МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Севастопольський національний технічний університет

О.С.ДОЦЕНКО

ПРАКТИКУМ
З ЗАГАЛЬНОЇ ТЕОРІЇ СТАТИСТИКИ

Рекомендовано
Міністерством освіти і науки України
як навчальний посібник
для студентів вищих навчальних закладів,
які навчаються за освітньо-професійною програмою бакалавра
з напряму підготовки “Прикладна статистика”

Севастополь 2010

ББК 60.6:65.051
Д 71
УДК 311.1:33(075)

Рецензенти:

д-р екон. наук, проф., заслужений діяч науки і техніки України, завідувач кафедрою статистики Державної академії статистики, обліку та аудиту Держкомстату України Н.О.Парфенцева;

д-р екон. наук, проф., заступник завідувача кафедрою статистики Київського національного економічного університету ім. Вадима Гетьмана З.П.Бараник;

канд. екон. наук, завідувач сектору методології вибіркових обстежень НТК статистичних досліджень Держкомстату України О.В.Гончар

Науковий редактор – канд. екон. наук, доцент кафедри фінанси і кредит Т.М. Одінцова.

Доценко О.С.
Д 71 Практикум з загальної теорії статистики: навч. посібник / О.С. Доценко. – Севастополь: Вид-во СевНТУ, 2010. – 204 с.: додат.
ISBN 978 – 966 – 2960 – 67 – 9

Практикум складено відповідно до типової учбової програми курсу “Статистика”. Містить стислий огляд основних понять і методів загальної теорії статистики як галузі статистичної науки. Розглядаються групування статистичних даних; способи розрахунку абсолютних, відносних і середніх величин; статистичні розподіли; вибіркове спостереження; ряди динаміки; індекси та їх використання в економіко-статистичних дослідженнях. Особливу увагу приділено статистичним методам аналізу динаміки і структури даних та багатомірному групуванню з застосуванням комп’ютерних технологій.

Доценко О.С.
Д 71 Практикум по общей теории статистики: учеб. пособие / О.С. Доценко. – Севастополь: Изд-во СевНТУ, 2010. – 204 с.: прил.
ISBN 978 – 966 – 2960 – 67 – 9

Практикум составлен в соответствии с типовой учебной программой курса «Статистика». Содержит краткий обзор основных понятий и методов общей теории статистики как отрасли статистической науки. Рассматриваются группировки статистических данных; способы расчета абсолютных, относительных и средних величин; статистические распределения; выборочное наблюдение; ряды динамики; индексы и их использование в экономико-статистических исследованиях. Особое внимание уделено статистическим методам анализа динамики и структуры данных и многомерному группированию с применением компьютерных технологий.

Рекомендовано Міністерством освіти і науки України як навчальний посібник для студентів вищих навчальних закладів, які навчаються за освітньо-професійною програмою бакалавра з напряму підготовки “Прикладна статистика” (№ 1/11-1414 від 09.03.10 р.)

ISBN 978 – 966 – 2960 – 67 – 9 ББК 60.6:65.051

ЗМІСТ

ВСТУП

“Загальна теорія статистики” є однією з основних дисциплін у системі економічної освіти. Вона є методологічною основою всіх галузевих статистик, яка формує професійний рівень сучасного економіста. Без знання загальних методів статистичної роботи неможливо здійснювати збирання, наукове оброблення і узагальнення інформації, що характеризує розвиток об’єкта. А без грамотно проведеного аналізу інформаційних характеристик неможливо реалізувати кінцеву мету статистичного дослідження, тобто отримати достовірні параметри щодо загальних тенденцій розвитку суспільних явищ і процесів, які є основою для ухвалення ефективних управлінських рішень.

Запропонований навчальний посібник розраховано на студентів і спеціалістів, які володіють базовими знаннями в обсязі програм економічних вузів з вищої математики, теорії імовірності, макро- і мікроекономіки, а також мають досвід прикладної роботи з пакетами елементарного комп’ютерного оброблення даних.

Практикум покликаний допомогти студентам – майбутнім економістам краще осмислити категорії статистичної науки, навчитися застосовувати наукові методи статистичного аналізу і бачити за статистичними показниками їхній конкретний зміст, а також оволодіти практичними навичками розв’язання конкретних задач різного типу в галузі соціально-економічної статистики на рівні підвищених вимог.

Навчальний посібник складається з трьох розділів і тринадцяти глав. Кожна глава містить два параграфи. У першому параграфі дано методичні вказівки для студентів, де розкрито основні категорії статистичної науки і показано методологію обчислення показників, використовуваних в аналітичній роботі, а також приклади розв’язання типових задач (крім глав 1 і 2). У другому параграфі представлено перелік занять і самостійних завдань для студентів, побудованих на умовних або на фактичних даних, взятих зі статистичних збірників і ЗМІ.

Практикум завершується переліком рекомендованої літератури і додатками, які містять окремі поняття і формули теорії імовірності, статистичні таблиці с даними, необхідними для розв’язання багатьох пропонованих у посібнику задач.

При викладенні матеріалу автор намагалася показати, що статистика не є нудною та складною наукою, що вона може принести задоволення за умови систематичної та послідовної роботи студентів над курсом протягом усього часу його проходження за навчальним планом.

РОЗДІЛ І

ВСТУП ДО ЗАГАЛЬНОЇ ТЕОРІЇ СТАТИСТИКИ

ГЛАВА 1

ПРЕДМЕТ, МЕТОДИ І ЗАВДАННЯ СТАТИСТИКИ

1.1
МЕТОДИЧНІ ВКАЗІВКИ

Тема цієї глави має велике значення не лише для курсу загальної теорії статистики, а й для всіх дисциплін загалом. У ній викладаються такі найважливіші питання статистичної науки, як предмет і метод статистики, основні поняття статистики, сучасні завдання статистики та її взаємозв'язок з іншими науками.

Вивчення теми повинно озброїти студента розумінням основ теорії статистики і статистичної методології.

Розглядаючи цей матеріал, важливо усвідомити необхідність залучення масових даних для об'єктивного пізнання дійсності, а також визначальну роль соціально-економічних категорій у статистичному дослідженні.

Студент має засвоїти такі найважливіші поняття загальної теорії статистики, як сукупність та її однорідність, одиниця сукупності, ознаки та їх класифікація, показники та їх класифікація, варіація, варіант, статистична закономірність. Адже без них неможливо обійтися у подальшому при вивченні як методів загальної теорії статистики, так і інших статистичних дисциплін, у яких застосовуються поняття, терміни, показники і формули загальної теорії статистики.

Предмет статистики. Пропонований курс присвячений статистиці соціально-економічних явищ. Автори більшості сучасних вузівських підручників зі статистики під статистикою розуміють науку, що має особливий предмет і методи пізнання та являє собою цілісну систему наукових дисциплін, яка охоплює загальну теорію статистики, економічну статистику та її галузі, соціально-демографічну статистику та її галузі.

Загальна теорія статистики є методологічною основою всіх галузевих статистик, оскільки вона розробляє найбільш загальні поняття і методи вивчення явищ і процесів у нерозривному зв'язку з їхньою якісною стороною, а також *досліджує* кількісне відображення закономірностей суспільного розвитку в конкретних умовах місця і часу.

Об'єкт конкретного *статистичного дослідження* називають сукупністю.

Предметом статистики є розміри і кількісні співвідношення якісно визначених соціально-економічних явищ, закономірності їхнього зв'язку і розвитку в конкретних умовах місця і часу.

У процесі вивчення цієї теми важливо усвідомити, що статистика як наука досліджує не окремі факти, а масові соціально-економічні явища і процеси, які складаються з багатьох окремих факторів, що мають як індивідуальні, так і загальні ознаки.

Методи статистики. Предмет статистики вивчають за допомогою сукупності прийомів, які утворюють методи статистики.

Важливо знати, що статистичне дослідження зазвичай складається з трьох взаємопов'язаних стадій, кожна з яких реалізується за допомогою відповідних методів, зумовлених змістом виконуваної роботи:

1) метод масового статистичного спостереження, який забезпечує загальність, повноту і представництво отриманої первинної інформації;

2) метод зведення і групування, який дозволяє піддавати систематизації і класифікації всю зібрану інформацію;

3) метод аналізу за допомогою узагальнених показників, який дозволяє характеризувати явища, що вивчаються. На цьому етапі виявляють взаємозв'язки і масштаби явищ, визначають закономірності їх розвитку, дають прогнозні оцінки.

Основні поняття статистики. Статистика оперує певними категоріями (поняттями), які відображають суттєві і загальні ознаки явищ і основні зв'язки між ними.

Сукупність – це множина одиниць (об'єктів, явищ), об'єднаних єдиною закономірністю, і таких, що варіюють (є відмінними) у межах загальної якості. Наприклад, сукупність підприємств, що виробляють однотипну продукцію, але відмінні за обсягами виробництва, трудовими і фінансовими ресурсами і т.і. Сукупності притаманна масовість одиниць.

Одиниці сукупності – це її неподільні первинні елементи, що відображають її якісну однорідність, тобто є носіями ознак. Наприклад, одиницями сукупності можуть бути фірми, акціонерні товариства, люди, сім'ї, вироби тощо.

Ознака – це властивість, характерна риса об'єкта або явища. Наприклад, така одиниця статистичної сукупності, як підприємство, має ознаки: обсяги виробленої і реалізованої продукції, витрати виробництва тощо. Окреме значення ознаки є *варіантом.*

Ознаки поділяються на:

- якісні (атрибутивні) – які не мають числового відображення і є смисловими поняттями (професія, стать і т.д.);
- кількісні – варіанти яких відображаються числовими значеннями (вік, стаж роботи і т.д.). Вони поділяються на:
  а) дискретні – такі, що мають точне значення;
  б) неперервні – такі, що мають значення у певних межах;
- альтернативні – варіанти яких можуть мати лише одне з двох протилежних значень (так або ні). (Наприклад, продукція може бути доброю або бракованою). Вони можуть бути і кількісними, і якісними;
- факторні – незалежні ознаки, що впливають на інші, пов'язані з ними ознаки;
- результативні – залежні ознаки, які змінюються під впливом факторних ознак (наприклад, кваліфікація і стаж роботи робітника – факторні ознаки, а продуктивність праці – результативна ознака).

Якісна *однорідність одиниць сукупності* – це схожість одиниць сукупності (об'єктів, явищ) за якимись суттєвими ознаками, хоча за іншими ознаками вони можуть різнитися. Наприклад, промислові підприємства певної сукупності (однієї галузі), поряд з якісною однорідністю (приналежністю до цієї галузі), мають відмінності щодо розміру основних фондів, обсягу виробництва, чисельності працівників тощо. Однорідність сукупності встановлюють у кожному конкретному статистичному дослідженні відповідно до його мети і пізнавальних завдань.

Варіація – це відмінності в значеннях певної ознаки у окремих одиниць сукупності (нижня межа – мінімальне значення, верхня межа – максимальне значення).

Показник – це узагальнена кількісна характеристика соціально-економічних явищ і процесів у їхній якісній визначеності в умовах конкретного місця і часу. Показник характеризує групу явищ і сукупність загалом. Саме цим він відрізняється від ознаки, яка характеризує індивідуальний об'єкт. Наприклад, середній розмір ощадного вкладу громадян – це показник; розмір вкладу окремої людини – ознака.

Показники бувають:

- кількісними (первинні показники) – характеризують загальний (сумарний) розмір (обсяг) явища (наприклад, обсяг товарообороту, чисельність робітників тощо);
- якісними (вторинні показники) – характеризують рівень явища у розрахунку на певну одиницю сукупності (наприклад, ціна за одиницю продукції, собівартість одиниці продукції тощо).

Статистична закономірність – це кількісна закономірність зміни у просторі і в часі масових суспільних явищ і процесів, які складаються з безлічі елементів. Це обумовлює її взаємозв'язок з законом великих чисел.

Сутність *закону великих чисел*: при сумуванні даних щодо достатньо великої кількості випадків (одиниць статистичної сукупності), відмінності між окремими одиницями досліджуваної маси випадків взаємно погашаються (взаємно врівноважуються), і в загальних середніх числах виступають суттєві, характерні риси і взаємозв'язки явища в цілому, тобто сукупна дія великої кількості випадкових факторів призводить до результату, який майже не залежить від випадку [16].

Завдання статистики. Виходячи з характеру і головних рис предмета статистики як науки, можна сформулювати такі пізнавальні завдання статистики. Це вивчення:

- рівня і структури масових соціально-економічних явищ і процесів;
- взаємозв'язків між масовими соціально-економічними явищами і процесами;
- динаміки масових соціально-економічних явищ і процесів.

Статистика виконує важливу роль у механізмі управління економікою. Основними завданнями соціально-економічної статистики є:

- збирання, оброблення, узагальнення і всебічний аналіз інформації про соціально-економічні явища, що відбуваються у регіонах і в країні загалом;
- розроблення і впровадження статистичної методології, що ґрунтується на міжнародних стандартах;
- забезпечення достовірності, об'єктивності, оперативності, стабільності, цілісності інформації;
- забезпечення доступності, гласності, відкритості зведених статистичних показників у межах законодавства.

1.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Сформулюйте визначення статистики як науки.

2. До яких видів (кількісного чи атрибутивного) належать такі ознаки:

а) кількість працівників на фірмі;

б) родинні зв'язки членів сім'ї;

в) стать і вік людини;

г) соціальне становище вкладників Ощадбанку;

д) кількість дітей у сім'ї;

е) роздрібний товарооборот торговельних мереж.

3. Вкажіть, які сукупності можна виділити у вузі для статистичного вивчення.

4. Якими кількісними і атрибутивними ознаками можна охарактеризувати сукупність студентів вузу.

5. Досліджується сукупність банків України. Якими кількісними і якісними ознаками можна її охарактеризувати?

6. Назвіть найсуттєвіші варіативні ознаки, що характеризують студентську групу.

7. Наведіть перелік показників, якими можна було б при статистичному обстеженні повно охарактеризувати такі явища:

а) населення;

б) споживчий ринок;

г) промисловість;

д) транспорт і зв'язок.

З цією метою використайте статистичні щорічники Держкомстату України.

8. Якими ознаками (дискретними чи неперервними) є:

а) чисельність населення країни;

б) кількість шлюбів і розлучень;

в) виробництво продукції легкої промисловості у грошовому вимірі;

г) процент виконання плану у вартісному вимірі;

д) кількість посадочних місць у літаку.

9. До яких видів (атрибутивного чи кількісного) належать такі ознаки:

а) тарифний розряд робітника;

б) бал успішності у навчанні;

в) форма власності;

г) національність;

д) шлюбний стан.

10. Використовуючи статистичні збірники, випишіть дані, що характеризують структуру:

а) використання грошових доходів населення;

б) виробничих інвестицій за галузями економіки.

11. Назвіть поняття і методи, використовувані в галузі статистичної науки – загальній теорії статистики.

12. Чому вивчення статистичної науки починається з загальної теорії статистики?

13. Назвіть методи статистичного дослідження.

14. Які ви знаєте статистичні збірники, що видаються в Україні?

15. Який існує взаємозв'язок між статистикою й іншими науками?

РОЗДІЛ 2

ОПИСОВА СТАТИСТИКА

ГЛАВА 2

СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

2.1
МЕТОДИЧНІ ВКАЗІВКИ

У цій темі розглядаються основні питання, що стосуються збирання первинних даних, які у подальшому підлягають систематизації та узагальненню.

Вивчаючи цю тему, необхідно зрозуміти:

а) сутність статистичного спостереження;

б) програмно-методологічні питання і організаційні питання плану спостереження;

в) форми, види і способи спостереження;

г) помилки спостереження і контроль даних.

З розвитком ринкових відносин ускладнюються зв'язки між суб'єктами ринку, що зумовлює нагальну потребу у вивченні впливу різних факторів на результати діяльності, соціальні наслідки, а також у прогнозуванні та узагальненнях як на макро-, так і на макрорівні. Найважливішим ресурсом в управлінні стає *статистична інформація* – статистичний матеріал (масові дані) про соціально-економічні явища, отриманий в ході статистичного спостереження.

Статистичне спостереження – це перша стадія статистичного дослідження, яка передбачає науково обґрунтоване збирання масових даних про суспільні явища і процеси, що вивчаються.

Статистичне спостереження може бути:

- первинним – реєстрація даних, що надходять безпосередньо від об'єкта, який їх продукує (наприклад, поточний облік кількості шлюбів у відділі з реєстрації громадянських актів);
- вторинним – збирання даних, які зареєстровані та оброблені (наприклад, банківські звіти, підсумки біржових торгів тощо).

Статистичні дані – це масові та системні характеристики соціально-економічних явищ і процесів (наприклад, обсяг експорту і імпорту, розмір капіталу акціонерних товариств, чисельність безробітних тощо). Вони повинні бути достовірними (тобто реально відображати становище), повними щодо обсягу і змісту, сучасними, зіставними у часі і просторі, науково обґрунтованими.

Необхідно чітко усвідомити, що статистичне спостереження є цілеспрямованим процесом і організується за планом.

План спостереження – це сукупність організаційних і програмно-методологічних питань спостереження.

Організаційні питання плану спостереження

1. Хто проводитиме спостереження (виконавці спостереження, наприклад, окремі органи чи відомства, такі як податкова інспекція, органи охорони здоров'я тощо).
2. Де проводитиметься спостереження (місце спостереження).
3. За допомогою чого проводитимуть спостереження (визначення матеріально-технічної бази).
4. Спосіб забезпечення точності результатів (система контролю і пробні дослідження).
5. Коли проводити спостереження (час і період спостереження).

Проектуючи статистичне спостереження, треба вирішити низку питань щодо часу його проведення. *Час спостереження* (об'єктивний час) – це час, якого стосуються дані спостереження. Якщо спостерігається процес, то визначають інтервал часу, протягом якого накопичуються дані. Наприклад, спостереження пасажиропотоку можна здійснювати протягом повного тижня по днях або по годинах протягом доби. Якщо спостерігають явище, то визначають критичний момент часу (момент часу, станом на який реєструються дані). Наприклад, при переписі населення встановлюють певний час (критичний момент), після якого всі зміни, що відбулися у чисельності населення при проведення спостереження, не беруться до уваги.

Програмно-методологічні питання плану спостереження

1. Встановлення мети спостереження.
2. Визначення об'єкта і одиниці спостереження.
3. Розроблення програми спостереження.
4. Вибір виду і способу спостереження.

Дуже важливо добре усвідомити головні поняття і визначення, які складають основний зміст програмно-методологічного розділу плану статистичного спостереження.

Мета статистичного спостереження полягає в отриманні достовірної інформації для виявлення закономірностей розвитку явищ і процесів. Мета визначає об'єкт статистичного спостереження.

Об’єкт статистичного спостереження – це сукупність суспільних явищ і процесів, що підлягають спостереженню (наприклад, населені пункти, міста, персонал фірми тощо). Інакше кажучи, об’єкт спостереження – це досліджувана статистична сукупність, що складається з окремих одиниць.

Одиниця спостереження – це первинний елемент об’єкта статистичного спостереження, який є носієм ознак, що підлягають реєстрації (наприклад, людина, факт, предмет, процес і т.д.). Визначаючи одиницю конкретного статистичного спостереження, треба якомога точніше її охарактеризувати, вказавши її специфічні риси, які дозволять легше відрізнити її від близьких до неї за видом одиниць інших об’єктів. Наприклад, при демографічних обстеженнях одиницею спостереження може бути людина, але може бути і сім’я; при бюджетному обстеженні – сім’я або домашнє господарство.

Програма спостереження – це перелік питань, на які слід отримати відповіді у результаті спостереження. Зміст питань залежить від мети спостереження і можливостей його проведення.

Програма спостереження спрямована на вироблення статистичного інструментарію.

Статистичний інструментарій – це набір статистичних формулярів, роз’яснень та інструкцій, що стосуються проведення спостережень.

Статистичний формуляр – це обліковий документ єдиного зразка, який містить адресну характеристику об’єкта спостереження і статистичні дані про нього (звіти, переписні листи, бланки документів і анкети). До статистичного формуляру надається інструкція.

Перелік можливих відповідей на поставлені питання називається *статистичним підказом.*

Форми спостереження:

- *звітність* – форма спостереження, за якої кожний об’єкт діяльності подає свої дані до органів статистики у вигляді документів або звітів спеціально затвердженої форми;
- *спеціально організоване спостереження* – форма спостереження, яка охоплює сфери життя, що знаходяться за межами звітності, наприклад:

  а) перепис – спостереження масових явищ з метою визначення їхнього розміру на певну дату (перепис населення);

  б) облік – спостереження, яке ґрунтується на даних опитування, огляду або документальних записів;

  в) опитування (або референдуми) – спостереження думок, мотивів, оцінок, що реєструються зі слів респондента;

  г) обстеження – спостереження окремих масових явищ на певну тематику, яке здійснюється періодично або одноразово (наприклад, неформальна зайнятість населення);

- *статистичний регістр* – список певних об’єктів спостереження з визначенням необхідних ознак, які постійно оновлюються (наприклад, регістр населення – поіменний список мешканців регіону з їхніми паспортними даними, який регулярно переглядається і оновлюється; регістр підприємств і організацій – список суб’єктів усіх видів економічної діяльності з зазначенням їхніх реквізитів і основних показників).

Види спостереження

1. *За ступенем охоплення одиниць спостереження*:

   а) суцільне – коли реєстрації підлягають усі без винятку одиниці сукупності, що вивчається (наприклад, перепис населення; збирання даних у формі звітності, що охоплює великі та середні підприємства різних форм власності; облік, референдум і таке інше);

   б) несуцільне – коли обстеженням охоплюються не всі одиниці досліджуваної сукупності, а тільки їхня частина, на основі якої можна отримати узагальнену характеристику всієї сукупності. До нього належить:

   - вибіркове спостереження – коли реєструється окрема частина сукупності, відібрана у випадковому порядку (див. главу 8);
   - обстеження основного масиву – коли реєструються дані щодо таких одиниць сукупності, за якими можна отримати адекватну (повну) характеристику об’єкта спостереження (наприклад, структуру вантажообороту можна вивчити, дослідивши лише найбільш крупні транспортні вузли; рейтинг банків регіону можна вивчити, обстеживши лише впливові банки);
   - монографічне спостереження – детальне обстеження окремих типових одиниць сукупності з метою їх ретельного вивчення (наприклад, обстеження роботи окремих підприємств, що перейшли у приватну власність);
   - анкетне спостереження (спостереження на основі опитувань) – це спостереження шляхом розповсюдження анкет (формулярів), що підлягають реєстрації. При цьому не всі розповсюджені (розіслані) формуляри, що підлягають реєстрації, повертаються з відповідями. У такий спосіб фірми отримують, наприклад, інформацію від покупців про свої товари;
   - моніторинг – спеціально організоване систематичне спостереження стану явищ і процесів, об’єктів сукупності. Використовується для стеження за соціальними індикаторами, а останнім часом набуває поширення моніторинг навколишнього середовища.

2. *За часом реєстрації фактів:*

   - поточне – систематичне спостереження фактів (наприклад, надходження платежів до банку);

- періодичне – спостереження через певні проміжки часу (наприклад, перепис населення);
- одноразове – у міру виникнення потреби в інформації (наприклад, перепис житлового фонду).

Способи спостереження:

- безпосередній облік даних – облік безпосередньо шляхом підрахунку, вимірювання, оцінювання тощо (наприклад, облік грошей у банку);
- документальний облік – облік за даними різних документів (наприклад, табель робочого часу, накладні, свідоцтво про народження тощо) –

– ці два види обліку забезпечують дуже високу достовірність даних;

- опитування – це спостереження, здійснюване експедиційним шляхом, самореєстрацією, кореспондентським шляхом або анкетним способом:

  а) кореспондентський спосіб – реєстрація фактів на місцях їх виникнення особами, які добровільно присилають їх у відповідні інстанції;

  б) самореєстрація – реєстрація фактів самими респондентами після попереднього інструктажу;

  в) експедиційний шлях – реєстрація фактів спеціально підготовленими фахівцями з обліку з одночасною перевіркою точності реєстрації.

Помилки спостереження. Найважливішим завданням спостереження є отримання якісних і достовірних даних. Його вирішення залежить від успішного виконання вимог, що висуваються до спостереження. Як показує практика, навіть за чітко організованого статистичного спостереження виникають похибки і помилки, що потребують виправлення.

Помилки статистичного спостереження – це розходження між дійсними значеннями ознак та їхньою величиною, отриманою в процесі спостереження. За джерелом походження помилки спостереження поділяють на такі:

1) навмисні (лиходійські) – спеціальне завищення або заниження конкретних значень ознаки;

2) ненавмисні:

- випадкові – пов’язані з неуважністю реєстратора при заповненні документації, з неточністю вимірювальних приладів;
- систематичні – пов’язані, наприклад, з округленням ознаки у більший або менший бік;

- помилки репрезентативності (представництва) – притаманні лише вибірковому спостереженню (наприклад, коли оброблена кількість даних є недостатньою для об'єктивного судження про всю сукупність) (див. главу 8).

Особливе значення при цьому має контроль отримуваних від спостереження даних.

Контроль даних – здійснюється для перевірки даних на зіставність.

Методи контролю

1. Рахунковий контроль – виконується з метою перевірки підсумків і розрахунку показників, а також виявлення помилок.
2. Логічний контроль – виконується з метою виявлення неправдоподібних випадків (помилок), шляхом зіставлення отриманих даних з іншими відомими ознаками, показниками.

2.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Що розуміють під статистичною інформацією?
2. Для чого і кому потрібна статистична інформація у сучасних умовах? Дайте визначення статистичного спостереження.
3. Які характерні риси притаманні статистичному спостереженню?
4. Які питання входять до плану спостереження?
5. Що є метою спостереження?
6. Що таке об'єкт спостереження і як він визначається?
7. Що таке одиниця спостереження?
8. Що таке програма спостереження?
9. У яких формах здійснюється спостереження?
10. На які види поділяється спостереження?
11. Назвіть основні програмно-методологічні питання статистичного спостереження.
12. Поставте мету, визначте об'єкт, одиницю спостереження, конкретні ознаки одиниць спостереження. Складіть програму спеціального статистичного спостереження.
13. Складіть перелік найсуттєвіших ознак таких одиниць статистичного спостереження:

а) житлового будинку (для житлового перепису);
б) вузу;
в) бібліотеки;
г) театру;
д) підприємства.

14. Окресліть ознаки, які, на вашу думку, слід реєструвати при проведенні:

а) обстеження промислової фірми з метою вивчення плинності робочої сили;

б) обстеження роботи міського транспорту з метою вивчення ролі різних його видів у перевезеннях пасажирів;

в) обстеження студентів вузу з метою вивчення їхнього бюджету часу.

15. Сформулюйте питання програми спостереження і складіть макет статистичного формуляру, а також стислу інструкцію до його заповнення для вивчення залежності успішності навчання від статі, віку, сімейного стану, житлових умов і громадської активності студентів вузу при проведенні статистичного обстеження станом на 1 лютого поточного року. Вкажіть, до якого виду належить це спостереження за часом, охопленням і способом отримання даних.

16. За допомогою логічного контролю виконайте перевірку таких відповідей на питання переписного листа перепису населення:

а) прізвище, ім'я, по батькові – Зайцева Олена Петрівна;
б) стать – чоловіча;
в) вік – 10 років;
г) перебуває у шлюбі на поточний час – так;
д) національність – українка;
е) рідна мова – українська;
ж) освіта – вища;
з) місце роботи – школа;
і) посада – вчителька.

У відповідях на які питання імовірніше за все зроблені помилкові записи? Чи можна виправити якийсь з них?

ГЛАВА 3

СТАТИСТИЧНЕ ЗВЕДЕННЯ І ГРУПУВАННЯ

3.1
МЕТОДИЧНІ ВКАЗІВКИ

Найважливішим етапом дослідження соціально-економічних явищ і процесів є систематизація первинних даних і отримання на цій основі зведеної характеристики об'єкта за допомогою узагальнених показників, що досягається шляхом зведення і групування первинного статистичного матеріалу.

У цій темі розглядаються питання, що стосуються змісту і особливостей зведення і групування. Вивчаючи цей матеріал, треба навчитися:

а) визначати вид групування;

б) групувати статистичні дані;

в) давати економічну інтерпретацію узагальненим за групами статистичним показникам.

Зведення – це наукове оброблення первинних даних з метою отримання узагальнених характеристик досліджуваного явища.

Зведення у вузькому сенсі – це підрахунок підсумків у групах і підгрупах та оформлення матеріалів у таблицях.

Зведення, яке проводиться без поділення одиниць сукупності на групи, називається простим (наприклад, коли підсумовують чисельність населення у районі без поділення на чоловіків і жінок). Зведення, при якому застосовують групування, називається складним.

Що стосується *статистичного зведення у широкому сенсі,* то його програма включає:

1) вибір групувальної ознаки;
2) визначення порядку формування груп;
3) розроблення системи статистичних показників;
4) розроблення макетів статистичних таблиць для представлення результатів зведення.

Організація і виконання зведення відбувається за таким планом:

1) визначення послідовності і строків виконання зведення;
2) виконання зведення;
3) викладення отриманих результатів.

При зведенні окремі одиниці об’єднують у групи за допомогою методу групування.

Групування – це розчленування досліджуваної сукупності на однорідні групи за певними суттєвими для них ознаками.

Залежно від поставленої мети і змісту матеріалів за допомогою групування вирішують такі завдання:

1) розроблення первинного статистичного матеріалу;
2) розрахунок зведених показників за групами;
3) порівняння зведених показників, аналіз причин відмінностей між групами, вивчення взаємозв’язку між ознаками.

Види групування та їхні особливості

- *Типологічне групування* – дозволяє виділяти явища за соціально-економічними типами (наприклад, групи секторів економіки за формами власності або населення за родом діяльності).

При його проведенні важливим є вибір групувальної ознаки, яка найповніше відображає сутність досліджуваного явища. Якщо ця ознака є атрибутивною (якісною), то число груп і їхнє найменування залежать від самої ознаки. Якщо ця ознака є кількісною, то при утворення груп необхідно визначити межі переходу кількісної ознаки у нову якість або у новий тип явища.

- *Структурне групування* – передбачає розподілення типів явищ, однорідних сукупностей на групи, що характеризують їхню структуру за якоюсь варіативною ознакою (наприклад, групи господарств за обсягом продукції, структура депозитів за строком їх розміщення).

Його зазвичай здійснюють на базі типологічного групування і використовують з метою певного керівництва. Це групування дозволяє дати характеристику внутрішньої структури явища. Приклад структурного групування, що характеризує масштаби і значущість окремих видів забудови на території міста, наведено у табл. 3.1.1.

Таблиця 3.1.1 – Структурне групування, що характеризує масштаби і значущість окремих видів забудови на території міста

Вид міської забудови	**Площа, га**	**% до підсумку**
Площа під забудову	3030	100,0
у тому числі:		
під промислову забудову	670	22,1
під житлову забудову	1920	63,4
під громадську забудову	315	10,4
під іншу забудову	125	4,1

- *Аналітичне групування,* яке дозволяє виявляти зв'язки і залежності між явищами (наприклад, групуючи велику кількість робітників за рівнем їхньої кваліфікації (факторна ознака) з зазначенням їхньої заробітної плати (результативна ознака), можна помітити пряму залежність заробітної плати від кваліфікації, тобто чим вищою є кваліфікація, тим вищою буде заробітна плата).

Його проводять на базі типологічного і структурного групування для вивчення причин виявлених закономірностей і причин, що обумовлюють послідовність зв'язків між факторними і результативними ознаками.

Залежно від ступеня складності масового явища і завдань аналізу групування можна виконувати за одним або декількома ознаками:

- групування, при якому групи утворюються за однією ознакою, називається *простим*;
- групування, при якому групи утворюються за двома або декількома ознаками (але не більш ніж за чотирма, оскільки надмірне подрібнення інформації ускладнює виявлення закономірностей за двома або декількома ознаками), називається *комбінованим*.

Послідовність проведення групування

1. Вибір групувальної ознаки або їхньої комбінації.
2. Визначення кількості груп і величини інтервалу групування.
3. Встановлення для конкретного групування набору тих показників, якими мають характеризуватися виділені групи.
4. Складання таблиці, у якій мають бути представлені результати групування.

Кількість груп (*n*) визначають за формулою американського вченого Стерджесса (стандартизований підхід):

$$n = 1 + 3{,}322\, lgN \,, \qquad \textbf{(3.1.1)}$$

де n – кількість груп;

N – кількість одиниць усієї сукупності.

Залежність між кількістю груп (n) і кількістю одиниць (N) сукупності за цією формулою є такою:

- якщо N=від 15 до 24 одиниць, то n=5;
- якщо N=від 25до 44 одиниць, то n=6;
- якщо N=від 45 до 89 одиниць, то n=7;
- у подальшому, зі зростанням кількості одиниць у 2 рази, кількість груп зростає тільки на одну (наприклад, N=від 90 до 179 одиниць, то n=8; якщо N=від 180 до 359 одиниць, то n=9 і так далі).

Інтервал – це розмір окремих груп або підгруп за кількісною ознакою. Інтервали поділяються на:

- відкриті – якщо вказана тільки одна межа: або верхня (до якогось числа), або нижня (вище якогось числа). Наприклад: до 750; вище 15;
- закриті – якщо вказана і верхня, і нижня межа. Вони можуть бути:
  а) рівні;
  б) нерівні – використовуються при проведенні типологічних групувань. Наприклад: розподілення підприємств за чисельністю працівників, чол.: до 100, 100–200, 200–300, 300–500, 500–1000, 1000 і більше.

Для групувань з рівними інтервалами (i) розмір інтервалу визначають за формулою:

$$i = \frac{X_{max} - X_{min}}{n}\,, \qquad \textbf{(3.1.2)}$$

де X_{max} – найбільше значення варіативної ознаки;
X_{min} – найменше значення варіативної ознаки;
n – кількість груп.

Приклад 3.1.1. Потрібно провести групування з рівними інтервалами за даними про рівень місячної зарплати робітників, яка коливається у межах від
600 до 750 грош.од. При цьому необхідно виділити 5 груп.

Розв’язок

Використовуючи формулу 3.1.2, отримуємо:

$$i = \frac{X_{max} - X_{min}}{n} = \frac{750 - 600}{5} = 30 .$$

Виділяємо 5 груп:
[600–630) – нижня межа включена, а верхня не включена, аби не потрапити до обох інтервалів одночасно,
[630–660),
[660–690),
[690–720),
[720–750].

Якщо, припустимо, вказано зарплату робітників до 600 грош.од., то інтервал формується (570–600). Якщо вказано зарплату вище 750 грош.од., то інтервал формується (750–780).

Верхня межа останнього, розрахованого за формулою 3.1.2 інтервалу не обов’язково відповідатиме максимальному значенню досліджуваної сукупності (як це вийшло у прикладі 3.1.1). Однак вона не повинна бути менше цього максимального значення, бо інакше не всі значення розподілу потраплять до останньої групи. Для запобігання такому варіанту необхідно правильно округляти (математично) розраховане значення інтервалу. Наприклад, якщо $i = 1{,}56$, то його можна округлити до 1,6 (знову ж таки, округляти до більшого значення слід дуже акуратно, інакше до останнього інтервалу може взагалі не потрапити жодне значення).

У типологічних групуваннях розмір інтервалу визначають залежно від соціально-економічних властивостей явищ.

У структурних групуваннях межі інтервалу і розмір груп визначаються або на підставі поступового укрупнення мінімальних груп (тобто складають ранжований – за зростанням або за зменшенням – ряд, потім здійснюють групування з мінімальним інтервалом, а потім ці інтервали укрупнюють), або з використанням стандартизованого підходу. Інтервал структурного групування визначається формулою:

$$i = \frac{X_{max} - X_{min}}{1 + 3{,}322\, lgN}, \qquad \textbf{(3.1.3)}$$

де знаменник – формула Стерджесса для визначення кількості груп n за стандартизованим підходом.

У аналітичному групуванні для визначення розміру інтервалу використовуються такі методи:
1) метод рівних інтервалів – з використанням формул 3.1.2 або 3.1.3;
2) метод рівних частот – інтервали обирають таким чином, аби число одиниць сукупності розподілялось в них рівномірно.

Розглянемо принцип групування сукупності з подальшою економічною інтерпретацією узагальнених показників за групами у прикладі 3.1.2.

Приклад 3.1.2. Є дані про роботу 25 підприємств однієї з галузей промисловості (табл. 3.1.2). Для вивчення залежності випуску продукції заводами від розміру основних фондів розподіліть усі заводи за вартістю всіх виробничих фондів на 5 груп з рівними інтервалами. Кожну групу заводів охарактеризуйте за: 1) числом заводів, 2) розміром основних фондів (усього і у середньому на один завод), 3) чисельністю працівників х (усього і у середньому на один завод), 4) фактичним випуском товарної продукції (усього і у середньому на одного працівника), 5) випуском продукції на 1 грн. основних фондів.

Таблиця 3.1.2 – Вихідні дані для досліджуваних підприємств

№	Середньорічна вартість ОФ, млн.грн.	Середньооблікова чисельність працівників, чол.	Фактичний випуск продукції, млн. грн
1	*2*	*3*	*4*
1	4,2	350	5,6
2	1,8	220	2,2
3	2,6	200	1,9
4	4,8	340	6,1
5	3,5	400	4,5
6	2,9	280	3,8
7	2,9	250	3,8
8	5,6	450	8,1
9	3,1	250	3,6
10	3,5	380	4,5
11	3,1	310	3,0
12	7,1	260	9,1
13	3,1	310	3,5
14	3,3	250	2,4

Продовження табл. 3.1.2

1	*2*	*3*	*4*
15	5,3	400	6,4
16	3,9	350	4,2
17	2,5	280	3,0
18	2,0	220	1,8
19	7,2	270	8,9
20	3,2	390	3,2
21	1,7	330	2,4
22	4,7	390	4,6
23	2,0	270	2,8
24	1,8	210	1,8
25	6,5	200	6.6

Розв’язок

Визначаємо інтервал груп підприємств за групувальною ознакою – середньорічною вартістю основних фондів (ОФ):
$i=(X_{max}–X_{min})/5=(7,2–1,7)/5=1,1$.
Розбиваємо на групи: [1,7–2,8), [2,8–3,9), [3,9–5,0), [5,0–6,1), [6,1–7,2]

Таблиця 3.1.3 – Перша група підприємств [1,7–2,8)

№	Середньорічна вартість ОФ, млн.грн.	Середньооблікова чисельність працівників, чол.	Фактичний випуск продукції, млн. грн.
2	1,8	220	2,2
3	2,6	200	1,9
17	2,5	280	3,0
18	2,0	220	1,8
21	1,7	330	2,4
23	2,0	270	2,8
24	1,8	210	1,8
Усього 7	14,4	1730	15,9

Таблиця 3.1.4 – Друга група підприємств [2,8–3,9)

№	Середньорічна вартість ОФ, млн.грн.	Середньооблікова чисельність працівників, чол.	Фактичний випуск продукції, млн.грн.
5	3,5	400	4,5
6	2,9	280	3,8
7	2,9	250	3,8
9	3,1	250	3,6

Продовження табл. 3.1.4

10	3,5	380	4,5
11	3,1	310	3,0
13	3,1	310	3,5
14	3,3	250	2,4
20	3,2	390	3,2
Усього 9	28,6	2820	32,3

Таблиця 3.1.5 – Третя група підприємств [3,9–5,0)

№	Середньорічна вартість ОФ, млн.грн.	Середньооблікова чисельність працівників, чол.	Фактичний випуск продукції, млн.грн.
1	4,2	350	5,6
4	4,8	340	6,1
16	3,9	350	4,2
22	4,7	390	4,6
Усього 4	17,6	1430	20,5

Таблиця 3.1.6 – Четверта група підприємств [5,0–6,1)

№	Середньорічна вартість ОФ, млн.грн.	Середньооблікова чисельність працівників, чол.	Фактичний випуск продукції, млн.грн.
8	5,6	450	8,1
15	5,3	400	6,4
Усього 2	10,9	850	14,5

Таблиця 3.1.7 – П’ята група підприємств [6,1–7,2]

№	Середньорічна вартість ОФ, млн.грн.	Середньооблікова чисельність працівників, чол.	Фактичний випуск продукції, млн. грн.
12	7,1	260	9,1
19	7,2	270	8,9
25	6,5	200	6,6
Усього 3	20,8	730	24,6

Результати за групами зведемо в узагальнену табл. 3.1.8 з метою виявлення закономірностей за показниками.

