За іншим способом фактичні значення χ^2 порівнюються з критичними для вірогідності $1-\alpha$ (де α - достатньо мала величина, наприклад 0,01; 0,05; 0,1) та числа ступенів свободи k . **Критичне значення $\chi^2_{1-\alpha}(k)$** - максимально можливе значення χ^2 за умови випадкового походження відхилень $f - f'$. Якщо

фактичне значення перевищує критичне $\chi^2 > \chi^2_{1-\alpha}(k)$, то відхилення між емпіричними та теоретичними частотами слід вважати істотними. В протилежному випадку, істотність відхилень залишається недоведеною.

Величина χ^2 залежить від числа груп, з яких складається сукупність. Тому значення вірогідності $P(\chi^2)$ та критичні значення χ^2 табульовані (див. додаток 11 та додаток 1 відповідно) в залежності від числа ступенів свободи k , яке є різницею між зменшеним на одиницю числом груп в сукупності й числом загальних характеристик теоретичного розподілу. Для нормального розподілу число ступенів свободи $k=m-3$, де m - число груп.

Для нашого прикладу при $\chi^2=1,53$ і $k=2$ визначимо, що $P(\chi^2 > 1,53) = 0,487$. Тобто, вірогідність отримати при двох ступенях свободи $\chi^2 > 1,53$ значно перевищує 0,05. Значить, відхилення фактичних частот від теоретичних можна вважати випадковими, а сам розподіл - нормальним.

Критичні значення χ^2 для вірогідності 0,9 (1-0,1) складають (додаток 1):

Число ступенів свободи	$\chi^2_{0,9}(k)$
2	4,6
3	6,3
4	7,8
5	9,2

В нашому прикладі критичне значення $\chi^2_{0,9}(2)=4,6$. Фактичне значення $\chi^2=1,53$ менше критичного, тобто з вірогідністю 0,9 можна затверджувати, що емпіричний розподіл відповідає теоретичному нормальному закону.

Критерій згоди Пірсона є дуже строгою математичною оцінкою ступеня згоди емпіричного та теоретичного розподілів. Звичайно він застосовується для кількісної оцінки збігу теоретичного й емпіричного розподілу в сукупностях, що містять не менше 50 ознак та при числі ознак в інтервалі не менше 5.

Визначення критерію λ засновано на зіставленні сум накопичених емпіричних та теоретичних частот й проводиться за формулою

$$\lambda = \frac{D}{\sqrt{n}} \quad (5.23)$$

де D - максимальне (по абсолютній величині) значення різниці між накопиченими емпіричними та теоретичними частотами,

$n = \sum f$ - обсяг сукупності або сума емпіричних частот (сума теоретичних частот може дещо відрізнятися від даної величини за рахунок точності обчислень).

Для нашого прикладу $\lambda = 0,281$ або $2,52 : \sqrt{80}$ (див. табл. 5.8).

Таблиця 5.8 - Розрахунок критерію λ .

Номер групи	Накопичені частоти		Відхилення $ S_f - S_f^* $
	емпіричні, S_f	теоретичні, S_f^*	
1	5	6,08	1,08
2	31	29,12	1,88
3	63	60,48	2,52
4	76	76,32	0,32
5	80	79,2	0,8

Далі за допомогою спеціальних таблиць (див. табл.5.9) визначається вірогідність близькості між емпіричним та теоретичним розподілом при знайденому значенні λ

Можливо також порівняння λ з критичним значенням для вірогідності $P(\lambda)$. Якщо критичне значення більше розрахункового, то емпіричний розподіл відповідає теоретичному.

Таблиця 5.9- Таблиця вірогідності для λ

λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
0,30	1,000	1,00	0,270	1,70	0,006
0,40	0,999	1,10	0,178	1,80	0,003
0,50	0,964	1,20	0,112	1,90	0,002
0,60	0,864	1,30	0,068	2,00	0,0007
0,70	0,711	1,40	0,040	2,10	0,0003
0,80	0,544	1,50	0,022	2,20	0,0001
0,90	0,393	1,60	0,012	2,31	0,0000

Як видно з таблиці 5.9 вірогідність близькості між теоретичним та емпіричним розподілами (або вірогідність того, що λ не перевищить 0,281) близька до одиниці.

В загальному випадку слід заздалегідь упевнитися в тому, що аналізований розподіл за природою своєї тотожний нормальному. Тільки при цій умові критерій Колмогорова дає об'єктивну оцінку ступеню їх близькості. Критерій згоди Пірсона χ^2 використовують для оцінки близькості емпіричних та теоретичних частот при аналізі як нормального, так й інших видів розподілів (наприклад тоді, коли випадкова величина має розподіл Пуассона).

Величини χ^2 і k використовуються також для обчислення критерію згоди Романовського

$$\frac{|\chi^2 - k|}{\sqrt{2k}} \quad (5.24)$$

Якщо величина цього критерію менше 3, розбіжність між емпіричним та теоретичним розподілом можна вважати неістотною. Якщо ж цей критерій виявляється більше 3, то слід вважати, що теоретичний розподіл не може служити моделлю для емпіричного, що вивчається.

Розглянуті вище критерії згоди дають загальну оцінку ступеню близькості емпіричного розподілу до нормального, але не містять ніякої інформації про характер розбіжностей між ними. Тому при істотних відхиленнях $f-f'$ аналіз розподілу слідує доповнювати характеристиками асиметрії та ексцесу.

Тема 6. Аналіз концентрації, диференціації та подібності розподілів.

1. Аналіз концентрації розподілів.
2. Аналіз диференціації та подібності розподілів.

1. Аналіз концентрації розподілів.

Дуже важливими у статистичному аналізі є характеристика нерівномірності розподілу певної ознаки між окремими складовими сукупності, а також оцінка концентрації значень ознаки в окремих її частинах (наприклад, розподіл майна чи доходів між окремими групами населення, кількості зайнятих між окремими галузями промисловості, площі сільськогосподарських угідь між окремими агрогосподарствами).

Так, наведені в табл. 6.1 дані про розподіл промислових підприємств регіону за вартістю основних виробничих фондів і за обсягами спожитої електроенергії свідчать про нерівномірне споживання електроенергії. До першої групи належить 20% підприємств, а частка спожитої електроенергії становить 4%. Натомість шоста група містить 3% підприємств, які споживають 46% електроенергії. На відхиленнях часток двох розподілів — за кількістю елементів сукупності d_j і обсягом значень ознаки D_j — ґрунтується оцінювання концентрації.

Таблиця 6.1- До розрахунку коефіцієнту концентрації

Вартість основних виробничих фондів, грн.	У % до підсумку		Модуль відхилення часток $\frac{1}{100} d_j - D_j $
	Кількість підприємств d_j	Спожито електроенергії D_j	
До 5	20	4	0,16
5 — 10	38	5	0,33
10 — 20	22	8	0,14
20 — 50	13	12	0,01
50 — 100	4	25	0,21
100 і більше	3	46	0,43
Разом	100	100	1,28

Якщо розподіл значень ознаки в сукупності рівномірний, то частки однакові — $d_j = D_j$, відхилення часток свідчать про певну концентрацію.

Коефіцієнт концентрації обчислюється як півсума модулів відхилень:

$$K = \frac{1}{2} \sum_{j=1}^m |d_j - D_j|.$$

Значення коефіцієнта коливаються в межах від нуля (рівномірний розподіл) до одиниці (повна концентрація). Чим більший ступінь концентрації, тим більше значення коефіцієнта K . У нашому прикладі $K = 1,28 : 2 = 0,64$, що свідчить про високий ступінь концентрації споживання електроенергії у промисловості регіону.

Коефіцієнти концентрації широко використовуються в регіональному аналізі для оцінювання рівномірності територіального розподілу виробничих потужностей, фінансових ресурсів тощо. За кожним регіоном визначається також *коефіцієнт локалізації*, який характеризує співвідношення часток.

$$L_j = \frac{D_j}{d_j} 100,$$

За даними табл. 6.2 коефіцієнти локалізації свідчать про нерівномірність купівлі (продажу) на душу населення і певною мірою про варіацію життєвого рівня населення різних регіонів.

Таблиця 6.2 – Коефіцієнти територіальної локалізації

Регіон	У % до підсумку		Коефіцієнти локалізації L_j , %
	Чисельність населення d_j	Обсяг товарообороту D_j	
А	30	34	113
В	50	42	84
С	20	24	120
Разом	100	100	*

Порівняння структур на основі відхилень часток доцільне в рядах з нерівними інтервалами, а також в атрибутивних рядах.

2. Аналіз диференціації та подібності розподілів.

За аналогією з коефіцієнтом концентрації обчислюється *коефіцієнт подібності (схожості) структур* двох сукупностей:

$$P = 1 - \frac{1}{2} \sum_{j=1}^m |d_j - d_K|.$$

Якщо структури однакові, $P = 1$; якщо абсолютно протилежні, $P = 0$. Чим більше схожі структури, тим більше значення P . За наведеними у табл. 6.3 даними про галузеву структуру зайнятості населення у двох країнах коефіцієнт подібності структур становить:

$$P = 1 - \frac{|36 - 25| + |24 - 42| + |40 - 33|}{2} = 1 - 0,18 = 0,82,$$

тобто розподіл зайнятих за галузями економіки відхиляється в середньому на 18 п. п.

Таблиця 6.3 – Галузева структура зайнятості населення

Країна	Структура зайнятих, %		
	Сільське господарство	Промисловість та будівництво	Сфера послуг
А	36	24	40
В	25	42	33

Структура будь-якої статистичної сукупності динамічна. Змінюються склад і технічний рівень виробничих фондів, вікова й професійна структура робітників, склад і якість залучених до виробництва природних ресурсів, асортимент і якість продукції, що виробляється, структура споживчого бюджету тощо. Зміна часток окремих складових сукупності свідчить про структурні зрушення. Так, за даними табл. 6.4 структура спожитого в регіоні палива (у перерахунку на умовне) змінилася: зменшились частки газу та мазуту, зросли частки вугілля та інших видів палива. **Інтенсивність структурних зрушень** оцінюється за допомогою середнього лінійного $\bar{l}_d$ або середнього квадратичного σ_d відхилень часток:

$$\bar{l}_d = \frac{\sum_{j=1}^m |d_{j1} - d_{j0}|}{m};$$

$$\sigma_d = \sqrt{\frac{\sum_{j=1}^m (d_{j1} - d_{j0})^2}{m}},$$

де d_{j0} та d_{j1} — частки відповідно базисного та поточного періоду;

m — число складових сукупності.

Таблиця 6.4 – Структура та структурні зрушення споживання палива по роках

Вид палива	1-й рік, d_0	2-й рік, d_1	Відхилення часток, $d_1 - d_0$	Модулі відхилень, $ d_1 - d_0 $	Квадрати відхилень, $(d_1 - d_0)^2$
Вугілля	29	42	13	13	169
Газ	23	16	- 7	7	49
Мазут	45	36	- 9	9	81
Інші види	3	6	+ 3	3	9
Разом	100	100	0	32	308

Лінійний коефіцієнт структурних зрушень становить $\bar{l}_d = \frac{32}{4} = 8$, тобто частки окремих видів палива змінилися в середньому на 8 п. п. Завдяки своїм математичним властивостям квадратичний коефіцієнт структурних зрушень дещо більший — $\sigma_d = \sqrt{\frac{308}{4}} = 8,8$ п. п.

Тема 7. Вибірковий метод

1. Суть вибіркового спостереження.
2. Оцінка генеральної дисперсії по виправленій вибірковій.
3. Вибіркові оцінки середньої та частки. Довірчий інтервал.
4. Різновиди вибірок.
5. Визначення обсягу вибірки та способи розширення вибіркових даних.

1. Суть вибіркового спостереження.

Вибіркове спостереження — такий вид несуцільного спостереження, при якому обстежуються не всі елементи сукупності, що вивчається, а лише певним чином дібрана їх частина. Сукупність, з якої вибирають елементи для обстеження, називається **генеральною**, а сукупність, яку безпосередньо обстежують, — **вибірковою**. Статистичні характеристики вибіркової сукупності розглядаються як оцінки відповідних характеристик генеральної сукупності.

Практика вибіркових спостережень досить різноманітна. Це обстеження домогосподарств, маркетингові дослідження, аудиторські перевірки великих фірм, вивчення громадської думки тощо. При обстеженні невеликої частини генеральної сукупності зменшуються помилки реєстрації, можна розширити й деталізувати програму обстеження. З іншого боку, вибіркове спостереження забезпечує економію матеріальних, трудових, фінансових ресурсів і часу.

При вивченні певного кола соціально-економічних явищ вибіркове спостереження єдино можливе. Це стосується передусім перевірки якості продукції (жирності молока, чистоти та вологості зерна, міцності пряжі тощо). Часом вибіркове спостереження поєднується із суцільним. Наприклад, при перепису населення кожна четверта одиниця спостереження дає докладнішу інформацію. Крім того, вибірковий метод використовують для прискореної обробки матеріалів суцільного спостереження та перевірки правильності даних переписів і одноразових обстежень.

При складанні вибірки можна поступати двома способами: після того, як об'єкт відібраний та над ним проведено спостереження, він може бути повернений або не повернений в генеральну сукупність. Відповідно:

- **повторною** називають вибірку, при якій відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність;
- **бесповторною** називають вибірку, при якій відібраний об'єкт в генеральну сукупність не повертається.

На практиці звичайно користуються неповторним випадковим відбором.

Крім того, вибірка повинна правильно представляти пропорції генеральної сукупності, тобто вибірка повинна бути **репрезентативною** (представницькою).

Але, оскільки вибіркова сукупність не точно відтворює склад генеральної сукупності, то й вибіркові оцінки не збігаються з відповідними характеристиками генеральної сукупності. Розбіжності між ними називають **помилками репрезентативності**: для середньої — це різниця між генеральною $\bar{x}_a$ та вибірковою $\bar{x}$ середніми, для частки — різниця між генеральною p і вибірковою d_1 частками, для дисперсії — між генеральною σ_a^2 та вибірковою σ^2 дисперсією тощо.

За причинами виникнення помилки репрезентативності поділяються на систематичні та випадкові. Систематичні помилки виникають, коли при формуванні вибіркової сукупності порушений принцип випадковості (упереджений вибір елементів, недосконала основа вибірки тощо). Ці помилки для всіх елементів сукупності однонапрямлені і призводять до зсувення результатів обстеження.

При організації вибіркового обстеження важливо уникнути систематичних помилок. Незсуненість — одна з вимог до будь-якої вибіркової оцінки.

Введемо позначення: хай θ^* - статистична (вибіркова) оцінка невідомого параметра θ теоретичного розподілу.

Незсуненою називають статистичну оцінку θ^* , математичне очікування якої дорівнює оцінюваному параметру θ при будь-якому обсязі вибірки, тобто

$$M(\theta^*) = \theta$$

Дотримання цієї вимоги не усуне помилок, проте помилки різних знаків зустрічатимуться однаково часто, тобто це гарантує від отримання систематичних помилок.

Вибіркова середня, знайдена за даними однієї вибірки, є певне число. Якщо витягувати інші вибірки того ж обсягу з генеральної сукупності, то вибіркова середня змінюватиметься від вибірки до вибірки. Тобто вибіркову середню можна розглядати як випадкову величину, а отже, можна говорити про розподіл вибіркової середньої та про числові характеристики цього розподілу (зокрема про математичне очікування й дисперсію вибіркового розподілу).

Розглядаючи вибіркову середню як випадкову величину, можна довести, що

$$M(\bar{X}_e) = \bar{x}_e$$

Тобто вибіркова середня є незсуненою оцінкою генеральної середньої.

Більшість значень вибірових середніх зосереджується навколо генеральної середньої, а отже, найбільшу ймовірність мають відхилення, близькі до нуля. Чим більше відхилення, тим менша його ймовірність. Для будь-якої ймовірності існує межа відхилень вибіркової середньої від генеральної. Використовуючи властивості нормального розподілу, для однієї конкретної вибірки можна визначити:

- помилки репрезентативності — середню та граничну для взятої ймовірності;
- ймовірність того, що помилка вибірки не перевищить допустимого рівня;
- обсяг вибірки, який забезпечить потрібну точність результатів для взятої ймовірності.

2. Оцінка генеральної дисперсії по виправленій вибірковій.

Нехай вимагається за даними вибірки оцінити невідому генеральну дисперсію σ_a^2 . Відомо, що вибіркова дисперсія є зсуненою оцінкою σ_a^2 , іншими словами, математичне очікування вибіркової дисперсії не рівно оцінюваній генеральній дисперсії. Його можна знайти за формулою

$$M(\sigma^2) = \frac{n-1}{n} \sigma^2 \quad (7.1)$$

Очевидно, що легко «виправити» вибіркову дисперсію так, щоб її математичне очікування дорівнювало генеральній дисперсії. Достатньо для цього помножити вибіркову дисперсію на дріб $n/(n-1)$. Зробивши це, отримаємо виправлену дисперсію (її звичайно позначають через s^2)

$$s^2 = \sigma^2 \frac{n}{n-1} \quad (7.2)$$

Перетворимо формулу (7.2)

$$s^2 = \frac{n}{n-1} \sigma^2 = \frac{n}{n-1} \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_e)^2}{n} = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_e)^2}{n-1}$$

Порівнюючи формулу для σ^2 і s^2 , бачимо, що вони відрізняються лише знаменниками. Очевидно, при достатньо великих значеннях n обсягу вибірки вибіркова та виправлені дисперсії розрізняються мало. На практиці користуються виправленою дисперсією, якщо приблизно $n < 30$.

3. Вибіркові оцінки середньої та частки. Довірчий інтервал

У статистиці використовують два типи оцінок параметрів генеральної сукупності — точкові та інтервальні.

Точкова оцінка — це значення параметра за даними вибірки: вибіркова середня $\bar{X}$ та вибіркова частка d_1 .

Інтервальною оцінкою називають інтервал значень параметра, розрахований за даними вибірки для певної ймовірності, тобто **довірчий інтервал**. Чим менший довірчий інтервал, тим точніша вибіркова оцінка.

Якщо знайдена за даними вибірки статистична характеристика θ^* служить оцінкою невідомого параметра θ , то чим менше абсолютна величина різниці $|\theta - \theta^*|$, тим точніше оцінка θ^* визначає параметр θ . Тобто, якщо $|\theta - \theta^*| < \Delta$, де $\Delta > 0$, то чим менше Δ , тим оцінка точніша. Таким чином, позитивне число Δ характеризує точність оцінки.

Інтервальні оцінки дозволяють встановити не тільки точність, а й надійність оцінок.

Надійністю (довірчою ймовірністю) оцінки параметру θ по θ^* називають вірогідність γ , з якою здійснюється нерівність $|\theta - \theta^*| < \Delta$.

$$P(|\theta - \theta^*| < \Delta) = \gamma \quad \text{або} \quad P(\theta^* - \Delta < \theta < \theta^* + \Delta) = \gamma$$

Звичайно надійність оцінки задається наперед (причому в якості γ беруть число, близьке до одиниці. Найбільш часто задають надійність, рівну 0,95; 0,99; 0,999).

Ймовірність появи випадкової помилки вибірки при достатньо великому її обсязі підкоряється закону нормального розподілу та визначається за формулою

$$\gamma = F(t) = 2\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

де γ - довірна ймовірність (надійність) оцінки (значення функції $\Phi(t)$ й $F(t)$ табульовані при різних значеннях t (додатки 5 і 9)).

Межі довірного інтервалу визначаються на основі точкової оцінки та граничної помилки вибірки $\Delta = t\mu$:

- для середньої

$$\bar{x} - t\mu \leq \bar{x}_2 \leq \bar{x} + t\mu ;$$

- для частки

$$d_1 - t\mu \leq p \leq d_1 + t\mu ,$$

де μ — стандартна (середня) помилка вибірки;

t — квантиль розподілу ймовірностей (коефіцієнт довіри, що відповідає ймовірності γ).

Під **середньою помилкою** вибірки розуміють таку розбіжність середньої вибіркової та генеральної сукупностей, яке не перевищує $\pm \sigma_2$.

Граничною помилкою вибірки прийнято вважати максимально можливу розбіжність $(\bar{x}_6 - \bar{x}_2)$.

Довірче число t показує, як співвідносяться гранична та стандартна помилки.

Як бачимо з рис. 7.1, з імовірністю 0,683 гранична помилка не вийде за межі стандартної $\Delta = \pm 1\mu$, з імовірністю 0,954 вона не перевищить $\pm 2\mu$, з імовірністю 0,997 — $\pm 3\mu$. На практиці найчастіше застосовують імовірність 0,954 (на рис. 7.1 незаштрихована частина площини).

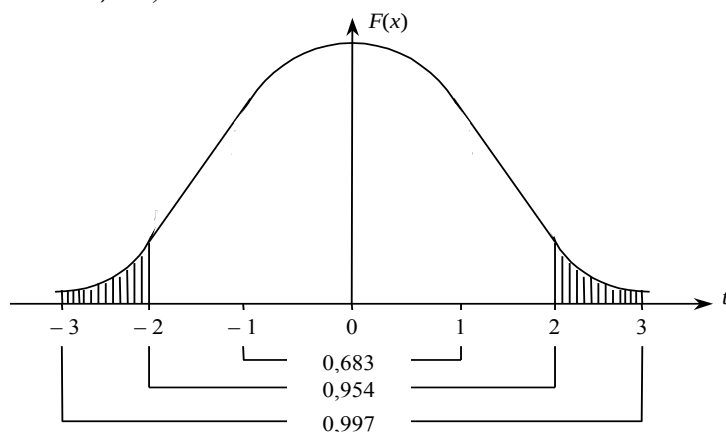


Рисунок 7.1 - Співвідношення ймовірностей та ширини довірчих меж

Величина стандартної помилки залежить від обсягу вибірки та при достатньо великому обсязі вибірки n її можна визначити за формулою

$$\mu = \sqrt{\frac{\sigma_2^2}{n}} = \frac{\sigma_2}{\sqrt{n}}$$

Проте у приведеній формулі σ_2 - показник коливання ознаки в генеральній сукупності, який звичайно невідомий. Але вище ми розглянули, що

$$\sigma_2^2 = \sigma^2 * \frac{n}{n-1}$$

причому при достатньо великих n можна прийняти

На підставі цього твердження у формулах для визначення середньої й граничної помилок замість генеральної дисперсії приймають значення вибіркової дисперсії.

При малих вибірках ($n < 30$), у розрахунках стандартних помилок використовують виправлені вибіркові оцінки та в формулу помилки вибірки вноситься поправка $n/(n-1)$. Тобто

$$\mu_{mv} = \sqrt{\frac{\sigma^2}{n-1}}$$

Квантилі t визначають за розподілом імовірностей Стюдента. У табл. 7.1 наведено деякі значення квантилів t розподілу Стюдента для ймовірності $Pk(t)$ і **числа ступенів свободи** $k = n - 1$. Ймовірність $Pk(t)$ – це довірча ймовірність, тобто ймовірність того, що помилка вибірки буде не більше заданої величини

$$\Delta_{\hat{\mu}} = t^* \mu_{\hat{\mu}}.$$

Для малих вибірок звичайно її значення приймають рівними 0,95, 0,99, 0,997. Іноді в таблицях дається рівень значущості $\alpha = 1 - Pk(t)$, відповідно рівний 0,05, 0,01, 0,003

Таблиця 7.1 - Квантилі t -розподілу Стюдента (витяг)

$k \backslash Pk(t)$	0,95	0,99	0,997
4	2,776	4,604	6,435
9	2,262	3,250	4,024
14	2,145	2,977	3,583
20	2,086	2,845	3,376

Таким чином, на основі табличної величини t та розрахованого значення $\mu_{\hat{\mu}}$ визначають довірчий інтервал

$$\bar{x}_v - \Delta_{mv} < \bar{x}_z < \bar{x}_v + \Delta_{mv}$$

В теорії вибіркового методу розглядаються два способи відбору: повторний та бесповторний.

Якщо генеральна сукупність обмежена й проводиться бесповторний відбір, то при розрахунку помилки вибірки слід враховувати частку вибіркової сукупності в генеральній $D = n/N$. При цьому розмір граничної помилки середньої величини та частки обчислюється по наступних формулах для відбору:

повторного

$$\Delta = t \sqrt{\frac{\sigma^2}{n}}$$

бесповторного

$$\Delta = t \sqrt{\frac{\sigma^2}{n} (1 - D)}$$

Якщо при вибіркового спостереженні вивчається альтернативна ознака, то **стандартна помилка** вибірки **альтернативної ознаки** визначається за формулою

$$\mu = \sqrt{\frac{pq}{n}}$$

де pq - дисперсія альтернативної ознаки в генеральній сукупності.