Таблиця 3.1.8 – Підсумкова

ГРУПА	Вартість ОФ, млн.грн.	Число заводів	Середньорічна вартість ОФ, млн.грн.		Середньооблікова чисельність працівників, чол.		Фактичний випуск продукції, млн. грн.		Випуск продукції на 1 грн. основних фондів (фондовіддача)
			усього	у середньому на 1 завод	усього	у середньому на 1 завод	усього	у середньому на 1 працівника, тис.грн.	
	А	1	2	3 =(2/1)	4	5 =(4/1)	6	7 =(6/4)	8 =(6/2)
1	1,7–2,8	7	14,4	2,1	1730	247	15,9	9,19	1,10
2	2,8–3,9	9	28,6	3,2	2820	313	32,3	11,45	1,13
3	3,9–5,0	4	17,6	4,4	1430	358	20,5	14,34	1,16
4	5,0–6,1	2	10,9	5,5	850	425	14,5	17,06	1,33
5	6,1–7,2	3	20,8	6,9	730	243	24,6	33,70	1,18
Усього		25	92,3	3,7	7560	302	107,8	14,26	1,17

За результатами табл. 3.1.8 стає очевидним, що зі зростанням середньорічної вартості основних фондів (графа 3) збільшується чисельність працівників у середньому на один завод (графа 5), за винятком п’ятої групи підприємств. У п’ятій групі відбувається різке зменшення середньооблікової чисельності працівників (з 425 до 243 чоловік) при одночасному збільшенні середньорічної вартості основних фондів (з 5,5 до 6,9 млн. грн. відповідно). Це може бути пов’язано з забезпеченням робітників на підприємствах цієї групи новою потужною технікою і заміною ручної праці автоматизованою (тому підприємства вже не потребують великої кількості працівників). Про це ж саме свідчить різке зростання фактичного випуску продукції у середньому на одного працівника (33,70 тис. грн. для п’ятої групи порівняно з 17,06 тис. грн. для четвертої групи підприємств). Однак фондовіддача у п’ятій групі значно знижена порівняно з четвертою групою підприємств. Тут можна припустити, що нові фонди, у які вкладено великі гроші, ще себе не окупили.

Багатомірні групування. Усе вищесказане про групування стосується групувань, здійснюваних на основі аналізу однієї чи двох ознак, але не більше чотирьох. Але якщо комбінація двох ознак дозволяє зберегти осяжність таблиці, то комбінація трьох або чотирьох ознак дає розмитий результат: адже навіть за виділення трьох категорій за кожною з групувальних ознак ми отримаємо 9 або 12 підгруп! Рівномірність розподілу одиниць за групами в принципі неможлива, звідки й отримуємо групи, до яких входять по 1-2 спостереження. Зберегти повноту опису груп і разом з тим подолати недоліки комбінованого групування дозволяють методи багатомірних групувань.

Такі групування виконують за великою кількістю ознак. Набір ознак утворює ознаковий простір. Кожній ознаці надається смисл координати: тобто якщо набір має n ознак, то кожний об'єкт розглядається як точка у n-мірному просторі. Задача зводиться до виділення згущення точок (груп об'єктів) у цьому просторі, які називаються *кластери*. Знаходження цих груп здійснюється методом кластерного аналізу. Більш докладно це питання розглядатиметься у 3 розділі (у главі 13).

3.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Які види зведення ви знаєте?
2. Що називається статистичним групуванням?
3. Які завдання вирішує статистика за допомогою методу групувань?
4. Дайте характеристику типологічних, структурних і аналітичних групувань. Які завдання вони вирішують?
5. Які бувають інтервали групувань і як точно окреслити їхні межі?
6. Які завдання вирішують багатомірні групування?
7. До яких групувальних ознак – атрибутивних або кількісних – належать: а) вік людини; б) національність; в) бал успішності навчання; г) доход співробітників фірми; д) форма власності?
8. Які з вказаних нижче групувань є типологічними:
а) групування населення за статтю;
б) групування населення, зайнятого у народному господарстві, за галузями;
в) групування капітальних вкладень на будівництво виробничого і невиробничого призначення;
г) групування підприємств громадського харчування за формами власності?
9. Користуючись формулою Стерджесса, визначте інтервал групування співробітників фірми за рівнем доходів, якщо загальна чисельність співробітників складає 20 чоловік, а їхній мінімальний і максимальний доход, відповідно, дорівнює 1000 і 3500 грн.

10. За звітний період є дані про роботу підприємств, що випускають однойменну продукцію. З метою вивчення залежності між виробленою продукцією та її собівартістю виконайте аналітичне групування заводів за обсягом виробленої продукції. Для цього утворіть 4 групи заводів з різними інтервалами. Кожну групу охарактеризуйте за:

1) кількістю заводів;
2) обсягом виробленої продукції – усього і в середньому на один завод;
3) загальною сумою витрат;
4) собівартістю одиниці продукції.

Складіть групові таблиці і зробіть висновки.

№	Фактично вироблено продукції, тис.грн.	Загальна сума витрат на виробництво продукції, тис.грн.
1.	659	420
2.	742	457
3.	570	360
4.	1250	659
5.	306	219
6.	101	593
7.	525	330
8.	774	477
9.	1267	665
10.	482	321
11.	561	361
12.	474	307
13.	276	321
14.	1012	182
15.	1155	570
16.	515	592
17.	144	330
18.	896	112
19.	1324	542
20.	702	700
21.	625	447
22.	840	403
23.	952	495
24.	874	575

11. Побудуйте аналітичне групування за даними задачі 10. У якості групувальної ознаки оберіть загальну суму витрат на виробництво продукції, тис. грн. Число груп знайдіть за формулою Стерджесса.

ГЛАВА 4

РЯДИ РОЗПОДІЛУ

4.1
МЕТОДИЧНІ ВКАЗІВКИ

У цій темі розглядаються питання, що стосуються змісту рядів розподілу, їхніх видів і найбільш популярних способів графічного відображення, а також правил побудови статистичних таблиць.

Після визначення групувальної ознаки і меж груп будується ряд розподілу.

Ряд розподілу – це впорядкований розподіл одиниць сукупності за певною варіативною ознакою. Інакше кажучи, це просте групування, в якому відома кількість одиниць у групах або питома вага кожної групи у сукупності.

Ряд розподілу має два основних елемента:
1) значення групувальної ознаки (X) (тобто варіант ознаки);
2) частота або частість груп.

Частота (f) – кількість окремих варіантів, тобто число, що показує, яку кількість разів або як часто зустрічаються ті чи інші варіанти. Сума всіх частот дорівнює загальній кількості одиниць.

Частість (w) – частота, виражена у частках одиниць або у процентах до підсумку

$$w = \frac{f}{\sum f}. \qquad \textbf{(4.1.1)}$$

Сума частостей дорівнює 1 або 100 % [9].

Види рядів розподілу:

- атрибутивні – тобто побудовані за атрибутивною (якісною) ознакою (за статтю, за національністю тощо);
- варіаційні – тобто побудовані за кількісною ознакою (вік, розмір доходу тощо).

Варіаційні ряди розподілу поділяються на:

а) ранжовані – містять перелік окремих одиниць сукупності у порядку збільшення або зменшення (табл. 4.1.1):

Таблиця 4.1.1 – Великі банки України, ранжовані за розмірами активів на 01.01.04 р. (за даними офіційного сайту Асоціації Банків України).

Банк	Активи, (млн. грн.)
АВАЛЬ	9923,493
ПРИВАТБАНК	9626,522

Продовження табл. 4.1.1

ОЩАДБАНК	5603,678
УКРСОЦБАНК	5157,091

б) дискретні (перервні) – у яких варіанти мають значення цілих чисел. Представляються у вигляді таблиць, що складаються з двох строк або графів: конкретних значень варіативної ознаки (варіанта) *x* і кількості одиниць сукупності з певним значенням ознаки (частоти) *f* (табл. 4.1.2):

Таблиця 4.1.2 – Показники вироблення продукції робітниками підприємства

Вироблення продукції на одного робітника, од. (*X*)	Число робітників, чол. (*f*)
20	16
30	22
35	33
40	14

Ряд розподілу за дискретною ознакою, який має обмежене число варіантів, будується у такій послідовності:

1) усі наявні варіанти записуються у порядку збільшення або зменшення, тобто виконується ранжування;
2) підраховуються частоти (повторюваність) кожного варіанта, які записуються напроти цього варіанта;

в) інтервальні – у яких варіанти дано у вигляді інтервалів, і вони можуть набувати в цих інтервалах будь-яких значень. Представляються у вигляді таблиці, що складається з двох граф або строк – інтервалів ознаки, варіація якої вивчається, і кількості одиниць сукупності, які потрапляють до певного інтервалу (частот), або часток цього числа від загальної кількості одиниць сукупності (частостей) (табл. 4.1.3):

Таблиця 4.1.3 – Показники продуктивності корів на фермі

Групи корів за надоєм за рік, л	Кількість корів (*f*)
3000–3400	43
3400–3800	71
3800–4200	102
4200–4600	64

Для інтервальних рядів розподілу до основних елементів додають дві додаткові характеристики:

1) абсолютна щільність розподілу (m_a) – частота, віднесена до ширини інтервалу:

$$m_a = \frac{f}{i}; \qquad \textbf{(4.1.2)}$$

2) відносна щільність розподілу (m_o) – частість, віднесена до ширини інтервалу:

$$m_o = \frac{w}{i}. \qquad \textbf{(4.1.3)}$$

Ці дві характеристики зазвичай використовують у варіаційних рядах з нерівними інтервалам для перетворення інтервалів. Це може бути необхідним при порівняльному оцінюванні даних, зібраних для різних сукупностей і оброблених різним способом.

Ряди розподілу можна перетворювати у *кумулятивні ряди*, побудовані за накопиченими (кумулятивними) частотами (*S*) з наростаючим підсумком (табл. 4.1.4).

Таблиця 4.1.4 – Показники вироблення продукції робітниками підприємства (за даними таблиці 4.1.2)

Вироблення продукції на одного робітника, од. (*X*)	Число робітників, чол. (*f*)	Кумулятивна частота (*S*)
20	16	16
30	22	38
35	33	71
40	14	85

Приклад 4.1.1. Є дані про розмір заробітної плати робітників комплексної бригади (табл. 4.1.5). Дослідити вихідні дані, застосувавши метод групувань з рівними інтервалами. Виконати розрахунок основних характеристик отриманого ряду розподілу.

Таблиця 4.1.5 – Дані про розмір заробітної плати 30 робітників комплексної бригади, грошових одиниць.

896	325	745	658	214	754	605	999	357	878
324	658	154	968	950	517	201	548	478	865
875	985	302	754	320	615	329	320	698	558

Етапи оброблення даних

1. Побудова інтервального варіаційного ряду.

Вихідні дані представляють у вигляді первинного ранжованого дискретного варіаційного ряду (табл. 4.1.6).

Таблиця 4.1.6 – Первинний ранжований дискретний варіаційний ряд

154	302	324	357	548	615	698	754	878	968
201	320	325	478	558	658	745	865	896	985
214	320	329	517	605	658	754	875	950	999

Використовуючи метод рівних групувань (інтервалів), переходять від дискретного варіаційного ряду до інтервального, визначивши кількість груп (інтервалів) за формулою Стерджесса (3.1.1):

$$n = 1 + 3{,}322\, lg 30 = 5{,}91 \approx 6\,.$$

З формули (3.1.2) $$i = \frac{999 - 154}{6} = 140{,}83\,.$$

Після розбиття діапазону значень варіативної ознаки на інтервали визначають кількість даних, що потрапили до кожного інтервалу. Для дискретного ряду, наведеного у табл. 4.1.6, інтервальний варіаційний ряд представлено у табл. 4.1.7.

Таблиця 4.1.7 – Групування робітників за заробітною платою

Групи робітників за заробітною платою, грош.од.	Значення заробітної плати, грош.од. (X_i)	Кількість робітників, чол. (*f*)
А	*1*	*2*
154,00–294,83	154, 201, 214,	3
294,83–435,66	302, 320, 320, 324, 325, 329, 357	7
435,66–576,49	478, 517, 548, 558	4
576,49–717,32	605, 615, 658, 658, 698	5
717,32–858,15	745, 754, 754	3
858,15–999,00	865, 896, 875, 878, 950, 968, 985, 999,	8
Усього	X	30

2. Розрахунок числових характеристик інтервального варіаційного ряду.
Використовуючи формули (4.1.1 – 4.1.3), розраховують частості і щільності розподілів за групами. Результати заносять у додаткову таблицю 4.1.8

Таблиця 4.1.8 (додаткова) – Групування робітників за заробітною платою

Групи робітників за заробітною платою	Кількість робітників, чол. (*f*)	Накопичена частота, чол. (*S*)	Частість, % (*w*)	Щільність абсолютна, (m_a)	Щільність відносна, (m_o)
А	*1*	*2*	*3*	*4*	*5*
154,00–294,83	3	3	10,0	0,0213	0,00071
294,83–435,66	7	10	23,3	0,0497	0,00166
435,66–576,49	4	14	13,3	0,0284	0,00095
576,49–717,32	5	19	16,7	0,0355	0,00118
717,32–858,15	3	22	10,0	0,0213	0,00071
858,15–999,00	8	30	26,7	0,0568	0,00189
Усього	30	Х	100,0	Х	Х

Графічне зображення рядів розподілу полегшує їхній аналіз і дозволяє судити про форму їхнього розподілу. Найчастіше їх зображують у вигляді гістограми, полігона або кумуляти.

Гістограму розподілу зазвичай застосовують для зображення інтервальних рядів. Для її побудови на осі абсцис позначають інтервали ознаки, а на осі ординат – чисельності одиниць сукупності. На відрізках, що зображують інтервали, будують прямокутники, площі яких пропорційні кількостям одиниць сукупності (рис. 4.1.1).

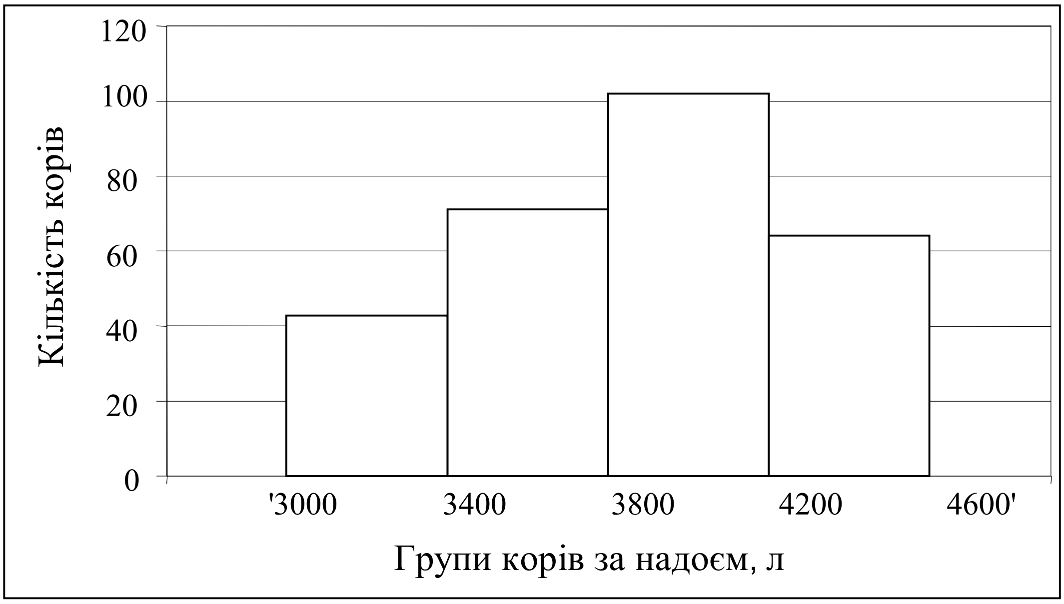


Рисунок 4.1.1 – Гістограма розподілу (за даними табл. 4.1.3)

Полігон будують для зображення як дискретних, так і інтервальних рядів. При його побудові на осі абсцис вказують або значення варіативної ознаки (для дискретного ряду), або значення середин інтервалів ознаки (для інтервального ряду); а на осі ординат – абсолютні або відносні кількості одиниць сукупності (частоти або частості) (рис. 4.1.2).

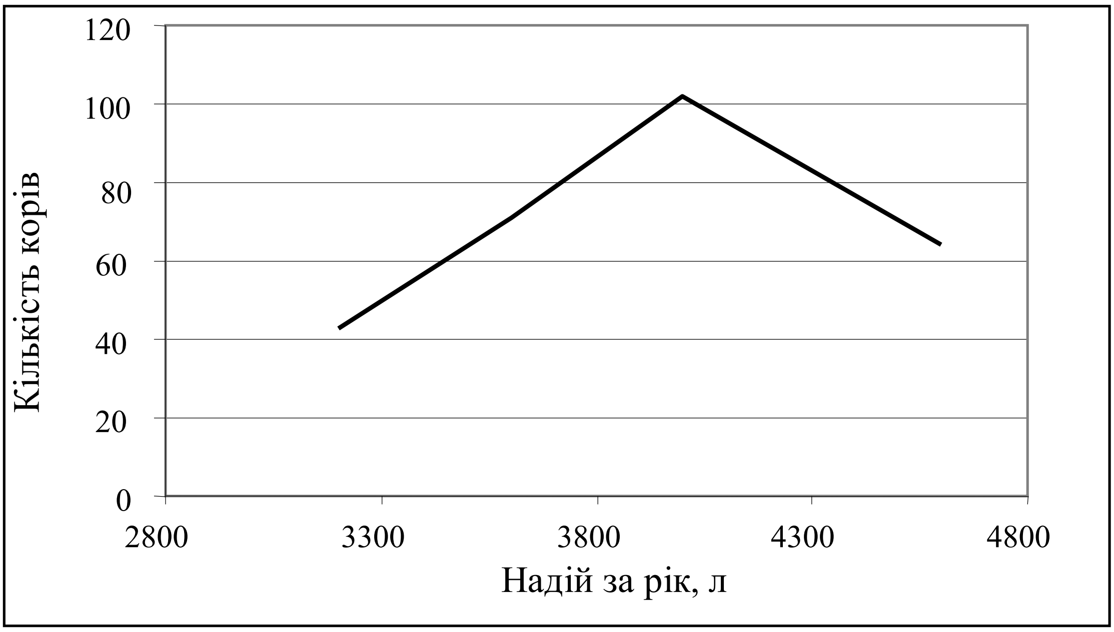


Рисунок 4.1.2 – Полігон розподілу (за даними табл. 4.1.3)

У багатьох випадках для зображення варіаційних рядів використовують кумулятивну криву (*кумуляту*). Для побудови кумуляти дискретного ряду розподілу значення варіативної ознаки вказують на осі абсцис, а на осі ординат вказують накопичені підсумки частот або частостей (рис. 4.1.3).

Накопичені частоти визначають шляхом послідовного сумування частот за групами. При побудові кумуляти інтервального ряду розподілу нижній межі першого інтервалу відповідає частота, що дорівнює нулю, а верхній межі – уся накопичена частота цього інтервалу; верхній межі другого інтервалу відповідає вся накопичена частота другого інтервалу, верхній межі третього інтервалу відповідає вся накопичена частота третього інтервалу і т.д.

Статистична таблиця – це форма найбільш раціонального викладення даних. Аналіз даних статистичної таблиці як метод наукового дослідження дозволяє виявити співвідношення і пропорції між групами явищ за однією або декількома ознаками, провести порівняльний аналіз, охарактеризувати типи соціально-економічних явищ, виявити характер і спрямованість взаємозв’язків і взаємозалежностей між різними ознаками, сформулювати висновки і визначити резерви розвитку досліджуваного явища, об’єкта, процесу.

Рисунок 4.1.3 – Кумулятивна крива (за даними таблиці 4.1.4)

Статистична таблиця містить зведену числову характеристику досліджуваної сукупності за однією або декількома суттєвими ознаками і являє собою комбінацію горизонтальних строк і вертикальних графів.

Статистичну таблицю заповнюють у два етапи:

1) будують макет таблиці;

2) макет заповнюють числовими даними.

Макет таблиці – це схема таблиці, у якій заповнені тільки заголовки без числових даних (рис. 4.1.4).

Підмет таблиці – це те, про що йдеться у таблиці, тобто окремі одиниці сукупності або групи (бокові заголовки).

Присудок таблиці – це система показників, що характеризують підмет (те, що розміщується у верхніх заголовках).

За побудовою підмета таблиці поділяються на:

- прості (перелікові) – містять простий перелік одиниць сукупності;
- монографічні – коли розглядається одна з одиниць сукупності, виділена за певною ознакою;
- групові – містять групи одиниць сукупності за однією ознакою;
- комбінаційні – містять групи одиниць сукупності за декількома ознаками.

За побудовою присудка розрізняють таблиці:

- з простою розробкою присудка – тобто показник, що його визначає, отримують шляхом простого сумування значень для кожної ознаки окремо, незалежно від інших;

- зі складною розробкою присудка – коли передбачається поділення ознаки, що його формує, на групи.

При простій розробці присудка кількість графів зростає в арифметичній прогресії, а при комбінованій (складній) – у геометричній прогресії.

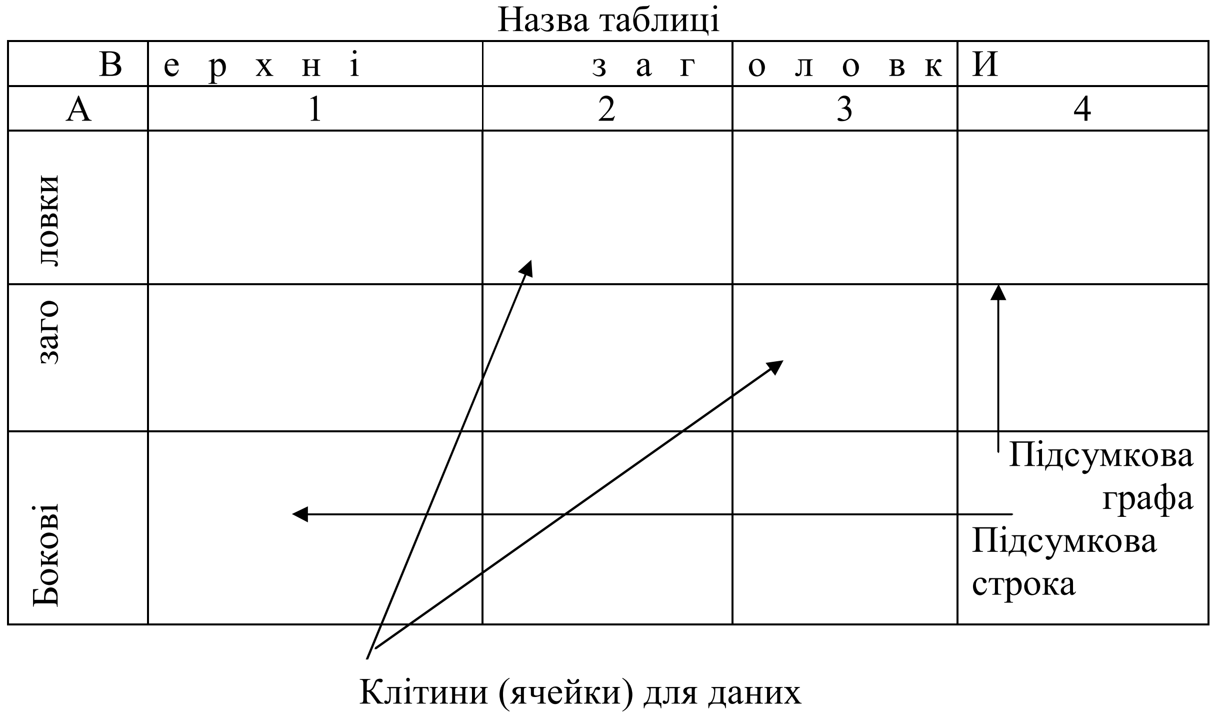


Рисунок 4.1.4 – Макет таблиці

Правила побудови таблиць

1. Загальний заголовок має відображати основний зміст таблиці, місце і час явища, а також одиницю вимірювання, якщо вона однакова для всіх клітин таблиці.
2. Спільні для різних показників назви мають бути винесені у бокові або верхні заголовки. У цих заголовках можуть також бути вказані одиниці вимірювання (або ж одиниці вимірювання виносяться окремою графою).
3. Мають використовуватися лише загальноприйняті скорочення.
4. Многозначні цифрові дані треба округлити, причому всі дані однієї графи (строки) треба наводити з однаковим ступенем точності.
5. Таблиця повинна бути замкненою, тобто мати підсумкові або середні для сукупності показники.
6. У таблиці не повинні бути порожні клітини. Треба використовувати такі позначки-скорочення:

   (–) явище відсутнє;

   (Х) клітина не підлягає заповненню;

   (…) немає даних;

(0,0) дані клітини знаходяться за межами точності, прийнятої у таблиці, графі або строчці.

7. Строки і графи доцільно нумерувати (назву підмета – літерою, назву присудка – цифрою).
8 Аналіз таблиці слід проводити від загального до конкретного. Спочатку аналізуються загальні підсумки таблиці, потім підсумки для груп, потім конкретні показники.
9. У разі необхідності додаткової інформації (роз’яснень до таблиці) можна давати примітки.

Приклад 4.1.2. Зразок оформлення заголовка, підмета і присудка статистичної таблиці (табл. 4.1.9):

Таблиця 4.1.9 – Динаміка обсягів зовнішньої торгівлі України за 2004 – 2008 рр., млн. грн.

Роки	Експорт	Імпорт	Зовнішньоторговельний оборот	Торговельний баланс
А	1	2	3 (гр.1+гр.2)	4 (гр.1–гр.2)
2004				
2005				
2006				
2007				
2008				
Усього				

4.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Що таке ряди розподілу і за якими ознаками вони можуть бути утворені?
2. Як поділяються варіаційні ряди розподілу?
3. Якою є методика побудови дискретних та інтервальних рядів розподілу?
4. Призначення і принципи побудови статистичних таблиць.
5. Є такі дані про успішність навчання 20 студентів групи зі статистики у літню сесію:
5, 4, 4, 4, 3, 2, 5, 3, 4, 4, 4, 3, 2, 5, 2, 5, 5, 2, 3, 3.
Побудуйте:
а) ряд розподілу студентів за балами оцінок, отриманих у сесію;
б) ряд розподілу студентів за рівнем успішності навчання, виділивши в ньому дві групи: неуспішні (2 бала); успішні (3 бала і вище);
в) вкажіть, яким видом ряду (варіаційним або атрибутивним) є кожен з цих двох рядів.

6. Визначте вид ряду розподілу за даними про розподіл робітників заводу за тарифним розрядом:

№ тарифного розряду	Число робітників	Питома вага, %
1	5	10
2	6	12
3	5	10
4	12	24
5	22	44
Усього	50	100

7. Відомі показники діяльності комерційних банків на 01.09.05 (за даними офіційного сайту Асоціації Банків України).

Банк	Активи	Зобов'я-зання	Капітал	Фінансовий результат
ПРИВАТБАНК	19422,25	17524,67	1941,6714	221,59939
АВАЛЬ	15637,83	14301,61	1425,2503	9,4198131
УКРСОЦБАНК	8993,138	8029,443	954,48231	55,188304
УКРСИББАНК	8141,595	7376,427	845,54122	23,490655
ОЩАДБАНК	8069,081	7707,873	346,77261	5,786874
РАЙФФАЙЗЕНБАНК УКРАЇНА	5052,41	4708,156	516,36715	11,879115
НАДРА	4389,942	4012,833	465,92595	11,262245
УКРПРОМБАНК	3606,933	2984,763	635,96426	13,290511
ФІНАНСИ І КРЕДИТ	3542,307	3132,251	494,53294	6,1359413
БРОКБІЗНЕСБАНК	3415,084	2969,31	503,09893	3,9480822
ФОРУМ	3072,318	2735,605	381,41909	16,94351
ПУМБ	2549,748	2080,333	453,70736	32,342607
ПРАВЕКС-БАНК	2101,588	1954,211	181,59586	6,135511
УКРГАЗБАНК	1971,41	1765,293	255,85944	5,8073449
КРЕДИТПРОМБАНК	1932,278	1609,715	323,7446	10,818874
ДОНГОРБАНК	1926,421	1660,11	268,16175	52,524653
ПІВДЕННИЙ	1878,922	1693,768	186,66753	14,400697
ХРЕЩАТИК	1874,991	1691,844	182,58816	10,209527
ВАБАНК	1793,509	1632,134	211,75749	7,4960286
МРІЯ	1726,623	1519,029	209,52476	21,318516

Побудуйте інтервальний ряд розподілу комерційних банків за величиною капіталу, виділивши не більше п'яти рівних інтервалів. Розрахуйте для кожного інтервалу суму активів, капіталу, зобов'язань і фінансового результату. Результати представте у табличному вигляді. За даними інтервального ряду розподілу банків за величиною капіталу побудуйте гістограму, полігон і кумулятивну криву.

8. За даними задачі **7** побудуйте інтервальний ряд розподілу банків за величиною активів. Величину інтервалу знайдіть за формулою Стерджесса. З'ясуйте, чи існує взаємозв'язок між цими показниками. Результати представте у табличному вигляді. За даними інтервального ряду розподілу банків за величиною активів побудуйте гістограму, полігон і кумулятивну криву.

9. За даними статистичних щорічників і періодичної печаті підберіть приклади таких видів таблиць: а) монографічної; б) перелікової;
в) групової; г) комбінаційної.

10. Складіть макети перелікових статистичних таблиць, у яких розробка підмета здійснена за принципами: а) видовим; б) територіальним;
в) часовим.

11. За даними статистичних щорічників і періодичної печаті підберіть приклади статистичних таблиць з варіантами розробки присудка: а) з простою розробкою присудка; б) зі складною розробкою присудка за двома ознаками.

12. За даними про розподіл кількості робітників однієї з галузей промисловості за тарифними розрядами (дані умовні) побудуйте полігон розподілу:

Тарифний розряд	1	2	3	4	5	6
частка робітників, %	4,3	12,1	20,6	32,4	24,0	6,6

ГЛАВА 5

СТАТИСТИЧНІ ПОКАЗНИКИ

5.1 МЕТОДИЧНІ ВКАЗІВКИ

Подальші теми курсу багато в чому ґрунтуються на теорії статистичних показників, яка посідає одне з провідних місць у курсі загальної теорії статистики.

При вивченні цієї теми необхідно приділити особливу увагу класифікації статистичних показників, а також принципам вибору їхньої форми залежно від наявних даних і поставленої задачі.

Соціально-економічні явища, що вивчаються статистикою, мають кількісну визначеність. Статистика вимірює і відображає її за допомогою особливих кількісних категорій, що називаються статистичними показниками.

Статистичний показник – це кількісна характеристика соціально-економічних явищ і процесів в умовах якісної визначеності. Всі використовувані у статистичній практиці показники за формою відображення класифікуються на:

- абсолютні;
- відносні;
- середні.

Абсолютними називають узагальнені показники, що характеризують розміри (рівні, обсяги) суспільних явищ у конкретних умовах місця і часу (наприклад, площа території країни, обсяг промислового виробництва, експлуатаційна довжина залізничних шляхів сполучення, кількість підприємств галузі тощо).

Абсолютні статистичні показники завжди є іменованими числами. Вибір одиниць вимірювання цих показників залежить від сутності явища, що вивчається, його фізичних і соціально-економічних властивостей, а також від мети дослідження. У найзагальнішій класифікації одиниць вимірювання вони зводяться до таких типів: натуральні (умовно-натуральні), грошові (вартісні).

Натуральними зазвичай називають такі одиниці вимірювання, які відображають величину предметів і т.п. у мірах ваги, довжини, об’єму, площі і т.п. відповідно до фізичних властивостей предметів. Наприклад, видобуток вугілля – в тонах, виробництво взуття – у мільйонах пар. А якщо облік тільки в одній одиниці вимірювання не дає достатнього уявлення про обсяг або розміри явища, воно обліковується в двох одиницях вимірювання. Наприклад, електромотори обліковуються у штуках та за потужністю, тканини – у погонних і квадратних метрах і таке інше.

До групи натуральних також входять умовно-натуральні вимірники, які використовуються тоді, коли якийсь продукт має декілька різновидів і загальний об’єм можна визначити тільки виходячи зі спільної для всіх різновидів споживчої властивості. Наприклад, різні види палива перераховують в умовне паливо, тракторний парк – у еталонні трактори, мило різних сортів – в умовне мило з 40-процентним вмістом жирних кислот, консерви різного об’єму – в еталонні банки.

Вищезгадані одиниці вимірювання мають обмежене застосування, оскільки вони придатні для сумування лише однотипних явищ.

Для характеристики різнорідних явищ використовуються грошові одиниці вимірювання. Їхнім недоліком є те, що з плином часу ціни на окремі товари і послуги змінюються і тому показники стають незіставними. Для усунення цього недоліку здійснюють переоцінювання показників у ціни одного і того ж самого періоду часу.

Відносний показник найчастіше є результатом ділення одного абсолютного показника на інший і виражає співвідношення між кількісними характеристиками соціально-економічних явищ і процесів. Наприклад, поділивши число жінок на чисельність населення країни, отримаємо відносну величину, що показує частку жінок у населенні країни.

Показник, з яким здійснюється порівняння, тобто знаменник відношення, називається базою порівняння, а той, який порівнюється – порівнюваним (іноді – поточним або звітними показником). Якщо базою обрано 1, то відносний показник буде виражений у вигляді коефіцієнта, якщо базою обрано 100 – то у процентах, а якщо 1000 – то у продециміле.

Залежно від розмірності порівнюваних показників обирають найбільш зручні форми їх відображення. Якщо порівнюваний показник набагато перевищує базу, то відношення краще виражати коефіцієнтом. Якщо показник порівняння не дуже сильно відрізняється від бази, то він виражається у процентах. Якщо порівнюваний показник дуже малий порівняно з базою, то його виражають у проміле. Проміле широко застосовують у демографічній статистиці, де народжуваність, смертність, шлюби і розлучення обчислюють на 1000 чоловік населення.

Часто при обчисленні відносних показників порівнюють не тільки абсолютні, а й відносні показники; а також не тільки однойменні, а й різнойменні показники.

Обрані для порівняння показники повинні бути зіставними:

- у межах одного і того ж самого місця і періодів часу з урахуванням сезонних коливань;
- для одного і того ж самого кола одиниць спостереження;
- за умовами і способами збирання даних первинного обліку та їх статистичного зведення;
- за методологією розрахунку;
- за одиницями вимірювання.

Залежно від призначення і сутності розрізняють такі види відносних показників:

- виконання плану;
- планового завдання;
- динаміки;
- структури;

- координації;
- інтенсивності;
- порівняння.

Відносний показник виконання плану *(коефіцієнт виконання плану* $Квп$):

$$Квп = \frac{Y_1}{Y_{пл}}, \qquad \textbf{(5.1.1)}$$

де Y_1 – фактично досягнутий рівень звітного періоду;

$Y_{пл}$ – запланований рівень на звітний період.

Цей показник відображає ступінь виконання плану за певний період часу.

Приклад 5.1.1. За планом на звітний період підприємство повинно випустити продукції на суму 500 грош.од., а фактично випустило на 550 грош.од. Знайти коефіцієнт виконання плану.

<u>Розв'язок</u>

$$Квп = \frac{Y_1}{Y_{пл}} = \frac{550}{500} \times 100 = 110 \ \%.$$

Отже, план виконано на 110 %, а перевиконання склало 10 %.

Відносний показник планового завдання *(коефіцієнт планового завдання* $Кпз$ *):*

$$Кпз = \frac{Y_{пл}}{Y_0}, \qquad \textbf{(5.1.2)}$$

де Y_0 – фактично досягнутий рівень за базисний період.

Цей коефіцієнт показує, яка планується зміна показників порівняно з базовим періодом.

Відносний показник динаміки *(коефіцієнт динаміки* $Кд$):

$$Кд = \frac{Y_1}{Y_0}. \qquad \textbf{(5.1.3)}$$

Цей показник відображає ступінь зміни явища у часі і характеризує швидкість цієї зміни або, інакше, темп розвитку.

Залежно від характеру бази порівняння розрізняють два види відносних показників динаміки:

1) зі змінною базою порівняння (або ланцюгові) – вони показують, як змінюється величина показника від одного періоду до іншого;

2) з постійною базою порівняння (або базисні) – вони характеризують поступове віддалення від базисного періоду, просування уперед від вихідної точки.
Вибір бази порівняння у кожному конкретному випадку залежить від поставленої задачі.

Між вищезгаданими показниками існує такий взаємозв'язок:

$$Кд = Квп \times Кпз. \quad \textbf{(5.1.4)}$$

Приклад 5.1.2. Планове завдання щодо обсягу виробництва на 2008 р. порівняно з фактичним обсягом виробництва за 2007 р. склало 105 %. Виконання плану складає 102 %. Знайти коефіцієнт динаміки.

Розв'язок

$Кд = Квп \times Кпз. = 1{,}02 \times 1{,}05 = 1{,}071$ (107,1 %).
Таким чином, у 2008 р. обсяг виробництва зріс порівняно з 2007 р. на 7,1 %.

Приклад 5.1.3. Планом передбачалось збільшення реалізації продукції на 10%, а фактично реалізовано на 15,5 % більше, ніж у базисному періоду. Визначити ступінь виконання плану для обсягу реалізації.

Розв'язок

$$Квп = \frac{Кд}{Кпз} = \frac{100 + 15{,}5}{100 + 10} \times 100 = 105{,}0 \ \%.$$

Таким чином, план з реалізації продукції перевиконано на 5,0 %.

Приклад 5.1.4. Планове завдання передбачало зниження собівартості за рік на 5%, а фактично вона знизилася на 8 %. Визначити виконання плану зі зниження собівартості.

Розв'язок

$$Квп = \frac{Кд}{Кпз} = \frac{100 - 8}{100 - 5} \times 100 = 96{,}8 \ \%.$$

Таким чином, план зі зниження собівартості перевиконано на 3,2 %.

Приклад 5.1.5. Планове завдання передбачало зниження собівартості за рік на 4%, а фактично вона знизилася лише на 2 %. Визначити виконання плану зі зниження собівартості.

Розв'язок

$$K\mathit{вп} = \frac{K\partial}{K\mathit{пз}} = \frac{100-2}{100-4} \times 100 = 102{,}1 \text{ \%}.$$

Таким чином, план зі зниження собівартості недовиконано на 2,1 %.

Відносний показник структури *(коефіцієнт структури $K\mathit{стр}$)* використовують для характеристики складу сукупності. Він показує, яку питому вагу (частку) у загальному підсумку складають окремі частини цілого:

$$K\mathit{стр} = \frac{\mathit{показник,\ що\ характеризує\ частину\ сукупності}}{\mathit{показник\ усієї\ сукупності\ загалом}}. \qquad \textbf{(5.1.5)}$$

Сума цих показників для сукупності завжди дорівнює або 100 % або 1 (якщо у коефіцієнтах).

Приклад 5.1.6. Розглянемо табл. 3.1.1.

Розраховані в останній графі табл. 3.1.1 проценти є відносними показниками структури (питомої ваги). Сума всіх питомих ваг дорівнює 100 %.

Відносний показник координації *(коефіцієнт координації $K\mathit{коорд}$)* характеризує співвідношення між частинами цілого. Одну зі складових частин цілого обирають базою порівняння і знаходять відношення до неї всіх інших частин:

$$K\mathit{коорд} = \frac{\mathit{показник, що\ характеризує\ i-ту\ частину\ сукупності}}{\begin{matrix}\mathit{показник, що\ характеризує\ частину\ сукупності,}\\ \mathit{обрану\ базою\ порівняння}\end{matrix}}. \qquad \textbf{(5.1.6)}$$

За допомогою відносних показників координації визначають, скільки одиниць певної частини цілого припадає на 1, на 100, на 1000 і т.д. одиниць іншої частини, обраною базою порівняння. До таких показників належать, наприклад, кількість хлопчиків на 100 дівчат серед новонароджених; кількість осіб, зайнятих у сфері обслуговування, на 100 осіб, зайнятих у сфері виробництва і т.д. У результаті визначають, у скільки разів ця частина більше базисної, або скільки процентів ця частина складає від базисної.

Приклад 5.1.7. Розглянемо табл. 3.1.1.

За даними цієї таблиці можна розрахувати, що на кожний гектар, відведений під громадську забудову, припадає 6,095 га під житлову забудову ($\frac{1920}{315}$) і 2,127 га під промислову забудову ($\frac{670}{315}$).

Відносний показник інтенсивності *(коефіцієнт інтенсивності Кінт)* характеризує ступінь поширення того чи іншого явища у певному середовищі. Він вимірює ступінь насиченості певного середовища цим явищем. Зазвичай це співвідношення між двома різнойменними величинами, наприклад, кількість голів крупної рогатої худоби на 100 га сільськогосподарських угідь:

$$Кінт = \frac{показник, що\ характеризує\ явище}{показник, що\ характеризує\ середовище\ поширення\ явища}. \quad \textbf{(5.1.7)}$$

Відносні показники інтенсивності, на відміну від інших видів відносних показників, є іменованими числами і часто мають подвійну розмірність тих абсолютних показників, співвідношення яких вони виражають. Відносний показник інтенсивності зазвичай розраховують тоді, коли абсолютна величина виявляється недостатньою для формулювання обґрунтованих висновків про масштаби явища, його розміри, насиченість, щільність розподілу. Наприклад, для визначення рівня забезпеченості населення легковими авто розраховують кількість авто, що припадає на 100 сімей; для визначення щільності населення розраховується кількість людей, що припадає на 1 км^2.

Різновидом відносних показників інтенсивності є відносні показники рівня економічного розвитку, які характеризують розміри виробництва різних видів продукції на душу населення, рівень споживання тощо. Вони обчислюються шляхом ділення річного виробництва або споживання певного виду продукції на середньорічну чисельність населення.

Відносний показник порівняння *(коефіцієнт порівняння Кпорівн)* являє собою відношення однойменних величин (абсолютних показників), що характеризують різні об’єкти або території за один і той самий період або момент часу:

$$Кпорівн = \frac{показник, що\ характеризує\ об'єкт\ А}{показник, що\ характеризує\ об'єкт\ Б}. \quad \textbf{(5.1.8)}$$

Наприклад, співвідношення між рівнями собівартості певного виду продукції, виготовленої на двох підприємствах; співвідношення між рівнями продуктивності праці у різних країнах (за однакової методики розрахунку).

Середня величина є найбільш поширеною формою статистичних показників, використовуваною у соціально-економічних дослідженнях. Вона є узагальненою кількісною характеристикою ознаки у статистичній сукупності у конкретних умовах місця і часу. Найбільш докладно про цю характеристику йтиметься у главі 6.

5.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Що таке абсолютний статистичний показник? Наведіть приклади.
2. Назвіть види статистичних показників. Наведіть приклади.
3. У яких одиницях вимірювання виражаються абсолютні статистичні показники? Наведіть приклади.
4. Що називається відносним статистичним показником?
5. Які умови правильного розрахунку відносних показників? Які види відносних показників ви знаєте? Наведіть приклади.
6. Вкажіть відносні показники інтенсивності:
а) на 1000 жінок припадає 895 чоловіків;
б) кількість новонароджених на 1000 мешканців складає 13,5.
7. Вкажіть відносні величини просторового порівняння:
а) забезпеченість житлом населення у регіоні А складає 19 м^2 на людину, в регіоні Б – 26 м^2 на людину;
б) вартість 1 м^2 житла у регіоні А у 2 рази вище, ніж у регіоні Б.
8. Вкажіть відносні величини координації:
а) на 1000 зайнятих у народному господарстві регіону припадає 175 з вищою освітою;
б) у розрахунку на 1000 осіб відповідної статі у шлюбі перебувають 730 чоловіків і 510 жінок.
9. Вкажіть відносні величини динаміки:
а) за минулий рік реальна заробітна плата підвищилась у 1,2 разів;
б) дефіцит бюджету зменшився на 5 %.
10. За 1 півріччя 2008 р. підприємства міста випустили фарби 1159 т, що склало 90,4 % фактичного випуску за 1 півріччя 2007 р. План на 1 півріччя 2008 р. виконано на 98,2 %. Визначити обсяг фактичного випуску фарби у 1 півріччі 2007 р. і обсяг планового завдання на 1 півріччя 2008 р.
11. Робітник виготовив у 2008 р. 10000 деталей, за норми 8000. У 2007 р. його виробіток склав 5500 деталей. Визначити відносні величини виконання плану і планового завдання.
12. Товарооборот магазину в серпні склав 237,5 тис. грн. за плану 250 тис. грн. У липні товарооборот склав 245,0 тис. грн. Визначити відносні показники виконання плану і динаміки.
13. Планом передбачалося підвищення продуктивності праці на 2 %, а фактично вона підвищилась на 6 %. Визначте перевиконання плану зі зростання продуктивності праці .
14. Планом передбачалося зниження собівартості одиниці продукції на 3%, а фактично вона була знижена на 5 % у порівнянні з базисним періодом. Визначте виконання плану зі зниження собівартості.
15. Вкажіть види відносних показників: а) у 2007 р. інвестиції у розвиток промисловості було збільшено порівняно з 2006 р. на 10 %;

б) капіталовкладення у сільське господарство склали 19 % усіх капіталовкладень в економіку району.

16. Вкажіть види відносних показників: а) на початок року в області на кожних 1000 чоловік міського населення припадало 594 сільського населення; б) на заводі на кожних 100 робітників припадає 1 службовець.

17. Планове завдання з випуску продукції складає 110%, показник динаміки – 105%. Визначте виконання плану.

18. Показник динаміки собівартості складає 98,3 %, план виконано на 102 %. Визначте планове завдання зі зниження собівартості, поясніть результат.

19. Планом передбачалося підвищити врожайність на 2 %, а фактично вона зросла на 3,1 % і склала 35,4 ц/га. Визначте виконання плану з підвищення врожайності.

20. Вкажіть види відносних показників: а) у Токіо на кожного мешканця припадає 4 м^2 паркової зони; б) у Парижі цей показник у 3 рази вище.

21. Вантажооборот підприємства склав у 2008 р. 140 млн. ткм. при плані 135 млн. ткм. У 2007 р. вантажооборот склав 128 млн. ткм. Визначте відносні показники виконання плану, планового завдання і динаміки. Зробіть висновки.

22. Вкажіть види відносних показників: а) 80 % науково-дослідницьких робіт у Японії фінансуються приватним сектором; б) у місті на кожних 100 чоловіків з вищою освітою припадає 90 жінок з такою ж освітою.

23. У 2006 р. у регіоні: а) на кожні 100 га посіву зернових культур припадало 3 га цукрового буряку; б) щоденно вироблялося 102,5 пари шкіряного взуття. Вкажіть види відносних показників.

24. Електробаланс народного господарства характеризується такими даними, млрд. кВтч: вироблено електроенергії – 156; спожито –128 (у тому числі: промисловістю – 72; сільським господарством – 19; транспортом – 14; іншими галузями – 23). Експорт електроенергії – 20.

Визначити: а) втрати електроенергії у мережі загального користування; б) частку втрат і частку експортованої електроенергії; в) структуру і співвідношення електроенергії, спожитої галузями народного господарства.

25. Є дані про виробництво паперу у регіоні. Розрахувати відносні показники динаміки зі змінною і постійною базою порівняння.

	2005 р.	2006 р.	2007 р.	2008 р.
Вироблено паперу, тис. т	3603	2882	2215	2771

ГЛАВА 6

СЕРЕДНІ ВЕЛИЧИНИ

6.1
МЕТОДИЧНІ ВКАЗІВКИ

У цій темі розглядаються питання, що стосуються класифікації середніх величин і принципів вибору їхньої конкретної форми залежно від поставленої задачі. При вивченні матеріалу особлива увага приділяється сутності і видам середніх величин; середній арифметичній, середній гармонічній; моді і медіані.

Середньою величиною називають показник, який дає узагальнену характеристику варіативної ознаки одиниць однорідної сукупності. Середня є величиною абстрактною, але при цьому відображає конкретні властивості всієї сукупності у вигляді одної величини.

Середня дозволяє

1. У вигляді одної величини дати узагальнену характеристику певної властивості (тобто замінити багату кількість різних значень ознаки середньою величиною, що характеризує всю сукупність).
2. Виявити загальні риси і усунути випадкові риси соціально-економічних явищ і процесів.
3. Визначити загальну закономірність (тенденцію), притаманну конкретному розподілу.
4. Встановити критерій для оцінювання рівня, досягнутого окремими одиницями сукупності.

Умови правильного розрахунку середніх

1. Індивідуальні величини, з яких обчислюються середні, повинні стосуватися однорідної сукупності, а їхня кількість повинна бути значною. Наприклад, якщо розрахувати середній рівень доходів службовців якогось району, то вийде фіктивний середній показник, оскільки для його обчислення використано неоднорідну сукупність, яка включає службовців підприємств різних типів (державних, орендних, акціонерних), а також органів державного управління, сфери науки, культури, освіти тощо.
2. Метод середніх величин повинен поєднуватися з методом групувань (тобто якщо ми маємо справу з прикладом п. 1, то для знаходження середнього рівня доходів необхідно виділити однорідні групи, для яких і обчислюються типові групові середні).

Середня, розрахована для сукупності загалом, називається *загальною середньою.* Середні, обчислені для кожної групи, називаються *груповими* *середніми.* Загальна середня відображає загальні риси явища, що вивчається, а групова середня дає характеристику розміру явища, на рівні окремої групи. Середні, застосовувані у статистиці, належать до класу *ступеневих.*

Загальна формула ступеневої середньої:

$$\overline{X} = \sqrt[\kappa]{\frac{\sum X^{\kappa}}{n}}, \qquad \textbf{(6.1.1)}$$

де $\overline{\boldsymbol{X}}$ – середня величина;
X – варіанти ознаки;
κ – показник ступеня;
n – число варіантів.

Найбільш поширені в економічній статистиці прості середні представлено у табл. 6.1.1.

Таблиця 6.1.1 – Форма простих середніх залежно від показника ступеня κ

Значення κ	Формула простої середньої	Вид середньої
-1	$\overline{X} = \dfrac{n}{\sum \frac{1}{X}}$	Середня гармонічна проста
0	$\overline{X} = \sqrt[n]{\Pi(X)}$, П – добуток	Середня геометрична проста
1	$\overline{X} = \dfrac{\sum X}{n}$	Середня арифметична проста
2	$\overline{X} = \sqrt{\dfrac{\sum X^2}{n}}$	Середня квадратична проста

Правило мажорантності середніх:

Чим вище показник ступеня κ, тим більшим буде розмір самої середньої,

тобто $\overline{\boldsymbol{X}}_{\text{гарм}} < \overline{\boldsymbol{X}}_{\text{геом}} < \overline{\boldsymbol{X}}_{\text{арифм}} < \overline{\boldsymbol{X}}_{\text{квадр}}$

Приклад 6.1.1. Нехай $X_1=2$, $X_2=5$,

тоді $\overline{X} = \dfrac{1+1}{\frac{1}{2}+\frac{1}{5}} = 2{,}67$ – середня гармонічна проста,

$$\overline{X} = \frac{2+5}{2} = 3{,}5$$ – середня арифметична проста,

$$\overline{X} = \sqrt{\frac{2^2+5^2}{2}} = 3{,}8$$ – середня квадратична.

Правильну характеристику окремих сукупностей можна отримати лише через використання певних видів середньої. Для встановлення виду середньої використовують її визначальну властивість – при заміні всіх значень варіантів середньою величиною обсяг ознаки не зміниться.

У багатьох випадках середню найзручніше визначити через вихідне (логічне) співвідношення середньої (ВСС):

$$ВСС = \frac{обсяг\ \ варіативної\ \ ознаки}{чисельність\ \ одиниць\ \ \ сукупності} \text{ – середня агрегатна.} \qquad (\mathbf{6.1.2})$$

Для кожного показника, використовуваного у соціально-економічному аналізі, можна розрахувати тільки одну середню агрегатну. Однак алгоритм побудови ВСС залежить від способу представлення вихідних даних. Розрахунок більшості конкретних статистичних показників оснований на використанні середньої агрегатної, середньої арифметичної і середньої гармонічної.

Середня арифметична проста розраховується за формулою:

$$\overline{X} = \frac{\sum X}{n}, \qquad (\mathbf{6.1.3})$$

і використовується тоді, коли дані представлені у вигляді первинного ряду розподілу, тобто не згруповані, або тоді, коли у варіаційному ряді розподілу всі частоти є однаковими.

Приклад 6.1.2. Визначити середній виробіток одного робітника за даними про виробіток кожного з п'яти цехів (дет.): 21, 23, 25, 19, 20.

Розв'язок

$$\overline{X} = \frac{21+23+25+19+20}{5} = 22 \text{ (дет.)}$$

Якщо окремі значення досліджуваної сукупності повторюються неоднакову кількість разів і представлені у вигляді ряду розподілу з нерівними частотами, то для розрахунку середньої використовують формулу **середньої арифметичної зваженої:**

$$\overline{X} = \frac{\sum Xf}{\sum f}, \qquad (\mathbf{6.1.4})$$

де f – частота окремих варіантів.

Множення варіанта на частоту називається зважуванням, а частота – вагою середньої. Крім частоти f вагою може бути і частість w. Також необхідно враховувати, що вага іноді може розраховуватися як добуток двох або навіть трьох значень.

Приклад 6.1.3. За даними табл. 4.1.2 визначити середній виробіток робітника підприємства.

Розв'язок

$$\overline{X} = \frac{\sum Xf}{\sum f} = \frac{20\times16+30\times22+35\times33+40\times14}{85} = 32 \text{ (од.)}$$

Обчислення середньої у дискретному ряді не потребує ніяких перетворень.

Обчислення середньої для інтервального варіаційного ряду потребує переходу від інтервального ряду до дискретного, і для цього значення інтервалів замінюють їхніми центральними значеннями, тобто напівсумою верхньої та нижньої межі кожного інтервалу. Якщо ряд має відкриті інтервалами, то такі інтервали необхідно закрити. Розмір першого відкритого інтервалу дорівнює розміру наступного закритого інтервалу. Останній відкритий інтервал дорівнює попередньому закритому інтервалу.

Приклад 6.1.4. За даними прикладу 4.1.1, табл. 4.1.7 визначити середню заробітну плату робітника.

Розв'язок

Оскільки ряд згруповано, необхідно застосувати формулу (6.1.4). Розв'яжемо задачу табличним способом (див. графу 3 табл. 6.1.2):

Таблиця 6.1.2 – Розрахункова

Групи робітників за розміром зарплати, грош. од.	Середина інтервалу, грош.од. (X)	Кількість робітників, чол. (f)	($X\times f$)
А	*1*	*2*	*3*
154,00–294,83	224,415	3	673,245
294,83–435,66	365,245	7	2556,715
435,66–576,49	506,075	4	2024,300
576,49–717,32	646,905	5	3234,525
717,32–858,15	787,735	3	2363,205
858,15–999,00	928,575	8	7428,600
Усього	Х	30	18280,590

$$\overline{X} = \frac{18280,590}{30} = 609,4 \text{ (грош.од.).}$$

Найважливіші властивості середньої арифметичної

1. Добуток середньої на суму частот дає суму добутків варіантів на частоти:
$$\overline{X}\sum f = \sum Xf \, .$$
2. Сума відхилень від середньої дорівнює нулю:
$$\sum(X - \overline{X}) = 0 \text{ – за простою середньою,}$$
$$\sum(X - \overline{X})f = 0 \text{ – за зваженою середньою.}$$
3. Якщо з усіх значень ознаки вирахувати одне і те ж саме число, то середня, відповідно, зменшиться або збільшиться на це ж саме число:
$$\frac{\sum(X \pm A)}{n} = \overline{X} \pm A \text{ – проста середня,}$$
$$\frac{\sum(X \pm A)f}{\sum f} = \overline{X} \pm A \text{ – зважена середня.}$$
4. Якщо всі значення ознаки зменшити або збільшити в *a* разів, то середня відповідно зменшиться або збільшиться у це ж саме число разів:
$$\frac{\sum \frac{X}{a} f}{\sum f} = \frac{\overline{X}}{a}, \qquad \frac{\sum (X \times a) f}{\sum f} = \overline{X} \times a \, .$$
5. Якщо всі частоти поділити або помножити на одне і те ж саме число, то середня не зміниться:
$$\frac{\sum X \frac{f}{d}}{\sum \frac{f}{d}} = \overline{X}, \qquad \frac{\sum X(fd)}{\sum fd} = \overline{X} \, .$$
6. Середня залежить не від розміру значення ваг або частот, а від співвідношення між ними, тобто від структури сукупності.

На підставі властивостей середньої арифметичної для розрахунку середньої в інтервальних рядах розподілу з рівними інтервалами застосовують *спосіб моментів* або спосіб відліку від умовного нуля. Цей спосіб значно зменшує витрати часу і праці для обчислення середньої величини. За цим методом всі варіанти зменшують на число A (де A – варіант, що має найбільшу і тому найбільш незручну для обчислення, частоту), а потім ці варіанти зменшують у i разів (де i – розмір інтервалу). У такий спосіб отримують новий варіаційний ряд розподілу з варіантами $X' = \frac{X - A}{i}$, а їхня нова середня арифметична M_1 є моментом першого порядку:

$$M_1 = \frac{\sum(\frac{X-A}{i})f}{\sum f}. \qquad \textbf{(6.1.5)}$$

Середня для первинного ряду розподілу, розрахована способом моментів, виражається формулою:

$$\overline{X} = M_1 i + A. \qquad \textbf{(6.1.6)}$$

Приклад 6.1.5. Знайти середню вартість основних виробничих фондів (ОВФ) способом моментів і класичним способом, використовуючи вихідні дані граф А і 1 табл. 6.1.3.

Таблиця 6.1.3 – Розрахункова

Групи підприємств за вартістю ОВФ, млн. грн.	Кількість підприємств, F	Середини інтервалів, X	$X \times f$	$\frac{X-A}{i}$	$(\frac{X-A}{i}) \times f$
А	*1*	*2*	*3*	*4*	*5*
До 16 (14–16)	2	15	30	-2	-4
16–18	6	17	102	-1	-6
18–20	10	19	190	0	0
20–22	4	21	84	1	4
Більше 22 (22–24)	3	23	69	2	6
Усього	25	Х	475	Х	0

Розв'язок

Задачу розв'яжемо табличним способом (див. графи 2, 3, 4, 5 табл. 6.1.3).

1. Класичний спосіб знаходження середньої арифметичної зваженої – з застосуванням формули 6.1.4:

$\overline{X} = \frac{\sum Xf}{\sum f}$=475/25=19 (млн. грн.).

2. Способом моментів – з застосуванням формул 6.1.5 и 6.1.6.

i=2, A=19.

$$M_1 = \frac{\sum(\frac{X-A}{i})f}{\sum f} = \frac{0}{25} = 0.$$

$\overline{X} = M_1 i + A = 0 \times 2 + 19 = 19$ (млн. грн.).

Середня гармонічна зважена застосовується тоді, коли відомі індивідуальні значення варіанта X і добутку $X \times f = Z$, які не дорівнюють одне одному:

$$\overline{X} = \frac{\sum Z}{\sum \frac{Z}{X}} \quad . \qquad (6.1.7)$$

Середня гармонічна проста розраховується за формулою:

$$\overline{X} = \frac{n}{\sum \frac{1}{X}} \quad , \qquad (6.1.8)$$

і застосовується тоді, коли $(X \times f)$ є однаковими.

Приклади розрахунку середніх відносних величин, виходячи з логічного співвідношення для середньої

- Розрахунок середнього процента виконання плану ($\overline{Квп}$):

а) за формулою середньої арифметичної зваженої:

$$\overline{Квп} = \frac{\sum Квп \times Y_{пл}}{\sum Y_{пл}}; \qquad (6.1.9)$$

б) за формулою середньої гармонічної зваженої:

$$\overline{Квп} = \frac{\sum Y_1}{\sum \frac{Y_1}{Квп}}. \qquad (6.1.10)$$

- Розрахунок середнього процента продукції вищої якості ($\overline{d}_{вя}$):

а) за формулою середньої арифметичної зваженої:

$$\overline{d}_{вя} = \frac{\sum d_{вя} \times Y_1}{\sum Y_1}; \qquad (6.1.11)$$

б) за формулою середньої гармонічної зваженої:

$$\overline{d}_{вя} = \frac{\sum Y_{вя}}{\sum \frac{Y_{вя}}{d_{вя}}}, \qquad (6.1.12)$$

де $Y_{вя}$ – обсяг випуску продукції вищої якості;

$d_{вя}$ – питома вага продукції вищої якості (гатунку) у загальному обсязі фактично виробленої продукції.

- Розрахунок середнього процента бракованої продукції ($\overline{d}_{бр}$):

а) за формулою середньої арифметичної зваженої:

$$\bar{d}_{бр} = \frac{\sum d_{бр} \times Y_1}{\sum Y_1}; \qquad \textbf{(6.1.13)}$$

б) за формулою середньої гармонічної зваженої:

$$\bar{d}_{бр} = \frac{\sum Y_{бр}}{\sum \frac{Y_{бр}}{d_{бр}}}, \qquad \textbf{(6.1.14)}$$

де $Y_{бр}$ – обсяг бракованої продукції;

$d_{бр}$ – питома вага бракованої продукції у загальному обсязі фактично виробленої продукції.

Приклад 6.1.6. Є такі дані про виробничі показники двох цехів підприємства за звітний період (табл. 6.1.4):

Таблиця 6.1.4 – Вихідна

№ фабрики	Фактичний випуск продукції, тис. грн., Y_1	Процент виконання плану, %, *Квп*	Частка стандартної продукції, %, $d_{сп}$
1	475,0	95,0	80,0
2	420,0	105,0	90,0

Обчисліть для двох цехів разом:
1) середній процент виконання плану випуску продукції;
2) середній процент стандартної продукції.

Вкажіть, який вид середньої треба застосувати для обчислення цих показників.

Розв'язок

1. Виходячи з формули 5.1.1, середній процент виконання плану випуску продукції для всіх цехів знаходимо за формулою:

$$\overline{Квп} = \frac{\sum \mathrm{Y}_1}{\sum Y_{пл}},$$

Сума фактичного випуску продукції (Y_1) відома, а для знаходження суми запланованого випуску продукції ($Y_{пл}$) необхідно з формули 5.1.1 виразити $Y_{пл}$ і підставити у формулу, представлену вище.

Таким чином: $$\overline{Квп} = \frac{475{,}0 + 420{,}0}{\frac{475{,}0}{0{,}95} + \frac{420{,}0}{1{,}05}} = \frac{895}{900} = 0{,}994 \ (99{,}4\,\%).$$

2. Середній процент стандартної продукції розраховується за такою формулою:

$$\bar{d}cn = \frac{\sum \text{Y}cn}{\sum Y_1},$$

де $\bar{d}cn$ – середній процент стандартної продукції;

Ycn – кількість стандартної продукції.

Тоді: $$\bar{d}cn = \frac{475{,}0 \times 0{,}80 + 420{,}0 \times 0{,}90}{475{,}0 + 420{,}0} = \frac{758{,}0}{895{,}0} = 0{,}847\ (84{,}7\%).$$

Для обчислення $\bar{K}вп$ використовувалась формула розрахунку середнього процента виконання плану за допомогою середньої гармонічної зваженої (6.1.7). Для обчислення $\bar{d}cn$ використовувалась формула розрахунку середнього процента продукції вищої якості за допомогою середньої арифметичної зваженої (6.1.4).

Мода і медіана характеризують особливості розподілу одиниць сукупності за величиною ознаки, що вивчається, але менше, ніж загальна середня, залежать від складу сукупності. Вони не залежать від крайніх значень сукупності і називаються розподільними, або структурними, або позиційними середніми.

Мода – це варіант дискретного ряду, який зустрічається найчастіше, тобто варіант, що має найбільшу частоту.

В інтервальному ряді розподілу приблизною модою вважають центральний варіант так званого *модального інтервалу*, тобто інтервалу, що має найбільшу частоту або частість. Тоді мода визначається так:

$$M_o = X_0 + i_m \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}, \qquad \textbf{(6.1.15)}$$

де X_0 – нижня межа модального інтервалу;

i_m – величина модального інтервалу;

f_m – частота модального інтервалу;

f_{m-1} – частота інтервалу, що передує модальному;

f_{m+1} – частота інтервалу, що слідує за модальним.

Медіана – це варіант, що знаходиться у середині ранжованого варіаційного ряду.

Якщо дискретний ряд має непарну кількість рівнів ряду, які не повторюються, то медіаною буде значення варіанта, що знаходиться у центрі ряду. А якщо ряд складається з парної кількості рівнів, медіана дорівнюватиме середньому з двох значень ознаки, розташованих у середині ряду.

Для визначення медіани у дискретному ряді за наявності частот необхідно підрахувати суму накопичених (кумулятивних) частот ряду. Сумування треба продовжувати до отримання такої накопиченої суми частот, яка перевищить половину обсягу сукупності або дорівнюватиме їй.

Напроти цієї кумулятивної частоти і знаходиться варіант ознаки, який є медіаною.

В інтервальному варіаційному ряді розподілу медіану визначають після попереднього знаходження *медіанного інтервалу*, тобто інтервалу, накопичена частота якого дорівнює напівсумі частот або перевищує її, за формулою:

$$M_e = X_0 + i_m \frac{\frac{1}{2}\sum f - S_{me-1}}{f_{me}}, \qquad \textbf{(6.1.16)}$$

де X_0 – початкове значення інтервалу, який містить медіану;
i_m – величина медіанного інтервалу;
$\sum f$ – сума частот ряду;
S_{me-1} – накопичена частота в інтервалі, що передує медіанному;
f_{me} – частота медіанного інтервалу.

Приклад 6.1.7. За даними прикладу 6.1.5 визначити моду і медіану в інтервальному ряді розподілу (табл. 6.1.5).

Таблиця 6.1.5 – Розрахункова

Групи підприємств за вартістю ОВФ, млн. грн.	Кількість підприємств, од., f	Накопичена (кумулятивна) частота, S
14–16	2	2
16–18	6	8
18–20	**10**	**18**
20–22	4	22
22–24	3	25
Усього	25	Х

Розв’язок

$$M_O = 18 + 2 \times \frac{10-6}{(10-6)+(10-4)} = 18{,}8 \text{ (млн. грн.)}$$

Тобто найчастіше зустрічаються підприємства з розміром ОВФ 18,8 млн. грн.

$$\frac{\sum f}{2} = \frac{25}{2} = 12{,}5; \quad Me = 18 + 2 \times (\frac{25}{2} - 8)/10 = 18{,}9 \text{ (млн. грн.)}$$

Це свідчить, що майже у половини з 25 підприємств вартість ОВФ менше 18,9 млн. грн., а в іншої частини – більше 18,9 млн. грн.

Мода і медіана, на відміну від середніх, є конкретними характеристиками, тобто їхні значення мають якийсь конкретний варіант у варіаційному ряді. Мода і медіана як правило відрізняються від значення середньої і дорівнюють їй лише у разі симетричного розподілу частот варіаційного ряду (тобто дозволяють оцінити асиметрію ряду).

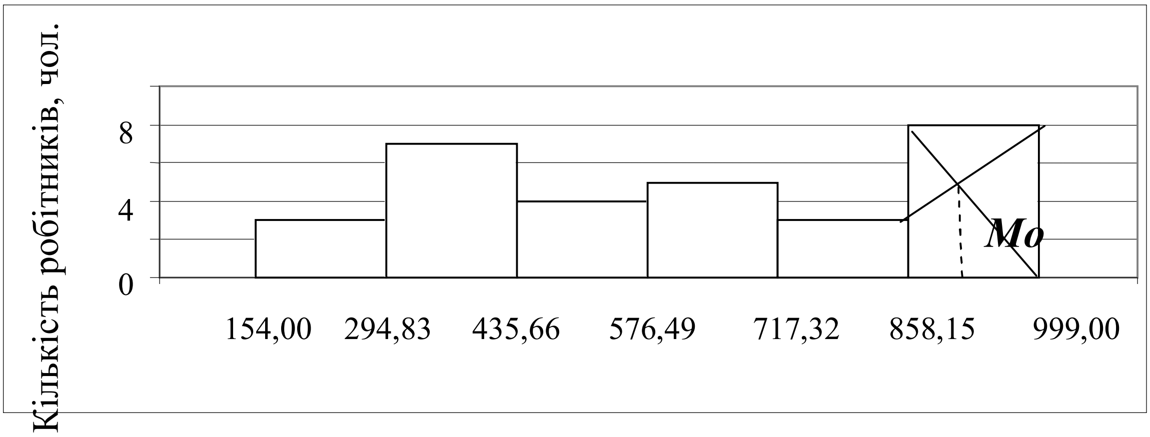
Приклад 6.1.8. За даними табл. 6.1.2 (А, 1, 2 графи) розрахувати моду і медіану і зобразити ці характеристики графічно (табл. 6.1.6)

Таблиця 6.1.6 – Розрахункова

Групи робітників за рівнем зарплати, грош. од.	Середина інтервалу, грош. од., *X*	Кількість робітників, чол., *F*	Кумулятивна частота, *S*
A	*1*	*2*	*3*
154,00–294,83	224,415	3	3
294,83–435,66	365,245	7	10
435,66–576,49	506,075	4	14
576,49–717,32	646,905	5	19
717,32–858,15	787,735	3	22
858,15–999,00	928,575	8	30
Усього	Х	30	Х

Розв’язок

Модальний інтервал – (858,15–999,00), оскільки частота цього інтервалу (f=8) є максимальною. Медіанний інтервал – (576,49–717,32), оскільки накопичена частота цього інтервалу вперше перевищує напівсуму всіх частот ряду розподілу ($S=19,\ 19 \geq \frac{30}{2}$).



Заробітна плата, грош.од.

Рисунок 6.1.1 – Графічне зображення моди інтервального ряду

$$M_o = 858{,}15 + 140{,}83 \times \frac{8-3}{(8-3)+(8-0)} = 912{,}3 \text{ (грош. од.).}$$

$$M_e = 576{,}49 + 140{,}83 \times \frac{15-14}{5} = 604{,}7 \text{ (грош. од.).}$$

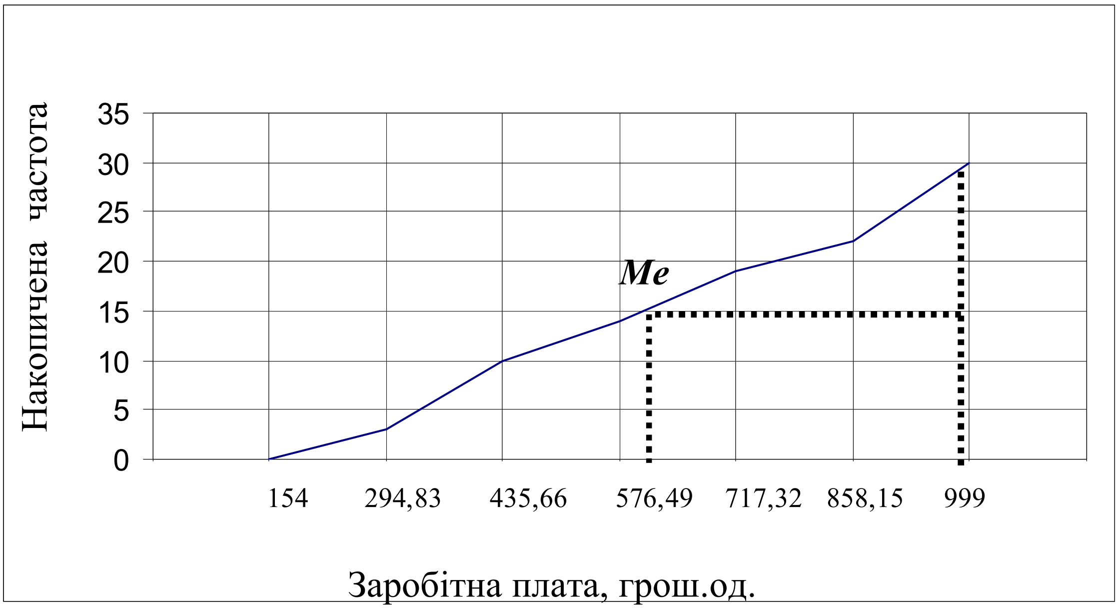


Рисунок 6.1.2 – Графічне зображення медіани інтервального ряду

6.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Середня, її сутність і визначення.
2. Умови правильного використання середніх.
3. Середня арифметична.
4. Властивості середньої геометричної. Метод моментів.
5. Середня гармонічна, область її застосування.
6. Розрахунок середньої з відносних величин.
7. Логічний смисл таких характеристик, як мода і медіана.
8. Продаж яблук на ринках двох міст характеризується такими даними:

Місто	Вересень		Грудень	
	Середня ціна, грош.од.	Кількість проданих яблук, тис.кг	Середня ціна, грош.од.	Кількість проданих яблук, тис.кг
А	0,70	40	1,20	20
Б	0,90	50	1,00	25

Розрахуйте для двох міст середню ціну яблук у вересні і грудні. Вкажіть вид середньої. Порівняйте отримані показники.

9. Є дані для підприємств, що випускають однойменну продукцію. Розрахуйте середній виробіток на одного робітника:

1) для цеху 1;
2) для цеху 2.

Поясніть застосування різних видів середніх величин.

№ бригади	Цех №1		№ Бригади	Цех №2	
	Виробіток продукції на одного робітника, од.	Кількість робітників		Виробіток продукції на одного робітника, од..	Обсяг усієї фактично виробленої продукції, од.
1	20	16	1	21	252
2	30	22	2	29	522
3	35	33	3	35	875
4	40	14	4	40	800

10. Є дані про чисельність робітників у промислових і будівельних бригадах одного з районів області, які перейшли на орендну форму у звітний період:

№ Групи	Промисловість		Будівництво	
	Кількість робітників у бригаді, чол.	Кількість бригад, од.	Кількість робітників у бригаді, чол.	Загальна кількість робітників у бригадах, чол.
1	15	120	19	950
2	18	150	23	1840

Розрахуйте середню кількість робітників у одній бригаді:

1) у промисловості;
2) у будівництві.

Вкажіть, який вид середньої треба застосувати для розрахунку цих показників. Порівняйте отримані дані.

11. Є дані про середньоденний товарооборот продавців магазинів роздрібної торгівлі двох торговельних мереж. Обчисліть середньоденний товарооборот продавця торговельної мережі: №1; №2. Вкажіть, у якій мережі вище середньоденний товарооборот одного продавця і який вид середньої треба застосувати для обчислення цих показників.

№ магазина	Торговельна мережа №1		Торговельна мережа №2	
	Середній товарооборот продавця, тис. грн.	Кількість продавців, чол.	Середній товарооборот продавця, тис. грн.	Загальний товарооборот, тис. грн.
1	16,0	54	15,5	930
2	18,0	60	19,0	1615

12. На підставі наведених нижче даних визначте, за який рік середня врожайність озимої пшениці у трьох бригад господарства вище і на скільки?

Бригада	Базисний рік		Звітний рік	
	Врожайність, ц з 1 га	Посівна площа, га	Врожайність, ц з 1 га.	Валовий збір, ц
1	31,8	250	33,6	6900
2	32,5	260	34,5	10350
3	35,0	290	37,4	11220

Які види середніх застосовувалися і чому?

13. Для одного з цехів є такі дані про строк служби станків:

Строк служби станків (років)	Кількість станків (шт.)
До 5 років	5
5–10	10
10–15	25
15 і більше	10

Обчисліть середній строк служби станків у цьому цеху.

14. Є такі дані для господарств одного з районів:

Групи господарств за врожайністю картоплі, ц/га	Кількість господарств	Посівна площа у середньому на одне господарство
До 100	6	90
100–140	15	110
140 і вище	6	120
Усього	27	Х

Визначте середній розмір посівної площі на 1 господарство і середню врожайність картоплі у районі.