Оскільки pq - невідомо, то на практиці її замінюють дисперсією вибіркової сукупності $d^*(1-d)$ й формула приймає вигляд

$$\mu = \sqrt{\frac{d(1-d)}{n}}$$

Гранична помилка вибірки для альтернативної ознаки визначається по формулах, аналогічних приведенням вище для кількісної ознаки.

В статистичному аналізі часто виникає потреба в порівнянні помилок вибірки різних ознак або однієї й тієї ж ознаки в різних сукупностях. Таке порівняння здійснюють за допомогою коефіцієнтів варіації вибірових оцінок (інша назва - відносна помилка вибірки)

$$V_{\Delta} = \frac{t\mu}{\bar{x}_e} * 100 \quad \text{або} \quad V_{\Delta} = \frac{tV_x}{\sqrt{n}} * 100$$

де $V_x = \sigma/\bar{x}_e$ - коефіцієнт варіації ознаки x .

Для альтернативної ознаки

$$V_{\Delta} = \frac{t\mu}{d} * 100$$

4. Різновиди вибірок.

Формування вибірки виконується за певними правилами. Передусім визначається основа вибірки. У сукупностях, які складаються з «фізичних» елементів, одиниця основи може репрезентувати або окремий елемент сукупності, або певне їх угруповання. Найбільш вживаними способами відбору є: простий випадковий, систематичний, типовий, серійний, багатоступеневий, багатofазний.

Застосування цих форм організації вибіркового спостереження викликається або міркуваннями здешевлення, або полегшення процесу спостереження, або особливим характером об'єкту спостереження, або відсутністю необхідної інформації для складання списку одиниць спостереження.

Простий випадковий відбір здійснюють за допомогою жеребкування або таблиць випадкових чисел.

Це класичний спосіб формування вибіркової сукупності. Проте він передбачає досить складну підготовку до формування вибірки. Так, для жеребкування на кожну одиницю

відбору необхідно підготувати відповідну фішку; при використуванні таблиць випадкових чисел всі елементи генеральної сукупності повинні бути пронумеровані. Для великих за обсягом сукупностей така робота недоцільна. Тому на практиці застосовують інші різновиди вибірок.

Якщо власне випадкова вибірка організовується як повторна, то її помилка визначається за формулою

$$\Delta = t^* \sqrt{\frac{\sigma^2}{n}} \quad (7.3)$$

При неповторному відборі помилка розраховується за формулою

$$\Delta = t^* \sqrt{\frac{\sigma^2}{n} (1 - D)} \quad (7.4)$$

де $D = \frac{n}{N}$ - частка вибіркової сукупності в генеральній.

Систематичний (механічний) відбір. Основа вибірки — упорядкована множина елементів сукупності. Крок інтервалу визначають розподілом обсягу сукупності N на передбачений обсяг вибірки n . Початковий елемент відбору визначають як випадкове число в середині першого інтервалу. Так, для частки вибірки $D = n/N = 0,05$ кроком інтервалу є число $N/n = 1/0,05 = 20$. Тобто, у вибірку повинен потрапити кожний двадцятий елемент. Якщо початковий елемент - 7, то 2-й буде $7 + 20 = 27$ й т.д.

Помилку систематичної вибірки визначають за формулою простої випадкової неповторної вибірки (формула 7.4).

Типовий (районуючий) відбір - це спосіб формування вибірки з урахуванням структури генеральної сукупності.

Якщо у складі генеральної сукупності присутні різні типи явищ, то у вибірці необхідно забезпечити рівномірне представництво різних частин (типів) явища. Тобто типовий відбір передбачає попередню структурування генеральної сукупності й незалежний відбір елементів в кожній складовій частині. Наприклад, при обстеженні бюджетів робітників виділяють групи робітників в окремих галузях промисловості.

Обсягом типової вибірки є сума обсягів типових груп n_j , тобто

$$n = \sum_{j=1}^m n_j \quad (\text{де } m - \text{число складових частин (груп, типів)}).$$

При визначенні помилки типової вибірки використовують середню з групових дисперсій

$$\bar{\sigma}^2 = \frac{\sum \sigma_j^2 n_j}{\sum n_j}$$

Якщо сформовані групи об'єднують «схожі елементи», а групові середні помітно відрізняються, то варіація ознаки в групах буде значна менша, ніж в цілому по сукупності. У такому разі $\bar{\sigma}^2 < \sigma^2$, а це означає, що помилка типової вибірки в порівнянні з простою випадковою або систематичною буде меншою.

Помилка типової вибірки дорівнює

$$\Delta = t^* \sqrt{\frac{\bar{\sigma}^2}{n} (1 - D)}$$

Щоб забезпечити більшу точність типової вибірки необхідно обґрунтувати:

- ознаку розділення сукупності на групи;
- число складових частин m ;
- обсяг типових груп n_j ;
- спосіб відбору.

а) При визначенні *ознаки розділення сукупності* на групи враховують наступне: варіація ознаки зменшується за умови, якщо ознака розбиття сукупності на групи корелює з ознакою, характеристика якої розглядається.

Відповідно до правила складання дисперсій

$$\bar{\sigma}^2 = \sigma^2 - \delta^2,$$

або

$$\bar{\sigma}^2 = \sigma^2 * (1 - \eta^2)$$

де η^2 - кореляційне відношення, яке визначає тісноту зв'язку між ознаками, які відображають причину й слідство.

Таким чином, розбиття сукупності на групи зменшує помилку вибірки на складову $1 - \eta^2$. Чим тісніше зв'язок між ознаками, тим помітніше зменшення помилки вибірки. Наприклад, при $\eta^2 = 0,50$ помилка вибірки зменшується удвічі.

б) Для визначення *обсягу типових груп* n_j та в цілому *обсягу вибіркової сукупності* n на практиці використовують різні способи.

Найпростіший з них той, коли всі m груп представлені однаковою кількістю елементів, тобто

$$n_j = \frac{n}{m}$$

Якщо чисельності груп дуже відрізняються, може виникнути ситуація, коли $n_j > N_j$ (де N_j - обсяг j -ї типової групи в генеральній сукупності).

Тому, частіше використовують пропорційний відбір, який передбачає однакове для всіх складових частин представництво, тобто частки d_j

однакові й число спостережень з j -ї типової групи) залежить від обсягу відповідної складової частини

$$n_j = d_j N_j = \frac{n}{N} N_j = n \frac{N_j}{N}$$

Оптимальним щодо мінімізації помилки вибірки (тобто, який враховує також й ступінь варіації ознаки в різних групах генеральної сукупності) є відбір пропорційний середньому квадратичному відхиленню

$$n_j = n \frac{N_j \sigma_j}{\sum N_j \sigma_j}$$

Різновидом типової вибірки є **метод квот**, коли обсяг часткових вибірок n_j встановлюється завчасно. Так, при вивченні громадської думки в інституті Геллапа інтерв'юєру встановлюють квоти, наприклад, обстежувати двох фермерів - чоловіків у віці 30-40 років, трьох жителів міста - у віці 20-30 років і ін.

Серійна вибірка. Одиниця основи вибірки — серія елементів. Серії складаються з одиниць, які пов'язані або територіально (райони, селища), або організаційно (фірми, акціонерні товариства). Вибіркова сукупність серій формується за схемами механічної або простої випадкової вибірки. Дібрана серія розглядається як одне ціле, обстеженню підлягають усі без винятку елементи серії. При визначенні **помилки вибірки** враховують міжсерійну (міжгрупову) варіацію:

- при повторному відборі серій

$$\Delta = t^* \sqrt{\frac{\delta^2}{s}}$$

- при неповторному відборі серій

$$\Delta = t^* \sqrt{\frac{\delta^2}{s} \left(1 - \frac{s}{S}\right)}$$

де S - загальне число серій в генеральній сукупності,

s - число відібраних серій,

δ^2 - міжсерійна (міжгрупова) дисперсія.

$$\delta^2 = (\sum n_j (\bar{x}_j - \bar{x})^2) / n$$

де $\bar{x}_j$ - групова середня групи (серії) j ;

n_j - обсяг групи (серії) j .

Помилка серійної вибірки буде меншою порівняно з помилкою простої випадкової або систематичної вибірки у випадку, якщо серії більш менш однорідні та варіація серійних середніх незначна. Збільшення міжсерійної варіації приводить до збільшення помилки вибірки.

Застосування того або іншого способу формування вибіркової сукупності залежить від мети вибіркового спостереження, можливостей його організації і проведення.

Іноді поєднуються різні способи відбору: механічний і серійний, розшарований і механічний, випадковий і серійний.

Таке поєднання можливе в рамках **багатоступеневої вибірки**. Ступенів може бути два, три й більше. Кожний із них має свою, відмінну від інших основу вибірки. Відповідно поділяються й одиниці вибірки: першого ступеня, другого і т. ін. Повнота охоплення основи й схема добору одиниць на різних ступенях різняться.

Наприклад, формування вибіркової сукупності сімей працівників, зайнятих в промисловості, засновано на таких ступенях:

перший - райони області, 40% які охоплюються вибіркою ($D_1 = 0,4$);

другий – промислові підприємства, відбирають до 20% їх кількості в тих районах, які відібрані на першому ступені ($D_2 = 0,2$);

третій - сім'ї працівників; обстежують 5% їх чисельності в тих районах, які відібрані на другому ступені ($D_3 = 0,05$).

Відбір елементів для безпосереднього обстеження здійснюють на останньому, третьому ступені формування вибіркової сукупності. Частка вибіркової сукупності в генеральній залежить від часток вибірок на всіх ступенях

$$D = 0,4 * 0,2 * 0,05 = 0,004.$$

Тобто, обстеженню підлягає 0,4% сімей області.

При обчисленні помилки такої вибірки враховуються помилки на всіх ступенях відбору. Якщо чисельність груп однакова, то стандартна помилка **багатоступеневої вибірки** визначається за формулою

$$\mu = \sqrt{\mu_1^2 + \frac{\mu_2^2}{n_1} + \frac{\mu_3^2}{n_1 n_2} + \dots}$$

де $\mu_1, \mu_2, \dots, \mu_j$ - стандартні помилки на різних ступенях

$n_1, n_2, \dots, n_j$ - чисельність вибірок на відповідних ступенях відбору (до багатокрапки визначена середня помилка триступінчатої вибірки).

Приведеною формулою можна користуватися й якщо групи неоднакові за обсягом, але загальна частка відбору на всіх ступенях постійна.

Якщо обстежують сукупність за двома й більше ознаками, які різняться варіацією, ефективною є **багатофазна вибірка**. Суть її в тому, що для різних ознак формуються вибіркової сукупності різного обсягу. На відміну від багатоступеневої вибірки багатофазна використовує для всіх ознак одну й ту саму основу вибірки, проте програма обстеження різна.

Вибіркові сукупності формуються поетапно — фазами. З генеральної сукупності утворюється первинна вибірка, а з первинної — підвибірка і т. д. На кожній наступній фазі обсяг підвибірки зменшується, а програма обстеження розширюється. Вибіркові оцінки кожної фази використовуються як додаткова інформація на наступних фазах, що підвищує точність результатів вибіркового обстеження.

При організації багатофазної вибірки можливі комбінації різних способів і видів вибірки. Багатофазна вибірка поєднується з багатоступеневою, а також із суцільним спостереженням.

Стандартна помилка при багатофазній вибірці розраховується на кожній фазі окремо відповідно до формул того способу відбору й виду вибірки, за допомогою якої формувалася її вибіркова сукупність.

В практиці досліджень останнім часом використовують **моментні спостереження**, які використовуються при вивченні процесів. Суть їх — у періодичній фіксації стану процесу на певні моменти часу, які вибирають за схемою випадкової або механічної вибірки (через певні інтервали часу).

Такі спостереження використовують, наприклад, при вивченні структури витрат робочого часу, характеристиці використання устаткування. **Помилку моментної вибірки** визначають за формулою **повторної випадкової** вибірки. Оцінюється значення альтернативної ознаки.

*Наприклад, при обстеженні використання виробничого устаткування на підприємстві в певний момент часу фіксують, працює **r-й** верстат чи ні (якщо ні, з'ясовують причини простою). Припустимо, що в цеху працює 10 верстатів та за 8-годинну зміну щопівгодини проводилася реєстрація використання верстатів. Було зроблено 160 записів ($16 \cdot 10$), в 144 випадках зафіксовано, що верстат працював, в 16 - не працював. Частка працюючих верстатів склала 0,9, дисперсія частки $0,9 \cdot 0,1 = 0,09$. З вірогідністю 0,954 ($t=2$) гранична помилка вибірки рівна*

$$\Delta_p = 2 * \sqrt{\frac{0,09}{160}} = 0,047 \text{ або } 4,7\%.$$

Тобто, частка працюючих верстатів за зміну складала не менше $90 - 4,7 = 85,3\%$.

Що стосується повноти охопту елементів сукупності, то моментне спостереження є суцільним, воно є вибірковим по відношенню до часу, оскільки охоплює не весь час роботи устаткування, а лише певні моменти.

5. Визначення обсягу вибірки та способи поширення вибіркових даних

У процесі проектування вибіркових спостережень визначають **мінімально достатній обсяг вибірки**, при якому вибіркові оцінки репрезентували б основні властивості генеральної сукупності. Занадто великий обсяг вибірки потребує зайвих витрат, а занадто малий призведе до збільшення помилки репрезентативності. Оптимальний обсяг вибірки розраховують:

- для повторного відбору

$$n = \frac{t^2 \sigma^2}{\Delta^2} \quad (7.5)$$

- для неповторного

$$n = \frac{t^2 \sigma^2 N}{\Delta^2 N + t^2 \sigma^2} \quad (7.6)$$

Якщо граничну помилку вибірки замінити коефіцієнтом варіації помилок, то

$$n = \frac{V_x^2 t^2}{V_A^2} \quad (7.7)$$

де V_x - коефіцієнт варіації ознаки x .

При вивченні альтернативної ознаки (частки p) обсяг вибірки визначається як

$$n = \frac{t^2 pq}{\Delta^2} \quad (7.8)$$

Визначення обсягу вибірки передбачає своєчасне обґрунтування допустимого розміру граничної помилки. Певні труднощі викликає відсутність даних про варіацію ознаки. Їх долають, використовуючи приблизні оцінки варіації, обчислені за наслідками аналогічних досліджень або попередніх пробних спостережень. Для альтернативної ознаки використовують максимальне значення дисперсії $\sigma^2 = 0,25$. ($\sigma^2 = pq = p(1-p) = 0,5 * 0,5 = 0,25$).

Наприклад, необхідно визначити чисельність вибірки, яка б забезпечила оцінку потенційної плинності працівників з точністю до 2% при вірогідності 0,954. Дані аналогічних досліджень свідчать, що коефіцієнт варіації плинності складає 30%.

Необхідний обсяг вибірки визначають за формулою (7.7). Відносна гранична помилка складає 2%, або 0,02; $t=2$. Таким чином

$$n = \frac{0,3^2 * 2^2}{0,02^2} = 900$$

тобто передбачену точність вибіркової оцінки при вірогідності 0,954 забезпечить вибірка сукупності, обсяг якої 900 елементів.

Часто одночасно вивчають декілька ознак. Якщо бажаний ступінь точності визначати для кожної ознаки, то в результаті буде отриманий ряд значень обсягу вибірки. Для їх узгодження використовують або максимальний обсяг n (тоді ряд ознак оцінюється «надточно»), або обсяг головної ознаки.

Кінцева мета будь-якого вибіркового спостереження — поширення його характеристик на генеральну сукупність. Для середньої та частки визначаються межі можливих їх значень у генеральній сукупності з певною ймовірністю —

довірчі межі. Якщо метою вибіркового обстеження є визначення обсягових показників генеральної сукупності — обсягів значень ознаки $\sum_{i=1}^N x_i$, то вибірка середня поширюється на генеральну сукупність **прямим перерахунком**:

$$\bar{x}N = \sum_{i=1}^N x_i$$

Наприклад, середня продуктивність праці робітників підприємства складає 25 нат.од./люд., а гранична помилка середньої при вірогідності 0,954 - 0,5 нат.од./люд. У такому разі, якщо загальна чисельність робітників на підприємстві складає 2000 чоловік, то можливий обсяг виробництва підприємства буде не меншим 49 тыс.нат.ед. (2000 (25-0,5)). Максимальний обсяг виробництва складе 51 тыс.нат.ед. (2000(25+0,5)).

Коли вибіркове спостереження проводиться з метою уточнення результатів суцільного спостереження, застосовується **метод коефіцієнтів**.

Наприклад, після перепису невстановленого устаткування (що проводиться через 2 роки) можна провести 10%-й вибіркового контролю, під час якого визначають відсоток недообліку невстановленого устаткування. Припустимо, відповідно до даних перепису, які потрапили у вибірку, кількість одиниць устаткування складає 200, а за даними вибірки - 205. Тобто, недооблік устаткування складає $(5/200)*100=2,5\%$. Цей коефіцієнт поширюють на генеральну сукупність. Дані переписи $N=10000$ коректуються - $10000*1,025=10250$ одиниць устаткування.

Тема 8. Статистичні методи вимірювання взаємозв'язків.

1. Види взаємозв'язків.
2. Метод аналітичних угруповань.
3. Регресійний аналіз.
4. Множинна регресія.
5. Рангова кореляція.

1. Види взаємозв'язків

Всі явища суспільного життя знаходяться у взаємозв'язку. Вивчення цих взаємозв'язків є передумовою їх подальшого використання при прогнозуванні явищ та показників. Визначальна мета вимірювання взаємозв'язків — виявити і дати їх кількісну характеристику..

По статистичній природі зв'язки розділяють на функціональні та стохастичні. Ознаки, які при вивченні зв'язків характеризують результат, називають результативними, а ознаки, що характеризують причину, називають факторними.

При **функціональному зв'язку** кожному можливому значенню факторної ознаки x відповідає чітко визначене значення результативної ознаки y . Для дослідження функціональних зв'язків в статистиці застосовують індексний метод.

Частіше в економічних дослідженнях зустрічаються **стохастичні залежності**. Вони виявляються тільки в середньому по значній кількості об'єктів (спостережень) та виявляються як узгодженість варіації двох чи більше ознак. У ланці зв'язку « $x \rightarrow y$ » кожному значенню ознаки x відповідає певна множина значень ознаки y , які утворюють так званий **умовний розподіл**. Стохастичний зв'язок, відбиваючи множинність причин і наслідків, виявляється в зміні умовних розподілів, що схематично ілюструє табл. 8.1.

Наприклад, збільшення фондоозброєності праці робітників дає різний приріст продуктивності праці на різних підприємствах навіть за інших приблизно рівних умов. Це пояснюється тим, наскільки оптимально поєднуються на даних підприємствах ці умови (чинники).

Таблиця 8.1 – Види взаємозв'язків та їх особливості

Факторна ознака x_i	Результативна ознака y за наявності зв'язку		
	функціонального	стохастичного	кореляційного
x_1	y_1	$y_1 y_2$	$\bar{y}_1$
x_2	y_2	$y_1 y_2 y_3$	$\bar{y}_2$
x_3	y_3	$y_2 y_3 y_4$	$\bar{y}_3$
...	...	...	...
x_m	y_m	$y_{m-1} y_m$	$\bar{y}_m$

Якщо умовні розподіли замінюються одним параметром — середньою $\bar{y}_i$ то такий зв'язок називають **кореляційним**. Отже, кореляційний зв'язок є різновидом стохастичного і виявляється зміною середніх умовних розподілів.

Кореляційною залежністю між двома змінними величинами називається така функціональна залежність, яка існує між значеннями однієї з них та груповими середніми іншої.

Кореляційний зв'язок виявляється не у кожному окремому випадку, а тільки при досить великому числі випадків.

Наявність стохастичного зв'язку можна виявити, скориставшись комбінаційним розподілом елементів сукупності. Такий розподіл наведено в табл. 8.2. Сукупність однорідних підприємств регіону поділено на групи за двома ознаками: x — продуктивністю одиниці устаткування (потужністю устаткування) і y — продуктивністю праці робітника.

Таблиця 8.2 – Комбінаційний розподіл підприємств за потужністю устаткування та продуктивністю праці працівників

Продуктивність одиниці устаткування, од./годину	Продуктивність праці робітника од. / люд.-зміну.						Середній рівень продуктивності праці робітника, од./люд.-зміну
	До 20	20—22	22—24	24—26	26 і більше	Разом	
До 300	9	7	1			17	20,0
300—500		8	27	5		40	22,9
500—700			6	15	4	25	24,8
700 і більше				8	10	18	26,1
По сукупності в цілому	9	15	34	28	14	100	23,5

Кожна група за потужністю устаткування характеризується своїм особливим розподілом підприємств за продуктивністю праці робітників. Це умовні розподіли. Порівняння умовних розподілів указує на тенденцію підвищення продуктивності праці зі зростанням потужності устаткування. Для кожного окремого підприємства така залежність може не виявитись через вплив інших факторів, тому певні межі варіації продуктивності праці характерні для кожної групи. Проте середній рівень продуктивності праці в кожній наступній групі підприємств вище ніж в попередній.

$$\begin{aligned}\bar{y}_1 &= \frac{19 \cdot 9 + 21 \cdot 7 + 23 \cdot 1}{17} = 20,0 ; \\ \bar{y}_2 &= \frac{21 \cdot 8 + 23 \cdot 27 + 25 \cdot 5}{40} = 22,9 ; \\ \bar{y}_3 &= \frac{23 \cdot 6 + 25 \cdot 15 + 27 \cdot 4}{25} = 24,8 ; \\ \bar{y}_4 &= \frac{25 \cdot 8 + 27 \cdot 10}{18} = 26,1\end{aligned}$$

Середні рівні продуктивності праці наведено в останній графі таблиці. Зростання групових середніх від групи до групи свідчить про наявність кореляційного зв'язку між показниками.

Важливою характеристикою кореляційного зв'язку є лінія регресії.

Лінія регресії у на х - це залежність, яка зв'язує середні значення ознаки у із значеннями ознаки х.

Лінія регресії може мати різні зображення:

- табличне
- аналітичне
- графічне.

Графіки мають ілюстративне значення.

На табличному та аналітичному зображенні засновано дві головні моделі кореляційного зв'язку - аналітичних угруповань та регресійна. Етапи їх побудови однакові, тобто:

- теоретичне обґрунтування моделі;
- оцінка лінії регресії;
- вимір тісноти зв'язку між ознаками, визначення ролі чинника x в зміні ознаки y ;
- перевірка істотності зв'язку (доказ не випадкового характеру виявлених закономірностей).

2. Метод аналітичних групувань

Метод аналітичних угруповань полягає в тому, що всі елементи сукупності групують за факторною ознакою x і в кожній групі обчислюють середні значення результативної ознаки y , тобто лінія регресії оцінюється лише в окремих крапках.

На **1-му** етапі аналізу вибирають факторну ознаку та визначають число груп й межі інтервалів (див. табл. 8.2).

На **2-му** етапі проводиться оцінка лінії регресії. В моделі аналітичного групування – це **емпірична лінія** регресії, яка представлена груповими середніми результативної ознаки $\bar{y}_j$, кожна з яких належить до відповідного інтервалу значень групувального фактора x_j .

За наведеними даними можна не лише стверджувати, що існує кореляційний зв'язок між факторною x і результативною y ознаками, а й визначити, як у середньому змінюється y зі зміною x на одиницю. **Ефекти впливу** x на y визначаються відношенням приростів середніх групових цих величин $\Delta y : \Delta x$.

Наприклад, у другій групі порівняно з першою потужність устаткування більша на 200 од./годину, а продуктивність праці на $22,9 - 20,0 = 2,9$ од./люд.-зміну. Звідси

$$\frac{\Delta y}{\Delta x} = \frac{2,9}{200} = 0,0145$$

Тобто, зі зростанням потужності устаткування на 100 од./год. продуктивність праці зростає в середньому на 1,45 од. / люд.-зміну (або при зростанні потужності на 1 од. продуктивність праці росте на 0,0145 од.)

Аналогічно розраховані ефекти у третій групі становлять 0,95, у четвертій — 0,65 од./люд.-зміну.

3-й етап - вимір щільності зв'язку - базується на правилі складання дисперсій.

Як зазначалося в темі 5, відхилення індивідуальних значень ознаки y від загальної середньої $\bar{y}$ можна подати як дві складові: $(y - \bar{y}) = (y - \bar{y}_j) + (\bar{y}_j - \bar{y})$. У методі аналітичних групувань - це відхилення індивідуальних значень y від групових середніх та відхилення групових середніх від загальної середньої.