15. Обчисліть середньорічний виробіток на одного продавця продовольчого магазину, якщо за цим показником магазини розподіляються таким чином:

Виробіток продавця, тис. грн.	Кількість магазинів	Частка продавців магазину в загальній кількості продавців, %
До 600	7	20
600–800	8	25
800–1000	15	50
1000 і більше	10	5
Усього	40	100

16. Є дані для двох груп заводів:

№ заводу	1 група		№ заводу	2 група	
	Фактичний випуск продукції, тис. грн.	Виконання плану, %		Планове завдання, тис. грн.	Виконання плану, %
12	26,25	105	21	24,0	100
23	24,5	98	33	26,8	105
31	24,25	100	38	20,0	95

Обчисліть середній процент виконання плану випуску продукції:

1) для 1 групи заводів;
2) для 2 групи заводів.

Які види середніх застосовувалися і чому?

17. Є такі дані про товарооборот продовольчих магазинів за звітний період у двох мікрорайонах міста:

Мікрорайон А			Мікрорайон В		
Магазини	Фактичний товарооборот, тис. грн.	Виконання плану, %	Магазини	План товарообороту, тис. грн.	Виконання плану, %
1	250,0	104.2	4	360,0	101,4
2	240,0	96,4	5	140,0	96,2
3	580,0	102,5	6	620,0	106,0

На підставі наведених даних визначте для кожного мікрорайону:

1) середній процент виконання плану роздрібного товарообороту продовольчими магазинами;
2) суму понадпланової реалізації товарів у звітному періоді.

   Обґрунтуйте використання формул середніх для розрахунку заданих показників.

18. Є такі дані про виробничі показники за звітний період для двох цехів підприємства:

№ цеху	Фактичний випуск продукції, тис. грн.	Виконання плану, %	Частка стандартної продукції, %
1	475,0	95,0	80
2	420,0	105,0	90

Обчисліть для двох цехів разом:

1) середній процент виконання плану випуску продукції;
2) середній процент стандартної продукції.

   Вкажіть, який вид середньої треба застосувати для обчислення цих показників.

19. Є дані про розподіл робітників за кількістю обслуговуваних станків:

						Усього
Кількість обслуговуваних станків:	4	5	6	7	8	30
Кількість робітників	25	38	82	40	15	200

Визначте моду і медіану.

20. Є дані про середньоденний виробіток продавців універмагу. Визначте середньоденний виробіток на одного продавця, моду і медіану денного виробітку.

	Денний виробіток, грн.				
	200–400	400–600	600–800	800–1000	1000–1200
Кількість продавців	5	40	80	50	25

21. Робота автоколони за день характеризується такими даними:

	Вантажооборот (т/км)				
	100–130	130–160	160–190	190–220	220–250
Кількість автомашин	5	10	15	12	8

Визначте моду і медіану для вантажообороту однієї автомашини.

22. Зі звітів трьох цехів фабрики видно, що плановий випуск продукції у звітному періоді, відповідно, склав 500 тис. грн., 648 тис. грн. і 441 тис. грн. План випуску продукції першим цехом виконано на 100 %, другим – на 108% і третім – на 98%. Процент продукції першого ґатунку за той самий період склав, відповідно, 78,5 %, 95,5 % и 85 % до фактичного випуску продукції.

На підставі цих даних визначте для трьох цехів фабрики:

1) середній процент виконання плану випуску продукції;
2) середній процент продукції першого ґатунку.

Вкажіть, які види середньої застосовувалися і чому.

РОЗДІЛ 3

АНАЛІТИЧНА СТАТИСТИКА

ГЛАВА 7

ПОКАЗНИКИ ВАРІАЦІЇ

7.1
МЕТОДИЧНІ ВКАЗІВКИ

Дослідження варіації в статистиці та соціально-економічних дослідженнях має важливе значення, оскільки величина варіації ознаки у статистичній сукупності характеризує її однорідність.

У статистичній практиці для вивчення і вимірювання варіації використовуються різні показники (міри) варіації залежно від поставлених задач. До них належать розмах варіації, середнє лінійне відхилення, дисперсія, середнє квадратичне відхилення і коефіцієнт варіації.

При вивченні питання про варіацію треба чітко уявляти умови, що породжують варіацію ознак, а також сутність і значення вимірювання варіації ознак. Слід також усвідомити, що вивчення варіації ознак суспільних явищ знаходиться у прямому зв’язку з групуваннями, зокрема з рядами розподілу. Дуже важливо навчитися легко обчислювати всі показники варіації.

Необхідність вимірювання варіації ознак. *Варіація* (коливаємість) – це відмінність у значеннях якоїсь ознаки у різних одиниць певної сукупності в один і той самий період часу.

Приклад 7.1.1. На фірмі працюють дві бригади по три людини з різним виробітком (табл.7.1.1):

Таблиця 7.1.1 – Розрахункова

1 бригада (деталей на день)	95	100	105	$\overline{X_1} = 100$
2 бригада (деталей на день)	75	100	125	$\overline{X_2} = 100$

У табл. 7.1.1 розраховано середній виробіток на одного робітника в обох бригадах. Вона є однаковою і складає $\overline{X}_1 = \overline{X}_2 = 100$ деталей. Однак коливаємість виробітку окремих робітників у першій бригаді значно менша, ніж в іншій. Тобто середня величина, яка дає узагальнену характеристику досліджуваної ознаки, не розкриває будову сукупності, не показує, як розташовуються навколо неї варіанти, чи зосереджуються вони поблизу середньої або значно відхиляються від неї. Тому потрібні ще й інші показники, що характеризують відхилення окремих значень від загальної середньої , тобто показники варіації.

Показники варіації:

1) доповнюють середні величини, за якими приховані індивідуальні значення ознак, оскільки середні не показують будову сукупності;
2) характеризують ступінь однорідності сукупності за певною ознакою;
3) характеризують межі варіації ознаки.

Співвідношення між показниками варіації характеризує взаємозв’язок між показниками.

Основні показники варіації

1. *Розмах варіації* (R) – це амплітуда коливань, тобто різниця між мінімальним X_{min} і максимальним X_{max} значеннями ознаки:

$$R = X_{max} - X_{min}\,. \qquad \textbf{(7.1.1)}$$

2. *Середнє лінійне відхилення* (d) – це середнє арифметичне абсолютних значень відхилень окремих варіантів від їхньої середньої:

$$\bar{d} = \frac{\sum\left|X - \bar{X}\right|}{n} \text{ – для незгрупованих даних,} \qquad \textbf{(7.1.2)}$$

$$\bar{d} = \frac{\sum\left|X - \bar{X}\right| \times f}{\sum f} \text{ – для згрупованих даних.} \qquad \textbf{(7.1.3)}$$

d – число іменоване, виражене в тих самих одиницях вимірювання, що варіанти і середнє, і дає абсолютну міру варіації. За його допомогою аналізують, наприклад, склад робітників, ритмічність виробництва, оборот зовнішньої торгівлі.

3. *Дисперсія* (σ^2) – (від лат. «розсіювання») – середній квадрат відхилень варіантів від їхньої середньої величини:

$$\sigma^2 = \frac{\sum (X - \bar{X})^2}{n} \text{ – для незгрупованих даних,} \qquad \textbf{(7.1.4)}$$

$$\sigma^2 = \frac{\sum (X - \bar{X})^2 \times f}{\sum f} \text{ – для згрупованих даних.} \qquad \textbf{(7.1.5)}$$

При розрахунку дисперсії не вказуються одиниці вимірювання.

Методи спрощеного розрахунку дисперсії:

а)
$$\sigma^2 = \bar{X}^2 - (\bar{X})^2 = \frac{\sum X^2 \times f}{\sum f} - \left(\frac{\sum X \times f}{\sum f}\right)^2; \qquad \textbf{(7.1.6)}$$

б) розрахунок дисперсії за методом моментів (у рядах з рівними інтервалами).

Для застосування цього методу необхідно знати дві *властивості дисперсії*:

- якщо всі значення ознаки зменшити або збільшити на одну і ту ж саму постійну величину *A*, то дисперсія від цього не зміниться;
- якщо всі значення ознаки зменшити або збільшити в одне і те ж саме число разів (*i* разів), то дисперсія, відповідно, зменшиться або збільшиться в i^2 разів.

Тоді, за способом моментів:

$$\sigma^2 = i^2 (M_2 - M_1^2), \qquad \textbf{(7.1.7)}$$

де:

$$M_1 = \frac{\sum (\frac{X - A}{i}) \times f}{\sum f} \text{ – момент 1-го порядку (див. формулу 6.1.5),}$$

$$M_2 = \frac{\sum \left(\frac{X - A}{i}\right)^2 \times f}{\sum f} \text{ – момент 2-го порядку.} \qquad \textbf{(7.1.8)}$$

Якщо відхилення від середньої величини у сукупності або занадто малі, або занадто великі, то доцільно обчислювати не дисперсію, а тільки середнє лінійне відхилення. Інакше кажучи, при піднесенні у квадрат цих відхилень (при розрахунку дисперсії) відмінності між варіантами або дедалі зменшуватимуться, або дедалі збільшуватимуться, що затушовує тенденцію розподілу.

4. *Середнє квадратичне відхилення* (σ) – дорівнює кореню квадратному з дисперсії:

$$\sigma = \sqrt{\sigma^2}\,. \qquad \textbf{(7.1.9)}$$

Тобто середнє квадратичне відхилення є узагальненою характеристикою розміру варіації ознаки у сукупності. Вона показує, на скільки у середньому відхиляються конкретні варіанти від їхнього середнього значення, є абсолютною мірою коливаємості ознаки і виражається в тих самих одиницях, що й варіанти.

Чим меншим є значення дисперсії і середнього квадратичного відхилення, тим більш однорідною (кількісно) буде сукупність і тим більш типовою буде середня величина.

5. *Дисперсія альтернативної ознаки* (σ_p^2 **)** (різновид дисперсії).

Вище було розглянуто приклади розрахунку показників варіації для кількісних ознак. Однак поряд з варіацією кількісних ознак може ставитися задача оцінювання варіації якісних ознак. Якщо є два взаємовиключних варіанти значень ознаки, це означає наявність альтернативної мінливості ознак.

Альтернативним ознакам присвоюють два кількісних значення (варіанта):

1 – мають цікаву для нас ознаку;

0 – не мають цікаву для нас ознаку.

У якості ваг присвоюються:

p – частка одиниць (частість варіанта), які мають певну ознаку;

q – частка одиниць (частість варіанта), які не мають певну ознаку:

$$p+q=1, \qquad q=1-p$$

Тоді за формулою 7.1.5:

$$\sigma_p^{\ 2} = \frac{\sum (X-\overline{X})^2 \times f}{\sum f} = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = p \times q = p \times (1\text{-}p),$$

де $\overline{X} = \dfrac{\sum Xf}{\sum f} = \dfrac{1 \times p + 0 \times q}{p+q} = p$ – середнє значення альтернативної ознаки.

Зазначимо, що $p+q$ не може бути більше 1, а тому $p \times q$ не може бути більше 0,25. Отже, максимальне значення дисперсії альтернативної ознаки становить 0,25, коли $p=q=0{,}5$. Середнє квадратичне відхилення альтернативної ознаки може мати значення від 0 до 0,5 [20].

Приклад 7.1.2. На 10000 людей населення району припадає 4500 чоловіків і 5500 жінок. Обчислити дисперсію.

Розв’язок

$$p = \frac{4500}{10000} = 0{,}45, \qquad q = \frac{5500}{10000} = 0{,}55, \qquad \boldsymbol{\sigma_p^{\ 2}} = p \times q = 0{,}45 \times 0{,}55 = 0{,}2475.$$

6. *Коефіцієнт варіації* (V) – відносний показник варіації, який дорівнює відношенню середнього квадратичного відхилення до середньої:

$$V = \frac{\sigma}{\overline{X}} \cdot 100\,\%. \qquad \textbf{(7.1.10)}$$

Необхідність застосування цього коефіцієнта у статистичній практиці виникає в разі непридатності для аналізу розглянутих вище показників абсолютної коливаємості ознак. Це має місце за необхідності:
а) порівняння варіацій різних ознак в одній сукупності (наприклад, вік робітників та їхня кваліфікація, стаж робітників і розмір їхньої заробітної плати тощо); б) порівняння коливаємості однієї і тієї ж самої ознаки у декількох сукупностях з різними середніми.

Якщо коефіцієнт варіації менше 33 %, це означає, що сукупність є однорідною, а середнє для неї є типовим.

Приклад 7.1.3. Є інформація про врожайність соняшника у господарствах області (табл. 7.1.2):

Таблиця 7.1.2 – Вихідна

Врожайність, ц з 1 га	Посівна площа, га
До 13	10
13–15	25
15–17	40
17–19	20
Вище 19	5
Усього	100

На основі даних табл. 7.1.2 обчислити коефіцієнт варіації та зробити висновки про сукупність.

Розв'язок

Таблиця 7.1.3 – Розрахункова

Врожайність, ц з 1 га	Посівна площа, га	Середина інтервалу, (Xi) ц з 1 га	$\frac{Xi-A}{i}$	$\frac{Xi-A}{i}\times f$	$(\frac{Xi-A}{i})^2\times f$
11–13	10	12	–2	–20	40
13–15	25	14	–1	–25	25
15–17	40	16	0	0	0
17–19	20	18	1	20	20
19–21	5	20	2	10	20
ИТОГО:	100	Х	Х	–15	105

Середню врожайність ($\bar{X}$) розрахуємо способом моментів, з використанням формул 6.1.5, 6.1.6.

Довжина інтервалу

$$i = 2 \text{ (ц з 1 га)}.$$

Варіант ряду з максимальною частотою:

$$A = 16 \text{ (ц з 1 га)}.$$

Момент першого порядку:

$$M_1 = \frac{-15}{100} = -0{,}15.$$

Отже: $\bar{X} = 2 \times (-0{,}15) + 16 = 15{,}70$ (ц з 1 га).

Далі розрахуємо дисперсію способом моментів, за формулами 7.1.7, 7.1.8.

Момент другого порядку:

$$M_2 = \frac{105}{100} = 1{,}05.$$

Отже:

$$\sigma^2 = 2^2 \times (1{,}05 - (-0{,}15)^2) = 4{,}11.$$

Коефіцієнт варіації визначимо за формулою 7.1.10:

$$V = \frac{\sqrt{4{,}11}}{15{,}70} \times 100\ \% = 12{,}91\ \%.$$

Оскільки розрахований коефіцієнт варіації менше 33%, то сукупність є кількісно однорідною, а величина середньої врожайності соняшника на 1 га посіву є типовою.

Правило складання дисперсій. Якщо сукупність розчленовано на групи за одним фактором, вивчати варіацію можливо шляхом обчислення і аналізу трьох видів дисперсії:

- загальної;
- міжгрупової;
- внутрішньогрупової.

1. *Загальна дисперсія* (σ^2) – вимірює варіацію ознаки для всієї сукупності під впливом усіх факторів, що зумовили цю варіацію, і обчислюється за формулою 7.1.5.

Фактори поділяються на випадкові і систематичні, тому варіація може бути: а) випадковою – зумовленою дією випадкових величин; б) систематичною – зумовленою дією постійних факторів.

2. *Міжгрупова дисперсія* (δ^2) – характеризує систематичну варіацію результативної ознаки, обумовлену впливом ознаки-фактора, покладеного в основу групування:

$$\delta^2 = \frac{\sum \left(\bar{X}_i - \bar{X}\right)^2 \times f}{\sum f}, \qquad \textbf{(7.1.11)}$$

де $\bar{X}_i$ – групові середні;

$\bar{X}$ – загальна середня;

f – кількість одиниць у групі.

3. *Внутрішньогрупова дисперсія* **(**σ_i^2**)** – це дисперсія в межах кожної групи, яка відображає випадкову варіацію, тобто частину варіації, обумовлену дією неврахованих факторів і незалежну від ознаки-фактора, покладеного в основу групування.

$$\sigma_i^2 = \frac{\sum (X - \bar{X}_i)^2 \times f_i}{\sum f_i}, \qquad \textbf{(7.1.12)}$$

де f_i – частота варіанта X_i.

На підставі внутрішньогрупової дисперсії для кожної групи можна визначити *загальну середню з внутрішньогрупових дисперсій*:

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 \cdot f_i}{\sum f_i}, \qquad \textbf{(7.1.13)}$$

де f_i – кількість одиниць у *і*-й групі.

Згідно з ***правилом складання дисперсій,*** загальна дисперсія дорівнює сумі середньої з внутрішньогрупової і з міжгрупової дисперсії:

$$\sigma^2 = \overline{\sigma_i^2} + \delta^2. \qquad \textbf{(7.1.14)}$$

У статистичному аналізі широко використовується *емпіричний коефіцієнт детермінації* (η^2) – показник, що відображає частку міжгрупової дисперсії у загальній дисперсії результативної ознаки і характеризує силу впливу групувальної ознаки на утворення загальної варіації:

$$\eta^2 = \frac{\delta^2}{\sigma^2}. \qquad \textbf{(7.1.15)}$$

За відсутності зв’язку між групувальною ознакою і загальною варіацією (тобто результативною ознакою) $\eta^2 = 0$, а за функціонального зв’язку $\eta^2 = 1$.

Емпіричне кореляційне відношення **(**η**)** – це корінь квадратний з емпіричного коефіцієнта детермінації (показник Пірсона).

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}. \qquad \textbf{(7.1.16)}$$

η набуває значень від 0 до 1.

Якщо зв'язок відсутній, то $\eta = 0$, тобто всі групові середні будуть рівними і міжгрупової варіації не буде.

Якщо зв'язок є функціональним, то кореляційне відношення дорівнюватиме одиниці. Тоді $\sigma^2 = \delta^2$, тобто внутрішньогрупової варіації не буде.

Чим ближче до одиниці значення кореляційного відношення, тим щільнішим, ближчим до функціонального буде зв'язок між ознаками.

Таблиця 7.1.4 – Сила зв'язку залежно від значення емпіричного кореляційного співвідношення η (співвідношенняя Чеддока)

Значення η	Сила зв'язку
0,1–0,3	слабка
0,3–0,5	помірна
0,5–0,7	помітна
0,7–0,9	щільна
0,9–0,99	надто щільна

Приклад 7.1.4: за даними прикладу 4.1.1 (таблиця 4.1.7) розрахувати міжгрупову, внутрішньогрупову і загальну дисперсію; коефіцієнти варіації, детермінації та кореляційне співвідношення Пірсона. Зробити висновки.

Розв'язок

Міжгрупову дисперсію δ^2 знайдемо табличним способом (див. табл. 7.1.5), скориставшись формулою 7.1.11.

Таблиця 7.1.5 – Розрахункові значення для обчислення міжгрупової дисперсії

Групи	X_i	X	f	$\overline{X}_i$	$\overline{X}_i - \overline{X}$	$(\overline{X}_i - \overline{X})^2$	$(\overline{X}_i - \overline{X})^2 f$
A	*1*	*2*	*3*	*4*	*5*	*6*	*7*
154,00-294,83	154, 201, 214	224,415	3	189,67	-419,73	176173,3	528519,9
294,83-435,66	302, 320, 320, 324, 325, 329, 357	365,245	7	325,29	-284,11	80718,5	565029,5
435,66-576,49	478, 517, 548, 558	506,075	4	525,25	-84,15	7081,2	28324,8
576,49-717,32	605, 615, 658, 658, 698	646,905	5	646,80	37,40	1398,8	6994,0
717,32-858,15	745, 754, 754	787,735	3	751,00	141,60	20050,6	60151,8
858,15-999,00	865, 896, 875, 878, 950, 968, 985, 999,	928,575	8	927,00	317,60	100869,8	806958,4
Усього	Х	Х	30	Х	Х	Х	1995978,4

У табл. 7.1.5 значення присудка такі:

X_i – значення заробітної плати робітників, які потрапили до *i*-ї групи;

X – середина інтервалу;

f – кількість робітників у групі;

$\overline{X}_i$ – середнє значення заробітної плати в *i*-й групі, розраховане за формулою 6.1.3;

$\overline{X}$ – середня величина заробітної плати для всієї сукупності, розрахована у прикладі 6.1.4:

$$\overline{X} = \frac{18280{,}590}{30} = 609{,}4 \text{ (грош.од.)}.$$

$$\delta^2 = \frac{1995978{,}4}{30} = 66532{,}6 \text{ (див. формулу 7.1.11)}.$$

Для розрахунку внутрішньогрупових дисперсій за групами скористаємося формулою 7.1.12:

$$\sigma_1^{\,2} = \frac{(154-189{,}67)^2\times 1 + (201-189{,}67)^2\times 1 + (214-189{,}67)^2\times 1}{3} = 664{,}2,$$

$$\sigma_2^{\,2} = \frac{(302-325{,}29)^2\times 1 + (320-325{,}29)^2\times 2 + \ldots + (357-325{,}29)^2\times 1}{7} = 231{,}3,$$

аналогічно розраховуємо:

$$\sigma_3^{\,2} = 972{,}7,$$

$$\sigma_4^{\,2} = 1126{,}2,$$

$$\sigma_5^{\,2} = 18{,}0,$$

$$\sigma_6^{\,2} = 2583{,}5.$$

Обчислимо середню з внутрішньогрупових дисперсій, використовуючи формулу 7.1.13:

$$\overline{\sigma_i^{\,2}} = \frac{664{,}2\times 3 + 231{,}3\times 7 + 972{,}7\times 4 + 1126{,}2\times 5 + 18\times 3 + 2583{,}5\times 8}{30} = 1128{,}6.$$

За правилом складання дисперсій (див. формулу 7.1.14):

$$\sigma^2 = 66532{,}6 + 1128{,}6 = 67661{,}2.$$

Використовуючи формули 7.1.9 і 7.1.10, знайдемо коефіцієнт варіації:

$$V = \frac{\sqrt{67661{,}2}}{609{,}4}\cdot 100\,\% = 42{,}7\,\%.$$

Частку варіації результативної ознаки під впливом факторної ознаки обчислимо за допомогою емпіричного коефіцієнта детермінації (див. формулу 7.1.15):

$$\eta^2 = \frac{66532{,}6}{67661{,}2} = 0{,}983 \text{ (або 98,3 \%).}$$

Щільність зв’язку між групувальною і результативною ознаками (показник Пірсона) (див. формулу 7.1.16):

$$\eta = \sqrt{\frac{66532{,}6}{67661{,}2}} = 0{,}99 .$$

Висновки:

1) коефіцієнт варіації більше 33 % (42,7 %), тому сукупність є кількісно неоднорідною, а величина середньої заробітної плати на одного робітника є нетиповою;
2) загальна дисперсія, що відображає сумарний вплив усіх можливих факторів (стаж роботи, кваліфікація, характер роботи, хвороби і т.д.) на загальну варіацію середньої зарплати всіх робітників бригади, дорівнює 67661,2;
3) міжгрупова дисперсія характеризує варіацію групових середніх, обумовлену відмінностями між групами робітників щодо кваліфікації, і дорівнює 66532,6;
4) внутрішньогрупові дисперсії показують варіації заробітної плати у кожній групі, обумовлені всіма можливими факторами, крім відмінностей у кваліфікації. Середня з внутрішньогрупових дисперсій дорівнює 1128,6;
5) емпіричний коефіцієнт детермінації дорівнює 98,3 %. Це означає, що на 98,3 % варіація у сукупності обумовлена відмінностями у кваліфікації (а, отже, і у заробітній платі) робітників і на 1,7 % – впливом решти факторів;
6) емпіричне кореляційне відношення дорівнює 0,99. Це означає, що сила зв’язку між кваліфікацією (групувальною ознакою) і заробітною платою (результативною ознакою) є досить щільною.

7.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Що таке варіація ознаки, від чого залежить її величина?
2. Що таке розмах варіації, за якою формулою він обчислюється?
3. Що таке середнє лінійне відхилення, у чому його недоліки як показника варіації?
4. Який показник називається дисперсією? Його різновиди.

5. Що таке середнє квадратичне відхилення?
6. У чому полягає сутність спрощеного розрахунку дисперсії?
7. Чому дисперсія і середнє квадратичне відхилення не завжди є достатніми для характеристики варіації ознаки у досліджуваних сукупностях?
8. Коефіцієнт варіації як показник, формула його обчислення і значення для економічного аналізу.
9. Що характеризує міжгрупова дисперсія і яка її формула?
10. Що таке внутрішньогрупові дисперсії, середня з внутрішньогрупових дисперсій і за якими формулами вони розраховуються?
11. Що таке правило складання дисперсій?
12. Що називається емпіричним коефіцієнтом детермінації, у чому його смисл?
13. Що називається емпіричним кореляційним відношенням, у чому його смисл?
14. За даними обстеження витрат часу робітників на оброблення однієї деталі отримано такий розподіл:

Витрати часу на одну деталь, хв.	Кількість робітників у % до підсумку
До 20	10
20–22	15
22–24	25
24–26	35
26–28	10
28–30	5

Обчисліть:
1) середні витрати часу на одну деталь;
2) середнє лінійне відхилення;
3) дисперсію;
4) середнє квадратичне відхилення;
5) коефіцієнт варіації.

15. Є такі дані про розподіл робітників за виробітком виробів за зміну:

Кількість виробів, шт.	Кількість робітників, чол.
До 60	10
60–70	20
70–80	20
80–90	15
90–100	5
Усього :	

Обчисліть:
1) застосовуючи спосіб моментів:
 а) середній виробіток за зміну одним робітником;
 б) дисперсію і середнє квадратичне відхилення;
2) коефіцієнт варіації.
Зробіть висновки.

16. Є такі дані про розподіл виробів за вагою:

Вага виробу, г.	Кількість виробів, шт.
До 200	4
200–205	10
205–210	60
210–215	20
Вище 215	6

Обчисліть:
1) застосовуючи спосіб моментів:
 а) середню вагу виробу;
 б) дисперсію і середнє квадратичне відхилення;
2) коефіцієнт варіації.
Зробіть висновки.

17. Є такі дані про розподіл заводів за вартістю готової продукції:

№	Групи заводів за вартістю готової продукції, млн. грн.	Кількість заводів
1	До 2	10
2	2–3	20
3	3–4	30
4	4–5	25
5	5–6	10
6	Вище 6	5
	Усього	

На підставі наведених даних обчисліть:
1) середню вартість готової продукції на один завод;
2) дисперсію і середнє квадратичне відхилення;
3) коефіцієнт варіації.
Зробіть висновки.

18 Є такі дані про виконання норм виробітку робітниками заводу:

Виконання норм, %	Кількість робітників, чол.
До 100	4
100–102	10
102–104	60
104–106	20
Вище 106	6

На підставі наведених даних обчисліть:
1) застосовуючи спосіб моментів:
  а) середній процент виконання норм виробітку;
  б) дисперсію і середнє квадратичне відхилення;
2) коефіцієнт варіації.
Зробіть висновки.

19. Є такі дані про годинну продуктивність праці робітників цеху:

Групи робітників за кількістю продукції, виробленої за годину одним робітником (шт.)	Кількість робітників	Середній виробіток на одного робітника (шт.)	Групова дисперсія
9–10	10	9,5	0,25
10–12	11	11,6	0,23
12–14	16	13,4	0,23
14–17	13	16,4	0,53
Усього	50	13,4	

Обчисліть загальну дисперсію годинної продуктивності праці робітників, застосовуючи правило складання дисперсій.

20. Є дані про розподіл робітників за загальним стажем роботи:

Групи робітників за стажем роботи (років)	Кількість робітників			
	Цех №1	Цех №2	Цех №3	Усього
0–5	30	70	120	220
5–10	50	50	180	280
10–15	80	40	110	230
15–20	100	20	30	150
20–25	30	10	40	80
25–30	10	10	20	40
Усього	300	200	500	1000

Обчисліть:
1) внутрішньогрупові дисперсії (для кожного цеху);

2) середню з внутрішньогрупових дисперсій;
3) міжгрупову дисперсію, загальну дисперсію.

ГЛАВА 8

ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

8.1
МЕТОДИЧНІ ВКАЗІВКИ

Ця тема є однією з центральних у курсі загальної теорії статистики. Це обумовлено, насамперед, взаємозв'язком між цією й іншими темами, особливо зі статистичним спостереженням, статистичними показниками, таблицями, графіками тощо. Вибіркове спостереження тісно пов'язане з курсами математичної статистики і теорії ймовірностей. Тому оволодіння теоретичним матеріалом, вміння правильно розв'язувати відповідні практичні задачі та інтерпретувати отримані результати є необхідною умовою успішного вивчення курсу загальної теорії статистики і пов'язаних з нею наук.

Формування набору завдань цієї глави обумовлено практичними питаннями, що потребують вирішення при організації вибіркового спостереження і аналізі його результатів. Такими питаннями є: сутність вибіркового спостереження, його достоїнства і недоліки; види випадкової вибірки і способи відбору; визначення величини помилки випадкової вибірки; побудова довірчих інтервалів для середньої і для частки; визначення необхідної чисельності вибірки; статистична гіпотеза.

Сутність вибіркового спостереження. Його достоїнства і недоліки.

Вибіркове спостереження – це таке спостереження, при якому характеристика всієї сукупності одиниць дається за деякою їхньою частиною, відібраною у випадковому порядку.

Уся сукупність одиниць, з якої здійснюється відбір, називається *генеральною сукупністю* (з кількістю одиниць N). Відібрана частина одиниць сукупності, тобто частина, яка потрапила до вибірки, називається *вибірковою сукупністю* (з кількістю одиниць n).

Частка одиниць генеральної сукупності, яка має альтернативну варіативну ознаку, називається *генеральною часткою* (P):

$$P = \frac{M}{N}, \qquad \textbf{(8.1.1)}$$

де M – кількість одиниць генеральної сукупності, які мають альтернативну ознаку.

Частка одиниць вибіркової сукупності, яка має задану альтернативну ознаку, називається *вибірковою часткою* (w):

$$w = \frac{m}{n}, \qquad \textbf{(8.1.2)}$$

де m – кількість одиниць вибіркової сукупності, які мають альтернативну ознаку.

Достоїнства вибіркового спостереження

1. Економія праці.
2. Оперативність.
3. Більш докладна програми спостереження
4. Менша кількість помилок реєстрації.

Вибіркове спостереження організується так само, як і суцільне (див. главу 2). Крім того, при вибірковому спостереженні вирішуються такі питання:

1) визначається частина сукупності, яка підлягає вибірковому спостереженню;
2) встановлюється спосіб проведення відбору частини сукупності;
3) виконується безпосередні відбір;
4) визначається, у який спосіб на основі результатів вибіркового спостереження отримати необхідні характеристики всієї сукупності.

Найважливішою умовою проведення вибіркового спостереження є правильний відбір одиниць сукупності, тобто:

- суто об'єктивний відбір, при якому кожна одиниця сукупності отримує рівну можливість потрапляння до вибірки;
- кількість відібраних одиниць сукупності має бути достатнім.

Недоліки вибіркового спостереження.

- можливість помилок реєстрації (притаманних будь-якому статистичному спостереженню);
- можливість помилок репрезентативності або представництва (притаманних лише несуцільному спостереженню).

Види відбору одиниць з генеральної сукупності

1. Індивідуальний – коли за один прийом відбирається одна одиниця.
2. Груповий – коли за один прийом відбираються декілька одиниць, тобто партія або серія.
3. Комбінований – передбачає поєднання двох попередніх видів відбору.

Схеми відбору

1. Безповторний – при якому кожна відібрана одиниця не повертається до генеральної сукупності, тобто не може двічі потрапити до вибірки.

2. Повторний – при якому одна і та ж сама одиниця сукупності може потрапити до вибірки декілька разів.

Способи відбору одиниць (найбільш поширені)

1. *Власно-випадковий* – коли спостереженням охоплюється частина сукупності, відібрана у випадковому порядку. При цьому для кожної одиниці заготовлюють жетон або білет з порядковим номером, а потім у випадковому порядку відбирають необхідну кількість жетонів. Може бути повторним і безповторним.
2. *Механічний* – коли вся генеральна сукупність розбивається на рівні за обсягом частини або групи за випадковою ознакою, що є нейтральною до досліджуваної. При цьому розмір інтервалу у генеральній сукупності дорівнює оберненій величині відносного обсягу вибірки (наприклад, обсяг вибірки дорівнює 2 %. Тоді відбирають і перевіряють кожну 50-ту одиницю, тобто $i = \dfrac{1}{0{,}02} = 50$). Для досить великої сукупності механічний відбір за точністю результатів є близьким до випадкового. Тому для визначення середньої помилки механічної вибірки використовують такі ж самі формули, як при власно-випадковому відборі. Механічний відбір може бути тільки безповторним.
3. *Типовий* – застосовується тоді, коли сукупність є неоднорідною за досліджуваною ознакою. Спочатку виконують групування досліджуваної сукупності на однорідні типові групи за суттєвою ознакою, від якої залежить досліджуваний показник. Потім з кожної групи власно-випадковим або механічним способом відбирають кількість одиниць, пропорційну питомій вазі кожної групи генеральної сукупності.

Приклад 8.1.1. Здійснити 15-процентний типовий відбір для робітників заводу, які працюють у чотирьох цехах (табл. 8.1.1, графи А и 1).

Таблиця 8.1.1– Розрахункова

№ Цеху	Кількість робітників	Частка робітників у % до підсумку	Робітники, відібрані до вибіркової сукупності
А	*1*	*2*	*3*
1	180	15 %	180×0,15=27
2	240	20 %	180×0,20=36
3	420	30 %	180×0,30=54
4	460	35 %	180×0,35=63
Усього	1200	100 %	180

За результатами табл. 8.1.1, n=180 чоловік – це 15 % від 1200. За такого відбору до вибірки потрапляють представники всіх типових груп у такій самій пропорції, як і в генеральній сукупності, що робить вибірку репрезентативною (представницькою).

4. *Серійний* – коли відбору підлягають не окремі сукупності, а їхні серії, групи або гнізда, відібрані власно-випадковим або механічним способом. Може бути безповторним і повторним.

За ступенем охоплення одиниць сукупності розрізняють *великі* і *малі* (менше 5 %) вибірки.

Визначення розміру помилки випадкової вибірки. Зведені показники для всієї сукупності ніколи не дорівнюватимуть показникам, розрахованим за даними вибіркового спостереження, через помилки спостереження, які бувають двох видів:

1. Помилки реєстрації (притаманні будь-якому статистичному дослідженню) (див. главу 2).
2. Помилки репрезентативності (притаманні лише вибірковому спостереженню):
   - випадкові – виникають через те, що вибіркова сукупності недостатньо відтворює генеральну сукупність через несуцільний характер спостереження;
   - систематичні – виникають внаслідок порушення принципу випадкового відбору.

При дотриманні принципу випадковості помилка вибірки залежить від чисельності вибірки, а також від ступеня варіації досліджуваної ознаки.

Середня помилка вибірки для середньої при власно-випадковому методі відбору ($\mu_{\tilde{x}}$)*:*

а) при повторному відборі:

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma_{\tilde{x}}^{2}}{n}}\,; \qquad \textbf{(8.1.3)}$$

б) при безповторному відборі:

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma_{\tilde{x}}^{2}}{n} \cdot \left(1 - \frac{n}{N}\right)}\,, \qquad \textbf{(8.1.4)}$$

де n – чисельність вибіркової сукупності;
N – чисельність генеральної сукупності;

$\frac{n}{N}$ – процент вибірки (наприклад, при 20-процентному відборі $\frac{n}{N}$=0,2);

σ^2 – вибіркова дисперсія.

Середня помилка вибірки для частки одиниць сукупності при власно-випадковому методі відбору (μ_p):

а) при повторному відборі:

$$\mu_p = \sqrt{\frac{\sigma_p{}^2}{n}} = \sqrt{\frac{w(1-w)}{n}}, \qquad \textbf{(8.1.5)}$$

де $\sigma_p^2 = w(1-w)$ – дисперсія заданої альтернативної ознаки вибіркової сукупності;

$w = \frac{m}{n}$ – частка одиниць, що мають задану альтернативну ознаку у вибірковому спостереженні;

б) при безповторному відборі:

$$\mu_p = \sqrt{\frac{\sigma_p{}^2}{n} \cdot \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{w(1-w)}{n}(1 - \frac{n}{N})}. \qquad \textbf{(8.1.6)}$$

У разі малої вибірки, тобто при $\frac{n}{N}$<0,05 (5%), множник $(1-\frac{\boldsymbol{n}}{\boldsymbol{N}}) \to 1$, тому на нього не зважають і для розрахунку середньої помилки використовують формули повторного відбору.

Вибіркові середні і частки поширюються на генеральну сукупність з урахуванням межі їхньої можливої помилки. У кожній конкретній вибірці розходження між вибірковою середньою $\tilde{X}$ (вибірковою часткою w) і генеральною середньою $\overline{X}$ (генеральною часткою P) може бути менше середньої помилки вибірки $\mu_{\tilde{x}}$ (μ_p), дорівнювати їй або перевищувати її. Причому кожне з цих розходжень має різну ймовірність (об'єктивну можливість виникнення події) [8]. Тому фактичне розходження між вибірковою середньою (часткою) і генеральною середньою (часткою) можна розглядати як деяку *граничну помилку*, пов’язану з середньою помилкою і гарантовану з певною ймовірністю Р.

Гранична помилка вибірки для середньої при власно-випадковому методі відбору **($\Delta_{\tilde{x}}$):**

$$\Delta_{\tilde{x}} = t\mu_{\tilde{x}}, \qquad \textbf{(8.1.7)}$$

де t – коефіцієнт довіри, який залежить від ймовірності, з якою гарантується гранична помилка вибірки. Згідно з теоремою Чебишева-Ляпунова (див. додаток 1):

якщо Р=0,683 $\to$ t=1;

P=0,954 $\rightarrow$ t=2;
P=0,997 $\rightarrow$ t=3.

Тобто, інакше кажучи, за ймовірності P=0,683 можна казати, що у 68,3 % випадків помилка репрезентативності не вийде за межі ($\pm\mu_{\tilde{x}}$).

Гранична помилка вибірки для частки за власно-випадкового методу відбору (Δ_p)**:**

$$\Delta_p = t\mu_p \,. \qquad \textbf{(8.1.8)}$$

Гранична помилка вибірки дозволяє визначити граничні значення характеристик генеральної сукупності та їхні *довірчі інтервали* (межі, у яких з заданим ступенем імовірності знаходитиметься невідома величина).

Алгоритм розрахунку довірчого інтервалу

- Для середньої:

1) $\tilde{X} = \frac{\sum Xf}{\sum f}$ – вибіркова середня (можливий розрахунок способом моментів за формулою 6.1.6);

2) $\sigma_{\tilde{x}}^{\ 2} = \frac{\sum (X-\tilde{X})^2 \times f}{\sum f}$ – дисперсія вибіркової середньої (можливий розрахунок способом моментів за формулою 7.1.7);

3) $\mu_{\tilde{x}} = \sqrt{\frac{\sigma_{\tilde{x}}^{\ 2}}{n} \times \left(1-\frac{n}{N}\right)}$ – якщо відбір є безповторним (можливий розрахунок для повторного відбору за формулою 8.1.3);

4) $\Delta_{\tilde{x}} = t\mu_{\tilde{x}}$ – гранична помилка вибіркової середньої;

5) $\tilde{X} - \Delta_{\tilde{x}} \le \overline{X} \le \tilde{X} + \Delta_{\tilde{x}}$ – довірчий інтервал; **(8.1.9)**

Або $\overline{X} = \tilde{X} \pm \Delta_{\tilde{x}}$ – граничне значення.

- Для частки:

1) $w = \frac{m}{n}$ – частка одиниць, що мають задану однаку у вибірковому спостереженні;

2) $\mu_p = \sqrt{\frac{w(1-w)}{n}(1-\frac{n}{N})}$ – якщо відбір є безповторним (можливий розрахунок для повторного відбору за формулою 8.1.5);

3) $\Delta_p = t\mu_p$;

4) $w - \Delta_p \le p \le w + \Delta_p$ – довірчий інтервал для частки; **(8.1.10)**

або $p = w \pm \Delta_p$ – граничне значення частки.

Помилки і межі генеральних характеристик за інших способів формування вибіркової сукупності визначаються на основі відповідних формул, які відображають особливості цих видів вибірки. Наприклад: у разі механічної вибірки застосовуються ті ж самі формули, що й для власно-випадкової; у а разі типової вибірки показником варіації є середня з внутрішньогрупових дисперсій $\overline{\sigma_i^2}$; при серійній вибірці – міжгрупова (міжсерійна) дисперсія δ^2 (див. главу 7). Крім того, в останньому випадку замість обсягу вибіркової сукупності *n* використовується показник кількості серій *r*.

Приклад 8.1.2. За даними прикладу 7.1.3 (табл. 7.1.2) за умови, що проведено п'ятипроцентне вибіркове обстеження 100 га посівів, відібраних випадковим способом (вибірка є безповторною), обчислити:
1) з імовірністю 0,997 граничну помилку вибіркової середньої та можливі межі, в яких очікується середня врожайність соняшника в області;
2) з імовірністю 0,954 граничну помилку вибіркової частки і межі питомої ваги посівних площ області з врожайністю від 15 до 19 центнерів з 1 га.

Розв'язок

Скористаємося даними табл. 7.1.3. і подальшими викладками прикладу 7.1.3 для подальших розрахунків.

Отже, способом моментів було розраховано:

$\overline{X} = 15{,}70$ (ц з 1 га),

$\sigma^2 = 4{,}11$.

Розрахунок середньої помилки вибірки для середньої величини за випадкового безповторного відбору представлено у формулі 8.1.4:

$$\mu_{\tilde{x}} = \sqrt{\frac{4{,}11}{100} \cdot (1 - 0{,}05)} = 0{,}198.$$

Граничну помилку вибірки для середньої за випадкового відбору знаходимо за допомогою формули 8.1.7. Оскільки ймовірність Р складає 0.997, коефіцієнт довіри t дорівнює 3.

Тоді:

$$\boldsymbol{\Delta}_{\tilde{x}} = 3 \times 0{,}198 = 0{,}594.$$

Таким чином, довірчі інтервали для середньої генеральної сукупності мають такі значення і визначаються за формулою 8.1.9:

$$15{,}70 - 0{,}594 \le \overline{X} \le 15{,}70 + 0{,}594 \text{ (ц с 1 га)};$$

$$15{,}106 \le \overline{X} \le 16{,}294 \text{ (ц с 1 га)}.$$

Висновок: з ймовірністю Р=0,997 можна гарантувати, що середня врожайність соняшника в області знаходиться у межах від 15,106 до 16,294 ц з 1 га.

Для визначення граничної помилки вибіркової частки і меж питомої ваги посівних площ області з врожайністю від 15 до 19 ц з 1 га необхідно визначити частку варіантів, які мають цю ознаку w.

$$w = \frac{40+20}{100} = 0{,}60.$$

Дисперсія вибіркової частки, розрахована за даними вибіркового спостереження (див. чисельник формули 8.1.5):

$$\sigma_p^2 = 0{,}60 \times (1-0{,}60) = 0{,}24.$$

Середня помилка вибірки для частки одиниць сукупності за безповторного відбору визначається за формулою 8.1.6:

$$\mu_p = \sqrt{\frac{0{,}24}{100}\cdot\left(1-0{,}05\right)} = 0{,}048.$$

Гранична помилка вибірки для частки одиниць сукупності за безповторного відбору розраховується за допомогою формули 8.1.8. Оскільки ймовірність Р=0,954, коефіцієнт довіри t дорівнює 2.

Тоді:

$$\Delta p = 2 \times 0{,}048 = 0{,}096.$$

Таким чином, довірчі інтервали питомої ваги посівних площ області з врожайністю від 15 до 19 ц з 1 га мають такі значення (див. формулу 8.1.10):

$$0{,}60 - 0{,}096 \le p \le 0{,}60 + 0{,}096;$$

$$0{,}504 \le p \le 0{,}696;$$

$$50{,}4\% \le p \le 69{,}6\ \%.$$

Висновок: з ймовірністю, рівною 0,954, можна гарантувати, що питома вага посівних площ області з врожайністю від 15 до 19 ц з 1 га знаходиться у межах від 50,4 % до 69,6 %.

Визначення необхідної чисельності вибірки. До початку вибіркового спостереження встановлюється необхідна чисельність вибірки *n*, яка з певною ймовірністю забезпечить задану точність результатів спостереження. Формули для визначення необхідної чисельності вибірки *n* легко можна отримати з формул граничних помилок вибірки, попередньо піднісши у квадрат обидві частини рівняння.

Отже:

- чисельність вибірки для середньої за повторного відбору:

оскільки $\Delta_{\tilde{x}} = t\sqrt{\frac{\sigma_{\tilde{x}}^2}{n}}$, то $\Delta^2_{\tilde{x}} = t^2\,\frac{\sigma_{\tilde{x}}^2}{n} \quad \Rightarrow$

$$n = \frac{t^2 \sigma^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}}; \qquad \textbf{(8.1.11)}$$

- чисельність вибірки для частки за повторного відбору:

  оскільки $\Delta_p = t\sqrt{\frac{w(1-w)}{n}}$, то, піднісши обидві частини у квадрат, виразимо n:

$$n = \frac{t^2 w(1-w)}{\Delta^2{}_p}. \qquad \textbf{(8.1.12)}$$

Аналогічно:

- чисельність вибірки для середньої за безповторного відбору:

$$n = \frac{t^2 \sigma^2_{\tilde{x}} N}{N\Delta^2{}_{\tilde{x}} + t^2 \sigma^2_{\tilde{x}}}; \qquad \textbf{(8.1.13)}$$

- чисельність вибірки для частки за безповторного відбору:

$$n = \frac{t^2 w(1-w)N}{N\Delta^2{}_p + t^2 w(1-w)}. \qquad \textbf{(8.1.14)}$$

Статистична гіпотеза – це певне припущення щодо властивостей генеральної сукупності, яке можна перевірити за даними вибіркового спостереження.

Гіпотеза, яку необхідно перевірити, формулюється як відсутність відмінностей між параметром генеральної сукупності (G) і заданою величиною (a) (це нульова гіпотеза). Її зміст записується таким чином:

H_o: $G=a$.

Кожній нульовій гіпотезі обов’язково протиставляють альтернативну гіпотезу на основі аналізу можливих помилкових рішень та їхніх наслідків:

H_a: $G>a$.

H_a: $G<a$.

H_a: $G\neq a$.

Якщо вибіркові дані суперечать H_o, її відхиляють, а якщо вибіркові дані узгоджуються з нею, то її не відхиляють.

Правило, за яким гіпотезу H_o відхиляють (тобто приймають альтернативну гіпотезу H_a) або не відхиляють, називається *статистичним критерієм.*

Для побудови статистичного критерію, що дозволяє перевірити деяку гіпотезу, необхідно

1. Сформулювати гіпотезу H_o, яка перевіряється, а також, водночас, альтернативну гіпотезу H_a .

2. Обрати рівень значимості α, який контролює допустиму ймовірність помилки 1-го роду (*рівень значимості* – це таке мале значення імовірності потрапляння критерію до критичної області за умови справедливості гіпотези, що виникнення цієї події можна розцінювати як наслідок суттєвого розходження між висунутою гіпотезою і результатами вибірки. Зазвичай α припускається рівним 0,05 або 0,01 [16].
3. Визначити область допустимих значень і критичну область (*область допустимих значень* – область, при потраплянні значення критерію до якої H_o не відхиляють; *критична область* – область, при потраплянні значення критерію до якої H_o відхиляють), а також критичне значення статистичного критерію за відповідною таблицею.
4. Розрахувати фактичне значення статистичного критерію.
5. Винести те чи інше рішення (тобто або відхилити гіпотезу, або не відхиляти її) на основі порівняння фактичного і критичного значення критерію.

При перевірці гіпотези за одним з критеріїв (або відхиляється, або не відхиляється) можливі *два помилкових рішення:*

1) неправильне відхилення нульової гіпотези (помилка або ризик 1-го роду);
2) неправильне прийняття нульової гіпотези (помилка або ризик 2-го роду).

Приклад 8.1.3. Нехай H_o буде такою: нове добриво не відрізняється від старого.
Тоді можливі висновки при перевірці гіпотез будуть такими (табл. 8.1.2):

Таблиця 8.1.2 – Ситуаційна

Рішення за критерієм	Фактично	
	H_o є вірною	H_o є невірною
H_o відхиляється	Помилка 1-го роду	Правильне рішення
H_o не відхиляється	Правильне рішення	Помилка 2-го роду

Приклад 8.1.4 (*перевірка гіпотези про величину середньої арифметичної*). Відомо, що середні витрати сировини на одиницю продукції складають при існуючому методі виробництва 2,8 умовних одиниць. Після внесення змін до існуючої технології виготовлення продукції, за результатами перевірки достатньо великої партії виробів середні витрати сировини на одиницю продукції склали 2,6 умовних одиниць. Середня помилка вибірки виявилася рівною 0,1. За рівня значимості α=0,05 визначити, чи дійсно застосування нового методу оброблення приводить до зниження матеріалоємності продукції.

Розв’язок

І. Варіант

1. Формулюємо нульову гіпотезу H_o.

H_o : між новим ($\overline{X_н}$) та існуючим ($\overline{X_с}$) методами немає суттєвих відмінностей з погляду на їхній вплив на матеріалоємність, тобто: $\overline{X_c} = \overline{X_н}$.

Формулюємо альтернативну гіпотезу H_a.

H_a : застосування нового методу оброблення приводить до зміни витрат сировини на одиницю продукції, тобто: $\overline{X_c} \neq \overline{X_н}$.

2. Обираємо рівень значимості α, який контролює допустиму ймовірність помилки (його дано за умови α=0,05).

3. Тоді ймовірність P потрапляння значення статистичного критерію в цю область:

$$P(\left|\overline{X_н} - \overline{X_c}\right| \geq t\mu_{\tilde{x}})=0{,}05,$$

де $\Delta_{\tilde{x}} = t\mu_{\tilde{x}}$ – гранична помилка вибіркової середньої (див. формулу 8.1.7);

$\left|\overline{X_н} - \overline{X_c}\right| > t\mu_{\tilde{x}}$ – критична область (потрапляння значення статистичного критерію в яку приводить до відхилення H_o).

4. За таблицями інтегральної функції Лапласа (які зазвичай знаходяться у додатках підручників і довідників зі статистики), де наводиться і нормована функція Лапласа:

$$\Phi(t)=(P\left|\tilde{X} - \overline{X}\right| \leq t\mu),$$

Визначаємо коефіцієнт довіри: за $\Phi(t)$=0,95 (де 0,95=1–0,05), t=1,96.

Таким чином, величина граничного розходження двох середніх з ймовірністю P=0,95 не повинна перевищувати $t\mu_{\tilde{x}}$=1,96$\times$0,1=0,196.

Тобто 2,8–0,196$\leq \overline{X_н} \leq$2,8+0,196;

або 2,604$\leq \overline{X_н} \leq$2,996 – фактичне значення статистичного критерію.

5. $\overline{X_н}$=2,6 потрапляє до критичної області $\Rightarrow$ H_o відхиляється.

Таким чином, у методах, які приводять до змін у матеріалоємності, існують відмінності.

ІІ. Варіант

1. H_o : $\overline{X_c} = \overline{X_н}$;

H_a : $\overline{X_н} < \overline{X_c}$, тобто застосування нового метода оброблення приводить до зниження витрат сировини на одиницю продукції.

2. Тоді критична область за рівня значимості α=0,05:

$\overline{X_н} < \overline{X_c} - t\mu_{\tilde{x}}$ – це область великих від’ємних відхилень (критична область).

Тобто $P(\overline{X_н} < \overline{X_с} - t\mu_{\bar{x}})=0{,}05$,

t=1,64; μ =0,1.

$\overline{X_н}$ =2,8–1,64×0,1=2,636

Нульова гіпотеза не заперечуватиметься, якщо середні витрати матеріалу на одиницю продукції будуть більше величини 2,636. Оскільки за новою технологією витрати сировини складають 2,6 умовних одиниць, то з ймовірністю Р=0,995 можна вважати, що нульова гіпотеза H_o має бути відхилена. Отже, застосування нової технології приводить до зниження витрат сировини на виготовлення продукції.

8.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Сутність вибіркового спостереження, його достоїнства і недоліки.
2. Випадковий відбір, причини необхідності його дотримання при організації вибіркового спостереження.
3. Середня помилка вибірки.
4. Сутність механічного способу відбору.
5. Типовий відбір.
6. Серійна помилка.
7. Запишіть довірчі інтервали для генеральної середньої з ймовірністю 0,954 і 0,997.
8. Як визначається гранична помилка при формуванні великої і малої вибірок?
9. Визначення необхідної чисельності вибірки.
10. Що таке статистична гіпотеза?
11. Дайте поняття критичної області при статистичній перевірці гіпотез.
12. З метою вивчення витрат часу на виготовлення деталі робітниками заводу проведено 10-процентру випадкову безповторну вибірку, в результаті якої отримано такий розподіл деталей за витратами часу:

Витрати часу на одну деталь, хв.	Кількість деталей, шт.
До 10	10
До 12	20
12–14	50
14–16	15
16 і більше	5
Усього	

На підставі цих даних обчисліть:
1) з ймовірністю 0,954 граничну помилку вибіркової середньої і можливі

межі, у яких очікуються середні витрати часу на виготовлення однієї деталі на заводі;
2) з ймовірністю 0,954 граничну помилку вибіркової частки і межі питомої ваги кількості деталей, витрати часу на виготовлення яких складають від 10 до 15 хвилин.

13. З метою контролю дотримання норм витрат сировини проведено вибіркове обстеження партії готової продукції. При механічному (безповторному) способі відбору 5 % виробів отримано такі дані про вагу обстежених одиниць:

Вага виробів, г.	Кількість зразків, шт.
До 100	22
100–110	76
110–120	215
120–130	69
130 і вище	18
Усього	

На підставі вибіркових даних обчисліть:
1) способом моментів:
а) середню вагу виробу;
б) середнє квадратичне відхилення;
2) коефіцієнт варіації;
3) з ймовірністю 0,997 можливі межі, у яких знаходиться середня вага виробів у всій партії;
4) з ймовірністю 0,954 можливі межі питомої ваги (частки) стандартної продукції в усій партії за умови, що до стандартної продукції належать усі вироби з вагою від 100 г до 130 г.

14. Для вивчення вікової структури робітників заводу станом на 1 липня було проведено 3-процентне вибіркове обстеження методом випадкового безповторного відбору. Результати обстеження показали такий розподіл робітників за віком:

Групи робітників за віком, років.	Кількість робітників, чол.
До 20	10
20–30	18
30–40	40
40–50	24
50 років і старше	8
Усього	

На підставі даних вибіркового обстеження обчисліть:

1) способом моментів:
а) середній вік робітника;
б) середнє квадратичне відхилення;
2) коефіцієнт варіації;
3) з ймовірністю 0,997 можливі межі середнього віку робітників заводу;
4) з ймовірністю 0,954 можливі межі частки робітників заводу, вік яких складає менше 20 років.

15. Методом випадкового (безповторного) відбору було опитано 5 % студентів про час, який вони витрачають на дорогу до інституту. В результаті отримано такий ряд розподілу:

Час, який витрачає студент на дорогу, хв.	Кількість студентів, чол.
До 20	20
20–30	8
30–40	30
50–60	10
Більше 60	5
Усього	

На підставі даних вибіркового обстеження обчисліть:
1) способом моментів:
а) середній час, який витрачається на дорогу до інституту;
б) середнє квадратичне відхилення;
2) з ймовірністю 0,997 можливі межі, в яких знаходиться середній час, що витрачається на дорогу всіма студентами інституту;
3) з ймовірністю 0,954 можливі межі частки студентів, які витрачають на дорогу до 30 хв.

16. З метою вивчення річного виробітку продукції робітниками підприємства громадського харчування проведено 1-процентну безповторну випадкову вибірку 100 працівників, у результаті якої отримано такий їх розподіл за виробітком продукції:

Групи працівників за річним виробітком продукції, тис. грн.	Кількість працівників
До 140	10
140–160	15
160–180	35
180–200	25
Вище 200	15
Усього	

На підставі цих даних обчисліть:
1) з ймовірністю 0,954 можливі межі, в яких очікується середній річний виробіток продукції працівниками підприємства;
2) з ймовірністю 0.997 граничну помилку вибіркової частки і межі питомої ваги працівників з річним виробітком продукції від 160 до 200 тис. грн.

17. При механічній 5-процентній вибірці було випробувано на розрив 1000 ниток з партії. Установлено, що середня міцність пряжи дорівнює 340 г при середньому квадратичному відхиленні 20. З ймовірністю 0,954 визначте межі, в яких знаходиться середня міцність пряжи в партії.

18. На заводі з кількістю робітників 10 тис. чоловік при механічній вибірці передбачається визначення частки робітників зі стажем роботи 20 років і більше. Якою повинна бути чисельність вибірки, аби з ймовірністю 0,954 помилка вибірки не перевищувала 0,05, якщо на основі попередніх обстежень відомо, що дисперсія дорівнює 0,2?

19. У місті N, де мешкає 15 тис. сімей, методом випадкового безповторного відбору передбачається визначити частку сімей з дітьми ясельного віку. Якою повинна бути чисельність вибірки, аби з ймовірністю 0,954 помилка вибірки не перевищувала 0,03, якщо на основі попередніх обстежень відомо, що дисперсія дорівнює 0,3?

ГЛАВА 9

СТАТИСТИЧНЕ ВИВЧЕННЯ ДИНАМІКИ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ (РЯДІВ ДИНАМІКИ)

9.1 МЕТОДИЧНІ ВКАЗІВКИ

Це найважливіша тема курсу загальної теорії статистики, оскільки завдання статистичного дослідження найчастіше полягає в аналізі розвитку тих чи інших явищ.

У цій главі розглядаються задачі, розв’язання яких дає можливість оволодіння правилами побудови і аналізу рядів динаміки для характеристики зміни соціально-економічних явищ у часі, виявлення основної тенденції та закономірностей їх розвитку. Це досягається відповідним обробленням рядів динаміки, аналізом зміни його рівнів, розрахунком аналітичних показників.

Сутність і види рядів динаміки. *Рядом динаміки* називають ряд розташованих у хронологічній послідовності числових значень статистичного показника, які характеризують зміну соціально-економічних явищ у часі.

Ряд динаміки включає два елементи:

- рівень ряду (y);
- момент часу (t).

Для правильної побудови важливо, щоб статистичні дані або показники були зіставними щодо одиниць вимірювання, кола охоплюваних об’єктів, цін, території, часу реєстрації, методології розрахунків.

Залежно від часу, що відображається у ряді динаміки, виділяють:

- інтервальні ряди динаміки – рівні яких характеризують величину досліджуваних показників або соціально-економічних явищ за певний період часу (наприклад, видобуток нафти у регіоні за рік, квартал, місяць);
- моментні ряди динаміки – рівні яких характеризують величину досліджуваних показників або стану явища на певні дати (моменти часу).

Відмінною рисою інтервальних рядів динаміки є можливість сумування їхніх рівнів. У результаті цього отримують так звані “накопичені підсумки”, які мають реальний зміст. Утім сума рівнів моментного ряду динаміки не має жодного реального змісту і “накопичені підсумки” для цих рядів не розраховуються.

Особливості розрахунку середніх рівнів ряду. *Середній рівень ряду* характеризує узагальнену величину його абсолютних рівнів.

Середній рівень для інтервальних і моментних рядів динаміки розраховується різними методами.

1. ***Для інтервальних рядів динаміки.***
   - З рівними інтервалами (за середньою арифметичною простою):

$$\bar{y} = \frac{\sum y}{n}, \qquad \textbf{(9.1.1)}$$

де n – кількість рядів динаміки.

Приклад 9.1.1. Розрахувати середньорічний випуск друкованих видань в Україні за даними табл. 9.1.1:

Таблиця 9.1.1 – Вихідні дані

Роки	2004	2005	2006	2007	2008
Друковані видання, тис. друк. од.	41,2	34,0	28,7	29,0	29,5

Розв'язок

$$\bar{y} = \frac{\sum y}{n} = \frac{41{,}2 + 34{,}0 + 28{,}7 + 29{,}0 + 29{,}5}{5} = \frac{162{,}4}{5} = 32{,}5 \text{ (тис. друк. од.) .}$$

- З нерівними інтервалами (за середньою арифметичною зваженою):

$$\bar{y} = \frac{\sum yt}{\sum t}, \qquad \textbf{(9.1.2)}$$

де t – ваги, тобто тривалість інтервалів часу між суміжними датами .

Приклад 9.1.2. Розрахувати середньооблікову кількість працівників комерційного банку за червень, якщо відомо, що з 1 по 15 число місяця в ньому працювали 20 осіб, з 16 по 25 червня – 27 осіб, з 26 по 30 червня – 30 осіб.

Розв'язок

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{20 \times 15 + 27 \times 10 + 30 \times 5}{30} = \frac{720}{30} = 24 \text{ (осіб).}$$

2. Для моментних рядів динаміки.

- З рівностоячими рівнями (за середньою хронологічною простою):

$$\bar{y} = \frac{\frac{1}{2} y_1 + y_2 + y_3 + \ldots + \frac{1}{2} y_n}{n-1} \; . \qquad \textbf{(9.1.3)}$$

Приклад 9.1.3. Є дані про залишки матеріалів на складі (табл. 9.1.2):

Таблиця 9.1.2 – Вихідні дані

Дата	T	01.01	01.02	01.03	01.04
Залишки (т)	y	48 (y_1)	40 (y_2)	36 (y_3)	52 (y_4)

Визначити середню величину залишків за кожний місяць і за квартал загалом.

Розв'язок

$$\bar{y}_{січ} = \frac{y_1 + y_2}{2} = \frac{48 + 40}{2} = 44 \text{(т)} .$$

$$\bar{y}_{лют.} = \frac{y_2 + y_3}{2} = \frac{40 + 36}{2} = 38 \text{(т)} .$$

$$\bar{y}_{берез.} = \frac{y_3 + y_4}{2} = \frac{36 + 52}{2} = 44 \text{(т)} .$$

$$\bar{y}_{кв} = \frac{\bar{y}_{січ.} + \bar{y}_{лют.} + \bar{y}_{берез.}}{3} = \frac{44 + 38 + 44}{3} = \frac{\frac{1}{2} y_1 + y_2 + y_3 + \ldots + \frac{1}{2} y_n}{n-1} = 42 \text{ (т) .}$$

- З нерівностоячими рівнями**:**

а) за середньою арифметичною зваженою (спрощена формула) (за наявності повних даних про всі зміни, що відбуваються):

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i}, \qquad \textbf{(9.1.4)}$$

де y_i – нові (змінені) рівні динамічного ряду;

t_i – тривалість проміжків часу, протягом яких нові рівні ряду не змінювалися.

Приклад 9.1.4. Розглянемо рух грошових коштів у гаманці студента за квітень:

на 1.04 – було у гаманці 30 грн.,

5.04 – надійшло 150 грн.,

8.04 – витрачено 100 грн.,

15.04 – отримав стипендію 530 грн.,

20.04 – купив харчі на 200 грн.

Визначити середній залишок грошових коштів у гаманці за квітень.

Розв’язок

Розв’яжемо задачу табличним способом (табл. 9.1.3):

Таблиця 9.1.3 – Розрахункова

Дата	Залишок грошових коштів, грн. (y_i)	Кількість днів, протягом яких залишок не змінювався (t_i)	$y_i \times t_i$
01.04	30	4	120
05.04	180	3	540
08.04	80	7	560
15.04	610	5	3050
20.04	410	11	4510
Усього	Х	30	8780

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i} = \frac{8780}{30} = 292{,}7\,(грн.)\,;$$

б) за середньою арифметичною зваженою (спрощена формула) (за відсутності повних даних про всі поточні зміни):

$$\bar{y} = \frac{\sum \bar{y}_i t_i}{\sum t_i}, \qquad \textbf{(9.1.5)}$$

де $\bar{y}_i$ – середні рівні динамічного ряду;

t_i – тривалість проміжків часу, протягом яких нові рівні ряду не змінювалися.

Приклад 9.1.5. Розглянемо зміну залишку матеріалу на складі протягом року (табл. 9.1.4):

Таблиця 9.1.4 – Вихідні дані

Дата	Залишок матеріалу, т
01.01	100 (y_1)
01.03	120 (y_2)
01.09	128 (y_3)
01.01 наступного року	110 (y_4)

Знайти середній залишок матеріалу на складі за рік.

<u>Розв'язок</u>

$$\bar{y}_{січень-лютий} = \frac{y_1 + y_2}{2} = \frac{100+120}{2} = 110 \text{ (т)}, \qquad t_1=2 \text{ місяці.}$$

$$\bar{y}_{березень-серпень} = \frac{y_2 + y_3}{2} = \frac{120+128}{2} = 124 \text{ (т)}, \qquad t_2=6 \text{ місяців.}$$

$$\bar{y}_{вересень-грудень} = \frac{y_3 + y_4}{2} = \frac{128+110}{2} = 119 \text{ (т)}, \qquad t_3=4 \text{ місяці.}$$

$$\bar{y} = \frac{\sum \bar{y}_i t_i}{\sum t_i} = \frac{110\times 2 + 124\times 6 + 119\times 4}{12} = 125 \text{ (т)}.$$

Аналітичні показники динаміки:

- абсолютні прирости, Δ;
- темпи зростання, T;
- темпи приросту, $\boldsymbol{T_{пр}}$;
- абсолютний значення одного процента приросту, A (приклад 9.1.6).

Якщо кожний наступний рівень ряду порівнюється з попереднім рівнем ряду, то отримують *ланцюгові показники.*

Якщо кожний наступний рівень ряду порівнюється з одним і тим самим початковим рівнем ряду, то отримують *базисні показники.*

Абсолютний приріст (зниження) Δ – це абсолютна зміна, що характеризує збільшення або зменшення рівня ряду за певний проміжок часу. Може бути додатною або від'ємною величиною, яка має розмірність рівня ряду.

$\Delta_Ц = y_i - y_{i-1}$ – ланцюговий абсолютний приріст, **(9.1.6)**

$\Delta_б = y_i - y_0$ – базисний абсолютний приріст. **(9.1.7)**

Наприклад, абсолютне збільшення обсягу виробництва за 2008 р. порівняно з 2007 р. склало 40 т, а порівняно з базисним 2004 р. – 121 т.

Або, наприклад, абсолютне зменшення обсягу виробництва у 2007 р. порівняно з 2006 г. склало 20 т (див. приклад 9.1.6).

Приклад 9.1.6. Потрібно провести аналіз динаміки обсягу виробництва одного з підприємств України за 2000–2008 рр. (табл. 9.1.5):

Таблиця 9.1.5 – Розрахункова

Дата	Обсяг виробництва, т *y*	Абсолютний приріст, т Δ		Темп зростання, % *T*		Темп приросту, % $T_{пр}$		Абсолютне значення 1% приросту, т A
		$\Delta_{ц}$	$\Delta_{б}$	$T_{ц}$	$T_{б}$	$T_{пр.ц}$	$T_{пр.б.}$	
A	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
2004	99 y_0	-	-	100,0	100,0	-	-	-
2005	160 y_1	61	61	161,6	161,6	61,6	61,6	0,99
2006	200 y_2	40	101	125,0	202,0	25,0	102,0	1,60
2007	180 y_3	–20	81	90,0	181,8	-10,0	81,8	2,00
2008	220 y_4	40	121	122,2	222,2	22,2	122,2	1,80

Темп зростання (зниження) Т – це інтенсивність зміни ряду динаміки, тобто відносна зміна рівня динамічного ряду за якийсь період. Це завжди тільки додатна величина, виражена або у процентах, або у коефіцієнтах (якщо темп зростання виражений у коефіцієнтах, його називають коефіцієнтом зростання).

$$T_{Ц} = \frac{y_i}{y_{i-1}} \times 100\ \% \text{ – ланцюговий темп зростання,} \qquad \textbf{(9.1.8)}$$

$$T_{б} = \frac{y_i}{y_0} \times 100\ \% \text{ – базисний темп зростання.} \qquad \textbf{(9.1.9)}$$

Інтерпретація показників за темпу зростання більше 1 (100 %) і менше 1 (100 %) суттєво різниться. Наприклад, для 2008 р. темп зростання обсягу виробництва порівняно з 2007 р. склав 122,2 % (або – якщо у коефіцієнтах – у 1,22 рази перевищив обсяги 2007 р.). Або, наприклад, обсяг виробництва у 2007 р. склав 90 % від обсягу виробництва у 2006 р.

Темп приросту (скорочення) $T_{пр}$ показує, на скільки процентів порівнюваний рівень більше або менше рівня, обраного базою порівняння. Тобто це зміна величини абсолютного приросту. Він може бути і додатним, і від'ємним. (Якщо темп приросту виражений у коефіцієнтах, його називають коефіцієнтом приросту).

$$T_{пр.Ц} = \frac{y_i - y_{i-1}}{y_{i-1}} \times 100\% = \frac{\Delta_Ц}{y_{i-1}} \times 100\% = T_Ц - 100\%, \qquad \textbf{(9.1.10)}$$

або $T_{пр.Ц} = T_Ц \;- 1$ (у коефіцієнтах).

$$T_{пр.б} = \frac{y_i - y_0}{y_0} \times 100\,\% = \frac{\Delta_б}{y_0} \times 100\,\% = T_б - 100\,\%\,, \qquad \textbf{(9.1.11)}$$

або $T_{пр.б} = T_б - 1$ (у коефіцієнтах).

У прикладі 9.1.6 (табл. 9.1.5) обсяг виробництва у 2008 р. порівняно з 2007 р. зріс на 22,2 %.

Абсолютне значення 1 % приросту А – це відношення абсолютного приросту до темпу приросту за той самий період часу у процентах. Він показує, яке абсолютне значення припадає на відносний показник – один процент приросту. Має розмірність рівня ряду.

$$A = \frac{y_i - y_{i-1}}{\dfrac{y_i - y_{i-1}}{y_{i-1}} \times 100} = \frac{\Delta_ц}{T_{пр.Ц.} \times 100} = \frac{y_{i-1}}{100}. \qquad \textbf{(9.1.12)}$$

Як видно з представленої формули, розрахунок цього показника має економічний смисл тільки в разі його розрахунку на ланцюговій основі. У прикладі 9.1.6, у 2008 р. абсолютне значення 1 % приросту складає 1,8 т.

Між розглянутими вище аналітичними показниками ряду динаміки існує такий взаємозв'язок

1. $$\sum \Delta_Ц = \Delta_{б.п}\,. \qquad \textbf{(9.1.13)}$$
2. $$\prod (T_Ц) = T_{б.п}\,. \qquad \textbf{(9.1.14)}$$

Наведені взаємозв'язки дозволяють виконувати перевірку точності розрахованих аналітичних показників ряду динаміки, а також суттєво спрощують розрахунки значень ряду динаміки: середнього абсолютного приросту, середнього темпу зростання і середнього темпу приросту:

- ***середній абсолютний приріст*** – це узагальнений показник швидкості змін у часі

$$\overline{\Delta} = \frac{\sum \Delta_Ц}{n - 1} = \frac{\Delta_{б.п}}{n - 1}. \qquad \textbf{(9.1.15)}$$

У прикладі 9.1.6 середньорічний абсолютний приріст дорівнює:

$\overline{\Delta} = \dfrac{121}{5 - 1} = 30{,}3$ (т);

- ***середній темп зростання*** (середньорічний темп зростання) – це зведений узагальнений показник інтенсивності, який характеризує, у скільки разів у середньому на одиницю часу змінюється рівень ряду динаміки:

$$\overline{T} = \sqrt[n-1]{\prod T_Ц} = \sqrt[n-1]{T_{б.п}} = \sqrt[n-1]{\frac{y_n}{y_0}}\,, \qquad \textbf{(9.1.16)}$$

де $\overline{T} = \sqrt[n-1]{\prod T_Ц}$ – середня геометрична, $\boldsymbol{T_Ц}$ виражений у коефіцієнтах. Так, середньорічний темп зростання обсягів виробництва у прикладі 9.1.6:

$$\overline{T} = \sqrt[5-1]{2{,}22} = \sqrt[5-1]{\frac{220}{99}} = 1{,}22 \quad \text{(або 122 \%)};$$

- ***середній темп приросту*** (середньорічний темп приросту) відображає середню відносну швидкість зміни рівня ряду за одиницю часу:

$$\overline{T_{пр.}} = \overline{T} - 1\ (у\ коефіцієнтах) = \overline{T} - 100\ (у\ процентах)\,. \qquad \textbf{(9.1.17)}$$

Додатна величина середнього темпу приросту у прикладі 9.1.6 (22%) свідчить про щорічний приріст обсягу виробництва.

На практиці часто доводиться виконувати зіставлення динаміки декількох явищ. Таке зіставлення інтенсивностей змін здійснюють:

1) при порівнянні динамічних рядів однакового змісту, але таких, що стосуються різних територій (об'єктів) (приклад 9.1.7);

2) при порівнянні рядів різного змісту, але таких, що характеризують один і той самий об'єкт.

Приклад 9.1.7. За даними табл. 9.1.6 про виробництво цементу (млн.т) у двох країнах визначити, у якій країні виробництво цементу налагоджено краще.

Таблиця 9.1.6 – Вихідні дані

Роки	2004	2005	2006	2007	2008
Країна А	66,0	72,4	95,2	122,0	128,0
Країна Б	35,6	65,1	66,5	65,0	67,0

Різні значення абсолютних рівнів наведених рядів динаміки ускладнюють виявлення особливостей виробництва цементу в країні А і країні Б. Ці особливості можна виявити шляхом приведення рядів динаміки до спільної (єдиної) основи (обравши постійною базою порівняння показники 2004 р.). Далі необхідно розрахувати базисні темпи приросту (відносно 2004 р.) (табл. 9.1.7):

Таблиця 9.1.7 – Базисні темпи зростання виробництва цементу в двох країнах, %

Роки	2004	2005	2006	2007	2008
Країна А	100,0	109,7 (72,4/66,0)	144,2 (95,2/66,0)	184,8 (122,0/66,0)	193,9 (128,0/66,0)
Країна Б	100,0	185,5 (65,1/35,6)	189,5 (66,5/35,6)	185,2 (65,0/35,6)	190,9 (67,0/35,6)

Зіставивши базисні темпи зростання виробництва цементу в країнах А і Б, отримаємо *коефіцієнт випередження*, тобто відносний показник, що характеризує випередження (якщо він більше 1) або відставання (якщо він менше 1) у розвитку країн:

$$K = \frac{T_{бпА}}{T_{бпБ}}, \qquad \textbf{(9.1.18)}$$

де $T_{бпА}$ и $T_{бпБ}$ – базисні темпи зростання двох різних об'єктів.

За даними табл. 9.1.7:

$K_{2006} = \frac{1{,}895}{1{,}442} = 1{,}314$ (більше 1, тобто у 2006 р. країна Б випереджала країну А за темпами зростання виробництва цементу більш ніж у 1,3 рази).

Виявлення основної тенденції ряду динаміки. Основною тенденцією розвитку (*трендом*) називається плавна і стала зміна рівня явища у часі, вільна від випадкових коливань.

Методи аналізу основної тенденції розвитку

1. Метод укрупнення інтервалів.
2. Метод ковзної середньої.
3. Метод аналітичного вирівнювання.

Метод укрупнення інтервалів оснований на укрупненні періодів часу, яких стосуються рівні ряду динаміки (при одночасному зменшенні кількості інтервалів) (приклад 9.1.8).

Приклад 9.1.8. Є ряд щомісячного обсягу виробництва, тис. грн (табл. 9.1.8).

Таблиця 9.1.8 – Вихідні дані

місяць	січень	лютий	березень	квітень	травень	червень	липень	серпень	вересень	жовтень	листопад	грудень
Обсяг виробництва	5,1	5,4	5,2	5,3	5,6	5,8	5,6	5,9	6,1	6,0	5,9	6,2

З табл. 9.1.8 видно, що відмінність у напрямках змін рівня ряду в окремі місяці ускладнює висновки щодо основної тенденції виробництва. Якщо відповідні місячні рівні об'єднати у квартальні і розрахувати середньомісячний випуск продукції за кварталами, тобто укрупнити інтервали, то розв'язання задачі спрощується (табл. 9.1.9).

Таблиця 9.1.9 – Розрахункова

Квартал	За квартал	У середньому за місяць
I	5,1+5,4+5,2=15,7	15,7÷3=**5,23**
II	5,3+5,6+5,8+16,7	16,7÷3=**5,57**
III	5,6+5,9+6,1=17,6	17,6÷3=**5,87**
IV	6,0+5,9+6,2=18,1	18,1÷3=**6,03**

Таким чином, після укрупнення інтервалів основна тенденція зростання виробництва стала очевидною.

Метод ковзної середньої. Сутність методу полягає у заміні абсолютних даних середніми арифметичними за певні періоди. Розрахунок середніх ведеться способом ковзання, тобто поступовим вилученням з обраного періоду ковзання першого рівня і включенням наступного(табл. 9.1.10)

Таблиця 9.1.10 – Вихідні дані і результати розрахунку ковзної середньої, ц/га

Рік	Фактичний рівень врожайності, ц/га	Ковзна середня		
		трирічна	п’ятирічна	чотирирічна
1995	15,4	-	-	
1996	14,0	(15,4+14,0+17,6)/3=15,7	-	15.6
1997	17,6	(14,0 +17,6+15,4)/3=15,7	14,7	(15,6+14,5)/2=15,1
1998	15,4	14,6	15,1	14.5
1999	10,9	14,6	15,2	
2000	17,5	14,5	17,1	і т.д.
2001	15,0	17,0	16,8	
2002	18,5	15,9	17,6	
2003	14,2	15,9	-	
2004	14,9	-	-	
Усього	153,4			

Інтервал ковзної середньої може бути парним і непарним. Якщо інтервал непарний (тобто має кількість членів, що входять до нього непарну кількість разів), то рухому суму і середню відносять до центрального значення інтервалу. Якщо інтервал парний, то спочатку обчислюють суму і середні по кількості рівнів, а потім обчислюють середню з середніх, тобто виконують центрування .