Розглянемо економічний сенс цих відхилень. Всі підприємства однієї групи мають приблизно однаковий рівень потужності устаткування, тобто на групову варіацію рівень продуктивності устаткування не впливає. Групові та середня з групових дисперсій характеризують варіацію результативної ознаки, пов'язану з варіацією всіх ознак, окрім потужності устаткування (наприклад, може враховувати вплив різної електроозброєності й таке інше). Міжгрупова дисперсія характеризує варіацію, пов'язану з варіацією групової ознаки (потужність устаткування). Тому міжгрупову дисперсію називають факторною, а середню з групових – залишковою. Взаємозв'язок цих дисперсій описується правилом складання дисперсій

$$\sigma^2 = \overline{\sigma^2} + \delta^2$$

Очевидно, значення факторної дисперсії буде тим більшим, чим сильніший вплив фактора x на y . Відношення міжгрупової (факторної) дисперсії до загальної розглядається як міра щільності зв'язку за даними аналітичного групування й називається **кореляційним відношенням**

$$\eta^2 = \frac{\delta^2}{\sigma^2}$$

де δ^2 — міжгрупова дисперсія, яка вимірює варіацію ознаки y під впливом фактора x ,
 σ^2 — загальна дисперсія.

Кореляційне відношення показує яка частка загальної варіації результативної ознаки y пояснюється варіацією групової ознаки x .

Застосуємо кореляційне відношення для оцінювання щільності зв'язку між показниками за даними таблиці 8.2. Розрахунки загальної та факторної дисперсій подано в табл. 8.3 та 8.4.

Згідно з розрахунками загальна дисперсія становить 5,19, факторна — 3,86:

$$\sigma^2 = \frac{\sum_{i=1}^{m_y} (y_i - \bar{y})^2 f_i}{\sum_{i=1}^{m_y} f_i} = \frac{519}{100} = 5,19$$

;

$$\delta^2 = \frac{\sum_1^{m_x} (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^{m_x} f_j} = \frac{386,6}{100} = 3,866.$$

Кореляційне відношення

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{3,866}{5,19} = 0,745$$

тобто 74,5% варіації продуктивності праці на підприємствах регіону пояснюється варіацією продуктивності одиниці устаткування.

Таблиця 8.3 – Розрахунок загальної дисперсії в аналітичному групуванні
($\bar{y} = 23,5$)

Продуктивність праці робітника, од./люд.-зміну	18—20	0—22	22—24	24—26	26—28	Разом
Кількість підприємств f_i	9	15	34	28	14	100
y_i	19	21	23	25	27	×
$y_i - \bar{y}$	− 4,5	−2,5	−0,5	1,5	3,5	×
$(y_i - \bar{y})^2 f_i$	182,25	93,75	8,5	63,0	171,5	519

Таблиця 8.4 – Розрахунок факторної дисперсії в аналітичному групуванні
($\bar{y} = 23,5$)

Продуктивність одиниці устаткування, од./годину	f_j	$\bar{y}_j$	$\bar{y}_j - \bar{y}$	$(\bar{y}_j - \bar{y})^2 f_j$
До 300	17	20,0	−3,5	208,25
300 — 500	40	22,9	−0,6	14,40
500 — 700	25	24,8	1,3	42,25
700 і більше	18	26,1	2,6	121,68
У цілому	100	23,5	×	386,58

4-й етап - перевірка істотності зв'язку Кореляційне відношення змінюється від 0 до 1. При $\eta^2 = 1$ міжгрупова дисперсія рівна загальній, тобто кожному значенню факторної ознаки відповідає одне значення результативної

ознаки й зв'язок між ознаками функціональний. При $\eta^2=0$ міжгрупова дисперсія дорівнює нулю.

Проте, умова $\eta^2>0$ не завжди є доказом наявності кореляційного зв'язку між ознаками. Відмінне від нуля кореляційне відношення може з'явитися й при випадковому розподілі сукупності на групи.

Перевірка істотності кореляційного зв'язку ґрунтується на порівнянні фактичного значення η^2 з **критичним**, яке є максимально можливим значенням кореляційного відношення, яке може виникнути випадково за відсутності кореляційного зв'язку. Якщо фактичне значення η^2 більше критичного, то зв'язок між ознаками не випадковий (істотний).

Критичне значення позначається $\eta_{1-\alpha}^2(k_1, k_2)$, тобто розподіл η^2 в таблицях критичних значень (макет таблиці наведено таблицею 8.5) надається для рівня істотності α (звичайно 0,05 або 0,01) та залежить від числа ступенів свободи міжгрупової k_1 та середньої з групових k_2 дисперсій.

$$k_1 = m - 1;$$

$$k_2 = n - m$$

де n, m - відповідно число одиниць сукупності та кількість груп.

Визначимо критичне значення кореляційного відношення для нашого прикладу: $k_1 = 4 - 1 = 3$ та $k_2 = 100 - 4 = 96$. Оскільки значення $k_2 = 96$ у таблиці критичних значень відсутнє, можна використати найближче до нього число $k_2 =$

100. Критичне значення $\eta_{0,95}^2(3,100) = 0,075$.

Розраховане кореляційне відношення $\eta^2 = 0,745$ значно перевищує критичне, а отже зв'язок між потужністю одиниці устаткування та продуктивністю праці робітника з імовірністю 0,95 визнається істотним.

Для перевірки істотності зв'язку використовують також функціонально пов'язану з η^2 характеристику *F-критерій* (критерій Фішера), який визначається за формулою:

$$F = \frac{\eta^2}{1 - \eta^2} * \frac{k_2}{k_1}$$

Перевірку значущості зв'язку з його допомогою здійснюють аналогічно описаному вище для кореляційного відношення. Результати перевірки будуть ідентичні.

Таблиця 8.5 - Критичні значення кореляційного відношення η^2 й коефіцієнта детерміації R^2 для рівня значущості $\alpha=0,05$

k_2	k_1								
	1	2	3	4	5	6	8	10	20
16	0,219	312	378	429	477	507	564	609	740
17	208	297	361	411	452	488	545	590	724
18	197	283	345	394	435	470	527	573	709
19	187	270	331	379	419	454	510	556	694
20	179	259	318	364	404	432	495	540	680

3. Регресійний аналіз

В регресійному аналізі оцінюється теоретична лінія регресії.

Теоретична лінія регресії описується певною функцією $Y = f(x)$, яку називають **рівнянням регресії**, а Y — теоретичним рівнем результативної ознаки.

Тобто оцінка лінії регресії здійснюється не в окремих точках, а в кожній точці інтервалу зміни факторної ознаки x . Розглянемо парну модель.

На **I-му** етапі - при невеликому обсязі сукупності доцільно будувати кореляційне поле, де кожному елементу сукупності відповідає крапка. Загальний вид кореляційного поля дозволяє зробити висновок відносно форми лінії регресії. При значному обсязі сукупності використовують групові середні.

Для відображення особливостей зв'язків конкретних явищ, статистика використовує різні за функціональним видом регресійні рівняння. Якщо зі зміною фактора x результат y змінюється більш-менш рівномірно, такий зв'язок описується лінійною функцією $Y = b_0 + b_1x$. Коли йдеться про нерівномірне співвідношення варіацій взаємозв'язаних ознак, застосовують нелінійні регресії, зокрема:

степеневу $Y = ax^b$;

гіперболічну $Y = b_0 + \frac{b_1}{x}$;

параболічну $Y = b_0 + b_1x + b_2x^2$

Найбільш поширеною функцією є лінійна $Y = b_0 + b_1x$. Параметр b_1 (**коефіцієнт регресії**) — величина іменована, має розмірність результативної ознаки і розглядається як **ефект впливу** x на y . Параметр b_0 — вільний член рівняння регресії, це значення y при $x = 0$. Якщо межі варіації x не містять нуля, то цей параметр має лише розрахункове значення

На **2-му** етапі визначають параметри рівняння методом найменших квадратів, основною умовою якого є мінімізація суми квадратів відхилень емпіричних значень у від **теоретичних** Y

$$\sum (y - Y)^2 = \min$$

Математично можна довести, що значення параметрів b_0 й b_1 , при яких мінімізується сума квадратів відхилень, визначаються із системи нормальних рівнянь:

$$\begin{cases} \sum y = nb_0 + b_1 \sum x \\ \sum xy = b_0 \sum x + b_1 \sum x^2 \end{cases}$$

Значення $\sum x, \sum y, \sum xy, \sum x^2$ розраховуються на основі фактичних вихідних даних.

Якщо форма зв'язку $Y = b_0 + b_1 x + b_2 x^2$, то для визначення його параметрів потрібно вирішити наступну систему рівнянь

$$\begin{cases} b_0 + b_1 \sum x + b_2 \sum x^2 = \sum y \\ b_0 \sum x + b_1 \sum x^2 + b_2 \sum x^3 = \sum yx \\ b_0 \sum x^2 + b_1 \sum x^3 + b_2 \sum x^4 = \sum yx^2 \end{cases}$$

Для визначення параметрів гіперболи необхідно вирішити наступну систему рівнянь:

$$\begin{cases} nb_0 + b_1 \sum (1/x) = \sum y \\ b_0 \sum (1/x) + b_1 \sum (1/x)^2 = \sum (1/x) y \end{cases}$$

Приклад. Обчислити параметри лінійної регресії за даними про місячний обсяг виробництва на підприємствах галузі (млн. нат. од.) та її собівартість (млн. грош. од.). Значення взаємозв'язаних ознак та необхідні для розрахунку параметрів величини наведено в табл. 8.6.

Таблиця 8.6 – До розрахунку параметрів лінійної регресії

Номер підприємства	Обсяг виробництва, млн. од., x ,	Валові витрати y , млн. грош. од.	xy	x^2
1	1,1	23	25,3	1,21
2	1,4	25	35,0	1,96
3	1,2	26	31,2	1,44
4	2,0	33	66,0	4,00
5	1,5	27	40,5	2,25
6	1,3	2,8	36,4	1,69
7	1,8	30	54,0	3,24
8	1,7	32	54,4	2,89
Разом	12,0	224	342,8	18,68

$$\sum x = 12; \quad \sum y = 224; \quad \sum xy = 342,8; \quad \sum x^2 = 18,68;$$

Тобто необхідно рішити систему рівнянь:

$$\begin{cases} 224 = 8b_0 + 12b_1 \\ 342,8 = 12b_0 + 18,68 b_1 \end{cases}$$

Рівняння регресії має вигляд

$$Y = 13,0 + 10,0x,$$

тобто кожна одиниця зростання обсягу виробництва дає приріст валових витрат в середньому 10 млн. грош. од. Якщо обсяг виробництва нульовий ($x = 0$), то підприємство несе постійні витрати в суммі 13,0 млн. грош. од.

Рівняння регресії відбиває закон зв'язку між x і y не для окремих елементів сукупності, а для сукупності в цілому; закон, який абстрагує вплив інших факторів, виходить з принципу «за інших однакових умов».

На **3-му** етапі відбувається визначення щільності зв'язку. Воно, як і в методі аналітичних угруповань, засновано на правилі складання дисперсій.

Але, так як в регресійному аналізі оцінками лінії регресії є її теоретичні значення, то відхилення індивідуальних значень ознаки y від загальної середньої $\bar{y}$ можна подати як наступні дві складові:

- відхилення від лінії регресії ($y - Y$);
- відхилення лінії регресії від середньої ($Y - \bar{y}$).

Відхилення $(Y - \bar{y})$ є наслідком дії фактора x , відхилення $(y - Y)$ — наслідком дії інших факторів.

Взаємозв'язок факторної та залишкової дисперсій описується правилом:

$$\sigma_y^2 = \delta_Y^2 + \sigma_e^2,$$

де σ_y^2 — загальна дисперсія ознаки y ;

δ_Y^2 - факторна дисперсія;
 σ_e^2 - залишкова дисперсія

$$\sigma_y^2 = \frac{1}{n} \sum_1^n (y - \bar{y})^2$$

$$\delta_Y^2 = \frac{1}{n} \sum_1^n (Y - \bar{y})^2$$

$$\sigma_e^2 = \frac{1}{n} \sum_1^n (y - Y)^2$$

де Y , y - відповідно теоретичні та фактичні значення результативної ознаки.

Значення факторної дисперсії δ_Y^2 буде тим більшим, чим сильніший вплив фактора x на y . Відношення факторної дисперсії до загальної розглядається як міра щільності кореляційного зв'язку і називається **коефіцієнтом детермінації** (аналогічний кореляційному відношенню):

$$R^2 = \frac{\delta_Y^2}{\sigma_y^2}.$$

Даний показник є універсальним. Його можна застосовувати при будь-якій формі залежності (лінійній або нелінійній).

Приведені формули факторної та залишкової дисперсій вимагають попереднього визначення (розрахунку) теоретичних значень Y для всіх елементів сукупності.

Щоб зменшити обсяг обчислень, на практиці використовують іншу розрахункову формулу факторної дисперсії (для лінійної регресії)

$$\sigma_Y^2 = \frac{1}{n} (b_0 \sum y + b_1 \sum xy) - (\bar{y})^2$$

Для параболічної регресії другого порядку факторна дисперсія визначається за формулою

$$\sigma_Y^2 = \frac{1}{n} (b_0 \sum y + b_1 \sum xy + b_2 \sum yx^2) - (\bar{y})^2$$

Коефіцієнт детермінації, як і кореляційне відношення, набуває значень від 0 до 1. Він характеризує ступінь близькості кореляційного зв'язку до суворо функціонального. При $R^2=0$ кореляційний зв'язок між y та x відсутній. При $R^2=1$ зв'язок між ознаками є функціональним (оскільки залишкова дисперсія дорівнює нулю).

Корінь квадратний з коефіцієнта детерміації називають **індексом кореляції**. Він також змінюється в межах від 0 до 1 й характеризує тісноту зв'язку, проте економічної інтерпретації не має.

$$R = \sqrt{R^2}.$$

Індекс кореляції збігається з абсолютною величиною **коефіцієнта кореляції** Пірсона. Позначається цей коефіцієнт символом r . Оскільки сфера його використання обмежується лінійною залежністю, то і в назві фігурує слово «лінійний». Обчислення лінійного коефіцієнта кореляції r ґрунтується на відхиленнях значень взаємозв'язаних ознак x і y від середніх.

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}$$

Значення r коливаються від -1 до 1 й характеризують не лише тісноту, але й напрям зв'язку. Позитивне значення r означає прямий зв'язок між ознаками, а негативне - зворотній. Прийнято вважати, що при $|r| = 0$ зв'язок відсутній, при $|r| < 0,3$ - зв'язок слабкий, $0,3 \leq |r| \leq 0,7$ - середній, $|r| > 0,7$ - сильний, $|r| = 1$ - функційний. На практиці застосовують різні модифікації наведеної формули коефіцієнта кореляції. Наприклад,

$$r = \frac{\overline{xy} - \bar{x} * \bar{y}}{\sigma_x \sigma_y}$$

де $\overline{xy} - \bar{x} * \bar{y} = m_{xy}$ - коваріація x по y .

Тому часто зустрічається запис

$$r_{y/x} = \frac{m_{xy}}{\sigma_x \sigma_y}$$

Параметри лінійного рівняння регресії можуть бути також визначені за допомогою рівняння (отриманого на основі використання методу найменших квадратів)

$$y - \bar{y} = r_{y/x} \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

де $\bar{y}, \bar{x}$ - відповідно середнє значення y й x за фактичними даними (даними вибірки)

$r_{y/x}$ - коефіцієнт кореляції y по x

$r_{y/x} \frac{\sigma_y}{\sigma_x} = k$ - коефіцієнт рівняння регресії.

4-й етап. Перевірку істотності зв'язку в регресійному аналізі здійснюють за допомогою тих же критеріїв й по тих же процедурах, що і в аналітичних угрупованнях. Ступені свободи залежать від числа параметрів рівняння регресії

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

де n , m - відповідно кількість спостережень (обсяг вибірки) й кількість параметрів в рівнянні зв'язку, враховуючи вільний член рівняння.

Для лінійної моделі $k_1 = 2 - 1 = 1$.

Значущість зв'язку можна перевірити за допомогою таблиць критичних значень коефіцієнта детермінації R^2 . Якщо фактичне значення R^2 перевищує критичне, це свідчить про істотність зв'язку між факторною та результативною ознаками.

У випадку, якщо використовується *F-критерій*, його фактичне значення визначають за формулою

$$F = \frac{R^2}{1 - R^2} \frac{k_2}{k_1}$$

У невеликих сукупностях коефіцієнт регресії схильний до випадкових коливань, тому слід визначати довірчий інтервал коефіцієнта регресії. Середня помилка коефіцієнта регресії визначається за формулою

$$\mu_b = \sqrt{\frac{\sigma_e^2}{\sigma_x^2 (n - m)}}$$

де σ_x^2 , - варіація факторної ознаки,

σ_e^2 - залишкова дисперсія,

$k = n - m$ - число ступенів свободи,

m — кількість параметрів рівняння регресії (для лінійної функції $m = 2$)

Величина граничної помилки залежить від довірчої ймовірності ($\Delta_B = t\mu_B$, де t - коефіцієнт довіри.)

Перевірити слід й його істотність. Коли зв'язок лінійний, істотність коефіцієнта регресії перевіряють за допомогою *t-критерію* (Ст'юдента), статистична характеристика якого визначається відношенням коефіцієнта регресії b до власної стандартної похибки μ_b , тобто $t = b/\mu_b$. Якщо розраховане значення перевищує критичне для двостороннього *t-критерію*, то гіпотеза про випадковий характер коефіцієнта регресії відхиляється, а отже, з визначеною ймовірністю вплив факторної ознаки на результативну визнається істотним.

Важливою характеристикою регресійної моделі є відносний ефект впливу фактора x на результат y — **коефіцієнт еластичності**:

$$K_E = b \frac{\bar{x}}{\bar{y}}.$$

Він показує, на скільки процентів у середньому змінюється результат у зі зміною фактора x на 1%.

4. Множинна регресія.

Економічні явища залежать від великої кількості факторів. Тому на практиці часто використовують рівняння множинної регресії, коли на величину результативної ознаки впливають два і більш фактори.

Одна з умов кореляційного аналізу - **однорідність досліджуваної інформації**. Критерієм однорідності інформації служать коефіцієнти варіації, які розраховуються по кожному факторному й результативному показнику.

Коефіцієнт варіації показує відносну міру відхилення окремих значень від середньоарифметичної. Він розраховується по формулі

$$V = \frac{\sigma}{\bar{x}} * 100\%,$$
$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Чим більше коефіцієнт варіації, тим відносно більший розкид об'єктів, що вивчаються.

Мінливість варіаційного ряду прийнято вважати за незначну, якщо варіація не перевищує 10%, середньою - якщо складає 10-12%, значною - якщо вона більше 20%, але не перевищує 33%. Якщо вона вище 33%, то це говорить про неоднорідність інформації і необхідності відкинути нетипові спостереження.

На підставі найвищого показника варіації можна визначити необхідний обсяг вибірки даних для кореляційного аналізу по наступній формулі

$$n = \frac{V^2 * t^2}{m^2}$$

де n - необхідний обсяг вибірки даних;

V - варіація %;

t - показник надійності зв'язку (відповідний довірчій вірогідності);

m - показник точності розрахунків % (для економічних розрахунків допускається помилка 5-8%).

Після відбору факторів і оцінки початкової інформації важливим завданням є моделювання зв'язку між факторним і результативним показником. На практиці найчастіше використовують багатофакторні лінійні моделі і моделі, які приводяться до лінійного вигляду відповідними перетвореннями, тобто

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m$$

або

$$\lg Y = b_0 + b_1 \lg x_1 + b_2 \lg x_2 + \dots + b_m \lg x_m$$

та інші.

У разі, коли важко обґрунтувати форму залежності, рішення задачі можна провести по різних моделях і порівняти отримані результати. Адекватність різних моделей фактичним залежностям перевіряється по критерію Фішера, величині коефіцієнта множинної кореляції і детермінації.

Рішення задачі багатфакторного кореляційного аналізу зазвичай проводиться на ЕОМ по типових програмах. Початкові дані вводяться в ЕОМ, яка розраховує матриці парних і приватних коефіцієнтів кореляції, рівняння множинної регресії, а також показники, за допомогою яких оцінюється надійність коефіцієнтів кореляції і рівняння зв'язку.

Коефіцієнти парної кореляції характеризують тісноту зв'язку між показниками з врахуванням взаємозв'язків факторів, що впливають на результативний показник.

Приватний коефіцієнт кореляції виражає тісноту зв'язку між двома ознаками при усуненні змін, викликаних впливом інших ознак. Для їх обчислення проводяться досить трудомісткі розрахунки.

Якщо приватний коефіцієнт кореляції, що характеризує взаємозв'язок факторів між собою, вище 0,85, то один з факторів необхідно виключити з моделі. Наявність лінійного зв'язку між двома факторами називають колінеарністю, а між декількома чинниками - мультиколінеарністю.

Величина коефіцієнтів кореляції є випадковою, залежною від обсягу вибірки. Значущість коефіцієнтів кореляції (зазвичай приватних r_{yx_i}) перевіряється по критерію Стюдента:

$$t = \frac{r}{\sigma_r}$$

де σ_r - середня квадратична помилка коефіцієнта кореляції, яка визначається по формулі

$$\sigma_r = \frac{1 - r^2}{\sqrt{n - 1}}$$

Якщо розрахункове значення t вище табличного, то можна зробити висновок про значущість коефіцієнта кореляції. Табличні значення t знаходять по таблиці значень критеріїв Стюдента. При цьому враховуються кількість мір свободи ($V = n - 1$) і рівень довірчої вірогідності (0,05 або 0,01).

Розрахунок рівняння регресії проводиться зазвичай кроковим способом. Спочатку враховується один фактор, який робить найбільш значущий вплив на результативний показник, потім другий, третій і далі послідовно. І на кожному кроці розраховуються рівняння зв'язку, множинний коефіцієнт кореляції і

детермінації, критерій Фішера й інші показники, що оцінюють надійність рівняння. Величина їх на кожному кроці порівнюється з попередньою. Якщо додавання наступних факторів не покращує оцінних показників зв'язку, то треба їх відкинути.

Розглянемо, яким чином відбувається розрахунок багатofакторного лінійного рівняння.

Параметр рівняння b_i називають **приватним коефіцієнтом регресії**. Він показує, як в середньому змінюється результативна ознака y із зміною факторної ознаки x_i на одиницю за умови, що інші факторні ознаки залишаються незмінними.

Для визначення параметрів $b_0, b_1 \dots$ необхідно скласти і вирішити систему нормальних рівнянь

$$\left\{ \begin{array}{l} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 + ... b_m \sum x_m \\ \sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + ... + b_m \sum x_1 x_m \\ \\ \sum x_m y = b_0 \sum x_m + b_1 \sum x_m x_1 + b_2 \sum x_m x_2 + ... + b_m \sum x_m^2 \end{array} \right.$$

При двох факторах лінійне рівняння регресії має наступний вигляд

$$Y = b_0 + b_1 x_1 + b_2 x_2$$

У цьому випадку система рівнянь набере вигляду

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{cases}$$

Приватні коефіцієнти регресії у багатофакторній моделі при відповідній змінній зазвичай відрізняються від коефіцієнтів при цієї змінній у парній моделі. Причина розбіжностей приватних і парних коефіцієнтів регресії полягає у взаємозв'язку факторних ознак x_1 і x_2 .

Тісноту зв'язку між результативною ознакою і сукупністю факторних ознак визначають за допомогою сукупного коефіцієнта детермінації

$$R_{Y_{X_1 X_2 \dots X_m}}^2 = \frac{\sigma_{Y_{X_1 X_2 \dots X_m}}^2}{\sigma_y^2}$$

де $\sigma_{Yx_1x_2 \dots x_m}^2$ - теоретична дисперсія багатofакторного рівняння регресії.

Сукупний коефіцієнт детермінації характеризує частку варіації результативної ознаки, яка лінійно пов'язана з варіацією включених в рівняння регресії факторних ознак.

Теоретичну дисперсію обчислюють за формулою

$$\sigma_{Y_{x_1 x_2 \dots x_m}}^2 = \frac{1}{n} (b_0 \sum y + b_1 \sum yx_1 + b_2 \sum yx_2 + \dots + b_m \sum yx_m) - (\bar{y})^2;$$

Перевірку значущості зв'язку здійснюють за допомогою F-критерію і коефіцієнта детермінації. При цьому

$$k_1 = m - 1;$$

$$k_2 = n - m.$$

Якщо фактичне значення перевищує критичне, то істотність зв'язку результативної ознаки з обома факторами доведена.

Як вже було вказано вище, коефіцієнти рівняння регресії показують кількісну дію кожного фактору на результативний показник при незмінності інших. Проте, вони мають різні одиниці виміру, що робить їх незіставними, якщо виникає питання про порівняльну силу дії факторів на результативний показник. Щоб привести їх в зіставний вигляд, розраховують стандартизовані коефіцієнти регресії або **бетта-коефіцієнти**.