Метод аналітичного вирівнювання.

Етапи аналітичного вирівнювання

1. Обґрунтування вибору типу лінії (залежить від мети дослідження, основаної на теоретичному аналізі явища і на графічному зображенні ряду динаміки).
2. Розрахунок значення параметрів цієї лінії за емпіричними даними:
   - рівняння прямої (форма тренду, найчастіше застосовувана на практичних заняттях з дисципліни “Статистика”) – використовується при прямолінійній залежності і описується формулою:

$$\hat{y}_{\text{t}} = a_0 + a_1 \cdot t\,, \qquad \textbf{(9.1.19)}$$

де $\hat{y}_{\text{t}}$ – рівень ряду динаміки, вирівняний по t (тобто це теоретичний рівень, розташований на шуканій прямій),

a_0 – початковий рівень тренду;

a_1 – середньорічний абсолютний приріст (як правило абстрагований від впливу випадкової коливаємості рівнів в окремі роки або на окремі дати, постійна складова тренду);

t – показник часу, який позначається порядковим номером від нижчого до вищого.

Задача полягає у тому, аби замінити фактичні рівні ряду (y) теоретичними $\hat{y}_{\text{t}}$, які обчислюються на основі рівняння 9.1.19.

Параметри (a_0 і a_1) для рівняння 9.1.19 визначаються способом найменших квадратів (додаток 1), який передбачає систему двох нормальних рівнянь:

$$\begin{cases} \sum y = n \cdot a_0 + a_1 \cdot \sum t \\ \sum yt = a_0 \cdot \sum t + a_1 \cdot \sum t^2 \end{cases}. \qquad \textbf{(9.1.20)}$$

Система спрощується, якщо t підібрати так, щоб $\sum t = 0$, тобто початок відліку часу перенести в середину досліджуваного періоду.

За непарної кількості рівнів ряду центральний інтервал (момент) обирається початковим, де показник часу t має значення 0. Попередні рівні ряду матимуть, відповідно, такі значення показників часу t: –1, –2, –3 і т.д. Наступні рівні ряду матимуть, відповідно, t: +1, +2, +3 і т.д.

За парної кількості рівнів ряду два центральних інтервали (моменти) набувають значень показників часу t: –1 і +1. Попередні рівні ряду будуть, відповідно, дорівнювати –3, –5 і т.д. Наступні рівні ряду, відповідно, дорівнюватимуть +3, +5 і т.д. (це рівнозначно зміні часу не в роках, а в півріччях).

Тоді:

$$\begin{cases} \sum y = n \cdot a_0 \\ \sum yt = a_1 \sum t^2 \end{cases} \Rightarrow \begin{cases} a_0 = \dfrac{\sum y}{n} \\ a_1 = \dfrac{\sum yt}{\sum t^2}. \end{cases} \quad \textbf{(9.1.21)}$$

Якщо розрахунки виконано правильно, то $\sum y = \sum \hat{y}_t$.

Окрім лінійної форми тренду найчастіше використовуються:

- рівняння параболи (параболічна форма тренду) – виражає прискорення або уповільнення змін рівнів ряду з постійним прискоренням:

$$\hat{y}_t = a_0 + a_1 \cdot t + a_2 \cdot t^2; \quad \textbf{(9.1.22)}$$

- рівняння гіперболи (гіперболічна форма тренду) – придатне для описання процесів, що мають граничне (найбільше) значення рівнів ряду:

$$\hat{y}_t = a_0 + a_1 \cdot \frac{1}{t}. \quad \textbf{(9.1.23)}$$

Важливо пам'ятати, що для цих випадків потрібні інші системи рівнянь для знаходження параметрів (a_0 і a_1) (їхнє розв'язання здійснюють тільки за допомогою обчислювальної техніки).

Приклад 9.1.9. У табл. 9.1.11 (графи А і 1) наведено вихідні дані про динаміку експорту цукру в регіоні за 2004–2008 рр. Виконати згладжування ряді динаміки, застосувавши спосіб аналітичного вирівнювання. Знайти очікуваний обсяг експорту цукру в 2009 р., коефіцієнт варіації і середнє квадратичне відхилення.

Розв'язок

Таблиця 9.1.11 – Розрахункова

Роки	Експорт цукру, тис.т, y	t	t^2	yt	$\hat{y}_t$	$y-\hat{y}_t$	$(y-\hat{y}_t)^2$
А	*1*	*2*	*3*	*4*	*5*	*6*	*7*
2004	37	–2	4	–74	36,0	1,0	1,00
2005	39	–1	1	–39	39,9	–0,9	0,81
2006	43	0	0	0	43,8	–0,8	0,64
2007	48	1	1	48	47,7	0,3	0,09
2008	52	2	4	104	51,6	0,4	0,16
Усього	219	0	10	39	219,0	Х	2,70

Для вирівнювання ряду динаміки по прямій використаємо рівняння 9.1.19. За способом найменших квадратів, система нормальних рівнянь (9.1.21) для цього прикладу матиме вигляд:

$$\begin{cases} a_0 = \dfrac{\sum y}{n} = \dfrac{219}{5} = 43{,}8 \\ a_1 = \dfrac{\sum yt}{\sum t^2} = \dfrac{39}{10} = 3{,}9 \end{cases}$$

Таким чином, $\hat{y}_t = 43{,}8 + 3{,}9t$ – рівняння тренду.

Отже, середній рівень експорту цукру в момент часу t=0 склав 43,8 тис. т і щорічно зростав у середньому на 3,9 тис. т. Визначаємо вирівняні значення $\hat{y}_t$ для кожного року (графа 5):

$\hat{y}_t$(2004)=43,8+3,9×(–2)=36,0 (тис. т);

$\hat{y}_t$ (2005)=43,8+3,9×(–1)=39,9 (тис. т);

і так далі.

Визначаємо очікуваний обсяг експорту цукру в 2009 р.:

$\hat{y}_t$ (2009)=43,8+3,9×3=55,5 (тис. т).

Середнє квадратичне відхилення (σ_t) визначаємо за формулою:

$$\sigma_t = \sqrt{\frac{\sum (y - \hat{y}_t)^2}{n}} = \sqrt{\frac{2{,}7}{5}} = 0{,}73 \text{ (тис. т).}$$

Тоді коефіцієнт варіації $V = \dfrac{\sigma_t}{\bar{\hat{y}}_t} = \dfrac{\sigma_t}{\bar{y}_t} = \dfrac{0{,}73}{43{,}8} \times 100\ \% = 1{,}7\ \%$.

σ_t і V характеризують, відповідно, абсолютну і відносну коливаємість рівнів ряду тренду під дією залишкових (випадкових) факторів. У прикладі 9.1.9 ці значення є малими, що характеризує підібране нами рівняння лінійного тренду як дуже близьке до початкових (емпіричних) значень часового ряду. Однак можливі й інші типи тренду, які можуть точніше характеризувати цей розподіл (наприклад, за формулами 9.1.22 або 9.1.23). Зробити висновки про якість підібраних трендів можна за допомогою порівнянь значень σ_t і V для кожного випадку. Де вони менше – той тренд краще**.**

Змикання рядів динаміки – це об’єднання двох або більше рядів, що характеризують зміну явища, в один ряд. Змикання необхідно тоді, коли рівні стають незівставними у зв’язку з територіальними, організаційними змінами, або змінами методології розрахунків, кола охоплюваних об’єктів тощо.

Приклад 9.1.10. Є дані (табл. 9.1.12), що характеризують загальний обсяг продукції промисловості в одному з регіонів, млн. грн. (у фактичних цінах). Привести дані до зіставного вигляду (зімкнути ряд динаміки).

Таблиця 9.1.12 – Вихідні дані

Роки	2002	2003	2004	2005	2006	2007	2008
Рівні продукції промисловості у старих кордонах регіону, млн. грн.	20,1	20,7	21,0	21,2	-	-	-
Рівні продукції промисловості у нових кордонах регіону, млн. грн.	-	-	-	23,8	24,6	25,5	27,2

Розв'язок

- ***I спосіб.***

Визначаємо коефіцієнт співвідношення між рівнями двох рядів:

$$\frac{23{,}8}{21{,}2}=1{,}12\,.$$

Помножуючи на цей коефіцієнт рівні першого ряду, робимо їх зіставними з рівнями другого ряду (табл. 9.1.13):

Таблиця 9.1.13 – Розрахункова

Роки	2002	2003	2004	2005	2006	2007	2008
Рівні продукції промисловості у старих кордонах регіону,млн.грн.	20,1	20,7	21,0	21,2	-	-	-
Рівні продукції промисловості у нових кордонах регіону,млн.грн.	$20{,}1\times1{,}12=$ $=22{,}5$	$20{,}7\times1{,}12=$ $=23{,}2$	$21{,}0\times1{,}12=$ $=23{,}5$	23,8	24,6	25,5	27,2

Таким чином, отримуємо зіставний ряд динаміки загального обсягу виробництва продукції промисловості (у фактичних цінах, за структурою і методологією відповідних років) в одному з регіонів (у нових кордонах, млн. грн.).

- ***II спосіб.***

Рівні продукції промисловості у рік, у якому відбулася зміна (тобто рівні 2005 р.), як до змін, так і після змін припускаються рівними 100 %.

Решта рівнів перераховуються в процентах відносно цих рівнів (до змін – відносно 21,2, а після змін – відносно 23,8). У результаті отримуємо зімкнений ряд (табл. 9.1.14):

Таблиця 9.1.14 – Розрахункова

Роки	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Загальний обсяг промислової продукції у нових кордонах регіону, (% до 2005 р.)	94,8	97,6	99,1	100,0	103,4	107,2	114,3

Сезонні коливання. Періодичні коливання, які мають певний і постійний період, рівний річному проміжку, називаються *сезонними коливаннями,* а динамічний ряд у такому разі називається *сезонним рядом динаміки.*

У статистиці існує кілька методів вивчення сезонних коливань. Найпростіший передбачає побудову спеціальних показників, які називаються індексами сезонності. *Індекс сезонності* – це процентне відношення фактичних (емпіричних) внутрішньогрупових рівнів до теоретичних (розрахункових) рівнів, які є базою порівняння. Сукупність цих показників відображає сезонну хвилю.

Індекси сезонності розраховуються трьома варіантами

1. Для ряду внутрішньорічної динаміки, в якому основна тенденція зростання є незначною або не спостерігається взагалі:

$$I_c = \frac{y_i}{\bar{y}_i} \times 100\%, \qquad \textbf{(9.1.24)}$$

де y_i – фактичні рівні ряду;

$\bar{y}_i$ – середній рівень ряду.

2. Для ряду, що містить дані за декілька років (за умови, що середньорічні рівні ряду змінюються незначно):

$$I_c = \frac{\bar{y}_i}{\bar{y}} \times 100\%, \qquad \textbf{(9.1.25)}$$

де $\bar{y}_i$ – середній рівень для кожного місяця (мінімум за три роки);

$\bar{y}$ – загальна середня за роки, що розглядаються.

3. Для ряду, який має тенденцію до зростання або зниження, то

$$I_c = (\sum \frac{y_i}{\hat{y}_t} \times 100\%) \div n, \qquad \textbf{(9.1.26)}$$

де $\hat{y}_t$ – вирівняні рівні ряду;

n– кількість років.

Приклад 9.1.11. Є дані про кількість зареєстрованих шлюбів за місяцями (табл. 9.1.15, графи А, 1). Знайти індекс сезонності і середнє квадратичне відхилення індексів сезонності від 100 %.

<u>Розв'язок</u>

Таблиця 9.1.15 – Розрахункова

Місяць	Кількість зареєстрований шлюбів, y_i	I_c, %	(I_c–100) %	$(I_c–100)^2$
А	*1*	*2*	*3*	*4*
Січень	776	97	–3	9
Лютий	768	96	–4	16
Березень	672	84	–16	256
Квітень	760	95	–5	25
Травень	648	81	–19	361
Червень	805	101	–1	1
Липень	868	109	9	81
Серпень	890	111	11	121
Вересень	979	122	22	484
Жовтень	832	104	4	16
Листопад	819	102	2	4
Грудень	763	98	–2	4
Усього	9600	1200	X	1378

Оскільки ряд є інтервальним з рівними інтервалами, то за формулою 9.1.1:

$$\overline{y_i} = \frac{\sum y_i}{n} = \frac{9600}{12} = 800 \text{ (шлюбів).}$$

Тоді I_c розраховуються з застосуванням формули 9.1.24 (графа 2).

Звідси $\overline{I_c} = \dfrac{\sum I_c}{12} = \dfrac{1200}{12} = 100$ (%).

Узагальненим показником сили коливаємості динамічного ряду внаслідок сезонності є середнє квадратичне відхилення σ_c індексів сезонності (у %) від 100 %:

$$\sigma_c = \sqrt{\frac{\sum (I_c - 100)^2}{12}} = 10{,}7 \text{ (\%).}$$

Порівняння середніх квадратичних відхилень, обчислених за рівні періоди, показує зрушення у сезонності. Так, зменшення σ_c свідчить про зменшення впливу сезонності на динаміку аналізованого показника.

Приклад 9.1.12. Є дані про виробництво яєць за три роки (табл. 9.1.16, графи А, 1, 2, 3). Розрахувати індекси сезонності.

Розв’язок

Таблиця 9.1.16 – Розрахункова

Місяць	Яйценосність, шт./міс.				I_c, %
	2006 р.	2007 р.	2008 р.	Середньомісячна	
А	*1*	*2*	*3*	*4*	*5*
Січень	10,2	9,7	11,8	10,6	57,6
Лютий	15,2	16,1	14,4	15,2	82,5
Березень	17,3	14,8	15,6	15,9	86,3
Квітень	19,4	22,7	16,5	19,5	105,9
Травень	21,2	25,4	29,1	25,2	136,8
Червень	26,1	28,2	25,2	26,5	143,9
Липень	28,3	25,8	23,5	25,6	140,6
Серпень	21,4	23,3	23,6	22,8	123,8
Вересень	22,1	20,7	18,2	20,3	110,2
Жовтень	14,6	15,2	16,3	15,4	83,6
Листопад	9,5	8,6	13,3	10,5	57,0
Грудень	12,4	12,9	14,6	13,3	72,2
Усього	217,7	223,4	222,1	221,1	1200,4
У середньому	18,14	18,61	18,51	18,42	$\sum 100$

Розрахунок індексів сезонності (графа 5) виконується з застосуванням формули 9.1.25. Середній індекс сезонності для 12 місяців має дорівнювати 100 %, і тоді сума індексів має складати 1200. У прикладі 9.1.12 це співвідношення дорівнює 1200,4 (невелика похибка є наслідком округлень). Аналіз даних табл. 9.1.16 дозволяє зробити такі висновки:

1) виробництво яєць характеризується різко вираженою сезонністю;

2) яйценосність в окремі місяці року відхиляється від середньомісячної на 44 %;

3) найменшою яйценосністю характеризується листопад (57%), а найбільшою – червень (143,9 %).

Екстраполяція та інтерполяція. *Екстраполяція* – це прогноз або подовження ряду, тобто розрахунок приблизних рівнів за межами досліджуваного ряду. Екстраполяцію рядів динаміки вивчають різними способами, наприклад, екстраполюють ряди динаміки вирівнюванням за аналітичними формулами. Знаючи рівняння для теоретичних рівнів і підставляючи в нього значення t за межами досліджуваного ряду, розраховують для t імовірнісні $\hat{y}_t$.

Інтерполяція – це спосіб визначення невідомих значень динамічного ряду, який полягає у наближеному відображенні закономірності, що встановилася, всередині певного відрізку часу. Проміжні рівні інтерполюють по середньому абсолютному приросту, середньому темпу зростання, абсолютному приросту за прилеглі роки, середній хронологічній.

На практиці результати прогнозованих явищ y_{np} зазвичай отримують не точковим (дискретним), а інтервальним оцінюванням. Для визначення інтервалів використовують формулу:

$$y_{np} = \hat{y}_{\text{t}} \pm t_\alpha S_{\hat{y}_t}\,, \qquad \textbf{(9.1.27)}$$

де t_α – коефіцієнт довіри за розподілом Ст’юдента,

$S_{\hat{y}_t} = \sqrt{\dfrac{\sum (y_{\text{i}} - \hat{y}_t)^2}{n-m}}$ – залишкове середнє квадратичне від тренду, скореговане за кількістю ступенів свободи (*n*-*m*),

n – кількість рівнів ряду,

m – кількість параметрів адекватної моделі тренду (для рівняння прямої *m*=2).

Імовірнісні межі інтервалу прогнозованого явища:

$$\hat{y}_{\text{t}} - t_\alpha S_{\hat{y}_t} \le y_{np} \le \hat{y}_{\text{t}} + t_\alpha S_{\hat{y}_t}\,. \qquad \textbf{(9.1.28)}$$

Приклад 9.1.13. Розрахувати прогнозовані довірчі інтервали очікуваного обсягу експорту цукру за даними прикладу 9.1.9.

<u>Розв’язок</u>

Якщо *n*=5, *m*=2 (a_0 і a_1), то число ступенів свободи дорівнює 5–2=3. Тоді за довірчої імовірності Р=0,95 (тобто за рівня значимості α=0,05), коефіцієнт довіри t_α =3,183 (який знаходять за таблицею, що міститься у будь-яких статистичних збірниках і має назву “Значення *t*-критерію Ст’юдента за рівня значимості 0,10; 0,05; 0,01”).

Тоді:

$\sum (y - \hat{y}_{\text{t}})^2$=2,70 (табл. 9.1.11, графа 7);

$$S_{\hat{y}_t} = \sqrt{\frac{\sum 2{,}7}{5-2}} = 0{,}95\,.$$

Знаючи точкову оцінку прогнозованого обсягу експорту цукру на 2009 р. ($\hat{y}_{\text{t}}$(2009)=55,5 (тис. т)), визначимо імовірнісні межі інтервалу:

$$55{,}5 - 0{,}95 \times 3{,}183 \le y_{np} \le 55{,}5 + 0{,}95 \times 3{,}183\,;$$

52,5 тис. т $\le y_{пр} \le$ 58,5 тис. т.

Таким чином, з імовірністю, рівною 0,95, можна стверджувати, що очікуваний обсяг експорту цукру в 2009 р. буде не менше 52,5 тис. т., але й не більше 58,5 тис. т.

9.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Дайте визначення ряду динаміки. З яких елементів він складається?
2. Які існують види рядів динаміки? Які динамічні ряди називаються моментними і чому їхні рівні не можна сумувати?
3. Які ряди статистичних величин називають інтервальними? Чому їхні рівні не можна сумувати? Наведіть приклади.
4. Назвіть умови правильної побудови динамічного ряду.
5. Як обчислюється середня для інтервального ряду? Наведіть приклади. Як обчислюється середня для моментного ряду? Наведіть приклади.
6. Що характеризують показники абсолютного приросту і як вони обчислюються?
7. Що таке темп зростання і як він обчислюється? Що показує абсолютне значення одного процента приросту і як воно обчислюється?
8. Чому дорівнює темп приросту?
9. Яким є взаємозв’язок між аналітичними показниками?
10. Способи розрахунку середніх величин аналітичних показників.
11. Що таке коефіцієнт випередження?
12 Як здійснюється змикання рядів динаміки? Коли це необхідно?
13. Якими найбільш поширеними статистичними методами здійснюється вивчення тренду в рядах динаміки?
14. Сутність методу укрупнення інтервалів.
15. Сутність методу ковзної середньої. Його достоїнства і недоліки.
16. Сутність методу аналітичного вирівнювання**.**
17. Як визначити тип рівняння тенденції динаміки?
18. Техніка вирівнювання ряду динаміки по прямій.
19. Що таке сезонні коливання?
20. Що таке екстраполяція та інтерполяція рядів динаміки?
21. Є дані про випуск продукції підприємством за місяцями (тис. грн.):

січ	лют	берез	квіт	трав	черв	лип	серп	вер	жовт	лист	груд
118	124	124	128	127	132	136	131	135	141	139	146

Виконати згладжування рядів динаміки, застосувавши 1) метод укрупнення інтервалів; 2) метод ковзної середньої (для трьох місяців).

22. Обсяг продукції, реалізованої промисловим підприємством характеризується такими даними, тис. грн.:

2003 р.	2004 р.	2005 р.	2006 р.	2007 р.	2008 р.
642	654	670	680	688	696

Визначити абсолютні прирости, темпи зростання і приросту за роками і до 1999 р., абсолютне значення 1 % приросту, середній рівень ряду, середній абсолютний приріст, середній темп зростання і приросту.

23. Динаміка кредитних ресурсів філії банку на початок кожного місяця звітного року характеризується такими даними (млн. грн.):

01.01	01.02	01.03	01.04	01.05	01.06	01.07
48	53	51	50	55	54	52

Визначити загальний обсяг кредитних ресурсів за кожний квартал і півріччя загалом.

24. У таблиці наведено дані про динаміку виробництва молока (млн. т) у регіоні за 2004-2008 рр.:

2004 р.	2005 р.	2006 р.	2007 р.	2008 р.
13,3	13,5	14,8	16,1	16,6

Виконати згладжування ряду динаміки, застосувавши спосіб аналітичного вирівнювання. Знайти середнє квадратичне відхилення і коефіцієнт варіації. Зробити висновки.

25. Використовуючи взаємозв’язок між показниками динаміки, визначити рівні виробництва електроенергії в регіоні і ті базисні показники динаміки, яких бракує у таблиці:

Роки	Виробництва електроенергії, млн. кВт год.	Базисні показники динаміки		
		Абсолютний приріст, млн. кВт год.	Темп зростання, %	Темп приросту, %
2003	1202			
2004				3,1
2005			107,6	
2006		124,0		
2007			113,7	
2008				17,9

26. Є такі відомості про середню вартість основних виробничих фондів заводу (тис. грн.):

	2003 р.	2004 р.	2005 р.	2006 р.	2007 р.	2008 р.
До переоцінювання	1,40	1,52	1,88	2,40	3,16	
Після переоцінювання					41,4	48,5

Вкажіть причину незіставності рівнів ряду динаміки для порівняльного аналізу. Приведіть рівні ряду до зіставного вигляду.

27. Є такі дані про рух матеріалу на складі бази за січень, т:

залишок на 01.01	40,0
05.01 надійшло від постачальників	160,0
08.01 відвантажено споживачам	100,0
10.01 надійшло від постачальників	30,5
13.01 відвантажено споживачам	60,5
15.01 надійшло від постачальників	50,0
20.01 відвантажено споживачам	90,0.

Визначте середній залишок матеріалу на складі за січень.

28. Є дані про чисельність робітників підприємства на початок кожного місяця звітного року:

01.01	01.02	01.03	01.04	01.05	01.06	01.07	01.08	01.09	01.10	01.11	01.12	01.01 наступного року
202	208	212	210	216	220	202	198	198	206	214	216	218

Обчисліть середню місячну чисельність робітників:

1) за кожний квартал;
2) за кожне півріччя;
3) за рік.

29. На 1 жовтня у бригаді числилося 20 робітників. З 6 жовтня до складу бригади увійшли ще 3 чоловіка, з 16 жовтня 2 чоловіка вибуло, а з 20 жовтня бригада поповнилася ще 4 робітниками. Знайти середньомісячну численність бригади.

30. Виробництво продукції малим підприємством за півріччя характеризується такими даними:

Місяці	Виробництво продукції, тис. грн.
Січень	435
Лютий	465
Березень	501
Квітень	536
Травень	587
Червень	643

Для аналізу динаміки виробництва продукції обчисліть:

1) абсолютні прирости, темпи зростання і темпи приросту за місяцями і до січня, абсолютне значення 1 % приросту (отримані показники представте у таблиці);
2) середньомісячне виробництво продукції;
3) середньомісячний абсолютний приріст;
4) середньомісячні темпі зростання і приросту.

Зобразіть ряд динаміки на графіку.

31. Є такі дані про реалізацію цукрового піску в продовольчих магазинах міста, т:

	Роки			
Місяць	2005	2006	2007	2008
Січень	108	130	150	160
Лютий	105	125	145	162
Березень	106	135	154	162
Квітень	130	150	170	164
Травень	110	140	154	168
Червень	120	147	160	178
Липень	132	160	200	210
Серпень	121	145	165	190
Вересень	106	136	152	180
Жовтень	125	150	160	172
Листопад	116	140	155	170
Грудень	136	145	158	164

На основі наведених даних:
1) визначте середньомісячні рівні реалізації цукру за кожний рік і, обчисливши базисні темпи зростання середніх рівнів, визначте характер загальної тенденції реалізації цукру в 2005 – 2008 рр.;
2) оцініть сезонні коливання реалізації цукру (методом аналітичного вирівнювання) за місяцями цього чотириріччя;
3) показники сезонної хвилі річного циклу реалізації цукру зобразіть графічно.
Зробіть короткі висновки.

32. Є такі дані про продаж буряку на ринках декількох міст, тис. т.:

	Роки			
Місяць	2005	2006	2007	2008
Січень	35	38	40	49
Лютий	30	35	38	45
Березень	34	37	40	48
Квітень	32	33	39	46
Травень	25	32	36	40
Червень	23	30	35	38
Липень	28	32	38	45
Серпень	35	38	45	53
Вересень	44	46	50	52
Жовтень	50	52	55	62
Листопад	46	48	54	56
Грудень	33	40	50	54

На основі наведених даних:
1) визначте середньомісячні рівні реалізації буряку за кожний рік і, обчисливши базисні темпи зростання середніх рівнів, визначте характер загальної тенденції реалізації буряку в 2005-2008 рр.;
2) оцініть сезонні коливання реалізації буряку (методом 12-місячної ковзної середньої з центруванням) за місяцями цього чотириріччя;
3) показники сезонної хвилі річного циклу реалізації зобразіть графічно.
Зробіть короткі висновки.

ГЛАВА 10

СТАТИСТИЧНИЙ АНАЛІЗ СТРУКТУРИ

10.1
МЕТОДИЧНІ ВКАЗІВКИ

Процеси і явища у промисловості, сільському господарстві, фінансах, торгівлі, демографії, соціальній і політичній сферах, що вивчаються статистикою, зазвичай характеризуються внутрішньою структурою, яка з часом може змінюватися. Саме тому вивчення структури і структурних зрушень посідає важливе місце в курсі теорії статистики.

Під структурою у статистиці розуміють сукупність одиниць, які мають усталені внутрішньогрупові зв'язки при збереженні основних ознак, які характеризують цю сукупність як одне ціле. Основні напрями вивчення структури включають:

1) характеристику структурних зрушень окремих частин сукупності за два і більше періодів;
2) узагальнену характеристику структурних зрушень на рівні сукупності;
3) оцінювання ступеня концентрації та централізації.

Розглянемо послідовно ці напрями дослідження.

Частинні показники структурних зрушень. Аналіз структури та її змін ґрунтується на відносних показниках структури – частках або питомих вагах, що являють собою співвідношення між величиною частин і цілого. При цьому як частинні, так і узагальнені показники структурних зрушень можуть відображати або "абсолютну" зміну структури в процентних пунктах або частках одиниці (лапки означають, що ці показники є абсолютними за методологією розрахунку, а не за одиницями вимірювання), або її відносну зміну в процентах або коефіцієнтах. До цих показників належать:

а) *"абсолютний" приріст питомої ваги i-ї частини сукупності* (Δ_{d_i}) – показує, на скільки процентних пунктів збільшилась або зменшилась певна структурна частина у *j*-й період порівняно з (*j*–1)-м періодом:

$$\Delta_{d_i} = d_{ij} - d_{ij-1}, \qquad \textbf{(10.1.1)}$$

де d_{ij} – питома вага (частка) *i*-ї частини сукупності в *j*-й період;

d_{ij-1} – питома вага (частка) *i*-ї частини сукупності в (*j*–1)-й період;

б) *темп зростання питомої ваги* (T_{d_i}) – це відношення питомої ваги i-ї частини в j-й період часу до питомої ваги тієї ж частини у попередній період:

$$T_{d_i} = \frac{d_{ij}}{d_{ij-1}} \times 100 \quad . \qquad \textbf{(10.1.2)}$$

Темпи зростання питомої ваги виражаються в процентах і завжди є додатними величинами;

в) *середній "абсолютний" приріст питомої ваги i-ї структурної частини* ($\overline{\Delta}_{d_i}$) – показує, на скільки процентних пунктів у середньому за якийсь період (день, тиждень, місяць, рік тощо) змінюється певна структурна частина:

$$\overline{\Delta}_{d_i} = \frac{d_{in} - d_{i1}}{n-1}, \qquad \textbf{(10.1.3)}$$

де n – кількість усереднених періодів.

Сума середніх "абсолютних" приростів питомих ваг всіх k структурних частин сукупності, так як і сума їхніх приростів за одним часовий інтервал, повинна дорівнювати нулю;

г) *середній темп зростання питомої ваги* ($\overline{T}_{d_i}$) – характеризує середню відносну зміну питомої ваги i-ї структурної частини за n періодів і розраховується або за формулою середньої геометричної:

$$\overline{T}_{d_i} = \sqrt[n-1]{T_{d_{i1}} \cdot T_{d_{i2}} \cdot T_{d_{i3}} \cdot \ldots \cdot T_{d_{in-1}}}\,, \qquad \textbf{(10.1.4)}$$

де підкореневий вираз є послідовним добутком ланцюгових темпів зростання питомої ваги за всі часові інтервали,

або за формулою:

$$\overline{T}_{d_i} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}} \cdot 100\,; \qquad \textbf{(10.1.5)}$$

д) *середня питома вага кожної i-ї частини за весь аналізований часовий інтервал* ($\overline{d_i}$):

$$\overline{d_i} = \frac{\sum_{j=1}^{n} X_{ij}}{\sum_{j=1}^{n}\sum_{i=1}^{k} X_{ij}} \cdot 100\,, \qquad \textbf{(10.1.6)}$$

де X_{ij} – величина i-ї структурної частини в j-й період часу в абсолютному вимірі;

k – кількість структурних частин.

Узагальнені показники структурних зрушень. Іноді досліднику необхідно оцінити структурні зміни в аналізованому соціально-економічному явищі за певний часовий інтервал, які характеризують рухомість або, навпаки, стабільність, сталість структури. Це зазвичай потрібно для порівняння динаміки однієї і тієї ж структури в різні періоди або декількох структур, що стосуються різних об’єктів. У другому випадку кількість структурних частин об’єктів не обов’язково буде однаковою.

Серед застосовуваних для цього узагальнених показників найбільш поширеними є такі:

а) *лінійний коефіцієнт “абсолютних” структурних зрушень*($\overline{\Delta}_{d_1-d_0}$), який відображає ту середню зміну питомої ваги (у процентних пунктах), яка сталася за аналізований часовий інтервал в усіх структурних частинах сукупності:

$$\overline{\Delta}_{d_1-d_0} = \frac{\sum_{i=1}^{k}\left|d_{ij}-d_{ij-1}\right|}{k}; \qquad \textbf{(10.1.7)}$$

б) *квадратичний коефіцієнт “абсолютних” структурних зрушень*($\sigma_{d_1-d_0}$):

$$\sigma_{d_1-d_0} = \sqrt{\frac{\sum_{i=1}^{k}\left(d_{ij}-d_{ij-1}\right)^2}{k}}. \qquad \textbf{(10.1.8)}$$

Лінійний і квадратичний коефіцієнти “абсолютних” структурних зрушень дозволяють отримати зведену оцінку швидкості зміни питомих ваг частин сукупності;

в) *квадратичний коефіцієнт відносних структурних зрушень* ($\sigma_{\frac{d_1}{d_0}}$) – відображає той середній відносний приріст питомої ваги (у процентах), який спостерігався в аналізований період:

$$\sigma_{\frac{d_1}{d_0}} = \sqrt{\sum_{i=1}^{k}\frac{\left(d_{ij}-d_{ij-1}\right)^2}{d_{ij-1}}\cdot 100}; \qquad \textbf{(10.1.9)}$$

г) *лінійний коефіцієнт “абсолютних” структурних зрушень за n періодів* ($\overline{\Delta}^{(n)}_{d_1-d_0}$) – може використовуватися як для порівняння динаміки двох або більше структур, так і для аналізу динаміки однієї і тієї ж структури протягом різних за тривалістю періодів часу:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{\sum_{i=1}^{k}\left|d_{in}-d_{i1}\right|}{k(n-1)}. \qquad \textbf{(10.1.10)}$$

Показники концентрації та централізації. Під концентрацією мається на увазі ступінь нерівномірності розподілу досліджуваної ознаки, не пов'язана ані з об'ємом сукупності, ані з чисельністю її окремих груп. Централізація означає зосередження об'єму/обсягу ознаки в окремих одиницях сукупності. До *показників концентрації* належать:

а) коефіцієнт Джині (G):

$$G = 1 - 2\sum_{i=1}^{k} d_{xi} d_{yi}^{H} + \sum_{i=1}^{k} d_{xi} d_{yi}, \qquad \textbf{(10.1.11)}$$

де d_{xi} – частка i-ї групи в загальному об'ємі/обсязі сукупності;

d_{yi} – частка i-ї групи в загальному об'ємі/обсязі ознаки;

d_{yi}^{H} – накопичена частка i-ї групи в загальному об'ємі/обсязі ознаки.

Чим ближче значення цього показника до 1 (100 %), тим вищим буде рівень концентрації; при нульовому значенні маємо рівномірний розподіл ознаки по всім одиницям сукупності;

б) коефіцієнт Лоренца (L):

$$L = \frac{\sum_{i=1}^{k} \left| d_{xi} - d_{yi} \right|}{2}. \qquad \textbf{(10.1.12)}$$

Коефіцієнт Лоренца змінюється в тих самих межах, що й коефіцієнт Джині;

в) узагальнений показник централізації (I_z):

$$I_z = \sum_{i=1}^{k} \left(\frac{m_i}{M} \right)^2, \qquad \textbf{(10.1.13)}$$

де m_i – значення ознаки i-ї одиниці сукупності;

M – об'єм ознаки всієї сукупності.

Максимального значення, рівного 1, цей коефіцієнт досягає лише тоді, коли сукупність складається тільки з однієї одиниці, на яку припадає весь об'єм ознаки; мінімальне значення коефіцієнта наближується до нуля, але ніколи його не досягає.

Приклад 10.1.1. Національним банком України розроблено класифікацію комерційних банків залежно від величини їхніх активів (табл. 10.1.1, графа А) [26]. У результаті отримано такі дані для сукупності з 70 банків України станом на 01.01.04 р., 01.01.05 р. і 01.01.06 р. (табл. 10.1.1, графи 1–3). Необхідно провести аналіз структури сукупності банків на різні моменти часу: розрахувати частинні та загальні показники структурних зрушень, оцінити ступінь концентрації та централізації; а також пояснити отримані результати.

Таблиця 10.1.1. – Вихідні показники

Групи комерційних банків за розмірами активів (класифікація НБУ), млн. грн.	Кількість банків		
	01.01.04	01.01.05	01.01.06
А	*1*	*2*	*3*
Вище 2500 (I)	7	8	14
Від 1300 до 2500 (II)	5	12	14
Від 400 до 1300 (III)	25	19	21
Нижче 400 (IV)	33	31	21

Оброблення даних

Розрахуємо показники частинних структурних зрушень, використовуючи формули 10.1.1 – 10.1.2 (табл. 10.1.2).

Таблиця 10.1.2. – Розрахункова

Групи, млн. грн.	Кількість банків, X_{ij}			Питома вага, у % до підсумку, d_{ij}			Δ_{d_i}, проц. пунктів		T_{d_i}, %	
	01.01.04 X_{i0}	01.01.05 X_{i1}	01.01.06 X_{i2}	01.01.04 d_{i0}	01.01.05 d_{i1}	01.01.06 d_{i2}				
А	*1*	*2*	*3*	*4*	*5*	*6*	*7 (гр.5 - Гр.4)*	*8 (гр.6 - гр.5)*	*9 (гр.5 / гр.4)*	*10 (гр.6 / гр.5)*
I	7	8	14	10,0	11,4	20,0	1,4	8,6	114,0	175,4
II	5	12	14	7,1	17,1	20,0	10,0	2,9	240,8	117,0
III	25	19	21	35,8	27,2	30,0	-8,6	2,8	76,0	110,3
IV	33	31	21	47,1	44,3	30,0	-2,8	-14,3	94,1	67,7
Усього	70	70	70	100,0	100,0	100,0	0	0	Х	Х

Як слідує з розрахункових даних табл. 10.1.2, станом на 01.01.05 р. найсуттєвіше в “абсолютному” вимірі змінилася питома вага банків з розміром активів від 1300 до 2500 млн. грн. (зросла на 10,0 проц. пунктів); станом на 01.01.06 р. різко знизилася питома вага банків з розміром активів нижче 400 млн. грн. (на 14,3 проц. пунктів). Про різке зростання к 2006 р. кількості банків з розміром активів вище 2500 млн. грн. свідчать і відносні показники (у 1,75 разів). Отже, к початку 2006 р. (моменту виходу з кризи 2004-2005 рр. у банківській сфері, пов'язаній з політичною і соціально-економічною ситуацією в Україні) відбулося нарощування активів банками.

Оскільки досліджувана нами структура представлена даними за більш ніж два періоди, то виникає необхідність у динамічному усередненні наведених вище показників, тобто у розрахунку середніх показників структурних зрушень, а саме: середнього "абсолютного" приросту питомої ваги i-ї структурної частини ($\overline{\Delta}_{d_i}$); середнього темпу зростання питомої ваги i-ї структурної частини за n періодів ($\overline{T}_{d_i}$), середньої питомої ваги кожної i-ї частини за весь досліджуваний інтервал ($\overline{d_i}$). Визначимо два перші показника за допомогою формул 10.1.3 – 10.1.5 (табл. 10.1.3).

Згідно з отриманими результатами, питомі ваги перших двох груп щорічно зростали (на 5 проц. пунктів або на 41,4 % для I групи банків; на 6,5 проц. пунктів або на 67,8 % для II групи банків). А питомі ваги двох останніх груп щорічно зменшувалися (на 2,9 проц. пунктів або на 8,5 % для III групи; на 8,6 проц. пунктів або на 20,2 % для IV групи).

Таблиця 10.1.3. – Розрахунок середніх: "абсолютного" приросту і темпу зростання питомої ваги кожної i-ї структурної частини

Групи, млн.грн.	Кількість банків, X_{ij}			Питома вага, у % до підсумку, d_{ij}			$\overline{\Delta}_{d_i}$ (середньорічний), проц. пунктів	$\overline{T}_{d_i}$ (середньорічний), %
	01.01.04 X_{i0}	01.01.05 X_{i1}	01.01.06 X_{i2}	01.01.04 d_{i0}	01.01.05 d_{i1}	01.01.06 d_{i2}		
А	*1*	*2*	*3*	*4*	*5*	*6*	*7* $\left(\frac{гр.6-гр.4}{3-1}\right)$	*8* $\sqrt[3-1]{\frac{гр.6}{гр.4}}$
I	7	8	14	10,0	11,4	20,0	5,0	141,4
II	5	12	14	7,1	17,1	20,0	6,5	167,8
III	25	19	21	35,8	27,2	30,0	-2,9	91,5
IV	33	31	21	47,1	44,3	30,0	-8,6	79,8
Усього	70	70	70	100,0	100,0	100,0	0	Х

Аналогічним чином за наявною інформацією можна розрахувати не тільки середньорічні, а й середньомісячні "абсолютні" прирости і темпи зростання питомої ваги кожної i-ї структурної частини. Для цього у формулах 10.1.3 – 10.1.5 замість n – кількість років, розглядають n –

кількість місяців.
Наприклад:
$\overline{\Delta}_{d_i} = \frac{d_{in} - d_{i1}}{n-1} = \frac{20,0-10,0}{24-1} = 0,4$ (проц. пунктів) – середньомісячний "абсолютний" приріст питомої ваги банків І групи за 2004-2005 рр., $\overline{T}_{d_i} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}}\cdot 100 = \sqrt[24-1]{\frac{20}{10}}\cdot 100 = \sqrt[23]{2}\cdot 100 =$ 103,1 (%) – середньомісячний темп зростання питомої ваги банків І групи за 2004-2005 рр.

Середню питому вагу кожної *і*-ї частини за весь цей інтервал визначимо за допомогою формули 10.1.6. (табл. 10.1.4).

Таблиця 10.1.4. – Проміжні дані для розрахунку $\overline{d_i}$

Групи, млн. грн.	Кількість банків			
	01.01.04	01.01.05	01.01.06	Усього
I (X_{1j})	7	8	14	29
II (X_{2j})	5	12	14	31
III (X_{3j})	25	19	21	65
IV (X_{4j})	33	31	21	85
Усього	70	70	70	210

Використовуючи дані таблиці 10.1.4, розрахуємо:

$\overline{d}_I = \frac{29}{210}\cdot 100 = 13,8$ (%) – середня питома вага банків з розміром активів вище 2500 млн. грн. у загальному об'ємі сукупності.

$\overline{d}_{II} = \frac{31}{210}\cdot 100 = 14,8$ (%) – середня питома вага банків з розміром активів від 1300 до 2500 млн. грн. у загальному об'ємі сукупності.

$\overline{d}_{III} = \frac{65}{210}\cdot 100 = 31,0$ (%) – середня питома вага банків з розміром активів від 400 до 1300 у загальному об'ємі сукупності.

$\overline{d}_{IV} = \frac{85}{210}\cdot 100 = 40,4$ (%) – середня питома вага банків з розміром активів нижче 400 млн. грн. у загальному об'ємі сукупності.

Отже, у середньому за два повні роки, що розглядаються, тобто 2004-5005 рр., найбільшу питому вагу складали банки з розміром активів нижче 400 млн. грн.

Розрахуємо узагальнені показники структурних зрушень.

Для розрахунку лінійного коефіцієнта "абсолютних" структурних зрушень за ці періоди (формула 10.1.7) скористаємося даними підсумків у графах 4 і 7 табл. 10.1.5.

Тоді:

$$\overline{\Delta}_{d_1-d_0} = \frac{22{,}8}{4} = 5{,}70 \text{ (проц. пунктів);}$$

$$\overline{\Delta}_{d_2-d_1} = \frac{28{,}6}{4} = 7{,}15 \text{ (проц. пунктів).}$$

Таблиця 10.1.5. – Розрахунок узагальнених показників структурних зрушень

Групи, млн.грн	Питома вага, у % до підсумку, d_{ij}			Розрахункові графи						
	01.01.04 d_{i0}	01.01.05 d_{i1}	01.01.06 d_{i2}	$\lvert d_{i1}-d_{i0}\rvert$	$(d_{i1}-d_{i0})^2$	$\frac{(d_{i1}-d_{i0})^2}{d_{i0}}$	$\lvert d_{i2}-d_{i1}\rvert$	$(d_{i2}-d_{i1})^2$	$\frac{(d_{i2}-d_{i1})^2}{d_{i1}}$	$\lvert d_{i2}-d_{i0}\rvert$
А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
I	10,0	11,4	20,0	1,40	1,96	0,20	8,60	73,96	6,49	10,00
II	7,1	17,1	20,0	10,00	100,00	14,08	2,90	8,41	0,49	12,90
III	35,8	27,2	30,0	8,60	73,96	2,07	2,80	7,84	0,29	5,80
IV	47,1	44,3	30,0	2,80	7,84	0,17	14,30	204,49	4,62	17,10
Усього	100,0	100,0	100,0	22,80	183,76	16,52	28,60	294,70	11,89	45,80

Отже, к початку 2005 р. питома вага окремих груп банків змінилася у середньому на 5,70 проц. пунктів. К початку 2006 р. "абсолютні" структурні зрушення помітно зросли (на 7,15 проц. пунктів). Цей висновок підтверджується квадратичними коефіцієнтами "абсолютних" структурних зрушень (необхідні проміжні розрахунки виконано в графах 5 і 8 табл. 10.1.5).

Використовуючи формулу 10.1.8, отримуємо:

$$\sigma_{d_1-d_0} = \sqrt{\frac{183{,}76}{4}} = 6{,}8 \text{ (проц. пунктів);}$$

$$\sigma_{d_2-d_1} = \sqrt{\frac{294{,}70}{4}} = 8{,}6 \text{ (проц. пунктів).}$$

Далі визначимо величину квадратичних коефіцієнтів відносних структурних зрушень, скориставшись підсумковими даними граф 6 і 9 табл. 10.1.5

Використовуючи формулу 10.1.9, отримаємо:

$$\sigma_{\frac{d_1}{d_0}} = \sqrt{16{,}52 \cdot 100} = 40{,}6 \text{ (\%);}$$

$$\sigma_{\frac{d_2}{d_1}} = \sqrt{11{,}89 \cdot 100} = 34{,}5 \quad (\%).$$

Розрахунки показують, що якщо к початку 2005 р. питома вага кожної групи банків у середньому змінилася трохи менше, ніж на половину своєї величини, то к початку 2006 р. – більш ніж на третину.

Використовуючи підсумкові дані графи 10 табл. 10.1.5 і формулу 10.1.10, знайдемо лінійний коефіцієнт "абсолютних" структурних зрушень протягом n періодів:

$$\bar{\Delta}^{(n)}_{d_1-d_0} = \frac{45{,}80}{4 \cdot 2} = 5{,}73 \quad \text{(проц. пунктів).}$$

Таким чином, за досліджуваний період середньорічна зміна кількості банків у всіх групах залежно від розміру активів склала 5,73 проц. пунктів.

Розрахуємо показники концентрації та централізації.

Визначимо ступінь концентрації кількості банків залежно від обсягу активів за даними табл. 10.1.6. Сукупність банків необхідно впорядкувати за групами: від групи з найменшими розмірами активів до групи з найбільшими розмірами активів (а інакше коефіцієнт може виявитися від'ємним). Дані графи 3 отримано з офіційного сайту Асоціації українських банків (АУБ), де щомісячно публікується інформація про обсяг активів кожного банку, що належить до банківської системи України.

Таблиця 10.1.6. – Розподіл банків за обсягами активів на 01.01.06

Групи, млрд.. грн.	Кількість банків	d_{xi}	Обсяг активів, млрд. грн.			$d_{xi} \cdot d_{yi}$	d^H_{yi}	$d_{xi} \cdot d^H_{yi}$	$\lvert d_{xi} - d_{yi} \rvert$
			У групі, млрд. грн.	% до підсумку	d_{yi}				
А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
IV	21	0,3	4,3	2,6	0,026	0,0078	0,026	0,0078	0,274
III	21	0,3	15,4	9,3	0,093	0,0279	0,119	0,0357	0,207
II	14	0,2	27,2	16,5	0,165	0,0330	0,284	0,0568	0,035
I	14	0,2	118,1	71,6	0,716	0,1432	1,000	0,2000	0,516
Усього	70	1,0	165,0	100,0	1,000	0,2119	Х	0,3003	1,032

Використовуючи формулу 10.1.11, визначимо коефіцієнт Джині:

G=1–2×0,3003+0,2119=0,6113 (або 61,13 %).

Використовуючи формулу 10.1.12, визначимо коефіцієнт Лоренца:

$$L = \frac{1{,}032}{2} = 0{,}516 \text{ (або 51,60 \%).}$$

Обидва коефіцієнти вказують на високий ступінь концентрації (нерівномірності розподілу) банків за розмірами активів.

Розрахуємо показник централізації банків за розмірами активів з використання формули 10.1.13 (табл. 10.1.7).

Таблиця 10.1.7. – Проміжна таблиця для розрахунку коефіцієнта централізації на 01.01.06

Групи, млрд. грн.	Кількість банків	Обсяг активів млрд. грн.		Частка одного банку у загальному обсязі активів
		У групі	У середньому на один банк	
А	*1*	*2*	*3 (гр.2:гр.1)*	*4 (гр.3:* $\sum$ *гр.2)*
I	14	118,1	8,4	0,051
II	14	27,2	1,9	0,012
III	21	15,4	0,7	0,004
IV	21	4,3	0,2	0,001
Усього	70	165,0	X	X

$$I_z = 14 \cdot 0{,}051^2 + 14 \cdot 0{,}012^2 + 21 \cdot 0{,}004^2 + 21 \cdot 0{,}001^2 = 0{,}04$$

Розрахована величина свідчить про низький ступінь концентрації.

10.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Дайте визначення структури.
2. Які основні напрями вивчення структури ви знаєте?
3. Назвіть частинні показники структурних зрушень та їхні одиниці вимірювання.
4. Назвіть узагальнені показники структурних зрушень, а також умови їх застосування і одиниці вимірювання.
5. Показник концентрації: економічний смисл, види.
6. Сутність показника централізації.
7. Відомі такі дані про обсяги кредитів, наданих комерційним банкам

України, млрд. грн.:

Кредити	01.01.08 р.	01.01.09 р.
Короткострокові	60554,4	92993,8
Довгострокові	3410,1	4776,6

Проведіть аналіз зміни структури наданих банкам кредитів, використовуючи показники “абсолютного” приросту і темпу зростання питомої ваги.

8. Зміна питомої ваги експорту окремих видів продукції у загальному обсязі їх виробництва в одній з країн характеризується такими даними:

Продукція	Питома вага експорту, %	
	2007 р.	2009 р.
Нафта	40,3	39,9
Газ природний	31,7	33,7
Нафтопродукти	33,1	32,3
Вугілля	8,9	11,3
Промислова деревина	22,8	28,4
Пиломатеріали	32,5	29,6
Фанера	76,8	79,9
Целюлоза	78,8	78,4
Папір газетний	65,1	67,1
Автомобілі легкові	29,7	24,6
Автомобілі вантажні	22,3	15,0

Визначте, для яких видів продукції частка експорту зазнала за рік найбільших “абсолютних” і відносних змін.

9. Структура грошових доходів населення однієї з країн характеризується такими даними, %:

Роки	2003	2004	2005	2006	2007	2008
Грошові доходи – усього	100	100	100	100	100	100
у тому числі:						
оплата праці	74,1	59,7	69,9	58,0	46,4	39,5
соціальні трансферти	13,0	15,5	14,0	17,2	17,4	16,7
доходи від власності, підприємницької діяльності тощо.	12,9	24,8	16,1	24,8	36,2	43,8

Проаналізуйте динаміку структури, розрахувавши ланцюгові і базисні темпи зростання питомої ваги кожного джерела доходів.

10. Грошові доходи населення однієї з країн в абсолютному вимірі за 2003-2008 рр. складали:

Роки	2003	2004	2005	2006	2007	2008
Грошові доходи, млн. грн.	383,2	830,5	7099,9	79949,0	364834,1	947770,5

Розрахуйте середні питомі ваги окремих джерел грошових доходів за 2003-2005 рр. і 2005-2008 рр., використовуючи дані таблиці і задачі **9**.

11. За даними задачі **9** розрахуйте лінійні коефіцієнти "абсолютних" структурних зрушень для кожного року, починаючи з 2005 р. У які роки структура грошових доходів населення зазнала найбільших і найменших змін?

12. Виробництво легкових автомобілів у одному з регіонів характеризувалося такими даними:

Виробник	Умовні роки			
	0 рік	1 рік	2 рік	3 рік
№ 1	101,4	95,8	67,9	40,6
№ 2	673,8	660,3	530,9	609,2
№ 3	69,0	105,7	118,2	118,7
№ 4	54,4	31,3	21,9	12,8
№ 5	7,5	5,3	6,1	8,6
№ 6	54,3	57,6	53,2	44,9
Усього	960,4	956,0	798,2	834,8

Порівняйте зміну структури виробництва автомобілів у 1 році відносно 0 року і в 3 році відносно 2 року, використовуючи квадратичний коефіцієнт відносних структурних зрушень.

13. Порівняйте розподіл грошових доходів населення за базисний і звітний роки за допомогою коефіцієнта Джині за такими даними:

20-процентні групи населення	Частка у сукупних доходах, %	
	Базисний рік	Звітний рік
Перша (з найменшими доходами)	5,3	5,5
Друга	10,2	10,2
Третя	15,2	15,0
Четверта	23,0	22,4
П'ята (з найбільшими доходами)	46,3	46,9
Усього	100,0	100,0

14. За даними бюджетних обстежень отримано такий розподіл населення однієї з областей України за рівнем доходів:

10-процентні групи населення	Частка у сукупних доходах, %
1 (з найменшими доходами)	2,7
2	4,6
3	6,3
4	8,4
5	9,8
6	11,5
7	12,9
8	13,3
9	14,6
10 (з найбільшими доходами)	15,9
Усього	100,0

Оцініть диференціацію доходів населення, використовуючи коефіцієнт Джині.

15. За даними задачі **12** зробіть зведене оцінювання зміни структури виробництва автомобілів на автозаводах у цілому за 4 роки, використовуючи лінійний коефіцієнт “абсолютних” структурних зрушень за *n* періодів.

16. На основі коефіцієнта Лоренца визначте ступінь концентрації безробітного населення у районах однієї з областей України:

Район	Питома вага населення району, %	
	У чисельності працездатного населення області	У чисельності безробітного населення області
1	12,4	8,9
2	18,0	24,3
3	6,2	5,9
4	9,4	5,4
5	16,8	19,2
6	13,5	23,9
7	23,7	12,4
Усього	100,0	100,0

17. За варіантами, наведеними у таблиці А, розрахувати частинні показники структурних зрушень; узагальнені показники структурних зрушень; показники концентрації та централізації.

Дати економічну інтерпретацію розрахованих показників:

варіант 1 – дані про розмір заробітної плати робітників комплексної бригади, грош. од.;
варіант 2 – дані про розмір капіталу комерційних банків, грош. од.;
варіант 3 – дані про розмір основних фондів підприємств, грош. од.;
варіант 4 – дані про чисельність працівників на підприємствах, чол.

Таблиця А – Дані у динаміці (за варіантами)

1варіант			2 варіант			3 варіант			4 варіант		
Умовні роки			*Умовні роки*			*Умовні роки*			*Умовні роки*		
0	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*
700	657	754	684	804	784	900	698	548	217	568	968
600	325	517	454	684	895	654	478	984	881	784	754
650	884	615	514	390	451	326	587	214	254	856	214
575	217	189	298	248	562	659	657	985	635	587	950
900	881	951	815	588	230	245	248	1102	474	935	320
920	254	750	507	986	124	578	935	687	366	684	754
830	635	862	387	573	325	956	487	956	956	247	517
711	474	700	806	684	658	845	351	548	548	1050	615
815	366	850	741	210	214	512	267	368	368	605	189
520	987	480	952	657	325	623	483	478	478	789	951
560	541	650	842	851	658	784	248	320	711	654	750
575	269	755	624	687	985	895	975	587	815	357	862
595	558	955	1248	324	745	653	620	951	520	245	700
630	654	360	368	806	154	245	820	623	268	1450	548
710	456	620	258	342	302	205	348	847	758	963	984
740	265	792	754	740	658	606	500	753	698	852	214
555	587	665	963	846	951	879	384	865	630	741	985
480	584	410	954	954	842	328	784	745	745	502	1102
390	369	763	458	1025	953	987	504	896	896	637	687
395	587	357	327	1179	865	586	685	324	324	437	956
470	457	458	475	1258	862	268	359	875	875	982	987
490	148	857	684	349	847	758	749	985	325	327	541
380	568	258	247	745	984	698	624	412	658	1367	269
500	784	269	985	210	652	473	1287	385	985	655	558
350	856	258	605	326	326	157	987	458	745	873	654
300	587	369	347	852	968	116	1147	125	154	953	456
888	935	879	750	548	754	998	889	563	302	209	265
905	457	655	1124	626	214	1245	685	325	658	330	587
			1250	887	950	368	658	569	587	1456	584
			879	953	320	1300	214	905			

18. За даними задачі **13** розрахуйте коефіцієнти Лоренца за базисний і звітний роки. Порівняйте результати розрахунків зі зробленими раніше висновками.

19. Забезпеченість населення умовного районну житлом характеризується такими даними:

Групи населення за рівнем житлових умов ($м^2$ загальної площі на людину)	Чисельність населення, тис. чол.	Загальна площа, тис. $м^2$
До 10	27,3	218,4
10–20	48,0	768,2
20–30	96,4	2113,6
30–40	32,3	1130,5
40 і більше	8,5	357,0

Визначте ступінь розшарування населення за рівнем житлових умов.

20. Використовуючи наведене нижче групування, розрахуйте показники централізації виробництва нафтопродуктів у двох областях і проаналізуйте отримані результати:

Групи підприємств за обсягом виробництва, тис. т.	Область А		Область Б	
	Кількість підприємств	Вироблено, тис. т	Кількість підприємств	Вироблено, тис. т
До 50	1	38,6	2	51,7
50–100	2	182,1	4	285,0
100 і більше	2	215,6	1	270,3

ГЛАВА 11

ІНДЕКСИ

11.1
МЕТОДИЧНІ ВКАЗІВКИ

У цій главі розглядається сутність класифікації та властивості індексів; пояснюються принципи їх побудови.

Вивчення цієї теми має ґрунтуватися на знанні попередніх розділів курсу, особливо теми “Статистичні показники” і “Статистичне вивчення динаміки соціально-економічних явищ (рядів динаміки)”.

Індекс (лат. index) – у перекладі з латинської означає “покажчик”, “показник”. *Індекс* – це відносний показник, що характеризує зміну рівня явища в часу, у просторі та у порівнянні з будь-яким еталоном. Індекс вимірюється в процентах або у коефіцієнтах.

Основним елементом індексного співвідношення є індексована величина. *Індексована величина* – це значення ознаки статистичної сукупності, зміна якої є об’єктом вивчення.

Завдання, вирішувані за допомогою індексів

1. Вимірювання змін складних явищ, тобто виконання синтетичної функції.
2. Визначення впливу окремих факторів на зміну динаміки складного явища, тобто виконання аналітичної функції.
3. Виконання порівнянь не тільки з минулим періодом, а й з іншою територією, а також з нормативами, планами, еталонами тощо.

Класифікація індексів

1. За змістом об’єктів, що вивчаються:
   а) індекси якісних показників (якісні показники характеризують рівень явища у розрахунку на ту чи іншу одиницю сукупності – наприклад: ціна одиниці продукції, виробіток робітника за одиницю часу, заробітна плата одного робітника тощо);
   б) індекси якісних показників (кількісні показники характеризують загальний розмір (або обсяг) того чи іншого явища – наприклад: фізичний обсяг промислової і сільськогосподарської продукції, фізичний обсяг роздрібного товарообороту, чисельність працівників, вартість фондів тощо).
2. За ступенем охоплення елементів сукупності:
   а) індивідуальні;
   б) загальні.
3. За методами розрахунку загальних індексів:
   а) агрегатні;
   б) середні.

Показники індексів (позначки індексованих величин)**:**

q – кількість, обсяг якогось продукту;
p – ціна одиниці продукції;
z – собівартість продукції (або витрати на одиницю часу);
t – витрати часу на виробництво одиниці продукції;

w –виробіток продукції;
pq – загальна вартість виробленої або проданої продукції певного виду (товарооборот, виручка);
zq – витрати на виробництво всієї продукції (витрати виробництва).

Індивідуальні індекси. Найпростішим показником, використовуваним в індексному аналізі, є індивідуальний індекс, який дає порівняльну характеристику окремих елементів сукупності і виражається співвідношенням цих елементів:

$$i_q = \frac{q_1}{q_0} \text{ – індивідуальний індекс обсягу;} \quad \textbf{(11.1.1)}$$

$$i_p = \frac{p_1}{p_0} \text{ – індивідуальний індекс ціни;} \quad \textbf{(11.1.2)}$$

$$i_z = \frac{z_1}{z_0} \text{ – індивідуальний індекс собівартості;} \quad \textbf{(11.1.3 а)}$$

$$i_z = \frac{z_A}{z_B} \text{ – (на різних підприємствах);} \quad \textbf{(11.1.3 б)}$$

$$i_w = \frac{w_1}{w_0} \text{ – індивідуальний індекс виробітку;} \quad \textbf{(11.1.4)}$$

$$i_t = \frac{t_1}{t_0} \text{ – індивідуальний індекс трудомісткості;} \quad \textbf{(11.1.5)}$$

і таке інше.

Підстрочний символ 1 вказує на величину звітного періоду; підстрочний символ 0 вказує на величину базисного періоду

Індивідуальні індекси зазвичай є відносними величинами динаміки. Тому вони можуть бути ланцюговими і базисними. Тобто між ними існує такий самий зв'язок, як і між ланцюговими і базисними темпами зростання (див. главу 9).

Обов’язковою умовою обчислення індивідуальних індексів є однорідність продукції та зіставність цін.

Приклад 11.1.1. У ІІІ кварталі ціна 1 л молока склала 4 грош.од.; у IV кварталі – 5 грош. од. Знайти індивідуальній індекс цін.

<u>Розв’язок</u>

Використовуючи формулу 11.1.2, отримуємо:

$i_p = \frac{5}{4} = 1{,}25$ (125 %) – ціна на молоко у IV кварталі зросла на 25 %.

Загальні (зведені**) індекси** відображають зміну всіх елементів складного явища (під складним явищем розуміють таку статистичну сукупність, окремі елементи якої безпосередньо не підлягають сумуванню – наприклад, фізичний обсяг продукції, яка включає різнойменні товари; ціни на різні групи продуктів і таке інше).

Методика розрахунку загальних індексів є складнішою, ніж індивідуальних, і залежить від характеру індексованих величин, а також від наявності вихідних даних і мети дослідження.

Загальні індекси поділяються на:
1) агрегатні;
2) середні.

Агрегатний індекс **–** це основна і найбільш поширена форма індексу. Це індекс, чисельник і знаменник якого є набором (агрегатом), інакше кажучи, сумою добутку індексованих величин на ваги індексів. Наприклад:

$$Ipq = \frac{\sum q_1 p_1}{\sum q_0 p_0}$$ – загальний індекс товарообороту, (**11.1.6**)

де чисельник – вартість продукції поточного (звітного) періоду в поточних цінах;

знаменник –вартість продукції базисного періоду в базисних цінах;

$$Izq = \frac{\sum z_1 q_1}{\sum z_0 q_0}$$ – загальний індекс витрат на виробництво,(**11.1.7**)

де чисельник – витрати на виробництво у поточному (звітному) періоді;

знаменник – витрати на виробництво у базисному періоді.

У якості спеціальних прийомів індексного методу через несумірність і неоднорідність об’єктів використовують коефіцієнти–сумірники, основою для яких є такі показники, як p, q, z, t.

Агрегатний індекс кількісного показника – фізичного обсягу (товарообороту), при усуненні (елімінуванні) впливу іншого фактора (ціни):

$$Iq = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$ – формула Ласпейреса, (**11.1.8**)

де p_0 – базисна (фіксована) ціна одиниці товару;

чисельник – умовна вартість продукції у звітному періоді, реалізованої за базисними цінами;

знаменник – вартість продукції у базисному періоді,

або:

$$Iq = \frac{\sum q_1 p_1}{\sum q_0 p_1}$$ – формула Пааше, (**11.1.9**)

де p_1 – звітна (фіксована) ціна одиниці товару;

чисельник – вартість продукції у звітному періоді;

знаменник – умовна вартість продукції у базисному періоді, реалізованої за цінами звітного періоду.

Прикладом агрегатного індексу кількісного показника може також бути загальний індекс фізичного обсягу (зважений за собівартістю), побудований за принципом Ласпейреса:

$$Iq = \frac{\sum q_1 z_0}{\sum q_0 z_0}, \qquad \textbf{(11.1.10)}$$

де чисельник – умовні витрати на виробництво продукції в звітному періоді за збереження її собівартості на базисному рівні.

Агрегатний індекс якісного показника – ціни при елімінуванні впливу іншого фактора (фізичного обсягу):

$$Ip = \frac{\sum p_1 q_1}{\sum p_0 q_1} \text{ – формула Пааше,} \qquad \textbf{(11.1.11)}$$

або:

$$Ip = \frac{\sum p_1 q_0}{\sum p_0 q_0} \text{ – формула Ласпейреса.} \qquad \textbf{(11.1.12)}$$

Економічний зміст індексів цін Ласпейреса і Пааше є різним:

- індекс цін Пааше показує, у скільки разів зріс (зменшився) у середньому рівень ціни на масу товару, реалізовану в звітному періоді, тобто фактичну економію (перевитрати) від зміни цін;
- індекс цін Ласпейреса показує, у скільки разів товари базисного періоду подорожчали (подешевшали) через зміни цін на них у звітному періоді (цей індекс застосовується у період інфляції або економічної кризи для обсягів продукції, які випали зі споживання через надмірне підвищення цін).

Прикладом агрегатного індексу якісного показника може також бути загальний індекс собівартості, побудований за принципом Пааше:

$$Iz = \frac{\sum z_1 q_1}{\sum z_0 q_1}. \qquad \textbf{(11.1.13)}$$

У вітчизняній статистиці зазвичай використовують таке правило побудови загальних індексів: індекси кількісних показників будуються за вагами якісних показників базисного періоду; індекси якісних показників будуються за вагами кількісних показників звітного періоду.

Керуючись цим правилом, при побудові індексів кількісних показників необхідно використовувати формулу Ласпейреса, а при побудові індексів якісних показників – формулу Пааше.

За допомогою індексів також розраховують абсолютні зміни показників. Для цього з чисельника відповідного індексу віднімають його знаменник:

- абсолютна зміна товарообороту:

$$\Delta_{qp} = \sum q_1 p_1 - \sum q_0 p_0 \text{ (грош.од.);} \qquad \textbf{(11.1.14)}$$

- абсолютна зміна фізичного обсягу:

$$\Delta_q = \sum q_1 p_0 - \sum q_0 p_0 \text{, (грош.од.);} \qquad \textbf{(11.1.15)}$$

- абсолютна економія (перевитрати) грошових коштів покупців:

$$\Delta_p = \sum p_1 q_1 - \sum p_0 q_1 \text{ (грош.од.),} \qquad \textbf{(11.1.16)}$$

де (+) результат характеризує абсолютні перевитрати грошових коштів покупців;

(–) результат характеризує абсолютну економію грошових коштів покупців.

Цей принцип застосовується і для індивідуальних індексів.

Якщо величини пов'язані одна з одною як множники у добутку, тобто мультиплікативно:

$$qp = q \times p, \qquad \textbf{(11.1.17 а)}$$

$$qz = q \times z, \qquad \textbf{(11.1.17 б)}$$

то в такому ж самому зв'язку знаходяться і індекси цих показників:

$$Ipq = Ip \times Iq\ , \qquad \textbf{(11.1.18 а)}$$

$$Izq = Iz \times Iq\ . \qquad \textbf{(11.1.18 б)}$$

Адитивний зв'язок при цьому є таким:

$$\Delta_{qp} = \Delta_q + \Delta_p\ , \qquad \textbf{(11.1.19 а)}$$

$$\Delta_{qz} = \Delta_q + \Delta_z\ , \qquad \textbf{(11.1.19 б)}$$

Найважливішою особливістю загальних індексів, побудова і розрахунок яких складають сутність індексного методу, є те, що вони мають синтетичні та аналітичні властивості. Синтетичні властивості загальних індексів полягають у тому, що вони відображають відносні зміни складних (різнотоварних) явищ, а їхні аналітичні властивості – у тому, що за допомогою індексного методу можна виявити вплив факторів на зміну досліджуваного показника.

Середньоарифметичні та середньогармонічні індекси. Будь-які загальні індекси можуть бути побудовані двома способами: як агрегатні і як середні з індивідуальних. Останні, у свою чергу, поділяються на середні арифметичні і середні гармонічні. Такі середньозважені індекси обчислюють тоді, коли наявна інформація не дозволяє розрахувати агрегатний індекс. При виборі ваг слід мати на увазі, що середній індекс повинен бути тотожнім агрегатному, який є основною формою індексу.

Приклад 11.1.2. Перетворення агрегатного індексу фізичного обсягу у середньоарифметичний індекс (табл. 11.1.1):

Таблиця 11.1.1 – Вихідні дані

Товарна група	Продано у I кварталі, тис. грн. $q_0 p_0$	Зміна кількості проданих товарів у II кварталі порівняно з I, % $i_q = \frac{q_1}{q_0}$
Картопля	320	–5 (95 %)
Борошно	350	+10 (110 %)

Знайти: загальний індекс фізичного обсягу.

Розв'язок

$Iq = \frac{\sum q_1 p_0}{\sum q_0 p_0}$ – агрегатна форма загального індексу.

Оскільки $q_1 = i_q \times q_0$,

то:

$$Iq = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}. \qquad \textbf{(11.1.20)}$$

(Формула 11.1.20 аналогічна формулі середньої арифметичної зваженої $\bar{X} = \frac{\sum Xf}{\sum f}$ з вагами $q_0 p_0$).

Тоді: $I_q = \frac{0{,}95 \times 320 + 1{,}1 \times 350}{320 + 350} = 1{,}028\ (102{,}8\%)$.

Приклад 11.1.3. Перетворення агрегатного індексу цін у середньогармонічний індекс (табл. 11.1.2)

Знайти загальний індекс цін.

Розв'язок

$Ip = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ – агрегатна форма загального індексу.

Оскільки $p_0 = \frac{p_1}{i_p}$,

то: $$Ip = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}. \qquad \textbf{(11.1.21)}$$

(Формула 11.1.21 аналогічна формулі середньої гармонічної зваженої $\overline{X} = \frac{\sum Z}{\sum \frac{Z}{X}}$).

Тоді: $I_p = \frac{470 + 650}{\frac{470}{1{,}04} + \frac{650}{0{,}94}} = 0{,}98\ (98\%)$.

Таблиця 11.1.2 – Вихідні дані

Товарна група	Продано у II кварталі, тис. грн. $q_1 p_1$	Зміна цін у II кварталі порівняно з I, % $i_p = \frac{p_1}{p_0}$
Шерсть	470	+4 (104 %)
Шовк	650	–6 (94 %)

Приклад 11.1.4. Є такі дані про товарооборот магазину (табл. 11.1.3). Обчислити індивідуальний і загальний індекси цін.

Таблиця 11.1.3 – Вихідні дані

Товарна продукція	Продано товарів у фактичних цінах, тис. грн.		Середня зміна цін у III кварталі порівняно з I кварталом, %
	I квартал	III квартал	
Овочі різні	540,0	550,0	–10
Фрукти різні	422,5	630,0	+5
Кондитерські вироби	336,0	348,5	Без змін

Розв’язок

Індивідуальні індекси цін:

1) індекси цін для овочів: $ip = 100 - 10 = 90\%$;

2) індекси цін для фруктів: $ip = 100 + 5 = 105\%$;

3) індекси цін для кондитерських виробів: $ip = 100 + 0 = 100\%$.

Загальний індекс розраховуємо за формулою (11.1.21):

$$Ip = \frac{1528{,}5}{\frac{550{,}0}{0{,}90} + \frac{630{,}0}{1{,}05} + \frac{348{,}5}{1{,}00}} = 0{,}98 \ (98\%).$$

Таблиця 11.1.4 – Розрахункова

Товарна продукція	Продано товарів у фактичних цінах, тис. грн.		Індивідуальний індекс цін, %
	I квартал	III квартал	
A	$p_0 q_0$	$p_1 q_1$	i_p
Овочі різні	540,0	550,0	90
Фрукти різні	422,5	630,0	105
Кондитерські вироби	336,0	348,5	100
Усього	X	1528,5	X

Висновок: товарооборот магазину внаслідок загального зниження цін на товарну продукцію скоротився на 2 % (100 % – 98 %).

Індекси середніх величин. На зміну середнього рівня впливають два фактора (див. главу 6):

1) зміна самих варіантів;

2) зміна питомої ваги або частки окремих варіантів у сукупності. Наприклад, середня продуктивність праці на підприємстві може зрости внаслідок її підвищення у окремих робітників, а також внаслідок збільшення частки робітників з більш високою продуктивністю праці у загальній чисельності робітників, що виробляють однойменну продукцію.

Ці зміни вивчають за допомогою індексів змінного складу, постійного (фіксованого) складу і структурних зрушень.

Індекс змінного складу – відображає вплив обох факторів на зміну середньої величини.

$$I_{ЗМ.С.} = \frac{\overline{X_1}}{\overline{X_0}} = \frac{\sum X_1 f_1}{\sum f_1} \div \frac{\sum X_0 f_0}{\sum f_0}. \qquad \textbf{(11.1.22)}$$

Індекс постійного (фіксованого) складу – показує, якою мірою зміна середньої величина обумовлена зміною окремих варіантів якісного показника за незмінної структури сукупності.

$$I_{ПОСТ.С.} = \frac{\sum X_1 f_1}{\sum X_0 f_1} = \frac{\sum X_1 f_1}{\sum f_1} \div \frac{\sum X_0 f_1}{\sum f_1}. \qquad \textbf{(11.1.23)}$$

Індекс структурних зрушень – характеризує зміну середньої в результаті зміни розподілу одиниць усередині сукупності, тобто в результаті змін у структурі сукупності.

$$I_{СТР.ЗРУШ.} = \frac{\sum X_0 f_1}{\sum f_1} \div \frac{\sum X_0 f_0}{\sum f_0}. \qquad \textbf{(11.1.24)}$$

Між цими індексами існує взаємозв'язок:

$$I_{ЗМ.С.} = I_{ПОСТ.С.} \times I_{СТР.ЗРУШ.} \qquad \textbf{(11.1.25)}$$

Приклад 11.1.5. Є такі дані (табл.11.1.5).

Таблиця 11.1.5– Вихідні дані

Товар	Звітний період		Базисний період		Індивідуальні індекси, %	
	Ціна за 1 кг, умов. од.	Кількість, кг	Ціна за 1 кг, умов. од.	Кількість, кг	Ціни	Фізичний обсяг реалізації
1	1,51	?	1,74	270,8	?	112,5
2	0,72	?	0,83	131,6	?	105,7
3	?	314,6	1,37	?	97,1	125,9

Визначте:
1) показники, яких бракує у таблиці;
2) зведені індекси цін, фізичного обсягу реалізації і товарообороту;
3) індекси змінного складу, фіксованого складу і структурних зрушень.

Розв'язок

1. Визначимо показники, яких бракує у таблиці, за допомогою формул 11.1.1 і 11.1.2.

Таблиця 11.1.6 – Розрахункова

Товар	Звітний період		Базисний період		Індивідуальні індекси, %	
	Ціна за 1 кг, умов.од. (p_1)	Кількість, кг (q_1)	Ціна за 1 кг, умов.од. (p_0)	Кількість, кг (q_0)	Ціни, (i_p)	Фізичний обсяг реалізації (i_q)
1	1,51	304,7	1,74	270,8	86,8	112,5
2	0,72	139,1	0,83	131,6	86,7	105,7
3	1,33	314,6	1,37	249,9	97,1	125,9

Оскільки $i_p = \frac{p_1}{p_0}$, то $p_1 = i_p \times p_0$, тому для товару "3" ціна за звітний період розраховується:

$$p_3 = 0{,}971 \times 1{,}37 = 1{,}33 \text{ (умов.од.)}.$$

Оскільки $i_q = \frac{q_1}{q_0}$, то $q_0 = \frac{q_1}{i_q}$, тому кількість товару “3” у базисному періоду розраховується:

$$q_3 = \frac{314,6}{1,259} = 249,88 \text{ (умов.од.).}$$

Аналогічно розраховуються всі інші невідомі показники.

2. Загальний індекс цін розраховуємо за формулою 11.1.11:

$$Ip = \frac{1,51\times 304,7 + 0,72\times 139,1 + 1,33\times 314,6}{1,74\times 304,7 + 0,83\times 139,1 + 1,37\times 314,6} = \frac{978,67}{1076,63} = 0,909 \ (90,9\%).$$

Загальний індекс фізичного обсягу визначимо за формулою 11.1.8:

$$Iq = \frac{304,7\times 1,74 + 139,1\times 0,83 + 314,6\times 1,37}{270,8\times 1,74 + 0,83\times 131,6 + 249,9\times 1,37} = \frac{1076,63}{922,78} = 1,167 (116,7\%).$$

Загальний індекс товарообороту визначимо за формулою 11.1.6:

$$Ipq = \frac{1,51\times 304,7 + 0,72\times 139,1 + 1,33\times 314,6}{1,74\times 270,8 + 0,83\times 131,6 + 1,37\times 249,9} = \frac{978,67}{922,78} = 1,061 \ \ (106,1\%).$$

Висновок: товарооборот підприємства зріс у звітному періоду порівняно з базисним на 6,1 % (106,1 % – 100 %), і це відбулося лише внаслідок зростання обсягу продажів у середньому на 16,7 %, а загальне зниження цін на 9,1% (100 % – 90,9 %) негативно позначилося на товарообороті трьох видів товару.

3. Індекс постійного (фіксованого) складу розрахуємо за формулою 11.1.23:

$$I_{ПОСТ.С} = \frac{1,51\times 304,7 + 0,72\times 139,1 + 1,33\times 314,6}{1,74\times 304,7 + 0,83\times 139,1 + 1,37\times 314,6} = \frac{978,67}{1076,63} = 0,909 \ (90,9\%).$$

Індекс структурних зрушень розрахуємо за формулою 11.1.24:

$$I_{СТР.ЗРУШ.} = \frac{1,74\times 304,7 + 0,83\times 139,1 + 1,37\times 314,6}{304,7 + 139,1 + 314,6} \div \frac{1,74\times 270,8 + 0,83\times 131,6 + 1,37\times 249,9}{270,8 + 131,6 + 249,9};$$

$$I_{СТР.ЗРУШ.} = \frac{1076,63}{758,40} \div \frac{922,78}{652,30} = 1,0035 \ (100,35\ \%).$$

Індекс змінного складу визначимо на підставі взаємозв’язку індексів, формула 11.1.25):

$$I_{ЗМ.С.} = 0,909\times 1,0035 = 0,9122 \ (91,22\%).$$

Висновок: у результаті сукупного впливу двох факторів, тобто ціни і кількості товару, середня ціна реалізації скоротилася у звітному періоді порівняно з базисним на 8,78 % (100 % – 91,22 %). Тільки внаслідок зниження цін на всі види товарів середня ціна реалізації зменшилася на 9,1 %, а внаслідок зміни структури продажів середня ціна продажу трохи збільшилась – на 0,35 %.