Бетта-коефіцієнти й коефіцієнти регресії зв'язані наступним чином:

$$\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}$$

Тобто змінні рівняння регресії виражаються в долях середнього квадратичного відхилення. Бетта-коефіцієнти показують, що якщо величина фактору збільшиться на одне середнє квадратичне відхилення, то відповідна залежна змінна збільшиться або зменшиться на долю свого середнього квадратичного відхилення.

Зіставлення бетта-коефіцієнтів дозволяє зробити висновок про порівняльну міру дії кожного фактору на величину результативного показника.

Аналогічно можна зіставити й коефіцієнти еластичності, які розраховуються по формулі:

$$K_E = b_i \frac{\bar{x}_i}{\bar{y}}$$

5. Рангова кореляція

Взаємозв'язок між ознаками, які можна зранжувати, передусім на основі бальних оцінок, вимірюється методами рангової кореляції. **Рангами** називають

числа натурального ряду, які згідно зі значеннями ознаки надаються елементам сукупності і певним чином упорядковують її. Ранжування проводиться за кожною ознакою окремо: перший ранг надається найменшому значенню ознаки, останній — найбільшому або навпаки. Кількість рангів дорівнює обсягу сукупності. Рангові оцінки щільності зв'язку доцільно використовувати для сукупностей невеликого обсягу.

Ранги, надані елементам сукупності за ознаками x і y , позначають відповідно Rx_j та Ry_j . Залежно від ступеня зв'язку між ознаками певним чином співвідносяться й ранги. При прямому функціональному зв'язку $Rx_j = Ry_j$, тобто відхилення між рангами $d_j = Rx_j - Ry_j = 0$, отже, і сума квадратів

$$\sum_{j=1}^m d_j^2 = 0$$

відхилень .

На відхиленнях між рангами базується розрахунок коефіцієнту рангової кореляції, який використовують при вимірі зв'язку між ознаками порядкової шкали. Його обчислюють за формулою Спірмена

$$\rho = 1 - \frac{6 \sum_{j=1}^n d_j^2}{n(n^2 - 1)}$$

де n - число елементів сукупності.

Цей коефіцієнт має такі самі властивості, як і лінійний коефіцієнт кореляції: змінюється в межах від -1 до $+1$, водночас оцінює щільність зв'язку та вказує на його напрям.

Для прикладу використаємо дані таблиці 8.7, в якій приведені ранги 10 ділянок підприємства відповідно їх місцю по кількості порушень вимог правил безпеки (ПБ) й кількості випадків травматизму.

Підставивши дані у формулу коефіцієнта Спірмена, отримаємо

$$\rho = 1 - \frac{6 * 10}{1000 - 10} = 0,939$$

Це свідчить про високий прямий зв'язок між порушеннями вимог ПБ й рівнем травматизму.

Таблиця 8.7 - Розрахунок коефіцієнта рангової кореляції

Ділянка підприємства	Ранги		d	d^2
	Порушення вимог ПБ	Травматизм		
1	7	5	+2	4
2	3	3	0	0
3	4	4	0	0
4	1	2	-1	1
5	8	9	-1	1
6	2	1	+1	1
7	6	7	-1	1
8	10	10	0	0
9	5	6	-1	1
10	9	8	+1	1
Разом	55	55	x	10

Перевіримо істотність зв'язку. Критичне значення коефіцієнта рангової кореляції для рівня значущості $\alpha=0,05$ й $n=10$ складає $\rho_{0,95}(10)=0,56$ (див. табл. 8.8). Фактичне значення ρ більше критичного, тобто наявність зв'язку між ознаками доведена. Відмітимо, що для зворотних зв'язків з критичним порівнюється абсолютна величина фактичного значення ρ .

Рангові коефіцієнти мають як переваги, так й недоліки порівняно з параметричними. Тут не потрібно дотримувати певних математичних передумов щодо розподілу ознак, зокрема, умови нормальності розподілу. Проте, оскільки використовуються не значення, а лише ранги ознак, втрачається певна інформація про взаємозв'язки.

Таблиця 8.8 – Критичні значення коефіцієнту рангової кореляції Спірмена при $\alpha=0,05$

Обсяг вибірки n	5	6	7	8	9	10	11	12
$\rho_{0,95}(n)$	0,90	0,83	0,71	0,64	0,60	0,56	0,53	0,50

Тема 9. Індексний метод.

1. Суть і функції індексів.
2. Методологічні основи побудови зведених індексів. Агрегатні індекси.
3. Середньозважені індекси. Індекси із змінними й постійними вагами.
4. Системи взаємозалежних індексів й визначення впливу окремих чинників.
5. Індекси середніх величин.

1. Суть і функції індексів.

Індекс є співвідношенням двох значень показника, який індексується: оціночного (поточного) і взятого за базу порівняння.

Тобто за статистичною природою індекс — це відносна величина, яка характеризує зміну соціально-економічного явища в часі чи просторі або ступінь відхилення значення показника від певного стандарту (нормативу, середньої).

Індекс, як і будь-який статистичний показник, поєднує в собі якісний та кількісний аспекти. Назва індексу відбиває соціально-економічний зміст показника, числове його значення — інтенсивність змін або ступінь відхилення (наприклад, індекс продуктивності праці в галузі — 1,15).

Методика розрахунку (модель) індексу залежить від мети дослідження, статистичної природи показника, ступеня агрегованості інформації.

За характером порівнянь (у часі, просторі, з певним стандартом) індекси поділяються на *динамічні, територіальні, міжгрупові*. *Динамічний індекс* характеризує інтенсивність динаміки; при його розрахунку базою порівняння є одне з попередніх значень показника. База порівняння ідентифікується підрядковою позначкою «0», поточне значення показника — «1».

При *просторових порівняннях* визначається ступінь відхилення значень показника у просторі — між об'єктами, країнами, регіонами, які ідентифікуються певними літерами; вибір бази порівняння довільний. *Міжгруповий індекс* характеризує відхилення від певного стандарту (еталонного, максимального чи мінімального значення) або від середнього рівня по сукупності в цілому.

За ступенем агрегованості інформації (обхватом елементів сукупності) індекси поділяються на *індивідуальні та зведені*. Вони позначаються відповідно символами *i* та *I*. *Індивідуальні індекси* характеризують співвідношення рівнів показника окремих елементів сукупності, *зведені* — певної множини елементів. У структурованій сукупності зведений індекс може бути *груповим* (субіндексом) або *загальним* (тотальним). Наприклад, динаміку обсягів окремих видів продукції промислового виробництва (чавун, електроенергія, верстати) характеризують індивідуальні індекси, окремих галузей промисловості (металургія, енергетика, машинобудування) — субіндекси, промисловості в цілому — загальний індекс.

Показник, динаміку чи співвідношення якого характеризує індекс, називають *індексованою величиною*, йому надається певний символ.

Наприклад, індивідуальний індекс фізичного обсягу продукції позначають i_q , зведений індекс цін — I_p (имволи p та q відповідають початковим літерам англійських слів **price** (ціна) та **quantity** (кількість)).

Якщо показник вимірює загальний розмір (обсяг) того або іншого явища його називають об'ємним (екстенсивним).

Якщо показник вимірює рівень явища або ознаки з розрахунку на одиницю сукупності, його називають якісним (інтенсивним). Наприклад, собівартість одиниці продукції, витрати робочого часу на одиницю продукції.

Неодмінною умовою обчислення індивідуальних індексів є порівнянність методики вимірювання чисельника та знаменника відношення, що являє собою індекс.

Вивчаючи динаміку індивідуальних явищ слід враховувати наступні **правила**:

1. Добуток ланцюгових індексів дорівнює кінцевому базисному

$$\frac{q_1}{q_0} \times \frac{q_2}{q_1} = \frac{q_2}{q_0}$$

2. Результат від ділення наступного базисного індексу на попередній дорівнює відповідному ланцюговому

$$\frac{q_2}{q_0} : \frac{q_1}{q_0} = \frac{q_2}{q_1}$$

3. Якщо добуток двох або декількох показників є новим показником, що має економічний зміст, то добуток індексів показників-співмножників дорівнює індексу їх добутку

$$i_{pq} = i_p i_q$$

4. Результат від ділення одиниці на індивідуальний індекс прямого показника дорівнює індивідуальному індексу, що характеризує зміну зворотного йому показника.

Щодо зведених індексів, то розрахунок кожного з них окремо передбачає вирішення низки методологічних питань. Основою побудови зведених індексів є агрегування, узагальнення інформації. За своєю формою зведені індекси розділяють на агрегатні й середньозважені. Особливу групу складають індекси середніх величин (індекси змінного та постійного складу, індекс структурних зрушень).

2. Методологічні основи побудови зведених індексів. Агрегатні індекси.

Зведений індекс характеризує зміну складних явищ під впливом різних чинників (співмножників). Він показує, як у середньому змінився показник по сукупності елементів. Наприклад, зміни загального товарообігу окремих видів продукції. Перш ніж будувати зведений індекс, слід привести різні види продукції до зіставного вигляду, тому що соціально-економічні явища та показники можуть бути порівнянними й непорівнянними (наприклад, не можна безпосередньо підсумовувати тонни залізної й марганцевої руди). Приведення до зіставного вигляду можна здійснити за допомогою таких співвимірників, як ціна, собівартість або трудомісткість одиниці продукції.

Наприклад, при вивченні динаміки показника товарообігу за два періода (поточний та базовий) використовують індекс:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

де $p_0 q_0$, $p_1 q_1$ - товарообіг окремих видів товарів відповідно в базисному й поточному періодах.

Щоб вивчити вплив одного з факторів-співмножників на зміну результативного показника, слід інший умовно прийняти незмінним.

Одним з важливих положень теорії індексів є класифікація факторів-співмножників, серед яких виділяють екстенсивні та інтенсивні.

Коли при побудові індексів необхідно один з чинників залишати фіксованим, дотримуються правила: **інтенсивні фактори-співмножники фіксуються на рівні базисного періоду, а екстенсивні - на рівні поточного.**

Незмінні співмножники при побудові індексів грають різну роль. Екстенсивний показник виступає в ролі вагів, інтенсивний - в ролі співвимірника.

Якщо який-небудь з інтенсивних показників позначити через x , а екстенсивний чинник - w , то в загальному вигляді всі зведені двохфакторні індекси матимуть вигляд:

$$I_x = \frac{\sum x_1 w_1}{\sum x_0 w_1}; \quad I_w = \frac{\sum x_0 w_1}{\sum x_0 w_0}; \quad I_{xw} = \frac{\sum x_1 w_1}{\sum x_0 w_0}$$

Перший - це зведений індекс інтенсивного показника (наприклад, цін, собівартості, трудомісткості).

Другий - це зведений індекс екстенсивного показника. У нім можуть бути використані різні співвимірники, пов'язані з індексованим екстенсивним показником (ціна, собівартість й ін.).

Третій індекс характеризує зміну складного явища за рахунок двох чинників.

Форму розрахунку приведених трьох індексів називають *агрегатною*. Такі індекси дозволяють порівнювати рівні складного явища, до складу якого входить низка різнорідних елементів.

3. Середньозважені індекси. Індекси із змінними й постійними вагами

В окремих випадках вихідні дані не дозволяють вивчити динаміку економічного явища на основі агрегатної форми індексу. Другою формою зведеного індексу є середньозважений з індивідуальних індексів. Використовують два види середніх — арифметичну та гармонічну. Вибір виду середньої ґрунтується на загальних засадах: середньозважений індекс має бути тотожним відповідному індексу агрегатної форми.

Наприклад, якщо необхідно охарактеризувати зміну екстенсивного показника в середньому по сукупності різнорідних елементів, можна використовувати середньоарифметичний зважений індекс

$$I_w = \frac{\sum i_w x_0 w_0}{\sum x_0 w_0}$$

де i_w - індивідуальний індекс;

$x_0 w_0$ - ваги.

Легко довести, що цей індекс тотожний зведеному агрегатному індексу. Замінивши i_w в чисельнику на співвідношення w_1/w_0 , отримаємо

$$I_w = \frac{\sum \frac{w_1}{w_0} x_0 w_0}{\sum x_0 w_0} = \frac{\sum x_0 w_1}{\sum x_0 w_0}$$

Для обчислення індексу за формулою середньоарифметичною зваженою використовуємо дані таблиці 9.1

Таблиця 9.1- Зміна вартості та фізичного обсягу реалізації вугілля

Тип вугільної продукції	Обсяг реалізації за базисний період, тис. грн.	Зміна фізичного обсягу реалізації в поточному періоді в порівнянні з базисним %
Рядове вугілля	5,4	-1,0
Збагачене вугілля	14,6	+3,5

Очевидно, в таблиці відсутні дані, необхідні для визначення загального індексу в агрегатній формі. Тому загальний індекс фізичного обсягу продукції визначимо за формулою

$$I_q = \frac{\sum i_q x_0 w_0}{\sum x_0 w_0} = \frac{0,99 * 5,4 + 1,035 * 14,6}{5,4 + 14,6} = \frac{20,403}{20} = 1,022, \text{ або } 102,2\%$$

Тобто, кількість реалізованого рядового та збагаченого вугілля збільшилася в звітному періоді в порівнянні з базисним в середньому на 2,2%.

Середньозважений зведений індекс інтенсивного показника можна обчислити за формулою середньогармоничного індексу

$$I_x = \frac{\sum x_1 w_1}{\sum \frac{x_1 w_1}{i_x}}$$

де i_x - індивідуальний або груповий індекс показника;

$x_1 w_1$ - ваги.

Тотожність цієї форми індексу агрегатною також доводиться простим арифметичним перетворенням.

Для обчислення зведеного індексу цін використовуємо дані таблиці 9.2.

Таблиця 9.2 - Обсяг реалізації і зміна цін на вугілля

Марка вугілля	Обсяг реалізації за поточний період, тис. грн.	Зміна цін в поточному періоді в порівнянні з базисним
ДГСШ	15,3	+2
ДГР	177,6	-4

По формулі середньогармоничного індексу

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} = \frac{15,3 + 177,6}{15,3 / 1,02 + 177,6 / 0,96} = 0,965 \quad \text{або } 96,5\%.$$

Тобто, ціни на вугілля знизилися в середньому на 3,5%.

Очевидно, вибір форми індексу в кожному випадку залежить від економічної суті показників та наявної інформації.

При вивченні явищ більш ніж за два періоди використовують ряди індексів з ланцюговою та базисною системами. Вибір системи залежить від мети дослідження. При побудові таких індексів виникає проблема вибору ваги або вимірника індексів, оскільки вони можуть бути постійними (зафіксованими на рівні одного періоду) або змінними. Найчастіше при рішенні цього питання дотримуються такого **правила**: індекси *інтенсивних* (якісних) показників визначають із *змінними вагами*, а індекси *екстенсивних* (кількісних) показників - з *постійними вимірниками*.

Якщо позначити послідовність періодів символами "0", "1", "2", "3", то можна побудувати такі індексні ряди:

- інтенсивного показника (базисні) із змінними вагами

$$I_x' = \frac{\sum x_1 w_1}{\sum x_0 w_1}; \quad I_x'' = \frac{\sum x_2 w_2}{\sum x_0 w_2}; \quad I_x''' = \frac{\sum x_3 w_3}{\sum x_0 w_3}$$

- екстенсивного показника (ланцюгові) з постійними вимірниками

$$I_w' = \frac{\sum x_0 w_1}{\sum x_0 w_0}; \quad I_w'' = \frac{\sum x_0 w_2}{\sum x_0 w_1}; \quad I_w''' = \frac{\sum x_0 w_3}{\sum x_0 w_2}$$

При обчисленні зведених індексів з постійними вагами (або співвимірниками) діють названі в пункті 9.1 чотири правила про взаємозв'язки між індивідуальними індексами. На зведені індекси із змінними вагами або співвимірниками ці правила не поширюються.

4. Системи взаємозалежних індексів й визначення впливу окремих чинників.

Індексний метод дозволяє оцінити вплив зміни окремих чинників на зміну результативного показника при мультиплікативній формі зв'язку, причому оцінку можна здійснити як у відносному, так й абсолютному вираженні. Це означає обчислити індекси факторних показників відповідно до системи взаємозв'язаних індексів.

У загальному вигляді всі зведені двохфакторні індекси зв'язані так:

$$I_{xw} = I_x I_w$$

Дійсно

$$\frac{\sum x_1 w_1}{\sum x_0 w_0} = \frac{\sum x_1 w_1}{\sum x_0 w_1} * \frac{\sum x_0 w_1}{\sum x_0 w_0}$$

Отже, можна визначити, на скільки відсотків змінився (збільшився або зменшився) результативний показник, у тому числі за рахунок зміни відповідно інтенсивного та екстенсивного показника.

Наприклад, для даних таблиці 9.3.

Таблиця 9.3 - Виробництво продукції та її собівартість

Вид продукції	Кількість виробленої продукції за період, грн.		Собівартість		Розрахункові дані, тис. грн.		
	Базисний q_0	Поточний q_1	Базисна z_0	Поточна z_1	$q_1 z_1$	$q_0 z_0$	$q_1 z_0$
А	400	446	184,0	180,0	80,3	73,6	82,1
В	260	280	74,5	72,0	20,2	19,4	20,9
С	380	300	60,4	63,0	18,9	22,9	18,1
Разом	х	х	х	х	119,4	115,9	121,1

Тоді $I_{zq} = I_z I_q = 0,986 * 1,045 = 1,03$.

Загальні витрати на виробництво в цілому виросли на 3%, при цьому за рахунок зниження собівартості одиниці продукції вони знизилися на 1,4%, а за рахунок збільшення кількості випущеної продукції зросли на 4,5%.

За допомогою методу можна визначити також абсолютний приріст результативного показника за рахунок зміни кожного чинника. Якщо мова йде про один вид продукції, то використовують систему індивідуальних індексів.

Загальний абсолютний приріст

$$\Delta_{xw} = x_1 w_1 - x_0 w_0$$

Його можна розкласти по чинниках (факторах)

$$\Delta_x = x_1 w_1 - x_0 w_1 = w_1 (x_1 - x_0);$$

$$\Delta_w = x_0 w_1 - x_0 w_0 = x_0 (w_1 - w_0).$$

Очевидно, що при такому методі розкладання абсолютного приросту по чинниках Δ_{xw} дорівнюватиме $\Delta_x + \Delta_w$

Розкладання абсолютного приросту по чинниках на основі зведених індексів здійснюють аналогічно індивідуальним індексам

$$\Delta_{xw} = \sum x_1 w_1 - \sum x_0 w_0$$

у тому числі

$$\Delta_x = \sum x_1 w_1 - \sum x_0 w_1;$$

$$\Delta_w = \sum x_0 w_1 - \sum x_0 w_0.$$

За даними табл. 9.3 витрати на виробництво трьох видів продукції в поточному періоді в порівнянні з базисним виросли на 3,5 тис.грн. (119,4 -115,9). За рахунок зниження собівартості одиниці продукції загальні витрати зменшилися на 1,7 тис. грн. (119,4 - 121,1), а за рахунок збільшення продукції – виросли на 5,2 тис. грн. (121,1 - 115,9), тобто 3,5 = (-1,7) + 5,2.

Слід враховувати, що теорія індексів не дає загального методу розкладання абсолютних відхилень узагальнювального показника по факторах, коли їх число більше двох.

5. Індекси середніх величин

Поряд зі зведеними, агрегатними індексами в статистичній практиці широко використовують індекси середніх величин (індекси середньої заробітної плати, середньої собівартості одиниці продукції тощо). При цьому середні визначаються, по-перше, на основі групових середніх, а, по-друге, по однойменній продукції, фізичний обсяг якої можна підсумовувати.

Як відомо, рівень середньої залежить від значень ознаки x_j і структури сукупності:

$$x = \frac{\sum_{j=1}^m x_j f_j}{\sum_{j=1}^m f_j} = \sum_{j=1}^m x_j d_j,$$

де f_j — частота;

d_j — частка j -ї складової сукупності.

Наприклад, середня продуктивність праці робітника по видобутку по виробничому об'єднанню (холдингу) залежатиме як від продуктивності праці на окремих підприємствах, що входять в об'єднання, так і від частки їх в загальній чисельності робітників по видобутку об'єднання.

Очевидно, що й динаміка середньої визначається цими факторами:

а) зміною значень ознаки x_j ,

б) структурними зрушеннями.

Вплив кожного з них на динаміку середньої оцінюється за допомогою системи індексів середніх величин: змінного й фіксованого складу, а також структурних зрушень.

Індексом змінного складу $I_{\bar{x}}$ називають індекс **середньої величини**, він відбиває не лише зміни значень ознаки x , а й зміни в структурі сукупності:

$$I_{\bar{x}} = I_{\bar{x}} = \frac{\sum x_1 w_1}{\sum w_1} : \frac{\sum x_0 w_0}{\sum w_0}$$

де x_0, x_1 - рівні усереднюваного показника,

w_0, w_1 - частоти (ваги) інтенсивного показника.

В **індексі фіксованого складу** I_x ваги постійні, тобто усувається вплив на динаміку середньої структурних зрушень. Його величина показує, як у **середньому** змінилися значення ознаки при незмінній, фіксованій структурі:

$$I_x = I_{\phi c} = \frac{\sum x_1 w_1}{\sum w_1} : \frac{\sum x_0 w_1}{\sum w_1} = \sum x_1 d_1 : \sum x_0 d_1.$$

$$I_x = I_{\phi c} = \frac{\sum x_1 w_1}{\sum x_0 w_1}$$

або

Індекс структурних зрушень I_d , навпаки, показує, як змінилася середня за рахунок структурних зрушень; значення ознаки x фіксуються на постійному рівні:

$$I_d = I_{cз} = \frac{\sum x_0 w_1}{\sum w_1} : \frac{\sum x_0 w_0}{\sum w_0} = \sum x_0 d_1 : \sum x_0 d_0.$$

Між індексами існує взаємозв'язок: $I_{zc} = I_{\phi c} I_{cz}$. Кожен з індексів-співмножників оцінює ступінь впливу відповідного чинника на середній рівень інтенсивного показника.

Обчислимо індекси та розглянемо їх економічний зміст на основі даних таблиці 9.4.

Таблиця 9.4 - Обсяг виробництва цеглини та його собівартість на двох заводах

Завод	Виробництво за період, тис. шт.		Собівартість 1 тис. шт. цеглини за період, тис. грош. од.	
	Базисний	Поточний	Базисний	Поточний
№ 1	400	300	14,0	13,5
№ 2	600	700	13,4	12,0
Разом	1000	1000	x	x

На першому заводі собівартість продукції знизилася на 3,6% ($=13,5/14,0=0,964$), на другому - на 11,4% ($=12/13,4=0,896$).

Для оцінки динаміки собівартості по двох заводах в сукупності використовуємо індекс змінного складу

$$I_z = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{13,5 * 300 + 12,0 * 700}{300 + 700} : \frac{14 * 400 + 13,4 * 600}{400 + 600} = \text{або } 91,3\% \\ = 12,45 : 13,64 = 0,913$$

Таким чином, середня собівартість по двох заводах знизилася на 8,7%. Це сталося, по-перше, оскільки знизилася собівартість на кожному заводі, а, по-друге, оскільки зросла питома вага в загальному виробництві заводу №2, що має нижчу собівартість.

Визначимо вплив кожного чинника. Індекс фіксованого складу рівний

$$I_{\phi c} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_1}{\sum q_1} = 12,45 : \frac{14 * 300 + 13,4 * 700}{700 + 300} = 12,45 : 13,58 = 0,917$$

або 91,7%.

Це означає, що за рахунок зміни собівартості на окремих заводах середня собівартість знизилася на 8,3%.

Індекс структурних зрушень визначимо за формулою

$$I_{cz} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = 13,58 / 13,64 = 0,996 \quad \text{або } 99,6\%.$$

Тобто, зміна структури виробництва відносно двох заводів привела додатково до зниження середньої собівартості на 0,4%.

Тема 10. Аналіз інтенсивності динаміки.

1. Суть і складові елементи динамічного ряду.
2. Характеристики інтенсивності динаміки.
3. Середня абсолютна та відносна швидкість розвитку.

1. Суть і складові елементи динамічного ряду

Суспільні явища безперервно змінюються. Аналіз соціально-економічного розвитку — одне з важливих завдань статистики. Інформаційною базою його слугують динамічні (часові, хронологічні) ряди.

Динамічний ряд — це послідовність чисел, які характеризують зміну того чи іншого соціально-економічного явища. Елементами динамічного ряду є перелік хронологічних дат (моментів) або інтервалів часу і конкретні значення відповідних статистичних показників, які називаються **рівнями** ряду.

При вивченні динаміки важливі не лише числові значення рівнів, а і їх послідовність. Як правило, часові інтервали між рівнями однакові (доба, декада, календарний місяць, квартал, рік). Узявши будь-який інтервал за одиницю, послідовність рівнів запишемо так: $Y_1, Y_2, Y_3, \dots, Y_n$, де n — число рівнів (довжина динамічного ряду)

Залежно від статистичної природи показника-рівня розрізняють динамічні ряди *первинні* й *похідні*, ряди абсолютних, середніх і відносних величин. За ознакою часу динамічні ряди поділяються на інтервальні та моментні. Рівень **моментного ряду** фіксує стан явища на певний момент часу t , наприклад кількість працюючих на початок року, студентів — на 1 вересня і т. д. В **інтервальному ряді** рівень — це агрегований результат процесу й залежить від тривалості часового інтервалу (виробництво електроенергії за рік). Зауважимо, що й похідні показники, обчислені на основі інтервальних рядів, на відміну від моментних залежать від протяжності часу (середньодобове чи середньорічне виробництво електроенергії на душу населення).

Передумовою аналізу динамічних рядів є порівнянність статистичних даних, які його формують, щодо одиниць вимірювання, методології обчислення показників, території, кола об'єктів та інших позицій. Порівнянність даних забезпечується на етапах їх збирання та обробки. Використовують також спеціальні прийоми зведення даних до порівнянного вигляду — «статистичні ключі» зімкнення динамічних рядів. Припустимо, помісячні рівні витрат сировини на виробництво продукції в I півріччі непорівнянні, оскільки у квітні змінився порядок обліку витрат (див. табл. 10.1). Подолати переривчастість ряду можна двома способами. Перший — спосіб відносних рівнів, коли за базу порівняння для кожного ряду беруть квітневий рівень. Два ряди відносних рівнів об'єднуються в один.

Другий спосіб ґрунтується на співвідношенні квітневих рівнів: $55 : 50 = 1,1$. Помноживши рівні першого ряду на цей коефіцієнт, дістанемо єдиний зімкнений (порівнянний) ряд динаміки за весь період (остання графа таблиці).

Таблиця 10.1 – Зімкнення динамічних рядів

Місяці	Обсяг витрат, т		Зімкнений ряд	
	Старий порядок реєстрації	Новий порядок реєстрації	відносних величин, %	абсолютних величин, т
Січень	40	—	80	44,0
Лютий	45	—	90	49,5
Березень	48	—	96	52,8
Квітень	50	55	100	55,0
Травень	—	58	105	58,0
Червень	—	60	109	60,0

2. Характеристики інтенсивності динаміки

Швидкість та інтенсивність розвитку різних суспільних явищ значно варіюють, що позначається на структурі відповідних динамічних рядів. Для оцінювання зазначених властивостей динаміки статистика використовує низку взаємозв'язаних характеристик. Серед них: абсолютний приріст, відносний приріст, темп зростання, інші.

Розрахунок характеристик динаміки ґрунтується на порівнянні рівнів ряду. При порівнянні певної множини послідовних рівнів база порівняння може бути постійною чи змінною. За постійну базу вибирається або початковий рівень ряду, або рівень, який вважається вихідним для розвитку явища, що вивчається. Характеристики динаміки, обчислені відносно постійної бази, називаються *базисними*. Якщо кожний рівень ряду y_t порівнюється з попереднім y_{t-1} , характеристики динаміки називаються *ланцюговими*. Схематично варіанти порівняння ілюструє рисунок 10.1.

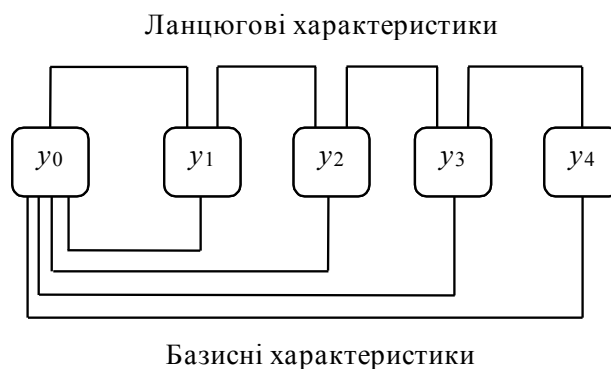


Рисунок 10.1. Схеми порівняння при обчисленні ланцюгових і базисних характеристик динаміки

Абсолютний приріст Δ_t характеризує абсолютний розмір збільшення (чи зменшення) рівня ряду y_t за певний часовий інтервал і обчислюється як різниця рівнів ряду:

базисний приріст $\Delta_t = y_t - y_0$;

ланцюговий приріст $\Delta_t = y_t - y_{t-1}$.

Знак «+», «-» свідчить про напрям динаміки.

Темп зростання k_t показує, у скільки разів рівень y_t більший (менший) від рівня, узятого за базу порівняння. Він являє собою кратне відношення рівнів:

базисний темп $k_t = \frac{y_t}{y_0}$,

ланцюговий темп $k_t = \frac{y_t}{y_{t-1}}$.

При збільшенні рівня $k_t > 1$, при зменшенні — $k_t < 1$. Темпи зростання виражаються як у коефіцієнтах, так і в процентах.

Ланцюгові Δ_t і k_t відображують відповідно абсолютну і відносну швидкість динаміки. Вони взаємозв'язані. Якщо подати $y_t = y_{t-1} + \Delta_t$, то

$$k_t = \frac{y_{t-1} + \Delta_t}{y_{t-1}} = 1 + \frac{\Delta_t}{y_{t-1}}.$$

Отже, при стабільному абсолютному прирості (стабільній абсолютній швидкості) темпи зростання зменшуватимуться. Стабільні темпи зростання можливі за умови прискорення абсолютної швидкості.

Величину $\frac{\Delta_t}{y_{t-1}}$ називають відносним прискоренням або **темпом приросту** і позначають символом T_t . Вона функціонально пов'язана з темпом зростання, але на відміну від останнього завжди виражається в процентах:

$$T_t = 100(k_t - 1).$$

Отже, темп приросту показує, на скільки процентів рівень y_t більший (менший) від бази порівняння.

Співвідношенням абсолютного приросту і темпу приросту визначається **абсолютне значення 1% приросту**. Нескладні алгебраїчні перетворення цього відношення показують, що воно становить соту частину рівня, узятого за базу порівняння:

$$A_t = \frac{\Delta_t}{T_t} = \frac{y_t - y_{t-1}}{100 \left(\frac{y_t - y_{t-1}}{y_{t-1}} \right)} = \frac{y_{t-1}}{100}$$

У таблиці 10.2 наведені всі розглянуті характеристики динаміки на прикладі виробництва умовної продукції за 3 роки. Очевидно, що ланцюгові й базисні характеристики динаміки взаємопов'язані:

а) сума ланцюгових абсолютних приростів дорівнює кінцевому базисному:

$$\sum_{t=1}^n \Delta_t = \sum_{t=1}^n (y_t - y_{t-1}) = y_n - y_0$$

У нашому прикладі: $12 + 9 = 21$ тис. нат.од.;

б) добуток ланцюгових темпів зростання дорівнює кінцевому базисному:

$$k_1 k_2 \dots k_n = \prod_{t=1}^n k_t = K_n = \frac{y_n}{y_0}$$

У нашому прикладі: $1,072 \cdot 1,051 = 1,127$ або $186 : 165 = 1,127$.

Щодо темпів приросту, то вони не мають таких властивостей, як абсолютні прирости чи темпи зростання. Ланцюгові й базисні темпи приросту співвідносяться через темпи зростання.

Таблиця 10.2 – Абсолютні та відносні характеристики динаміки

Порядковий номер року, t	Обсяг виробництва y_t , тис.нат.од.	Абсолютний приріст, тис.нат.од.		Темп зростання		Темп приросту, %		Абсолютне значення 1% приросту, нат.од.	
		Ланцюговий	базисний	Ланцюговий	базисний	Ланцюговий	базисний	Ланцюговий	базисний
0	165	—	—	—	1,0	—	—	—	—
1	177	12	12	1,072	1,072	7,2	7,2	1,65	1,65
2	186	9	21	1,051	1,127	5,1	12,7	1,77	1,65

Якщо швидкість розвитку в межах періоду, що вивчається, неоднакова, порівнянням однойменних характеристик швидкості вимірюється прискорення чи уповільнення динаміки. На базі абсолютних приростів оцінюються **абсолютне та відносне прискорення**. Абсолютне — це різниця між абсолютними приростами: $\delta_t = \Delta_t - \Delta_{t-1}$. Прискорення характеризується додатною величиною $\delta_t > 0$, уповільнення — від'ємною $\delta_t < 0$. Відносне прискорення обчислюється порівнянням абсолютних приростів

Обчислимо характеристики прискорення на прикладі табл. 10.2:

- $\delta_t = 9 - 12 = -3$ тис. т. Знак «мінус» свідчить про уповільнення динаміки.
- темп уповільнення абсолютної швидкості: $9 : 12 = 0,75$.

Порівняння темпів зростання дає **коефіцієнт прискорення (уповільнення)** відносної швидкості розвитку. Для наочності та зручності їх тлумачення дільником є більший за значенням темп зростання.

У нашому прикладі коефіцієнт уповільнення відносної швидкості динаміки $1,072 : 1,051 = 1,02$.

У статистичному аналізі порівнюється також інтенсивність динаміки в різних рядах. Відношення темпів зростання $k' : k''$ називають **коефіцієнтом випередження**. За допомогою останнього порівнюють відносну швидкість динамічних рядів однакового змісту по різних об'єктах (регіони, країни тощо) або різного змісту по одному об'єкту. Наприклад, за 3 роки фондоозброєність праці в одній галузі зросла на 50%, в іншій — на 25%. Коефіцієнт випередження темпу зростання фондоозброєності праці в першій галузі порівняно з другою становить $1,50 : 1,25 = 1,20$.

Можна порівняти динаміку фондоозброєності та продуктивності праці в кожній галузі. Якщо фондоозброєність зросла на 25%, а продуктивність праці — на 37,5%, то коефіцієнт випередження зростання продуктивності праці становить $1,375 : 1,250 = 1,10$.

Щодо темпів приросту, то співвідношення їх використовують лише для взаємозв'язаних показників x і y . Таке співвідношення називають **емпіричним коефіцієнтом еластичності**

$$\gamma = T_y : T_x ;$$

Він показує, на скільки процентів змінюється y зі зміною x на 1%. Наприклад, ціна на товар А зросла на 2%, а попит зменшився на 4%. Цінова

еластичність попиту на цей товар $\gamma = \frac{-4}{+2} = -2$, тобто зі зростанням цін на 1% попит на товар зменшується на 2%.

3. Середня абсолютна та відносна швидкість розвитку

З плином часу варіюють рівні динамічних рядів та обчислені на їх основі абсолютні прирости та темпи зростання. Постає потреба узагальнення притаманних динамічному ряду властивостей, визначення типових характеристик розвитку. Такими характеристиками є середні величини. Зауважимо, що динамічна середня буде типовою характеристикою лише за умови однорідності ряду, коли причинний комплекс формування закономірностей розвитку більш-менш стабільний.

Середні рівні використовують насамперед для узагальнення коливних рядів. Наприклад, при аналізі динаміки виробництва оперують не річними, а більш сталими середньорічними показниками за певні періоди. Середні рівні необхідні також при обчисленні похідних показників. Наприклад, виробництво продукції на одного працюючого. Обсяг продукції — інтервальний показник, а

кількість працюючих — моментний. Щоб забезпечити порівнянність цих показників, слід обчислити середньорічну кількість працюючих.

Метод обчислення середнього рівня динамічного ряду залежить від статистичної структури показника. В інтервальному ряді абсолютних величин, рівні якого динамічно адитивні, використовується середня арифметична проста:

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t,$$

де n — число рівнів ряду.

У моментному ряді, за припущення про рівномірну зміну показника між датами, середня розраховується як півсума значень на початок і кінець періоду:

$$\bar{y} = \frac{y_0 + y_n}{2}.$$

Якщо в моментному ряді $n > 2$ і між суміжними датами однакові інтервали, розрахунок виконується за формулою **середньої хронологічної**:

$$\bar{y} = \frac{\frac{y_1 + y_n}{2} + \sum_{t=2}^{n-1} y_t}{n-1}.$$

(розрахунок такої середньої наведено підрозд. 4.4.)

У моментних рядах з різними інтервалами між датами розраховується середня арифметична зважена:

$$\bar{y} = \frac{1}{\sum D_t} \sum_{t=1}^m y_t D_t,$$

де D_t — інтервал часу між датами,
 m — кількість інтервалів.

Середній абсолютний приріст (абсолютна швидкість динаміки) обчислюється діленням загального приросту за весь період на довжину цього періоду у відповідних одиницях часу (рік, квартал, місяць тощо):

$$\bar{\Delta} = \frac{y_n - y_0}{n} = \frac{\sum_{t=1}^n \Delta_t}{n}.$$

Наприклад, 2010 року автомобільним транспортом області перевезено 2072 тис. т вантажів, 2013 року — 2126 тис. т. Середьорічний приріст цього показника за 2010—2013 рр. становить $\bar{\Delta} = (2126 - 2072) : 3 = 18$ тис. т.

При обчисленні середнього темпу зростання враховується правило складних процентів, за якими змінюється відносна швидкість динаміки (нагромаджується приріст на приріст). Тому **середній темп зростання**

обчислюється за формулою середньої геометричної з ланцюгових темпів зростання:

$$\bar{k} = \sqrt[n]{k_1 \cdot k_2 \dots k_n} = \sqrt[n]{\prod_{t=1}^n k_t},$$

де n — кількість темпів зростання за однакові інтервали часу.

Наприклад, за останні 3 роки зростали тарифи на автоперевезення, при цьому темпи зростання становили: 1-й рік - 1,03; 2-й рік - 1,08; 3-й рік - 1,05. Середньорічний темп зростання

$$\bar{k} = \sqrt[3]{1,03 \cdot 1,08 \cdot 1,05} = 1,053 \text{ або } 105,3 \, \%.$$

Урахувавши взаємозв'язок ланцюгових і базисних темпів зростання, формулу середньої геометричної можна записати так:

$$\bar{k} = \sqrt[n]{K_n} = \sqrt[n]{\frac{y_n}{y_0}}.$$

Скажімо, вартість споживчого кошика за три роки зросла на 12,5%. Середньорічний темп зростання становить

$$\bar{k} = \sqrt[3]{1 + 0,125} = \sqrt[3]{1,125} = 1,04.$$

Тобто щороку споживчий кошик дорожчав у середньому на 4%.

Розрахунок можна виконувати за допомогою логарифмів: $\lg \bar{k} = \frac{1}{n} \sum \lg k_t$ або $\lg \bar{k} = \frac{1}{n} (\lg y_n - \lg y_0)$. Наприклад, прямі іноземні інвестиції в галузь 1996 року становили 172 млн дол. США, 1999 року — 313,7. Десяткові логарифми, відповідно, 2,2355 і 2,4965. Звідси $\lg \bar{k} = \frac{1}{3} (2,4965 - 2,2355) = 0,087$. Потенціювання дає $\bar{k} = 1,222$.

Отже, середній темп зростання можна обчислити на основі:

- ланцюгових темпів зростання k_t ;
- кінцевого (за весь період) темпу зростання K_n ;
- кінцевого y_n і базисного y_0 рівнів ряду.

При інтерпретації середньої абсолютної чи відносної швидкості динаміки необхідно вказувати часовий інтервал, до якого належать середні, та часову одиницю вимірювання (рік, квартал, місяць, доба тощо).

Тема 11. Аналіз тенденцій розвитку та коливань.

1. Характеристика основної тенденції розвитку
2. Оцінка коливань та сталості динаміки

1. Характеристика основної тенденції розвитку

Будь-який динамічний ряд у межах періоду з більш-менш стабільними умовами розвитку виявляє певну закономірність зміни рівнів — **загальну тенденцію**. Одним рядам притаманна тенденція до зростання, іншим — до зниження рівнів.

Тенденція – це певний напрям розвитку, тривала еволюція, яка набуває вигляду більш-менш плавної траєкторії.

Нерідко ряди динаміки через коливання рівнів не виявляють чітко вираженої тенденції. Щоб виявити й схарактеризувати основну тенденцію, застосовують різні способи згладжування та аналітичного вирівнювання динамічних рядів.

Суть **згладжування** полягає в укрупненні інтервалів часу та заміні первинного ряду рядом середніх по інтервалах. У середніх взаєморівноважуються коливання рівнів первинного ряду, внаслідок чого тенденція розвитку вирізняється чіткіше.

Серед методів статистичного описування тенденцій найпростішим є **метод плинних (ковзних) середніх**, коли первинні рівні динамічного ряду замінюються середніми по інтервалах. Кожний наступний інтервал утворюється з попереднього зрушенням на один рівень.

Оскільки середня $\bar{y}_j$ належить до середини інтервалу, то доцільно формувати інтервали з непарного числа рівнів первинного ряду. У разі парного числа рівнів необхідна додаткова процедура центрування (усереднення кожної пари значень $\bar{y}_j$).

Ряд ковзних середніх коротший за первинний на $(m - 1)$ рівнів, що потребує уважного ставлення до вибору ширини інтервалу m . На практиці, як правило, застосовують непарні інтервали ($m = 3, 5, 7$). Якщо первинному ряду динаміки притаманна певна періодичність коливань, то інтервал згладжування має бути рівним або кратним періоду коливань.

Плинну середню r -го інтервалу обчислюють за формулою

$$\bar{y}_r = \bar{y}_{r-1} + \frac{y_{r+p} - y_{r-(p+1)}}{2p + 1}$$

Значення p визначається з умови $m = 2p + 1$

Порядок згладжування методом ковзної середньої розглянемо на прикладі динамічного ряду даних, які наведено у таблиці 11.1. Ширина інтервалу згладжування $m = 3$. Первинний ряд складається із семи рівнів, ряд

ковзних середніх — з п'яти, тобто на два рівні коротший ($7 - (3 - 1)$),
 $p = (3-1)/2=1$

Таблиця 11.1 – Розрахунок ковзних середніх

Порядковий номер року	y_t	Ковзна середня $\bar{y}_j$	Розрахунок $\bar{y}_j$
1	23,8	—	—
2	19,1	21,6	$(23,8 + 19,1 + 21,9) : 3 = 21,6$
3	21,9	22,2	$21,6 + (25,6 - 23,8) : 3 = 22,2$
4	25,6	24,0	$22,2 + (24,5 - 19,1) : 3 = 24,0$
5	24,5	26,2	$24,0 + (28,5 - 21,9) : 3 = 26,2$
6	28,5	26,9	$26,2 + (27,7 - 25,6) : 3 = 26,9$
7	27,7	—	—

Тобто, перше значення ковзної середньої обчислюється як арифметична проста, кожне наступне визначається на основі попередньої середньої та коригуючого доданка. Наприклад:

$$\bar{y}_1 = \frac{23,8 + 19,1 + 21,9}{3} = 21,6 ;$$

$$\bar{y}_2 = 21,6 + \frac{25,6 - 23,8}{3} = 22,2 \quad \text{і так далі}$$

У згладженому ряді трирічних ковзних середніх усунено первинні коливання даних й чітко виявляється систематичне підвищення їх рівня.

Метод ковзних середніх застосовують також для попередньої обробки дуже коливних динамічних рядів; можливе подвійне згладжування.

При **аналітичному вирівнюванні динамічного ряду** фактичні значення y_t замінюються обчисленими на основі певної функції $Y = f(t)$, яку називають **трендовим рівнянням** (t — змінна часу або порядковий номер періоду, Y — теоретичний рівень ряду).

Вибір типу функції ґрунтується на теоретичному аналізі суті явища, яке вивчається, і характері його динаміки. Зазвичай перевага надається функціям, параметри яких мають чіткий економічний зміст і вимірюють абсолютну чи відносну швидкість розвитку. Суттєвою підмогою при виборі функцій є аналіз ланцюгових характеристик інтенсивності динаміки.

1. Якщо ланцюгові абсолютні прирости (абсолютна швидкість) відносно стабільні, не мають чіткої тенденції до зростання чи зменшення, вирівнювання ряду виконується на основі лінійної функції: $Y_t = a + bt$.

2. При стабільному прирості абсолютної швидкості доцільна парабола 2-го ступеня $Y_t = a + bt + ct^2$

3. Якщо відносно стабільними є ланцюгові темпи приросту, то найбільш адекватною такому характеру динаміки є експонента $Y_t = ab^t$ (у зазначених функціях a — рівень ряду при $t = 0$, параметр b характеризує швидкість динаміки)

Параметри трендових рівнянь визначають методом найменших квадратів. Згідно з умовою мінімізації суми квадратів відхилень фактичних рівнів ряду Y_t від теоретичних $\bar{Y}_t$ параметри визначаються розв'язуванням системи нормальних рівнянь. Для лінійної функції вона записується так:

$$\begin{cases} na + b\sum t = \sum y, \\ a\sum t + b\sum t^2 = \sum yt. \end{cases}$$

Система рівнянь спрощується, якщо початок відліку часу ($t = 0$) перенести в середину динамічного ряду. Тоді значення t , розміщені вище середини, будуть від'ємними, а нижче — додатними. При непарному числі членів ряду (наприклад, $n = 5$) змінній t надаються значення з інтервалом одиниця: $-2, -1, 0, 1, 2$; при парному: $-2,5, -1,5, -0,5, 0,5, 1,5, 2,5$. В обох випадках $\sum t = 0$, а система рівнянь набирає вигляду

$$\begin{cases} na = \sum y, \\ b\sum t^2 = \sum yt. \end{cases}$$

Отже, $a = \frac{\sum y}{n}$, $b = \frac{\sum yt}{\sum t^2}$. Значення $\sum t^2$ можна визначити за формулами:

- для непарного числа членів ряду

$$\sum t^2 = \frac{n(n^2 - 1)}{12};$$

- для парного числа членів ряду

$$\sum t^2 = \frac{n(n^2 - 1)}{3}.$$

Розраховані значення надають можливість записати рівняння $Y_t = a + bt$, яке дозволяє визначити теоретичні рівні показника Y_t .

Продовження виявленої тенденції за межі ряду динаміки називають **екстраполяцією тренду**. Це один із методів статистичного прогнозування, передумовою використання якого є незмінність причинного комплексу, що формує тенденцію. Прогнозний, очікуваний рівень Y_{t+v} залежить від бази прогнозування та періоду упередження v .

Метод екстраполяції дає точковий прогноз. На практиці, як правило, визначають довірчі межі прогнозного рівня $Y_{t+v} \pm ts_p$, де s_p — стандартна похибка прогнозу, t -квантиль розподілу Стюдента (див. тему 8).

Порядок обчислення параметрів лінійної функції розглянемо на прикладі динамічного ряду видобутку нафти в регіоні (див. табл. 11.2).

Таблиця 11.2 – Динаміка видобутку нафти

Рік	y_t , млн. т	Δ_t	Змінна часу t	$y_t t$	$Y_t = 74,5 + 3,8t$
1	63,5	—	–3	–190,5	63,1
2	66,8	3,3	–2	–133,6	66,9
3	71,0	4,2	–1	–71,0	70,7
4	74,3	3,3	0	0	74,5
5	76,9	2,6	1	76,9	78,3
6	82,2	5,3	2	164,4	82,1
7	86,8	4,6	3	260,4	85,9
Разом	521,5	$\times$	0	106,6	521,5

Ланцюгові абсолютні прирости динамічного ряду практично стабільні, тому тенденцію можна описати лінійною функцією. Оскільки довжина ряду $n = 7$, то $\sum t^2 = 7(7^2 - 1) : 12 = 28$. Параметри трендового рівняння становлять:

$$a = \sum y_t : n = 521,5 : 7 = 74,5;$$

$$b = \sum y_t t : \sum t^2 = 106,6 : 28 = 3,8.$$

Лінійний тренд має вигляд $Y_t = 74,5 + 3,8t$, тобто середній рівень видобутку нафти становить 74,5 млн т, середньорічний приріст видобутку — 3,8 млн. т.

В останній графі таблиці для кожного року наведено теоретичні рівні Y_t , тобто очікувані рівні видобутку нафти в t -му році, зумовлені дією основних чинників розвитку галузі: для 1-го року. $Y_1 = 74,5 + 3,8(-3) = 63,1$ млн т, для 2-го року: $Y_2 = 74,5 + 3,8(-2) = 66,9$ млн т і т. д. Суми фактичних рівнів $\sum y_t$ і розрахованих за лінійним трендом теоретичних рівнів $\sum Y_t$ однакові: $\sum y_t = \sum Y_t = 521,5$ млн т.

Продовжимо виявлену тенденцію. Припускаючи, що умови, в яких формувалась тенденція видобутку нафти, найближчим часом не зміняться, визначимо прогноз, при якому базою прогнозування є теоретичний рівень останнього року, період упередження $v = 2$. Очікуваний через два роки. видобуток нафти досягне 93,5 млн т:

$$Y_{t+v} = 85,9 + 3,8 \cdot 2 = 93,5.$$

2. Оцінка коливань та сталості динаміки

Фактичні рівні динамічних рядів під впливом різного роду чинників варіюють, відхиляючись від основної тенденції розвитку. В одних рядах коливання мають систематичний, закономірний характер, повторюються через певні інтервали часу, в інших — не мають такого характеру і тому називаються **випадковими**. У конкретному ряду можуть поєднуватися систематичні та випадкові коливання.

Найпростішою оцінкою систематичних коливань є **коефіцієнти нерівномірності**, які обчислюються відношенням максимального і мінімального рівнів динамічного ряду до середнього. Чим більша нерівномірність процесу, тим більша різниця між цими двома коефіцієнтами.

Наприклад, споживання питної води за добу становить 7200 м^3 , у середньому за годину $7200 : 24 = 300 \text{ м}^3$. Найбільший рівень споживання води в період від 20 до 21 години — 381 м^3 , найменший — у період від 2 до 3 год — 165 м^3 .

Коефіцієнти нерівномірності такі:

$$K_{\max} = 381 : 300 = 1,27;$$

$$K_{\min} = 165 : 300 = 0,55.$$

Амплітуда коливань у розмірі 72 пункти $[100 (1,27 - 0,55)]$ свідчить про істотну нерівномірність споживання води протягом доби.

Окремим соціально-економічним процесам притаманні внутрішньорічні, сезонні піднесення і спади. Наприклад, нерівномірне завантаження транспорту, коливання попиту на товари тощо. **Сезонні коливання** виявляються і аналізуються на основі рядів щомісячних або щоквартальних даних.

Характер сезонних коливань описується «**сезонною хвилею**», яку утворюють індекси сезонності. У динамічних рядах, які не виявляють чіткої тенденції розвитку, **індекси сезонності** є відношенням фактичних місячних (квартальних) рівнів y_t до середньомісячного (середньоквартального) за рік $\bar{y}$, %:

$$I_c = 100 \frac{y_t}{\bar{y}}.$$

Оскільки сезонні коливання з року в рік не лишаються незмінними, виявити сталу сезонну хвилю можна за допомогою середніх індексів сезонності за кілька років:

$$\bar{I}_c = \frac{1}{n} \sum_{1}^n I_c,$$

де n — число років.

Для порівняння інтенсивності сезонних коливань використовуються узагальнюючі характеристики варіації індексів сезонності:

- амплітуда коливань $R_t = I_{\max} - I_{\min}$;

- середнє лінійне відхилення $\bar{l}_c = \frac{1}{12} \sum_{i=1}^{12} |I_c - 100|$;

- середнє квадратичне відхилення $\sigma_t = \sqrt{\frac{1}{12} \sum_{i=1}^{12} (I_c - 100)^2}$.

Для порівняння інтенсивності коливань двох або більше процесів існує коефіцієнт варіації V_t , який обчислюється відношенням середнього лінійного або середнього квадратичного відхилення до середнього рівня динамічного ряду.

Протилежна коливності властивість – **сталість**. Мірою сталості служить різниця $1 - V_t$. Чим ближчий цей коефіцієнт до 1, тим вище сталість динамічного ряду.

Порядок обчислення сезонної хвилі розглянемо на прикладі споживання електроенергії комунальним господарством регіону (див. табл. 11.3). Сезонну хвилю споживання електроенергії можна представити графічно (див. рис. 11.1).

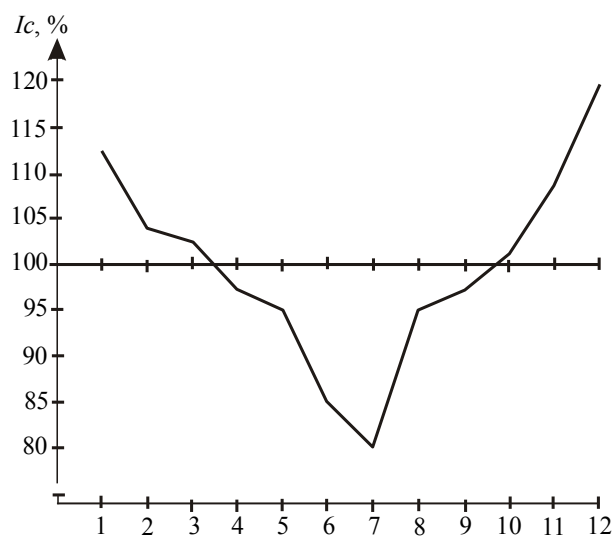


Рисунок 11.1 - Сезонна хвиля споживання електроенергії

Середньомісячний обсяг споживання $\bar{y} = 1848 : 12 = 154$ млн квт · год. Індекси сезонності коливаються від 121,4% у грудні $[(187 : 154)100]$ до 80,5% у липні $[124 : 154)100]$. Амплітуда сезонних коливань становить $R_t = 121,4 - 80,5 = 40,9$ п. п.

Середнє квадратичне відхилення у динамічному ряду споживання електроенергії:

$$\sigma_t = \sqrt{\frac{138914}{12}} = 10,8$$

Таблиця 11.3 – Щомісячна динаміка споживання електроенергії

Місяць року	Спожито електроенергії, y_t , млн квт · год	Індекс сезонності I_c , %	$I_c - 100$	$(I_c - 100)^2$
Січень	172	111,7	11,7	136,89
Лютий	161	104,5	4,5	20,25
Березень	158	102,6	2,6	6,76
Квітень	151	98,0	-2,0	4,00
Травень	147	95,5	-4,6	20,25
Червень	130	84,4	-15,6	243,36
Липень	124	80,5	-19,5	380,25
Серпень	146	94,9	-5,1	26,01
Вересень	149	96,8	-3,2	10,24
Жовтень	155	100,6	0,6	0,36
Листопад	168	109,1	9,1	82,81
Грудень	187	121,4	21,4	457,96

Якщо спостерігається тенденція розвитку, попередньо проводиться згладжування чи вирівнювання динамічного ряду, визначаються теоретичні рівні для кожного місяця (кварталу) року, а індекс сезонності обчислюється як відношення фактичних рівнів ряду y_t до теоретичних $\hat{Y}_t$, тобто

$$I_c = 100 \frac{y_t}{\hat{Y}_t}.$$

Дія випадкових причин зумовлює відхилення фактичних рівнів y_t від скоригованих трендів $\hat{Y}_t$.

Абсолютною мірою випадкових коливань є **середнє квадратичне відхилення** s_e , яке обчислюється на основі залишкової дисперсії:

$$s_e = \sqrt{s_e^2}$$

$$s_e^2 = \frac{1}{n - m} \sum_1^n (y_t - \hat{Y}_t)^2$$

Поряд з абсолютною мірою випадкових коливань використовують відносну — **коефіцієнт варіації**:

$$V_t = 100 \frac{s_e}{\bar{y}},$$

де $\bar{y}$ — середній рівень динамічного ряду.

Розрахунок сезонної хвилі за наявності тенденції подано в таблиці 11.4 на прикладі щоквартальної динаміки продажу безалкогольних напоїв (млн дкл).

Тенденція ряду описується рівнянням $Y_t = 48,2 + 1,445 t$, де t змінюється в межах від $t_1 = -5,5$ до $t_n = +5,5$.

Таблиця 11.4 – Тренд і сезонні коливання продажу безалкогольних напоїв

Рік	Квартал	Млн дкл, y_t	Тренд Y_t	Індекс сезонності	Тренд, скоригований на сезонність, $Y_t \times \hat{I}_t = \hat{Y}_t$
1-й	1	24,4	40,3	0,606	27,7
	2	52,6	41,7	1,262	51,0
	3	60,4	43,1	1,401	59,0
	4	34,0	44,6	0,763	32,1
2-й	1	32,7	46,0	0,711	31,6
	2	56,2	47,5	1,184	58,0
	3	67,3	48,9	1,377	66,9
	4	36,2	50,4	0,719	36,1
3-й	1	37,8	51,8	0,730	35,6
	2	65,3	53,3	1,225	65,1
	3	73,1	54,7	1,337	74,9
	4	38,4	56,1	0,689	40,4
Разом	×	578,4	578,4	12	578,4

Середньозважені індекси сезонності (для яких ваги — середньорічні обсяги продажу) становлять:

Для першого кварталу

$$\hat{I}_1 = \frac{60,6 \cdot 42,85 + 71,1 \cdot 48,1 + 73,0 \cdot 53,65}{1446} = 68,7.$$

Аналогічно розраховані індекси для другого кварталу $\hat{I}_2 = 122,2$; для третього $\hat{I}_3 = 136,9$; для четвертого $\hat{I}_4 = 72,0$.

Скоригований на сезонність тренд наведено в останній графі таблиці 12.4. Для першого кварталу 1-го року.

$$\hat{Y}_t = 40,3 \cdot 0,687 = 27,7.$$

Відхилення фактичних рівнів y_t від скоригованих трендів $\hat{Y}_t$ зумовлено дією випадкових причин.

Залишкова дисперсія продажу безалкогольних напоїв становить (див. табл. 11.5)

$$s_e^2 = \frac{1}{n - m} \sum_{t=1}^n (y_t - \hat{Y}_t)^2 = \frac{35,76}{12 - 2} = 3,576$$

Звідки

$$s_e = \sqrt{3,576} = 1,89.$$

Таблиця 11.5 – До розрахунку залишкової дисперсії

t	y_t	$\hat{Y}_t$	$y_t - \hat{Y}_t$	$(y_t - \hat{Y}_t)^2$
-5,5	24,4	27,7	-3,3	10,89
-4,5	52,6	51,0	1,6	2,56
-3,5	60,4	59,0	1,4	1,96
-2,5	34,0	32,1	1,9	3,61
-1,5	32,7	31,6	1,1	1,21
-0,5	56,2	58,0	-1,8	3,24
0,5	67,3	66,9	0,4	0,16
1,5	36,2	36,1	0,1	0,01
2,5	37,8	35,6	2,2	4,84
3,5	65,3	65,1	0,2	0,04
4,5	73,1	74,9	-1,8	3,24
5,5	38,4	40,4	-2,0	4,00
Разом	578,4	578,4	0	35,76

Щодо випадкових коливань продажу безалкогольних напоїв, то

$$V_e = 100 \frac{1,89}{48,2} = 3,92\%$$

Як відомо, різницю $100 - V_e$ використовують для оцінки сталості динаміки. У розглянутому прикладі ця різниця наближається до 100%, що свідчить про сталий характер тенденції і сезонних коливань реалізації безалкогольних напоїв.

Тема 12. Подання статистичних даних.

1. Статистичні таблиці.
2. Основні елементи статистичних графіків.
3. Класифікація графіків.

1. Статистичні таблиці.

Статистичними таблицями вважають тільки такі, які містять результати аналізу соціально-економічних явищ та процесів (на відміну від допоміжних розрахункових таблиць, таких як логарифмічні).

В статистичних таблицях розрізняють підмет та присудок. Підметом таблиці є така сукупність, ті об'єкти або їх частини, які характеризуються низкою кількісних показників. Показники, які характеризують сукупність, є її присудком.

Залежно від побудови підмета таблиці розділяють на три групи - прості, групові, комбінаційні.

В **простих таблицях** підмет містить перелік об'єктів, періодів, дат або інших ознак, які вивчаються. *Наприклад, при вивченні кількості студентів в окремих країнах, підметом таблиці є перелік країн, а присудком - кількість студентів.*

Групові таблиці відрізняються тим, що в підметі розміщуються групи елементів сукупності по одній ознаці. Найпростішим видом групових таблиць є ряди розподілу.

В підметі **комбінаційних таблиць** групи по одній ознаці розділяються на підгрупи по інших ознаках. Іноді в комбінаційних таблицях групи по одній ознаці розміщують в підметі, а по іншому - в присудку.

Розробка присудка таблиці буває простою та комбінованою. При простій розробці присудка показники, що характеризують підмет, розташовуються паралельно, при комбінованій - представляються в комбінації.

Розробка присудка не змінює вид таблиці, при цьому таблиця все також може бути простою, груповою або комбінаційною. Вид таблиці залежить тільки від представлення підмета.

Обов'язковими атрибутами таблиці є загальний та внутрішні заголовки. Загальний заголовок повинен коротко та чітко характеризувати її зміст. В бічних заголовках розкривається зміст підмета, у верхніх - зміст присудка. Сукупність горизонтальних рядків та вертикальних стовпців без представлення числових даних утворює *макет таблиці*.

При складанні таблиць необхідно дотримуватися наступних правил:

1. Таблиця по можливості повинна бути невеликою за розмірами.

2. Всі заголовки повинні формулюватися чітко, коротко, змістовно.

3. Якщо число показників присудка велике, їх необхідно пронумерувати.

При цьому стовпці, в яких приведений перелік об'єктів або груп, позначають заголовними буквами алфавіту, а стовпці з показниками присудка - арабськими цифрами.

4. Якщо немає відомостей про величину явища, то у відповідній клітці записують "Не має відомостей" або проставляють багатокрапку; відсутність явища позначають тире (-).

5. Кількісні показники в межах одного стовпця повинні приводитися з однаковою точністю (наприклад, до десятих, сотих).

6. Таблиці повинні бути закритими, тобто з підсумковими результатами; виключенням є аналітичні таблиці, в яких розрахунок підсумків не обов'язковий.

2. Основні елементи статистичних графіків.

Статистичний графік – це спосіб наочного подання і викладення статистичних даних за допомогою геометричних знаків та інших графічних засобів з метою їх узагальнення й аналізу.

Основна відмінність статистичних графіків від інших видів графічних зображень полягає в тому, що предметом зображення завжди є статистичні дані, цифрові показники, які дістають у результаті статистичного дослідження масових явищ і процесів, що характеризують ту чи іншу їх особливість, рису. З цієї причини графіки промислових процесів (наприклад, виплавки сталі), графіки відпусток і таке інше не належать до статистичних.

Графічне зображення статистичних даних здійснюється за допомогою геометричних площинних знаків – крапок, ліній, площин, фігур і різних їх комбінацій.

Поле графіка – це простір, у якому розміщуються геометричні або інші графічні знаки.

Розмір поля залежить від призначення графіка. Здебільшого графіки бувають прямокутної форми. Відношення сторін формується за правилом з назвою «золотий переріз», коли висота прямокутника відноситься до його основи, як основа до висоти плюс основа. У цілих числах це виражається як 3:5, 5:8, 8:13 і т.д.

Графічний образ – це сукупність геометричних або графічних знаків, за допомогою яких відображуються статистичні дані. Вони є основою графіка.

Серед геометричних знаків важливе значення мають крапки, за допомогою яких відображують значення ознак окремих елементів сукупності або певних груп. Важливе місце займають також площинні геометричні фігури та їх частина: кола, прямокутники, квадрати, півкола, сектори та ін. Ці фігури застосовують для зображення абсолютних і відносних розмірів явищ і складових частин з метою їх порівняння.

При побудові графіків використовують і негеометричні знаки-символи, які повинні відповідати змісту явищ, що відображуються (наприклад, силуети людських фігур для зображення населення).

Просторові орієнтири використовують для визначення порядку розміщення графічних знаків у полі графіка.

У *прямокутній системі координат* положення будь-якої точки визначається довжиною двох перпендикулярів, опущених з цієї точки на вісь абсцис і ординат.

Полярну систему координат будують навколо певної точки, яку називають центром обертання або полюсом (прикладом є рисунок 12.1). У полярній системі координат можна побудувати координатну сітку у вигляді концентричних кіл з центром у полюсі.

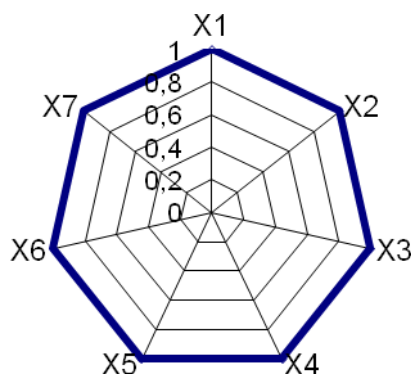


Рисунок 12.1 – Діаграма індикаторів конкурентоспроможності підприємства

У полярній системі координат положення будь-якої точки визначається двома координатами, одна з яких являє собою відстань певної точки від полюса

(*полярний радіус*), а друга – кут між полярною віссю і прямою, яка з'єднує полюс з цією точкою (*полярний кут*)

Масштабні орієнтири – масштаб, масштабна шкала і масштабний знак застосовують для визначення розмірі геометричних та інших графічних знаків.

Масштабні знаки – це знаки-еталони, за допомогою яких зображують статистичні величини у виглядів квадратів, кругів, силуетів тощо. Їх використовують для визначення розмірі і співвідношень величин, зображених на графіку.

Експлікація графіка – це словесні пояснення його змісту і основних елементів, яка включає в себе загальний заголовок графіка, підписи вздовж масштабних шкал, пояснювальні надписи, які розкривають зміст окремих елементів графічного образу.

3. Класифікація графіків.

Графіки, які використовують для зображення статистичних даних, дуже різноманітні. За загальним призначенням графіки поділяють на аналітичні, ілюстративні й інформаційні. За функціонально-цільовим призначенням виділяють графіки групувань і рядів розподілу, графіки рядів динаміки, графіки взаємозв'язку і графіки порівняння. Графіки рядів розподілу і групувань поділяють на графіки за номінальною ознакою і графіки варіаційних рядів, серед останніх виділяють графіки дискретних рядів і графіки інтервальних рядів розподілу.

Основною формою графічного образу розподілу за номінальною ознакою є площинні фігури, в основу яких покладено принцип пропорційності площ фігур. Найпростішим їх видом є одностовпчикові чи однострічні діаграми. Побудова такої діаграми ґрунтується на поділі неширокого прямокутника на стільки елементарних прямокутників, скільки різновидів ознаки має сукупність. Висоту основного прямокутника вибирають довільно, а висота елементарних прямокутників, а отже, і площі їх, мають бути пропорційні частотам або часткам відповідних складових частин сукупності. Графіки слід супроводжувати позначками різновидів номінальної ознаки, а також значеннями частот (часток). При цьому використовують різну штриховку або колір.

Для зображення комбінаційного розподілу за двома номінальними ознаками використовують серійні багатостовпчикові чи багатострічкові діаграми. Принцип їх побудови полягає в тому, що кожний різновид однієї з групувальних ознак зображується стовпчиками, ширина яких однакова, а висота (або довжина) різні пропорційно частотам (часткам) відповідно до різновидів номінальної ознаки. Стовпчики і стрічки розміщують серіями відповідно до різновидів другої групувальної ознаки (див. рис. 12.2 та 12.3)

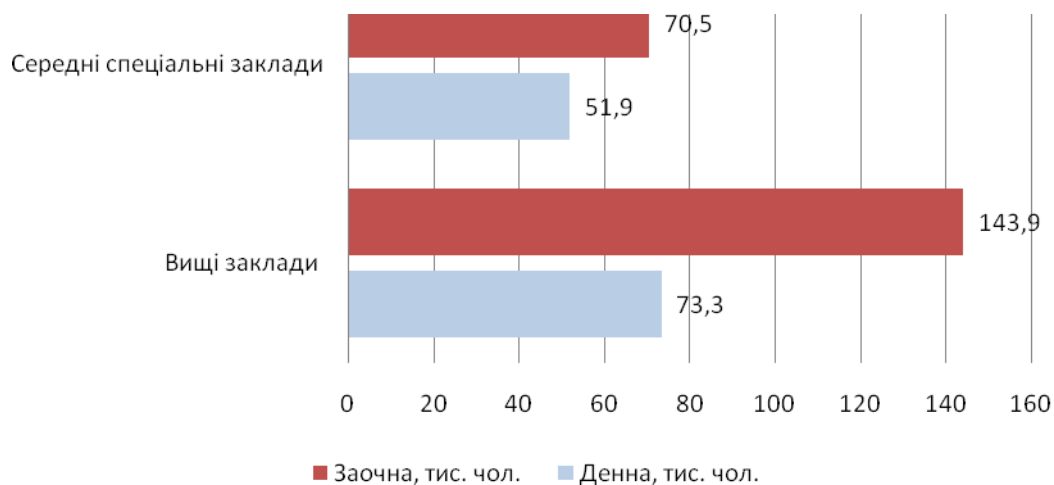


Рисунок 12.2 – Серійно-стрічкова діаграма випуску фахівців вищими та середніми спеціальними закладами області

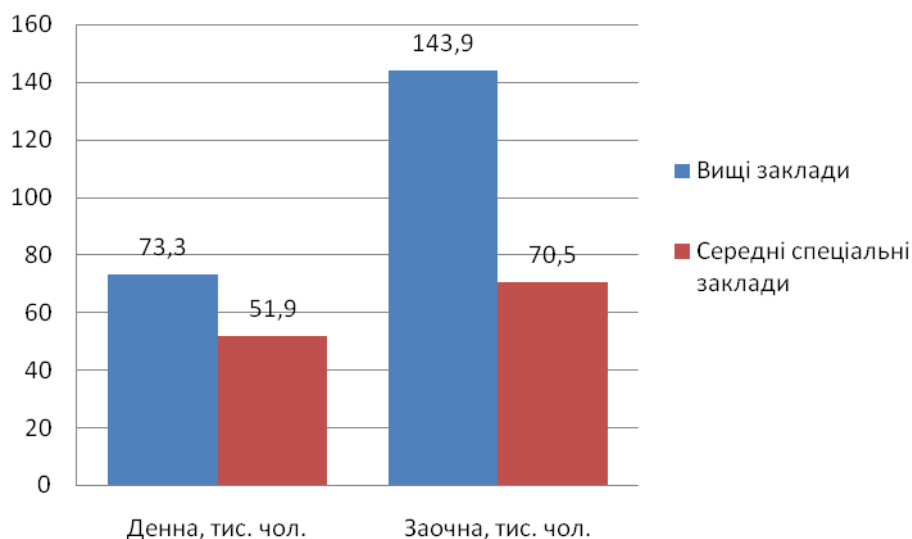


Рисунок 12.3 – Серійно-стовпчикова діаграма випуску фахівців вищими та середніми спеціальними закладами області

Секторні діаграми являють собою графічні зображення на площині круга. Круг ділять на стільки секторів, скільки різновидів має номінальна ознака. Площі секторів, розміри яких залежать від величини центральних кутів, пропорційні часткам відповідних різновидів номінальної ознаки. Секторні діаграми, як правило будують на основі часток, виражених у відсотках. Кожний сектор діаграми, що зображує один з різновидів ознаки, виділяють штрихами або кольорами (див. рис. 12.4)



Рисунок 12.4 - Точність прогнозу банкрутства підприємства за допомогою моделі

Для побудови графіків розподілу за варіаційними рядами застосовують, як правило, прямокутну систему координат, при цьому на осі абсцис відкладають значення варіюючої ознаки, а на осі ординат – відповідні частоти або частки, чи щільності розподілу.

Для графічного зображення дискретного ряду застосовують **полігон розподілу** (див. рис. 12.5).

Для його побудови на осі абсцис відзначають крапки, відповідні величині варіантів значень ознаки, з них відновлюються перпендикуляри, довжина яких відповідає частоті цих варіантів по прийнятому масштабу на осі ординат. Вершини перпендикулярів в послідовному порядку з'єднуються відрізками прямих. Для замикання полігону крайні вершини з'єднуються з крапками на осі абсцис, віддаленими на один розподіл в прийнятому масштабі від $X_{\max}$ і $X_{\min}$.



Рисунок 12.5 – Полігон розподілу

Для графічного зображення інтервальних варіаційних рядів застосовується **гістограма** (див. рис. 12.6).

Вона будується так, що на осі абсцис відкладаються рівні відрізки, які в прийнятому масштабі відповідають величині інтервалів варіаційного ряду. На відрізках будують прямокутники, площі яких пропорційні частотам (або частостям) інтервалу.

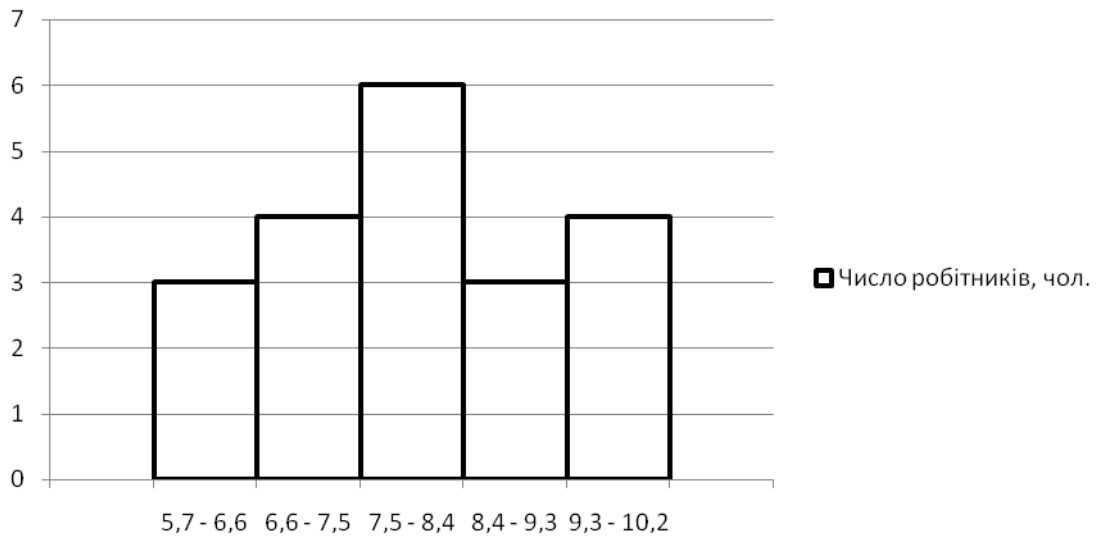


Рисунок 12.6 - Гістограма розподілу робітників за денної заробітної платою (в грн.)

Ряд розподілу з нерівними інтервалами також зображається у вигляді гістограм, але її побудова ґрунтується на щільності частоти.

Щільність частоти – це кількість елементів сукупності, яка доводиться на одиницю ширини інтервалу груповочної ознаки.

Наприклад, по кількості робочих місць 400 підприємств знаходиться в інтервалі від 800 до 1000. Результат від ділення $400:(1000-800)=2$ є щільністю розподілу. При побудові гістограми для ряду з нерівними інтервалами необхідно по осі ординат наносити показники щільності інтервалів.

Гістограма може бути перетворена в полігон розподілу, якщо середини верхніх сторін прямокутників з'єднати відрізками прямих. Дві крайні точки прямокутників замикають по осі абсцис на середини інтервалів, в яких частоти (частості) рівні нулю.

Збільшення числа груп інтервального ряду приводить до зменшення величини інтервалу. При цьому число сторін полігону розподілу ростиме, й ламана лінія матиме тенденцію перетворюватися на плавну криву, яку називають **кривою розподілу**.

У ряді випадків для зображення варіаційних рядів використовується кумулятивна крива (кумулята або кумулятивний полігон). Для її побудови треба розрахувати накопичені частоти або частості. Накопичені частоти показують, скільки одиниць сукупності мають значення ознаки не більше, ніж дане значення, й визначаються послідовним підсумовуванням частот інтервалів (рисунок 12.7 за даними таблиці 12.1).

Таблиця 12.1 - Розподіл робітників за розміром заробітної плати

Середня денна зарплата, грн. X_i	Число робітників f_i	Накопичені частоти, Si
5,7 - 6,6	3	3
6,6 - 7,5	4	7
7,5 - 8,4	6	13
8,4 - 9,3	3	16
9,3 - 10,2	4	20
Разом	20	

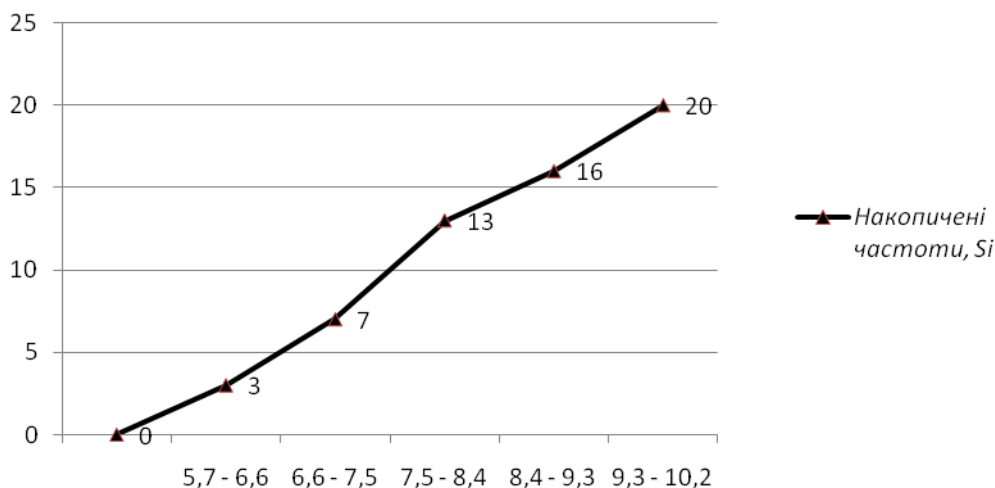


Рисунок 12.7 - Кумулята інтервального варіаційного ряду

Графіки динаміки найчастіше будують у прямокутній системі координат, на осі абсцис якої розміщують шкалу характеристик часу, на осі ординат – рівні ряду динаміки або характеристики швидкості зростання (абсолютний приріст, темпи зростання чи приросту). Основним видом графічного зображення ряді динаміки є **лінійна діаграма**. При її побудові слід правильно вибирати співвідношення масштабів на осі абсцис й осі ординат, бо це впливає на конфігурацію графічного образу. Крім того, шкала на осі ординат має починатися з нуля, інакше діаграма буде неправильно відображати характер розвитку (див. рис. 12.8). При великих значеннях рівнів ряду динаміки з початковим нульовим рівнем на осі ординат можна зробити розрив шкали.

Якщо рядів динаміки більше двох, то їх слід привести до однієї основи, замінивши базисними темпами зростання. Графічне зображення будь-якого ряду динаміки дозволяє виявити тенденцію розвитку, визначити функціональний вид трендової моделі.

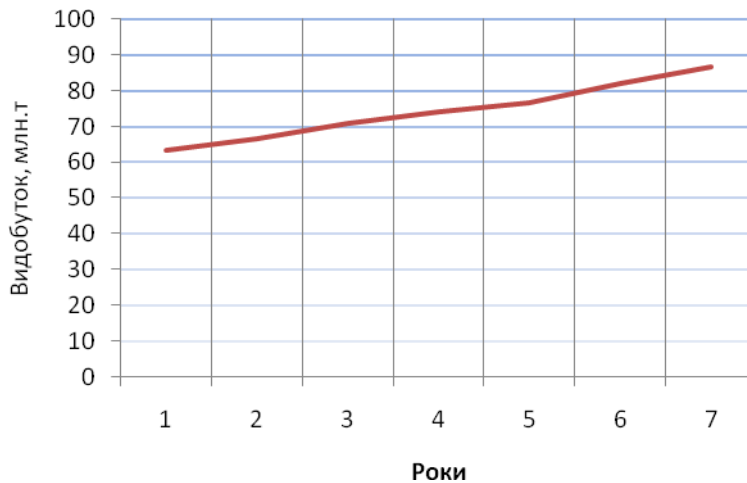


Рисунок 12.8 – Динаміка видобутку нафти

Лінійні діаграми широко застосовують для зображення індексів сезонності, сукупність яких утворює сезонну хвилю (графічне зображення сезонних коливань). З її допомогою можна також зобразити кумулятивний ряд динаміки. Конфігурація кутуляти ряду динаміки залежить від характеру розвитку явища, що вивчають (наприклад при стабільних темпах зростання вона майже рівна, при спадаючих має випуклість наверх й т.д.).

Для зображення сезонних коливань (та будь яких явищ, які періодично змінюються) використовують **радіальні діаграми**, які будують на основі полярної системи координат, у якої віссю абсцис є коло, а ординатами – радіуси цього кола. Ці діаграми можна будувати двома способами залежно від того, що взято за пункт відліку (полюс) – центр чи окружність кола.

При вивченні динаміки використовують також **стовпчикові й стрічкові діаграми**. Вони дають можливість відобразити зміну в часі рівнів, обсягу і розмірів явищ. Крім того вони придатні для зображення рядів динаміки з нерівними інтервалами часу. Техніка їх побудови така сама, як для рядів розподілу.

Графіки порівняння використовують для порівняння статистичних даних. Для цього можна використовувати різні види діаграм. Якщо одноіменні показники відносяться до різних об'єктів або територій, найбільш доцільним є стовпчикові і стрічкові діаграми, у яких стовпчики (стрічки) мають однакову ширину для кожного об'єкту, а висота стовпчиків чи довжина стрічок мають бути пропорційними величинам, що вивчають.

При значних коливаннях величин одноіменного показника застосовують кругові або квадратні діаграми.

ЛІТЕРАТУРА

1. Статистика: Підручник / А.В.Головач, А.М.Єріна та ін.; За ред. А.В.Головача. - К.: Вища шк., 1993. - 623 с.
2. Статистика: Підручник/С. С. Герасименко, А. В. Головач, А. М. Єріна та ін.; За наук. ред. д-ра екон. наук С. С. Герасименка. — 2-ге вид., перероб. і доп. — К.: КНЕУ, 2000. — 467 с.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. Учеб. пособие для ВТУЗов. Изд. 5-е, перераб. и доп. - М.: Высш. школа, 1977. - 479 с.
4. Карасев А.И. Теория вероятностей и математическая статистика: Учебн. для экон. специальностей вузов. - 4-е изд. - М.: Статистика., 1979. - 279 с.
5. Венецкий И.Г., Кильдишев Г.С. Теория вероятностей и математическая статистика: Учебн. пособие для экон. специальностей вузов. - М.: Статистика, 1975. - 264 с.

ДОДАТКИ

Додаток 1.

Критичні значення статистичних характеристик критеріїв χ^2 и λ ($k=m-3$)

$k \setminus 1-\alpha$	0.01	0.05	0.10	0.90	0.95	0.99
Квантилі χ^2 – розподілу						
1	0.02	0.04	0.02	2.71	3.84	6.64
2	0.02	0.10	0.21	4.61	5.99	9.21
3	0.12	0.35	0.58	6.25	7.82	11.35
4	0.30	0.71	1.06	7.78	9.94	13.28
5	0.55	1.15	1.61	9.24	11.07	15.09
6	0.87	1.64	2.20	10.65	12.59	16.81
7	1.24	2.17	2.83	12.02	14.07	18.48
8	1.65	2.73	3.49	13.36	15.51	20.09
9	2.09	3.33	4.17	14.68	16.92	21.66
10	2.66	3.94	4.87	15.99	18.81	23.21
11	3.05	4.58	5.58	17.28	19.68	24.73
12	3.57	5.23	6.30	18.55	21.03	26.22
13	4.11	5.89	7.04	19.81	22.36	27.69
14	4.66	6.57	7.79	21.06	23.69	29.14
15	5.23	7.26	8.55	22.31	26.00	30.58
16	5.81	7.96	9.31	23.54	26.30	32.00
18	7.02	9.39	10.87	25.09	28.87	34.81
20	8.26	10.85	12.44	28.41	31.41	37.57
22	9.54	12.34	14.04	30.81	33.92	40.29
26	12.20	15.38	17.29	35.56	38.89	45.64
30	14.95	18.49	20.60	40.26	43.77	50.89
35	18.51	22.47	24.80	46.06	49.00	57.34
40	22.16	26.51	29.05	51.81	55.76	63.69
50	29.71	34.76	37.69	63.17	67.51	76.15
Квантилі розподілу Колмогорова						
λ	0.44	0.52	0.57	1.23	1.36	1.63

Критичні значення коефіцієнтів асиметрії та ексцесу

Обсяг вибірки	Асиметрії при $1 - \alpha$		Ексцесу при $1 - \alpha$			
	0.95	0.99	0.99	0.95	0.05	0.01
50	0.533	0.787	4.92	4.01	2.13	1.95
100	389	567	40	3.77	35	2.18
150	321	464	14	66	45	30
200	280	403	3.98	57	51	37
250	251	360	87	51	55	42
300	230	329	79	47	59	46
350	213	305	72	44	62	50
400	200	285	67	41	64	52
500	179	255	60	37	67	57
600	163	233	54	34	70	60
700	151	215	50	31	72	62
800	142	202	46	29	74	65
900	134	190	43	28	75	66
1000	127	180	41	26	76	68
1200	116	165	37	24	78	71
1400	107	152	34	22	80	72
1600	100	142	32	21	81	74
1800	095	134	30	20	82	76
2000	090	127	28	18	83	77
0500	080	114	25	16	85	79
3000	073	104	22	15	86	81
4000	064	090	19	13	88	83
5000	057	081	17	12	89	85

Критичні значення кореляційного відношення й коефіцієнту детермінації R^2 .

а) Рівень істотності $\alpha = 0,05$

$k_2 \setminus k_1$	1	2	3	4	5	6	8	10	20
3	0.771	865	903	924	938	947	959	967	983
4	658	776	832	856	887	902	924	937	967
5	569	699	764	806	835	854	885	904	948
6	500	632	704	751	785	811	847	871	928
7	444	575	651	702	739	768	810	839	908
8	399	527	604	657	697	729	775	807	887
9	362	488	563	618	659	692	742	777	867
10	332	451	527	582	624	659	711	749	847
11	306	420	495	550	593	628	682	722	828
12	283	394	466	521	564	600	655	696	809
14	247	348	417	471	514	550	607	650	773
16	219	312	378	429	477	507	564	609	740
18	197	283	345	394	435	470	527	573	709
20	179	259	318	364	404	432	495	540	680
22	164	238	294	339	377	410	466	511	653
24	151	221	273	316	353	385	440	484	628
26	140	206	256	297	332	363	417	461	605
28	130	193	240	279	314	344	396	439	583
30	122	182	227	264	297	326	373	419	563
32	115	171	214	250	282	310	360	401	544
34	108	162	203	238	268	296	344	384	526
36	102	153	192	226	256	282	329	368	509
38	097	146	184	218	245	271	316	355	493
40	093	139	176	207	234	259	304	342	479
50	075	113	143	170	194	216	254	288	416
60	063	095	121	144	165	184	218	249	368
80	047	072	093	110	127	142	170	196	298
100	038	058	075	090	103	116	140	161	251
120	032	049	063	075	087	098	119	137	217
200	019	030	038	046	053	060	073	086	139
400	010	015	019	023	027	031	038	044	074

б) $\alpha = 0,01$

$k_2 \setminus k_1$	1	2	3	4	5	6	8	10	20
3	0.919	954	967	975	979	982	987	989	994
4	841	900	926	941	951	958	967	973	986
5	765	842	879	901	916	928	943	953	974
6	696	785	830	859	879	894	915	929	961
7	636	732	784	818	842	860	887	904	946
8	585	684	740	778	806	827	858	879	931
9	540	641	700	741	771	795	829	854	914
10	501	602	663	706	738	764	802	829	898
11	467	567	629	673	707	734	775	805	882
12	437	536	598	643	678	707	750	782	865
14	388	482	544	590	626	656	703	738	834
16	342	438	498	544	581	612	660	698	803
18	315	401	459	504	541	572	622	669	774
20	288	369	426	470	506	537	588	627	746
22	265	342	396	440	475	506	557	597	720
24	246	319	371	413	448	478	529	569	695
26	229	298	349	389	423	453	503	543	672
28	214	280	329	368	401	430	480	520	650
30	201	264	311	349	381	410	458	498	630
32	190	250	295	332	364	391	438	479	611
34	179	237	280	316	347	374	420	459	592
36	170	226	267	302	332	358	403	443	574

38	162	215	255	289	318	344	389	426	558
40	155	206	244	277	305	310	374	412	542
50	125	168	201	229	254	276	315	351	475
60	106	142	171	196	218	238	273	305	423
80	080	109	132	151	169	186	215	242	345
100	065	088	107	123	138	152	177	201	292
120	054	074	090	104	117	129	151	171	253
200	033	045	055	064	072	080	094	106	165
400	016	023	028	033	037	041	049	056	086

Додаток 4.

Критичні значення F-критерію.

$k_1 \backslash k_2$	1	2	3	4	5	6	8	10	20
а) Уровень значимости $\alpha = 0,05$									
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	242.0	248.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.39	19.44
3	10.13	9.55	6.28	9.12	9.01	8.94	8.84	8.78	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.63	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.34	3.15
9	5.12	4.26	3.89	3.63	3.48	3.37	3.23	3.13	2.93
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.97	2.77
11	4.82	3.98	3.59	3.63	3.20	3.09	2.95	2.86	6.65
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.76	2.54
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.60	2.39
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49	2.28
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41	2.19
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.12
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16	1.93
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.12	1.84
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	2.04	1.75
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.90	1.65
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.83	1.57
б) Уровень значимости $\alpha = 0,01$									
1	4052	4999	5403	5625	5764	5859	5981	6056	6208
2	98.49	99.00	99.17	99.25	99.30	99.33	99.36	99.40	99.45
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.23	26.69
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.54	14.02
5	16.26	13.27	12.02	11.39	10.97	10.67	10.27	10.05	10.55
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.87	7.39
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.62	6.15
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.82	5.36
9	10.56	8.02	6.99	6.42	6.02	5.80	5.47	5.26	4.80
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.85	4.41
11	9.65	7.20	6.22	5.64	5.32	5.07	4.74	4.54	4.10
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.30	3.86
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.94	3.51
16	8.58	6.23	5.29	4.77	4.44	4.20	3.89	3.96	3.25
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.51	3.07
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.37	2.94
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.98	2.55
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.80	2.37
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.63	2.20
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.47	2.03
∞	6.84	4.60	3.78	3.32	3.02	2.80	2.51	2.32	1.87

Таблиця значень функції $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{u^2}{2}} du$

t	Φ(t)	t	Φ(t)	t	Φ(t)	t	Φ(t)
0,00	0,0000	0,24	0,0948	0,48	0,1844	0,72	0,2642
0,01	0,0040	0,25	0,0987	0,49	0,1879	0,73	0,2673
0,02	0,0080	0,26	0,1026	0,50	0,1915	0,74	0,2703
0,03	0,0120	0,27	0,1064	0,51	0,1950	0,75	0,2734
0,04	0,0160	0,28	0,1103	0,52	0,1985	0,76	0,2764
0,05	0,0199	0,29	0,1141	0,53	0,2019	0,77	0,2794
0,06	0,0239	0,30	0,1179	0,54	0,2054	0,78	0,2823
0,07	0,0279	0,31	0,1217	0,55	0,2088	0,79	0,2852
0,08	0,0319	0,32	0,1255	0,56	0,2123	0,80	0,2881
0,09	0,0359	0,33	0,1293	0,57	0,2157	0,81	0,2910
0,10	0,0398	0,34	0,1331	0,58	0,2190	0,82	0,2939
0,11	0,0438	0,35	0,1368	0,59	0,2224	0,83	0,2967
0,12	0,0478	0,36	0,1406	0,60	0,2257	0,84	0,2995
0,13	0,0517	0,37	0,1443	0,61	0,2291	0,85	0,3023
0,14	0,0557	0,38	0,1480	0,62	0,2324	0,86	0,3051
0,15	0,0596	0,39	0,1517	0,63	0,2357	0,87	0,3078
0,16	0,0636	0,40	0,1554	0,64	0,2389	0,88	0,3106
0,17	0,0675	0,41	0,1591	0,65	0,2422	0,89	0,3133
0,18	0,0714	0,42	0,1628	0,66	0,2454	0,90	0,3159
0,19	0,0753	0,43	0,1664	0,67	0,2486	0,91	0,3186
0,20	0,0793	0,44	0,1700	0,68	0,2517	0,92	0,3212
0,21	0,0832	0,45	0,1736	0,69	0,2549	0,93	0,3238
0,22	0,0871	0,46	0,1772	0,70	0,2580	0,94	0,3264
0,23	0,0910	0,47	0,1808	0,71	0,2611	0,95	0,3289
0,96	0,3315	1,37	0,4147	1,78	0,4625	2,36	0,4909
0,97	0,3340	1,38	0,4162	1,79	0,4633	2,38	0,4913
0,98	0,3365	1,39	0,4177	1,80	0,4641	2,40	0,4918
0,99	0,3389	1,40	0,4192	1,81	0,4649	2,42	0,4922
1,00	0,3413	1,41	0,4207	1,82	0,4656	2,44	0,4927
1,01	0,3438	1,42	0,4222	1,83	0,4664	2,46	0,4931
1,02	0,3461	1,43	0,4236	1,84	0,4671	2,48	0,4934
1,03	0,3485	1,44	0,4251	1,85	0,4678	2,50	0,4938
1,04	0,3508	1,45	0,4265	1,86	0,4686	2,52	0,4941
1,05	0,3531	1,46	0,4279	1,87	0,4693	2,54	0,4945
1,06	0,3554	1,47	0,4292	1,88	0,4699	2,56	0,4948
1,07	0,3577	1,48	0,4306	1,89	0,4706	2,58	0,4951
1,08	0,3599	1,49	0,4319	1,90	0,4713	2,60	0,4953
1,09	0,3621	1,50	0,4332	1,91	0,4719	2,62	0,4956
1,10	0,3643	1,51	0,4345	1,92	0,4726	2,64	0,4959
1,11	0,3665	1,52	0,4357	1,93	0,4732	2,66	0,4961
1,12	0,3686	1,53	0,4370	1,94	0,4738	2,68	0,4963
1,13	0,3708	1,54	0,4382	1,95	0,4744	2,70	0,4965
1,14	0,3729	1,55	0,4394	1,96	0,4750	2,72	0,4967
1,15	0,3749	1,56	0,4406	1,97	0,4756	2,74	0,4969
1,16	0,3770	1,57	0,4418	1,98	0,4761	2,76	0,4971
1,17	0,3790	1,58	0,4429	1,99	0,4767	2,78	0,4973
1,18	0,3810	1,59	0,4441	2,00	0,4772	2,80	0,4974
1,19	0,3830	1,60	0,4452	2,02	0,4783	2,82	0,4976
1,20	0,3849	1,61	0,4463	2,04	0,4793	2,84	0,4977
1,21	0,3869	1,62	0,4474	2,06	0,4803	2,86	0,4979
1,22	0,3883	1,63	0,4484	2,08	0,4812	2,88	0,4980
1,23	0,3907	1,64	0,4495	2,10	0,4821	2,90	0,4981
1,24	0,3925	1,65	0,4505	2,12	0,4830	2,92	0,4982
1,25	0,3944	1,66	0,4515	2,14	0,4838	2,94	0,4984

1,26	0,3962	1,67	0,4525	2,16	0,4846	2,96	0,4985
1,27	0,3980	1,68	0,4535	2,18	0,4854	2,98	0,4986
1,28	0,3997	1,69	0,4545	2,20	0,4861	3,00	0,49865
1,29	0,4015	1,70	0,4554	2,22	0,4868	3,20	0,49931
1,30	0,4032	1,71	0,4564	2,24	0,4875	3,40	0,49966
1,31	0,4049	1,72	0,4573	2,26	0,4881	3,60	0,499841
1,32	0,4066	1,73	0,4582	2,28	0,4887	3,80	0,499928
1,33	0,4082	1,74	0,4591	2,30	0,4893	4,00	0,499968
1,34	0,4099	1,75	0,4599	2,32	0,4898	4,50	0,499997
1,35	0,4115	1,76	0,4608	2,34	0,4904	5,00	0,499997
1,36	0,4131	1,77	0,4616				

Додаток 6

Квантили t - розподілу Стьюдента. Критичні точки розподілу Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0, 10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
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	3,30	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	4,41
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	2,40	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	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
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
	Уровень значимости α (односторонняя критическая область)					

Таблиця значень функції $P(\lambda) = 1 - K(\lambda) = P(D \geq \lambda) = 1 - \sum_{k=-\infty}^{+\infty} (-1)^k e^{-k^2 \lambda^2}$

λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
0,29	1,00000	0,76	0,6104	1,23	0,0970	1,70	0,0052	2,17	0,0002
0,30	0,99999	0,77	0,5935	1,24	0,0924	1,71	0,0058	2,18	0,0001
0,31	0,99993	0,78	0,5770	1,25	0,0579	1,72	0,0054	2,19	0,0001
0,32	0,99995	0,79	0,5505	1,26	0,0535	1,73	0,0050	2,20	0,0001
0,33	0,99991	0,80	0,5441	1,27	0,0794	1,74	0,0047	2,21	0,0001
0,34	0,99993	0,81	0,5280	1,28	0,0755	1,75	0,0044	2,22	0,0001
0,35	0,99997	0,82	0,5120	1,29	0,0717	1,73	0,0041	2,23	0,0001
0,35	0,99995	0,83	0,4962	1,30	0,0581	1,77	0,0038	2,24	0,0001
0,37	0,99992	0,84	0,4806	1,31	0,0543	1,78	0,0035	2,25	0,0001
0,38	0,99987	0,85	0,4653	1,32	0,0613	1,79	0,0033	2,26	0,0001
0,39	0,99981	0,85	0,4503	1,33	0,0552	1,80	0,0031	2,27	0,0001
0,40	0,99972	0,87	0,4355	1,34	0,0551	1,81	0,0029	2,28	0,0001
0,41	0,99950	0,88	0,4209	1,35	0,0522	1,82	0,0027	2,29	0,0001
0,42	0,99945	0,89	0,4057	1,35	0,0495	1,83	0,0025	2,30	0,0001
0,43	0,99925	0,90	0,3927	1,37	0,0469	1,84	0,0023	2,31	0,000046
0,44	0,99903	0,91	0,3791	1,38	0,0444	1,85	0,0021	2,32	0,000042
0,45	0,9874	0,92	0,3657	1,39	0,0420	1,86	0,0020	2,33	0,000038
0,45	0,9540	0,93	0,3527	1,40	0,0397	1,87	0,0019	2,34	0,000035
0,47	0,9800	0,94	0,3399	1,41	0,0375	1,85	0,0017	2,35	0,000032
0,48	0,9753	0,95	0,3275	1,42	0,0354	1,89	0,0015	2,36	0,000030
0,49	0,9700	0,96	0,3154	1,43	0,0335	1,90	0,0015	2,37	0,000027
0,50	0,9539	0,97	0,3036	1,44	0,0315	1,91	0,0014	2,38	0,000024
0,51	0,9572	0,98	0,2921	1,45	0,0293	1,92	0,0013	2,39	0,000022
0,52	0,9497	0,99	0,2809	1,45	0,0282	1,93	0,0012	2,40	0,000020
0,53	0,9415	1,00	0,2700	1,47	0,0266	1,94	0,0011	2,41	0,000013
0,54	0,9325	1,01	0,2594	1,48	0,0250	1,95	0,0010	2,42	0,000016
0,55	0,9228	1,02	0,2492	1,49	0,0236	1,93	0,0009	2,43	0,000014
0,55	0,9124	1,03	0,2392	1,50	0,0222	1,97	0,0009	2,44	0,000013
0,57	0,9013	1,04	0,2295	1,51	0,0209	1,98	0,0003	2,45	0,000012
0,58	0,8895	1,05	0,2202	1,52	0,0197	1,99	0,0007	2,46	0,000011
0,59	0,8772	1,05	0,2111	1,53	0,0185	2,00	0,0007	2,47	0,000010
0,60	0,8543	1,07	0,2024	1,54	0,0174	2,01	0,0006	2,48	0,000009
0,61	0,8508	1,08	0,1939	1,55	0,0164	2,02	0,0006	2,49	0,000003
0,62	0,8358	1,09	0,1857	1,56	0,0154	2,03	0,0005	2,50	0,0000075
0,63	0,8222	1,10	0,1777	1,57	0,0145	2,04	0,0005	2,55	0,0000044
0,64	0,8073	1,11	0,1700	1,58	0,0133	2,05	0,0004	2,60	0,0000025
0,65	0,7920	1,12	0,1625	1,59	0,0127	2,06	0,0004	2,65	0,0000016
0,66	0,7754	1,13	0,1555	1,60	0,0120	2,07	0,0004	2,70	0,0000010
0,67	0,7504	1,14	0,1485	1,61	0,0112	2,08	0,0004	2,75	0,0000005
0,68	0,7442	1,15	0,1420	1,62	0,0105	2,09	0,0003	2,80	0,0000003'
0,69	0,7278	1,16	0,1356	1,63	0,0098	2,10	0,0003	2,85	0,00000018
0,70	0,7112	1,17	0,1294	1,64	0,0092	2,11	0,0003	2,90	0,00000010
0,71	0,6945	1,18	0,1235	1,65	0,0086	2,12	0,0002	2,95	0,00000006
0,72	0,6777	1,19	0,1177	1,66	0,0081	2,13	0,0002	3,00	0,00000003
0,73	0,6509	1,20	0,1122	1,67	0,0076	2,14	0,0002		
0,74	0,6440	1,21	0,1070	1,68	0,0071	2,15	0,0002		
0,75	0,6272	1,22	0,1019	1,69	0,0066	2,15	0,0002		

Таблиця значень функції $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

t	0	1	2	3	4	5	6	7	8	9
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	8661	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
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	0644	0632	0620	0608	0596	0584	0573	0562	0551
2.0	0540	0529	0519	0503	0198	0188	0178	0168	0459	0449
2.1	0440	0431	0422	0113	0104	0396	0387	0379	0371	0363
2.2	0855	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	0203	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	0090	0093	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
3.0	0014	0043	0012	0010	0039	0038	0037	0036	0035	0034
4,0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

Таблиця значень функції $\gamma = 2\Phi(t) = F(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$

t	F(t)	t	F(t)	t	F(t)	t	F(t)
0,00	0,00000	0,30	0,23582	0,60	0,45149	0,90	0,63188
01	00798	31	24344	61	45814	91	63718
02	01596	32	25103	62	46474	92	64243
03	02393	33	25860	63	47131	93	64763
04	03191	34	26614	64	47783	94	65278
05	03988	35	27366	65	48431	95	65789
06	04784	36	28115	66	49075	96	66294
07	05581	37	28862	67	49714	97	66795
08	06376	38	29605	68	50350	98	67291
09	07171	39	30346	69	50981	99	67783
0,10	0,07966	0,40	0,31084	0,70	0,51607	1,00	0,68269
11	08759	41	31819	71	52230	01	68750
12	09552	42	32552	72	52848	02	69227
13	10348	43	33280	73	53461	03	69699
14	11134	44	34006	74	54070	04	70166
15	11924	45	34729	76	54675	1,05	70628
16	12712	46	35448	76	55275	06	71086
17	13499	47	36164	77	55870	07	71538
18	14285	48	36877	78	56461	08	71986
19	15069	49	37587	79	57047	09	72429
0,20	0,15852	0,50	0,38292	0,80	0,57629	1,10	0,72867
21	16633	51	38995	81	58206	11	73300
22	17413	52	39694	82	58778	12	73729
23	18191	53	40389	83	59346	13	74152
24	18967	54	41080	84	5990J	14	74571
25	19741	55	41768	85	60468	15	74986
26	20514	56	42452	86	61021	16	75395
27	21284	57	43132	87	61570	17	75800
28	22052	58	43809	88	62114	18	76200
29	22818	59	44481	89	62653	19	76595
1,20	0,76986	1,55	0,87886	1,90	0,94257	2,25	0,97555
21	77372	56	88124	91	94387	26	97618
22	77754	57	88358	92	94514	27	97679
23	78130	58	88589	93	94639	28	97739
24	78502	59	88817	94	94762	29	97798
25	78870	1,00	0,89040	95	94882	2,30	0,97855
26	79233	61	89260	96	95000	31	97911
27	79592	62	89477	97	95116	32	97966
23	79945	63	89690	98	95230	33	98019
29	80295	64	89899	99	95341	34	98072
1,30	0,80640	65	90106	2,00	0,95450	35	98123
31	80980	66	90309	01	95557	36	98172
32	81316	67	90508	02	95662	37	98221
33	81648	68	90704	03	95764	38	98269
34	81975	69	90897	04	95865	39	98315
35	82298	1,70	0,91087	05	95961	2,40	0,98360
36	82617	71	91273	06	96060	41	98405

Продовження додатку 9

t	F (t)	t	F (t)	t	F (t)	t	F (t)
37	8293	72	9145	07	96155	42	98448
38	8324	73	9163	08	96247	43	98490
39	8351	74	9181	09	96338	44	98531
1,40	0,8384	75	9198	2,10	0,9642	45	98571
41	84146 84439	76	92159	11	96514	46	98611
42		77	92327	12	96599	47	98649
43	84728	78	92491	13	96683	48	98686
44	85013	79	92655	14	96765	49	98723
45	85294	1,80	0,92814	15	96844	2,50	0,98758
46	85571	81	92970	16	96923	51	98793
47	85844	82	93124	17	96999	52	98826
48	86113	83	93275	18	97074	53	98859
49	86378	84	93423	19	97148	54	98891
1,50	0,86639	85	93569	2,20	0,97219	55	98923
51	86696	86	93711	21	97289	56	98953
52	87149	87	93852	22	97358	57	98983
53	87398	88	93989	23	97425	58	99012
54	87644	89	94124	24	97491	59	99040
2,60	0,99068	2,95	0,99682	3,30	0,99903	3,65	0,99974
61	99095	96	99692	31	99907	60	99975
62	99121	97	99702	32	99910	67	99976
63	99146	98	99712	33	99913	68	99977
64	99171	99	99721	34	99916	69	99978
65	99195	3,00	0,99730	35	99919	3,70	0,99978
66	99219	01	99739	36	99922	71	99979
67	99241	02	99747	37	99925	72	99980
68	99263	03	99755	38	99928	73	99981
69	99285	04	99763	39	99930	74	99982
2,70	0,99307	05	99771	3,40	0,99933	75	99982
71	99327	06	99779	41	99935	76	99983
72	99347	07	99786	42	99937	77	99984
73	99367	08	99793	43	99940	78	99984
74	99386	09	99800	44	99942	79	99985
75	99404	3,10	0,99806	45	99944	3,80	0,99986
76	99422	11	99813	46	99946	81	99986
77	99439	12	99819	47	99948	82	99987
78	99456	13	99825	48	99950	83	99987
79	99473	14	99831	49	99952	84	99988
2,80	0,99489	15	99837	3,50	0,99953	85	99988
81	99505	16	99842	51	99955	86	99989
82	99520	17	99848	52	99957	87	99989
83	99535	18	99853	53	99958	88	99990
84	99549	19	99858	54	99960	89	99990
85	99563	3,20	0,99863	55	99961	3,90	0,99990
86	99576	21	99867	56	99963	91	99991
87	99590 99602	22	99872	57	99964 99966	92	99991
88	99615	23	99876	58	99967	93	99992
89	0,99027	24	99880	59	0,99968	94	99992
2,90	99639 99650	25	99885	3,60	99969 99971	95	99992

Значення $P_k(t)$ $S(t)$ в розподілі Стюдента

$t \setminus n$	1	2	3	4	5	6—7	8-10	11—15	16-25	25—30	∞
0.0	500	500	500	500	500	500	500	500	500	500	500.000.0
0.1	532	535	537	537	538	538	539	539	539	539	539.827.8
0.2	563	570	573	574	576	576	578	578	578	578	579.259.7
0.3	593	606	608	610	612	613	615	616	616	616	617.911.4
0.4	621	636	642	645	647	649	651	652	653	654	655.421.7
0.5	648	667	674	678	681	683	685	687	689	689	691.462.5
0.6	672	695	705	710	713	715	718	721	722	724	725.746.9
0.7	694	723	733	739	742	742	749	752	754	756	758.036.3
0.8	745	746	759	766	770	774	778	781	783	785	788.144.6
0.9	733	768	783	790	795	800	804	808	811	813	815.939.9
1.0	750	789	804	813	818	823	828	832	835	838	841.344.7
1.1	765		812	813	839	814	850	854	858	860	864.333.9
1.2	779		842	834	858	864	870	874	878	881	884.930.3
1.3	791		858	852	875	883	887	892	896	889	903.199.5
1.4	800		872	868	890	896	902	907	912	915	919.243.3
1.5			885	883	903	909	916	921	925	928	933.192.8
1.6			896	896	915	921	928	933	937	940	945.200.7
1.7			906	908	925	932	938	943	948	951	955.434.5
1.8			915	918	934	941	947	952	956	959	964.069.7
1.9			923	927	942	948	955	960	964	907	971.283.4
2.0				935	949	955	962	967	970	973	977.249.9
2.1				942	955	961	967	972	976	978	982.135.6
2.2				948	960	966	972	977	980	982	986.096.6
2.3				954	969	971	977	981	984	986	989.275.9
2.4				958	969	975	980	984	987	989	991.802.5
2.5				963	973	978	983	987	989	991	993.790.3
2.6				976	976	981	986	989	991	993	995.338.8
2.7				970	979	983	988	991	993	995	996.533.8
2.8				973	981	985	990	993	995	996	997.444.9
2.9				978	983	987	991	994	996	997	998.134.2
3.0				980	985	989	993	995	997	997	998.650.1

t \ n	1	2	3	4	5	6—7	8-10	11-15	16—25	25-30	∞
3,1	901	955	973	982	987	990	994	996	997	998	999.032.4
3,2	904	967	975	984	988	991	995	997	998	998	999.312.9
3,3	906	960	977	985	989	992	995	997	998	999	999.516.6
3,4	909	962	979	986	990	993	996	998	998	999	999.663.1
3,5	911	964	980	988	991	993	997	998	999	—	999.767.4
3,6	914	965	982	989	992	994	997	998	—	—	999.840.9
3,7	916	967	983	990	993	995	998	999	—	—	999.892.2
3,8	918	969	984	990	994	996	998	999	—	—	999.927.7
3,9	920	970	985	991	994	996	998	999	—	—	999.951.9
4,0	922	971	986	992	995	997	998	--	—	—	999.968.3
4,1	924	973	987	993	995	997	999	-	—	—	999.979.3
4,2	926	974	988	993	996	998	999	—	—	—	999.986.7
4,3	927	975	988	994	996	998	999	—	—	—	999.991.5
4,4	929	976	989	994	996	998	—	—	—	—	999.994.6
4,5	930	977	989	995	997	998	—	--	—	—	999.996.6
4,6	932	978	990	995	997	998	—	—	—	—	999.997.9
4,7	933	979	991	995	997	999	—	—	—	—	999.998.7
4,8	935	980	991	996	998	999	—	—	—	—	999.999.2
4,9	936	980	992	996	998	999	—	—	—	—	999.999.5
5,0	937	981	992	996	998	999	—	—	—	—	999.999.7
5,1	938	982	993	996	998	—	—	—	—	—	999.999.8
5,2	940	982	993	997	998	—	—	—	—	—	999.999.9
5,3	941	983	993	997	998	—	—	—	—	—	999.999.9
5,4	942	984	994	997	998	—	—	—	—	—	—
5,5	943	984	994	997	999	—	—	—	—	—	—
5,6	943	984	994	997	999	—	—	—	—	—	—
5,7	945	985	995	998	999	—	—	—	—	—	—
5,8	946	986	997	998	999	—	—	—	—	—	—
5,9	947	986	995	998	999	—	—	—	—	—	—
6,0	947	987	995	998	—	—	-	—	—	—	—
6,1	948	987	996	999	—	—	—	—	—	—	—

Таблиця ймовірностей $P(\chi^2)$

	k									
χ^2	1	2	3	4	5	6	7	8	9	10
1	0,3173	0,6065	0,8013	0,9098	0,9626	0,9856	0,9948	0,9982	0,9994	0,9998
2	1574	3679	5724	7318	8491	9197	9598	9810	9915	9963
3	0833	2231	3916	5578	7000	8088	8850	9344	9643	9814
4	0455	1353	2615	4060	5494	6767	7798	8571	9114	9473
5	0254	0821	1718	2873	4159	5438	6600	7576	8343	8912
6	0143	0498	1116	1991	3062	4232	5398	6472	7399	8153
7	0081	0302	0719	1359	2206	3208	4289	5366	6371	7254
8	0047	0183	0460	0916	1562	2381	3326	4335	5341	6288
9	0027	0111	0293	0611	1091	1736	2527	3423	4373	5321
10	0016	0067	0186	0404	0752	1247	1886	2650	3505	4405
11	0009	0041	0117	0266	0514	0884	1386	2017	2757	3575
12	0005	0025	0074	0174	0348	0620	1006	1512	2133	2851
13	0003	0015	0046	0113	0234	0430	0721	1119	1626	2237
14	0002	0009	0029	0073	0156	0296	0512	0818	1223	1730
15	0001	0006	0018	0047	0104	0203	0360	0591	0909	1321
16	0001	0003	0011	0030	0068	0138	0251	0424	0669	0996
17	0000	0002	0007	0019	0045	0093	0174	0301	0487	0744
18	--	0001	0004	0012	0029	0062	0120	0212	0352	0550
19	--	0001	0003	0008	0019	0042	0082	0149	0252	0403
20	--	0000	0002	0005	0013	0028	0056	0103	0179	0293
21	--	--	0001	0003	0008	0018	0038	0071	0126	0211
22	--	--	0001	0002	0005	0012	0025	0049	0089	0151
23	--	--	0000	0001	0003	0008	0017	0034	0062	0107
24	--	--	--	0001	0002	0005	0011	0023	0043	0076
25	--	--	--	0001	0001	0003	0008	0016	0030	0053
26	--	--	--	0000	0001	0002	0005	0010	0020	0037
27	--	--	--	--	0001	0001	0003	0007	0014	0026
28	--	--	--	--	0000	0001	0002	0005	0010	0018
29	--	--	--	--	--	0001	0001	0003	0006	0012
30	--	--	--	--	--	0000	0001	0002	0004	0009

χ^2	k									
	11	12	13	14	15	16	17	18	19	20
1	0,9999	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
2	9985	0,9994	0,9998	0,9999	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
3	9907	9955	9979	9991	0,9996	0,9998	0,9999	1,0000	1,0000	1,0000
4	9699	9834	9912	9955	9977	9989	9995	0,9998	0,9999	1,0000
5	9312	9580	9752	9858	9921	9958	9978	9989	9994	0,9997
6	8734	9161	9462	9665	9797	9881	9932	9962	9979	9989
7	7991	8576	9022	9347	9576	9733	9835	9901	9942	9967
8	7133	7851	8436	8893	9238	9489	9665	9786	9867	9919
9	6219	7029	7729	8311	8775	9134	9403	9597	9735	9829
10	5304	6160	6939	7622	8197	8666	9036	9319	9539	9682
11	4433	5289	6108	6860	7526	8095	8566	8944	9238	9462
12	3626	4457	5276	6063	6790	7440	8001	8472	8856	9161
13	2933	3690	4478	5265	6023	6728	7362	7916	8386	8774
14	2330	3007	3738	4497	5255	5987	6671	7291	7837	8305
15	1825	2414	3074	3782	4514	5240	5955	6620	7226	7764
16	1411	1912	2491	3134	3821	4530	5238	5925	6573	7166
17	1079	1496	1993	2562	3189	3856	4544	5231	5899	6530
18	0816	1157	1575	2068	2627	3239	3888	4557	5224	5874
19	0611	0885	1243	1649	2137	2687	3285	3918	4568	5218
20	0453	0671	0952	1301	1719	2202	2742	3328	3946	4579
21	0334	0504	0729	1016	1368	1785	2263	2794	3368	3971
22	0244	0375	0554	0786	1078	1432	1847	2320	2843	3405
23	0177	0277	0417	0603	0841	1137	1498	1906	2373	2888
24	0127	0203	0311	0458	0651	0895	1194	1550	1962	2424
25	0091	0148	0231	0346	0499	0698	0947	1249	1605	2014
26	0065	0107	0170	0259	0380	0540	0745	0998	1302	1658
27	0016	0077	0124	0193	0287	0415	0581	0790	1047	1353
28	0032	0055	0090	0142	0216	0316	0449	0621	0834	1094
29	0023	0039	0065	0104	0161	0239	0345	0484	0660	0378
30	0016	0028	0047	0076	0119	0180	0263	0374	0518	0099

Продовження додатку 11

χ^2	k								
	21	22	23	24	25	26	27	28	29
1	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0,9999	0,9999	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
6	9994	9997	0,9999	0,9999	1,0000	1,0000	1,0000	1,0000	1,0000
7	9981	9990	9995	9997	0,9999	0,9999	1,0000	1,0000	1,0000
8	9951	9972	9984	9991	9995	9997	0,9999	0,9999	1,0000
9	9892	9933	9960	9976	9986	9992	9995	9997	0,9999
10	9789	9863	9913	9945	9967	9980	9988	9993	9996
11	9628	9747	9832	9890	9929	9955	9972	9983	9990
12	9396	9571	9705	9799	9866	9912	9943	9964	9977
13	9086	9332	9520	9661	9765	9840	9892	9929	9954
14	8696	9015	9269	9466	9617	9730	9813	9872	9914
15	8230	8622	8916	9208	9414	9573	9694	9784	9850
16	7696	8159	8553	8881	9148	9362	9529	9658	9755
17	7111	7631	8093	8187	8818	9091	9311	9486	9622
18	6490	7060	7575	8080	8424	8758	9035	9261	9443
19	5851	6453	7012	7520	7971	8364	8700	8981	9213
20	5213	5830	6419	6968	7468	7916	8308	8645	8929
21	4589	5207	5811	6387	6926	7420	7863	8253	8591
22	3995	4599	5203	5793	6357	6887	7374	7813	8202
23	3440	4017	4603	5193	5776	6329	6850	7330	7765
24	2981	3472	4038	4616	5194	5760	6303	6815	7289
25	2472	2971	3503	4058	4624	5190	5745	6278	6782
26	2064	2517	3009	3532	4076	4631	5186	5730	6255
27	1709	2112	2500	3015	3559	4093	4638	5182	5717
28	1402	1757	2158	2600	3079	3585	4110	4644	5179
29	1140	1149	1303	2901	2639	3111	3609	4125	4651
30	0920	1185	1491	1848	2243	2676	3142	3632	4140

Функція нормального розподілу $F(x) = \frac{1}{2} + \Phi(t)$ $\left(F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \right)$

t	0	1	2	3	4	5	6	7	8	9
0,0	0,500	504	508	512	516	520	524	528	532	536
0,1	540	544	548	552	556	560	564	567	571	575
0,2	580	583	587	591	595	599	603	606	610	614
0,3	618	622	626	629	633	637	641	644	648	652
0,4	655	659	663	666	670	674	677	681	684	688
0,5	691	695	698	702	705	709	712	716	719	722
0,6	726	729	732	736	739	742	745	749	752	755
0,7	758	761	764	767	770	773	776	779	782	785
0,8	788	791	794	797	800	802	805	808	811	813
0,9	816	819	821	824	826	829	831	834	837	839
1,0	841	844	846	849	851	853	855	858	860	862
1,1	864	867	869	871	873	875	877	879	881	883
1,2	885	887	889	891	893	894	896	898	900	901
1,3	903	905	907	908	910	911	913	915	916	918
1,4	919	921	922	924	925	926	928	929	931	932
1,5	933	934	936	937	938	939	941	942	943	944
1,6	945	946	947	948	950	951	952	953	954	954
1,7	955	956	957	958	959	960	961	962	962	963
1,8	964	965	966	966	967	968	969	969	970	971
1,9	971	972	973	973	974	974	975	976	976	977
2,0	977	978	978	979	979	980	980	981	981	982
2,1	982	983	983	983	984	984	985	985	985	986
2,2	986	986	987	987	987	988	988	988	989	989
2,3	989	990	990	990	990	991	991	991	991	992
2,4	992	992	992	992	993	993	993	993	993	994
2,5	994	994	994	994	994	995	995	995	995	995
2,6	995	995	996	996	996	996	996	996	996	996
2,7	997	997	997	997	997	997	997	997	997	997
2,8	997	998	998	998	998	998	998	998	998	998
2,9	998	998	998	998	998	998	998	999	999	999
3,0	999									

Критичні значення коефіцієнтів рангової кореляції ($\alpha = 0,05$)

n	Коефіцієнт	
	Спірмена	Кендала
5	0,900	0,800
6	0,829	0,739
7	0,714	0,619
8	0,643	0,571
9	0,600	0,500
10	0,564	0,467
12	0,506	0,433
14	0,456	0,394
16	0,425	0,363
18	0,399	0,338

ЗМІСТ

Передмова	3
Тема 1. Методологічні засади статистики.....	4
Тема 2. Статистичне спостереження	8
Тема 3. Зведення і групування статистичних даних	15
Тема 4. Узагальнюючі статистичні показники.....	20
Тема 5. Аналіз рядів розподілу.....	34
Тема 6 Аналіз концентрації, диференціації та подібності розподілів.....	52
Тема 7. Вибірковий метод.....	55
Тема 8. Статистичні методи вимірювання взаємозв'язків.....	68
Тема 9. Індексний метод.....	87
Тема 10. Аналіз інтенсивності динаміки.....	96
Тема 11. Аналіз тенденцій розвитку та коливань.....	103
Тема 12. Подання статистичних даних.....	111
Література	120
Додатки	121

Учбове видання

КУРС ЛЕКЦІЙ

з нормативної навчальної дисципліни циклу природничо-наукової та
загальноєкономічної підготовки

СТАТИСТИКА

для студентів денної (заочної; денно-заочної) форми навчання

Напрямок(и) підготовки: 6.030504 Економіка підприємства – ЕГП, ЕПЕК, ЕПМ,
6.030507 Маркетинг - МПР

Укладач:

Мізіна Олена Вікторівна