Системи індексів. Часто в ході економічного аналізу зміну індексованих величин вивчають не за два, а за кілька послідовних періодів. Тому виникає необхідність побудови індексних систем.

Залежно від бази порівняння індекси бувають:

1) базисні;

2) ланцюгові.

***Для індивідуальних індексів*:**

Позначивши чотири послідовних періоди підстрочними значеннями 0, 1, 2, 3, обчислюємо:

базисні індекси: $i_{p1/0} = \frac{p_1}{p_0}$, $i_{p2/0} = \frac{p_2}{p_0}$, $i_{p3/0} = \frac{p_3}{p_0}$;

ланцюгові індекси: $i_{p1/0} = \frac{p_1}{p_0}$, $i_{p2/1} = \frac{p_2}{p_1}$, $i_{p3/2} = \frac{p_3}{p_2}$.

Взаємозв'язок: $i_{p3/2} = i_{p3/0} \div i_{p2/0}$.

Для агрегатних індексів (наприклад, для тих, які використовуються у вітчизняній статистиці)**:**

базисні індекси (зі змінними вагами):

$$Ip_{1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \quad Ip_{2/0} = \frac{\sum p_2 q_2}{\sum p_0 q_2}, \quad Ip_{3/0} = \frac{\sum p_3 q_3}{\sum p_0 q_3};$$

базисні індекси (з постійними вагами):

$$Iq_{1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \quad Iq_{2/0} = \frac{\sum q_2 p_0}{\sum q_0 p_0}, \quad Iq_{3/0} = \frac{\sum q_3 p_0}{\sum q_0 p_0};$$

ланцюгові індекси (зі змінними вагами):

$$Ip_{1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \quad Ip_{2/1} = \frac{\sum p_2 q_2}{\sum p_1 q_2}, \quad Ip_{3/2} = \frac{\sum p_3 q_3}{\sum p_2 q_3};$$

ланцюгові індекси (з постійними вагами):

$$Iq_{1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \quad Iq_{2/1} = \frac{\sum q_2 p_0}{\sum q_1 p_0}, \quad Iq_{3/2} = \frac{\sum q_3 p_0}{\sum q_2 p_0}.$$

Ряди агрегатних індексів з постійними вагами мають перевагу – зберігається взаємозв'язок між ланцюговими і базисними індексами. Наприклад:

$\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_2 p_0}{\sum q_1 p_0} \times \frac{\sum q_3 p_0}{\sum q_2 p_0} = \frac{\sum q_3 p_0}{\sum q_0 p_0}$ – ряд агрегатних індексів фізичного обсягу.

У рядах агрегатних індексів якісних показників, які будуються зі змінними вагами (наприклад, ряд цін Пааше), перемноження ланцюгових індексів не дає базисний індекс, тобто:

$$\frac{\sum p_1q_1}{\sum p_0q_1} \times \frac{\sum p_2q_2}{\sum p_1q_2} \times \frac{\sum p_3q_3}{\sum p_2q_3} \neq \frac{\sum p_3q_3}{\sum p_0q_3}.$$

Втім у статистичній практиці часто виникає необхідність визначення динаміки цін за тривалий період часу на основі ланцюгових індексів цін зі змінними вагами. Тоді для отримання приблизного базисного (підсумкового) індексу ланцюгові індекси цін перемножують, заздалегідь знаючи, що у такому розрахунку допускаються помилки.

11.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. Дайте визначення індексу в статистиці.
2. Які задачі розв'язують за допомогою індексів?
3. Що характеризують індивідуальні індекси? Наведіть приклади.
4. У чому полягає сутність загальних індексів?
5. Для чого необхідно ділення на індекси кількісних і якісних показників і яка система зважування прийнята в теорії індексів?
6. Як обчислюється агрегатний індекс вартості продукції (товарообороту)? Напишіть формулу.
7. Коли виникає потреба перетворення індексу фізичного обсягу в середній арифметичний і середній гармонічний індекс? Яким чином відбуваються такі перетворення?
8. Як обчислюють агрегатні індекси цін і собівартості (Пааше і Ласпейреса)? Напишіть їхні формули.
9. Коли виникає необхідність перетворення агрегатного індексу цін в середній арифметичний і середній гармонічний індекс? Покажіть на прикладі.
10. Який варіант агрегатних індексів якісних показників використовують при розрахунку індексу споживчих цін і чому?
11. Який індекс називається індексом змінного складу, як він обчислюється і що характеризує?
12. Який індекс називається індексом постійного (фіксованого) складу, як він обчислюється і що характеризує?
13. Що характеризує індекс структурних зрушень і як він обчислюється?
14. Який взаємозв'язок існує між індексами змінного, постійного складу і структурних зрушень?
15. Як будуються базисні і ланцюгові індекси і який зв'язок існує між ними?
16. За даними про ціни на яблука визначте показники, яких бракує у таблиці.

Місяць	Ціна за 1кг., грош.од.	Індивідуальні індекси цін	
		Ланцюгові	Базисні
Липень	?	-	1,0
Серпень	0,61	?	?
Вересень	?	0,63	0,47

17. У чому виражається взаємозв'язок між індексами цін, фізичного обсягу і товарообороту?

18. Є такі дані:

товар	Звітний період		Базисний період		Індивідуальні індекси, %	
	ціна за 1 кг, грош.од.	кількість, ц	ціна за 1 кг, грош.од.	кількість, ц	ціна за 1 кг, грош.од.	Фізичний обсяг реалізації
1	1,61	?	1,84	270,8	?	112,5
2	0,82	?	0,83	131,6	?	105,7
3	?	314,6	1,37	?	96,8	125,9

Визначте:

1) показники, яких бракує у таблиці;

2) загальні індекси цін, фізичного обсягу і товарообороту.

19. Є такі дані:

Вид Станка	Собівартість виробництва 1шт. продукції, тис. грн.		Вироблено продукції, тис.шт.		Індивідуальні індекси, %	
	базисний період	звітний період	базисний період	звітний період	собівартість	фізичний обсяг
1	1,32	?	1,27	1,31	0,92	?
2	2,39	2,46	?	2,96	?	0,71
3	?	3,07	0,72	?	0,99	2,92

Визначте:

1) показники, яких бракує у таблиці;

2) загальні індекси цін, фізичного обсягу виробленої продукції і витрат на виробництво.

20. Є дані про реалізацію овочів і ціни.

На підставі наведених даних визначити:

1) індивідуальні індекси цін і фізичного обсягу;

2) загальний індекс цін;

3) загальний індекс фізичного обсягу;
4) загальну суму економічного ефекту, яку отримало населення при купівлі цих продуктів у звітному періоді за зміненими цінами.
Показати взаємозв'язок між індексами, обчисленими в пунктах 2 і 3.

Овочі	Базисний період		Звітний період	
	кількість, т	середня ціна за 1кг, грн.	кількість, т	середня ціна за 1кг, грн.
Картопля	130	0,30	220	0,16
Капуста	440	0,27	570	0,20
Цибуля	160	0,30	240	0,25

21. Є такі дані про товарооборот магазину:

Товарна продукція	Продано товарів у фактичних цінах, тис. грн.		Середня зміна цін у II кварталі порівняно з I кварталом, %
	I квартал	II квартал	
Молочні продукти	540,0	550,0	–10
Овочі	422,5	630,0	+5
Кондитерські вироби	336,0	348,5	без змін

Обчислити індивідуальний і загальний індексі цін.
22. На швейній фабриці є дані про витрати на виробництво продукції та зміну її собівартості:

Вид виробів	Загальні витрати на виробництво виробів, тис. грн.		Зниження (–) або збільшення (+) собівартості виробу в звітному періоді порівняно з базисним, %
	Базисний період	Звітний період	
Блузи	2700	2780	+8
Спідниці	3100	3120	без змін
Піджаки	5800	6600	–8

Обчислити:
1) індивідуальні індекси собівартості і продукції;
2) загальні індекси собівартості, продукції і витрат на виробництво продукції;
3) на підставі взаємозв'язку індексів обчислити індекс фізичного обсягу продукції;

4) обчислити суму економії, отриману в результаті зниження собівартості.
23. Є такі дані про реалізацію фруктів на ринках міста:

Види товару	Товарооборот у фактичних цінах, тис. грн.		Процент зміни кількості товарів у звітному періоді порівняно з базисним, %
	Базисний період	Звітний період	
Айва	350	440	–10
Груші	300	320	+2
Абрикоси	670	700	Без змін

Обчислити:
1) індивідуальні індекси фізичного обсягу товарообороту;
2) загальні індекси фізичного обсягу товарообороту у зіставних цінах і товарообороту в фактичних цінах;
3) загальний індекс цін на підставі взаємозв’язку індексів.

24. Є такі дані:

Вид виробу	Трудомісткість, чол.-год.		Випуск продукції, тис. шт.	
	Базисний період	Звітний період	Базисний період	Звітний період
1	1,56	1,43	98,9	102,1
2	1,97	1,96	20,1	22,6
3	1,98	1,19	300,7	294,5

Обчисліть індекси:
1) продуктивності праці;
2) трудових витрат.
25. Є дані про товарооборот овочів на сільськогосподарському ринку:

Товар	Товарооборот у цінах відповідного періоду, тис. грн.	
	Серпень	Вересень
Морква	5,1	11,0
Буряк	0,77	2,16
Картопля	0,72	5,61

Обчисліть середній процент зміни цін, якщо відомо, що індекс фізичного обсягу реалізації цих товарів склав 213 %.

26. Є такі дані про продаж трикотажних виробів:

Вид виробу	Стара ціна грн. за 1 шт.	Нова ціна грн. за 1 шт.	Товарооборот після зниження цін, тис. грн.
1	19,7	16,0	1,42
2	21,0	17,9	2,95
3	30,5	28,6	1,81

Оцініть:
1) середнє зниження цін на цю групу товарів;
2) розмір економії покупців від зниження цін.

27. Середнє зниження на групу товарів у липні порівняно з червнем склало 10,5 %, а у серпні порівняно з липнем – 12 %. Визначте, як змінився фізичний обсяг реалізації товарів у серпні порівняно з червнем, якщо товарооборот за цей період зріс у 2,1 рази (середня зміна цін визначалась за допомогою ланцюгових індексів з вагами серпня).

28. Є дані для сільськогосподарського ринку:

Товар	Товарооборот за серпень, тис. грн.	Зниження цін у серпні порівняно з червнем, %
Капуста	56,1	37,2
Цибуля	51,2	37,7
Буряк	109,7	32,6

Визначте:
1) індекс цін;
2) індекс фізичного обсягу реалізації з урахуванням того, що товарооборот у серпні зріс на 42 % порівняно з червнем.

29. Сума витрат на виробництво продукції у 2007 р. склала 48,3 млн. грн. У 2008 р. ці витрати зросли до 50,2 млн. грн. Собівартість одиниці продукції в 2007 р. була на 2,5 % вище її собівартості у 2008 р. Індекс трудових витрат (витрат робочого часу) у цей період склав 0,98. Визначте, на скільки зросла продуктивність праці.

30. Продуктивність видобування вугілля на шахтах регіону зросла у звітному періоді порівняно з базисним на 3,2 %. Витрати робочого часу на виробництво продукції зросли при цьому на 1 %. Визначте, як змінилася собівартість 1 т вугілля, якщо витрати на його виробництво (витрати виробництва) залишились незмінними.

ГЛАВА 12

СТАТИСТИЧНЕ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКУ МІЖ ЯВИЩАМИ

12.1
МЕТОДИЧНІ ВКАЗІВКИ

Дослідження об'єктивних зв'язків між явищами є важливим завданням статистики. Володіння інформацією про показники, що відображають характер і силу цих зв'язків, дозволяє управляти соціально-економічними процесами передбачати їхній розвиток.

Соціально-економічні явища є результатом одночасної дії багатьох причин. При вивчення цих явищ необхідно виявляти головні причини, абстрагуючись від другорядних.

Етапи статистичного вивчення зв'язків

1. Кількісний аналіз явища, пов'язаний з аналізом його природи, методами економічної теорії, соціології, конкретної економіки.
2. Побудова моделі зв'язку (ґрунтується на методах статистики: групування, середніх величин, таблиць тощо).
3. Інтерпретація результатів (пов'язана з якісними особливостями досліджуваного явища).

Серед багатьох форм зв'язків найважливішим є причинний зв'язок, що визначає всі інші форми зв'язку. Сутність причинності полягає в обумовленості одного явища іншим. Для виникнення наслідку потрібні фактори, що його визначають, тобто причина і умова.

Необхідна обумовленість явищ багатьма факторами називається детермінізм. Об'єктами дослідження при статистичному вимірюванні зв'язків зазвичай є детермінованість наслідку факторами (причиною і умовами).

Ознака, що характеризує наслідок, називається результативною (*Y*). Ознаки, що характеризують причини, – факторними (*X*).

Зв'язок між явищами та їхніми ознаками класифікують

1. За характером.
2. За напрямком.
3. За аналітичним виразом (за формою).
4. За кількістю факторів, що взаємодіють.
5. За ступенем щільності зв'язку.

За характером. Між різними явищами та їхніми ознаками необхідно насамперед виділити два типи зв'язків: *функціональний* (жорстко детермінований) і *стохастичний* (стохастично детермінований).

Модель функціонального зв'язку можна представити рівнянням:

$$y_i = f(x_i), \qquad \textbf{(12.1.1)}$$

де y_i – результативна ознака (i=1,…., n);

$f(x_i)$ – відома функція зв'язку між результативною і факторною ознаками;

x_i – факторна ознака.

Стохастичний зв'язок – при якому кожному значенню X відповідають кілька значень Y (наприклад, рівень продуктивності праці робітників стохастично пов'язаний з комплексом факторів: кваліфікацією, стажем, рівнем автоматизації виробництва, здоров'ям робітників, їхнім настроєм тощо).

Модель стохастичного зв'язку можна представити рівнянням:

$$\hat{y}_i = f(x_i) + \varepsilon_i, \qquad \textbf{(12.1.2)}$$

де $\hat{y}_i$ – розрахункове значення результативної ознаки;

$f(x_i)$ – частина результативної ознаки, яка сформувалась під дією врахованих відомих факторних ознак (однієї або декількох), які знаходяться у стохастичному зв'язку з ознакою;

ε_i – частина результативної ознаки, яка сформувалась внаслідок дії неконтрольованих або неврахованих факторів, а також зміни ознак, які неодмінно супроводжуються деякими випадковими помилками.

Окремим випадком стохастичного зв'язку є *кореляційний зв'язок* (взаємозв'язок між випадковими величинами), за якого кожному значенню X відповідає середнє значення Y. (Наприклад, залежність врожайності від якості ґрунту: на врожайність крім якості ґрунту впливають такі невраховані фактори, як погода, якість насіння, рельєф місцевості, якість вбирання врожаю). Кореляційний зв'язок проявляється не в кожному окремому випадку, а загалом у сукупності.

За напрямком. Залежно від напрямку дії функціональні та стохастичні зв'язки можуть бути прямими і оберненими. Прямий зв'язок має місце, коли напрямок зміни результативної ознаки відповідає напрямку зміни ознаки-фактора; обернений зв'язок – коли значення результативної ознаки змінюються у протилежному напрямку порівняно зі зміною факторної ознаки.

За аналітичним виразом (за формою) зв'язки можуть бути:

а) прямолінійні (лінійні):

$$\hat{y}_x = a_0 + a_1 \cdot x \text{ – пряма;} \qquad \textbf{(12.1.3)}$$

б) криволінійні:

$$\hat{y}_x = a_0 + a_1 \cdot x + a_2 \cdot x^2 \text{ – парабола;} \qquad \textbf{(12.1.4)}$$

$$\hat{y}_x = a_0 + a_1 \cdot \frac{1}{x} \text{ – гіпербола;} \qquad \textbf{(12.1.5)}$$

і таке інше.

За кількістю факторів, що взаємодіють:

а) однофакторні (прості) – зазвичай називаються парними, оскільки розглядається пара ознак, тобто Y завжди пов’язана з однією X (наприклад, кореляційний зв'язок між прибутком і продуктивністю праці);

б) багатофакторні (множинні) – коли ознака Y пов’язана з двома і більше ознаками X (наприклад, кореляційний зв'язок між продуктивністю праці і рівнем організації праці, автоматизації виробництва, кваліфікації робітників, виробничим стажем й іншими факторними ознаками).

За ступенем щільності зв’язку (табл. 12.1.1)**:**

Таблиця 12.1.1 – Співвідношення Чеддока

Значення емпіричного кореляційного співвідношення η	Сила зв’язку
0,1–0,3	Слабка
0,3–0,5	Помірна
0,5–0,7	Помітна
0,7–0,9	Щільна
0,9–0,99	надто щільна

При η=0 – зв'язок між ознаками відсутній;
η=1 – зв'язок між ознаками є функціональним.

Найпростіші методи вивчення стохастичних зв’язків

1. Метод зіставлення двох паралельних рядів.
2. Метод аналітичних групувань.
3. Кореляційний аналіз.
4. Регресійний аналіз.
5. Непараметричні методи.

Метод зіставлення двох паралельних рядів **–** фактори, що характеризують результативну ознаку, розташовують у порядку збільшення або зменшення, а потім простежують зміну величини результативної ознаки, що дозволяє встановити наявність зв’язку і його напрямок.

Метод аналітичних групувань **–** виконують групування одиниць сукупності за факторною ознакою і для кожної групи обчислюють середнє значення результативної ознаки. Потім зіставляють зміни результативної ознаки по мірі зміни факторної ознаки і виявляють напрямок, характер та щільність зв'язку за допомогою емпіричного кореляційного відношення $\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$ (див. главу 7).

Кореляційний аналіз – його завдання полягає у вимірюванні щільності зв'язку і оцінюванні факторів, що впливають на результативну ознаку.

Регресійний аналіз – вибір форми моделі (рівняння регресії), встановлення ступеня впливу незалежних змінних на залежну змінну і визначення розрахункових значень залежної змінної (функції регресії).

Аби результати кореляційного аналізу знайшли практичне застосування і дали бажаний результат, повинні виконуватися певні *вимоги щодо відбору об'єкта дослідження і ознак-факторів*:

1) однорідність тих одиниць, які вивчаються методами кореляційного аналізу;

2) необхідність кількісного оцінювання однорідності досліджуваної сукупності за комплексом ознак (наприклад, розрахунок відносних показників варіації);

3) достатня кількість спостережень;

4) незалежність включених до дослідження факторів, оскільки наявність щільного зв'язку між ними свідчить про те, що вони характеризують одні і ті ж самі сторони досліджуваного явища і значною мірою дублюють один одного;

5) припущення про нормальний характер розподілу досліджуваних ознак (додаток 2).

Остання умова пов'язана з застосуванням методу найменших квадратів при розрахунку параметрів кореляції (тільки за нормального розподілу метод найменших квадратів дає оцінку параметрів, яка відповідає принципам максимальної правдоподібності (додаток 1)). Однак на практиці часто стикаються з тими чи іншими відхиленнями від початкових умов, але й тоді метод найменших квадратів дає непогані результати.

Однофакторний лінійний кореляційний і регресійний аналіз. Найбільш розробленою у теорії статистики є методологія парної кореляції (вплив X на Y), яка передбачає однофакторний кореляційний і регресійний аналіз. Тип рівняння між X і Y тут визначається графічним або логічним (встановлення напрямку зв'язку) шляхом. Складність тут полягає у тому, що серед багатьох функцій необхідно знайти таку, яка краще ніж інші відображає реальні зв'язки між аналізованими ознаками.

Якщо зв'язок є лінійним (тобто формула 12.1.3):

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

де $\hat{y}_x$ – теоретичні значення результативної ознаки, отримані з рівняння регресії;

a_0, a_1 – параметри рівняння регресії;

a_0 – показує усереднений вплив неврахованих факторів на результативну ознаку;

a_1 – коефіцієнт регресії, який показує, на скільки зміниться у середньому значення результативної ознаки за зміни факторної ознаки на одиницю її власного виміру. Знак a_1 вказує напрямок цієї зміни.

Для зручності інтерпретації параметра a_1 використовують *коефіцієнт еластичності.* Він показує середні зміни результативної ознаки (у процентах) за зміни факторної ознаки на 1 % і обчислюється за формулою:

$$K_{ел.} = a_1 \frac{\bar{x}}{\bar{y}} \ (\%). \qquad \textbf{(12.1.6)}$$

Оцінювання параметрів рівняння регресії здійснюється методом найменших квадратів МНК ($\sum (y - \hat{y}_x)^2 \to min$) (додаток 1).

Система нормальних рівнянь знаходження параметрів лінійної парної регресії методом найменших квадратів має вигляд:

$$\begin{cases} \sum y = n \cdot a_0 + a_1 \cdot \sum x \\ \sum x \cdot y = a_0 \cdot \sum x + a_1 \cdot \sum x^2 \end{cases}. \qquad \textbf{(12.1.7)}$$

Визначивши значення a_0, a_1 і підставивши їх у рівняння зв’язку (12.1.3), знаходимо значення $\hat{y}_x$, яке залежить тільки від заданого значення x.

Приклад 12.1.1. Розглянемо побудову однофакторного рівняння регресії залежності продуктивності праці (y) від стажу роботи (x) за даними табл. 12.1.2.

Дані 1 і 2 граф таблиці ранжовані за стажем роботи. Зіставлення даних цих паралельних рядів показує, що зі зростанням ознаки-фактора x майже завжди зростає результативна ознака y. Отже, між ними існує майже пряма залежність, для уточнення якої можна використовувати і графічний метод.

Таким чином, робимо висновок, що зв'язок між факторами виражається лінійним рівнянням регресії:

$$\hat{y}_x = a_0 + a_1 \cdot x.$$

Обчислимо параметри рівняння регресії, використовуючи систему рівнянь 12.1.7, для чого обчислимо у таблиці невідомі значення за допомогою вихідної інформації (див. графи 3, 4, 5, 6 табл. 12.1.2).

Таблиця 12.1.2 – Розподіл робітників бригади за виробітком і стажем роботи

Вихідні дані		Розрахункові значення			
Стаж роботи, роки x	Денний виробіток робітника, шт. y	x^2	y^2	xy	$\hat{y}_x$
1	*2*	*3*	*4*	*5*	*6*
1	4	1	16	4	4,6
2	5	4	25	10	5,2
3	6	9	36	18	5,8
4	7	16	49	28	6,4
5	7	25	49	35	7,0
6	8	36	64	48	7,6
7	8	49	64	56	8,2
8	9	64	81	72	8,8
9	10	81	100	90	9,4
10	9	100	81	90	10,0
$\sum x$=55	$\sum y$=73	$\sum x^2$=385	$\sum y^2$=565	$\sum xy$=451	$\sum \hat{y}_x$=73,0

Отримуємо систему:

$$\begin{cases} 73 = 10 \times a_0 + a_1 \times 55 \\ 451 = a_0 \times 55 + a_1 \times 385, \end{cases}$$

Звідки

$$a_1 \approx 0,6,$$

$$a_0 \approx 4,0.$$

Отже, регресійна модель розподілу виробітку за стажем для цього прикладу може бути записана у вигляді конкретного простого рівняння регресії:

$$\hat{y}_x = 4,0 + 0,6 \cdot x.$$

Підставляючи значення x (з графи 1 табл. 12.1.2) в отриману формулу, визначаємо значення $\hat{y}_x$ (6 графа табл. 12.1.2).

Це рівняння характеризує залежність середнього рівня виробітку робітників бригади від стажу роботи, стосовно якого можна зробити висновок: за збільшення стажу роботи на 1 рік денний виробіток робітника у середньому зростає на 0,6 штуки. Такою є економічна інтерпретація параметрів регресії.

Щодо коефіцієнта еластичності, то, розрахувавши його:

$$K_{ел.} = a_1 \frac{\bar{x}}{\bar{y}} = 0,6 \times \frac{5,5}{7,3} = 0,45,$$

можна дати іншу економічну інтерпретацію отриманого рівняння регресії – зі збільшенням стажу роботи на 1 % слід очікувати підвищення продуктивності праці у середньому на 0,45 %.

Правильність розрахунку рівняння регресії можна перевірити порівнянням сум $\sum y = \sum \hat{y}_x$ (при цьому можливе деяке розходження внаслідок округлення розрахунків і суті МНК) [9].

Для визначення щільності кореляційного зв'язку використовують такі показники

1. Емпіричне кореляційне відношення – показник Пірсона (див. главу 7, формула 7.1.16)

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

де:

- загальна дисперсія σ^2 – вимірює варіацію ознаки для всієї сукупності під впливом усіх факторів, що обумовлюють цю варіацію, і розраховується за формулою:

$$\sigma^2 = \frac{\sum (y - \bar{y})^2 \times f}{\sum f}, \qquad \textbf{(12.1.8)}$$

де y – результативна ознака;

- міжгрупова дисперсія δ^2 – характеризує систематичну варіацію результативної ознаки, обумовлену впливом ознаки-фактора, покладеної в основу групування.

$$\delta^2 = \frac{\sum (\bar{y}_i - \bar{y})^2 \times f}{\sum f}, \qquad \textbf{(12.1.9)}$$

де $\bar{y}_i$ – групові середні результативної ознаки;

$\bar{y}$ – загальна середня результативної ознаки;

f – кількість одиниць у групі.

η набуває значень від 0 до 1 (див. співвідношення Чеддока).

2. Теоретичне кореляційне відношення:

$$\eta_m = \sqrt{\frac{\delta^2}{\sigma^2}}, \qquad \textbf{(12.1.10)}$$

де $\delta^2 = \frac{\sum (\hat{y}_x - \bar{y})^2}{n}$ – дисперсія в ряду вирівняних (теоретичних) значень результативного показника $\hat{y}_x$;

$\sigma^2 = \frac{\sum (y - \bar{y})^2}{n}$ – дисперсія в ряду фактичних значень y.

Підставивши значення дисперсій у формулу для теоретичного кореляційного співвідношення 12.1.10, отримаємо:

$$\eta_m = \sqrt{\frac{\sum(\hat{y}_x - \bar{y})^2}{\sum(y - \bar{y})^2}} \,. \qquad \textbf{(12.1.11)}$$

За відсутності зв'язку $\eta_m{}^2$ =0, а за функціонального зв'язку $\eta_m{}^2$ =1. Усі співвідношення Чеддока зберігаються.

Чим ближчим до одиниці є значення кореляційного відношення, тим щільнішим і ближчим до функціональної залежності буде зв'язок між ознаками.

Приклад 12.1.2 (продовження прикладу 12.1.1). Розрахувати теоретичне кореляційне співвідношення за даними табл. 12.1.2.

Розв'язок

За даними прикладу 12.1.1, використовуючи формулу 12.1.11 і беручи до уваги, що $\bar{y}$ =7,3, розраховуємо за допомогою таблиці 12.1.3:

Таблиця 12.1.3 – Розрахункові значення

$y-\bar{y}$	$(y-\bar{y})^2$	$\hat{y}_x-\bar{y}$	$(\hat{y}_x-\bar{y})^2$	$y-\hat{y}_x$	$(y-\hat{y}_x)^2$
1	*2*	*3*	*4*	*5*	*6*
–3,3	10,89	–2,7	7,29	–0,6	0,36
–2,3	5,29	–2,1	4,41	–0,2	0,04
–1,3	1,69	–1,5	2,25	0,2	0,04
–0,3	0,09	–0,9	0,81	0,6	0,36
–0,3	0,09	–0,3	0,09	0,0	0,00
0,7	0,49	0,3	0,09	0,4	0,16
0,7	0,49	0,9	0,81	–0,2	0,04
1,7	2,89	1,5	2,25	0,2	0,04
2,7	7,29	2,1	4,41	0,6	0,36
1,7	2,89	2,7	7,29	–1,0	1,00
Усього	32,10	Х	29,70	Х	2,40

Значення графи 5 табл. 12.1.3 – це залишки ε_i ($\varepsilon_i = y - \hat{y}_x$).

Найбільші від'ємні залишки (наприклад, –0,6; –1,0) визначають робітників, що відстають і потребують найбільшої уваги. Найбільші додатні залишки (наприклад, 0,6; 0,4) визначають передових робітників, які забезпечують найбільше підвищення середнього виробітку.

За таблицею 12.1.3:

$$\eta_m = \sqrt{\frac{29{,}70}{32{,}10}} = \sqrt{0{,}925} = 0{,}962 .$$

Отримане значення теоретичного кореляційного відношення свідчить про можливу наявність доволі щільної прямої залежності між ознаками, що розглядаються.

Теоретичний коефіцієнт детермінації $\eta_m{}^2$, який розраховується як квадрат теоретичного кореляційного відношення, дорівнює:

$$\eta_m{}^2 = 0{,}962^2 = 0{,}925 \ \ (92{,}5\,\%)$$

Це означає, що 92,5 % загальної варіації вибірки у бригаді, що вивчається, обумовлено варіацією факторної ознаки – стажу робітників (і тільки 7,5 % загальної варіації не можна пояснити зміною стажу роботи).

3. Лінійний коефіцієнт кореляції – застосовується при лінійній формі рівняння.

Для практичних розрахунків за малої кількості спостережень, $n \le (20 \div 30)$, лінійний коефіцієнт кореляції зручніше обчислювати за такою формулою:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} . \qquad \textbf{(12.1.12)}$$

Значення лінійного коефіцієнта кореляції є важливим для дослідження соціально-економічних явищ і процесів, розподіл яких є близьким до нормального. Він набуває значень в інтервалі: $-1 \le r \le +1$. Від’ємні значення вказують на обернений зв'язок, додатні – на прямий зв'язок. За r=0 лінійний зв'язок відсутній. Чим ближчим є коефіцієнт кореляції за абсолютною величиною до одиниці, тим щільнішим буде зв'язок між ознаками. За $r = \pm 1$ зв'язок є функціональним.

За даними табл. 12.1.2 розрахуємо r.

$$\frac{\sum x \sum y}{n} = \frac{55 \cdot 73}{10} = 401{,}5 .$$

$$\sum x^2 - \frac{(\sum x)^2}{n} = 385 - 302{,}5 = 82{,}5 .$$

$$\sum y^2 - \frac{(\sum y)^2}{n} = 565 - \frac{73^2}{10} = 32{,}1 .$$

$$r = \frac{451 - 401{,}5}{\sqrt{82{,}5 \cdot 32{,}1}} = \frac{49{,}5}{51{,}46} \approx 0{,}962 .$$

Квадрат лінійного коефіцієнта кореляції r^2 називається *лінійним коефіцієнтом детермінації.* Ступінь щільності зв’язку повністю відповідає теоретичному кореляційному відношенню.

Значення η_m і r відповідатимуть одне одному тільки за наявності прямолінійного зв’язку. Невідповідність цих значень свідчить, що зв'язок між досліджуваними ознаками не прямолінійний, а криволінійний. Встановлено, що якщо різниця квадратів $\eta_m{}^2$ і r^2 не перевищує 0,1, то гіпотезу про прямолінійну форму зв’язку можна вважати підтвердженою. У нашому прикладі ці два значення відповідають одне одному, отже, зв'язок між виробітком робітників та їхнім стажем є прямолінійним.

Перевірка адекватності регресійної моделі і оцінювання значимості коефіцієнта кореляції. Оскільки кореляційний і регресійний аналіз проводиться для обмеженої за обсягом сукупності (тобто вибіркової сукупності), то показники регресії і кореляції можуть бути викривлені дією випадкових факторів. Для практичного використання моделей регресії важливою є їхня адекватність (тобто їхня відповідність фактичним статистичним даним) [9].

За чисельності об’єктів аналізу до 30 одиниць виникає *необхідність перевірки значимості* (суттєвості) кожного коефіцієнта регресії, яку здійснюють за допомогою t-критерію Ст’юдента (для простої лінійної регресії) *за такими етапами* (див. главу 8 “Статистична гіпотеза”):

1. Формулювання гіпотези H_o (коефіцієнти регресії є випадковими) водночас із формулюванням альтернативної гіпотези H_a (коефіцієнти регресії є невипадковими).

2. Вибір рівня значимості α (зазвичай припускається рівним 0,05 або 0,01), який контролює допустиму імовірність помилки (чисельної міри ступеня об’єктивної можливості цієї помилки [8]), а також обчислення розрахункових значень *t-критерію:*

для параметра a_0:

$$t_{a_0} = |a_0| \frac{\sqrt{n-2}}{\sigma_{зал.}}, \qquad \textbf{(12.1.13)}$$

для параметра a_1

$$t_{a_1} = |a_1| \frac{\sqrt{n-2}}{\sigma_{зал.}} \sigma_x, \qquad \textbf{(12.1.14)}$$

де n – обсяг вибірки;

$\sigma_{зал.}$ – залишкове середньоквадратичне відхилення результативної ознаки від вирівняних $\hat{y}_x$;

σ_x – середнє квадратичне відхилення факторної ознаки від загальної середньої:

$$\sigma_{зал.} = \sqrt{\sum (y - \hat{y}_x)^2 / n}, \qquad \textbf{(12.1.15)}$$

$$\sigma_x = \sqrt{\sum (x - \bar{x})^2 / n}. \qquad \textbf{(12.1.16)}$$

(Для прикладу 12.1.1, за даними розрахункових таблиць:

$$\sigma_x = \sqrt{\frac{\sum x^2}{n} - (\frac{\sum x}{n})^2} = \sqrt{\frac{385}{10} - (\frac{55}{10})^2} = 2{,}87,$$

$$\sigma_{зал.} = \sqrt{\frac{2{,}4}{10}} = 0{,}49,$$

$$t_{a_0} = 4 \times \frac{\sqrt{10-2}}{0{,}49} = 23{,}1,$$

$$t_{a_1} = 0{,}6 \times \frac{\sqrt{10-2}}{0{,}49} \times 2{,}87 = 9{,}94 \text{).}$$

3. Визначення області допустимих значень і критичної області. При цьому обчислені значення t_{a_0} і t_{a_1} порівнюють з критичним $t_{табл}$, яке визначають за таблицею Ст'юдента ("Значення t-критерію Ст'юдента за рівня значимості 0,10, 0,05, 0,01", яка зазвичай міститься у додатках підручників і довідників зі статистики) з урахуванням прийнятого рівня значимості α і числа ступенів свободи $v=n-2$ (число ступенів свободи – це кількість варіантів, які можуть набувати значень, функціонально не пов'язаних одне з одним). (Для прикладу 12.1.1, за v=8; α=0,05; $t_{табл}$=2,306; $t_{розр} > t_{табл}$).

4. Винесення конкретного рішення (тобто або відхилення гіпотези, або її прийняття) на основі порівняння фактичного і критичного значень критерію.

Параметр тут визнають значимим (суттєвим) за умови, якщо $t_{розр} > t_{табл}$. Тоді майже неймовірно, що знайдені значення параметрів обумовлені тільки випадковими факторами (тобто відхиляється нульова гіпотеза і приймається альтернативна)

Оцінювання значимості коефіцієнта кореляції здійснюють за допомогою t-критерію Ст'юдента за аналогічними етапами (див. вище). Різниця полягає лише в тому, що:

$$t_{розр} = r\sqrt{\frac{n-2}{1-r^2}}, \qquad \textbf{(12.1.17)}$$

де (n–2) – число ступенів свободи за заданого рівня значимості α і обсягу вибірки n.

Якщо $t_{розр} > t_{табл}$, то коефіцієнт кореляції є значимим.

(Для прикладу 12.1.1, $t_{розр} = 0{,}962 \times \sqrt{\frac{8}{1-0{,}925}} = 9{,}93$.

За v=8; α=0,05; $t_{табл}$=2,306; $t_{розр} > t_{табл}$).

У нашому прикладі коефіцієнт кореляції є значимим, що свідчить про суттєвість зв’язку між виробітком і стажем роботи, а побудована регресійна модель є адекватною, тобто висновки, отримані за результатами малої вибірки, можна з достатньою імовірністю поширити на всю генеральну сукупність.

Багатофакторний кореляційний і регресійний аналіз. Як відомо, явища суспільного життя складаються під впливом не одного, а багатої кількості факторів, між якими існують складні взаємозв’язки. Багатофакторний кореляційний і регресійний аналіз дозволяє оцінити міру впливу на досліджуваний результативний показник кожного з включених до моделі (рівняння) факторів за фіксованого положення (на середньому рівні) решти факторів, а також за будь-яких можливих поєднань факторів з певним ступенем точності знайти теоретичне значення цього показника. Важливою умовою тут є відсутність функціонального зв’язку між факторами. Окрім цього до рівняння не можна включати фактори, пов’язані один з одним сильніше, ніж з результативним показником.

Математична задача формулюється таким чином. Потрібно знайти аналітичний вираз, що найкращим чином відображає встановлений теоретичним аналізом зв'язок між незалежними і результативними ознаками, тобто функцію (див. формулу 12.1.2):

$$\hat{y}_i = f(x_{1,} x_{2,\dots} x_\kappa) + \varepsilon_i \, .$$

Побудова і статистичний аналіз двохфакторної лінійної моделі (трьохмірної регресії):

$$\hat{y}_{x_1 x_2} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 \, , \qquad \textbf{(12.1.18)}$$

де $\hat{y}_{x_1 x_2}$ – розрахункові значення залежної змінної (результативної ознаки);

$x_{1,}\, x_2$ – незалежні ознаки (факторні ознаки);

$a_{0,}\, a_{1,}\, a_2$ – параметри рівняння.

Параметри рівняння множинної регресії, як і у разі парної регресії, знаходять способом найменших квадратів.

В результаті отримуємо систему нормальних рівнянь:

$$\begin{cases} a_0 n + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x^2{}_1 + a_2 \sum x_1 x_2 = \sum y x_1 \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x^2{}_2 = \sum y x_2 \, . \end{cases} \qquad \textbf{(12.1.19)}$$

Параметри цієї системи можна знайти, наприклад, методом Гаусса. Вручну можна виконувати побудову і аналіз тільки двох- або трьохфакторних моделей. Якщо число факторів *К* більше трьох, то всі розрахунки виконуються на комп’ютері.

Побудову і аналіз трьохмірної регресійної моделі розглянемо на конкретному прикладі.

Приклад 12.1.3. За вибірковими даними, представленими в табл. 12.1.4, про виробіток деталей за зміну 20 робітниками цеху, потрібно виявити залежність продуктивності праці (y) від двох факторів: простоїв всередині зміни (x_1) і кваліфікації робітників (x_2).

Теоретичний аналіз вихідних даних дозволяє встановити наявність причинно-наслідкового зв’язку між факторними ознаками (простоями всередині зміни і кваліфікацією робітників) і результативним показником – продуктивністю праці.

Таблиця 12.1.4 (фрагмент) – Стохастичний зв'язок між продуктивністю праці, простоями всередині зміни і кваліфікацією робітників

Порядковий номер робітника	Простої всередині зміни, хв. x_1	Кваліфікація робітника (тарифний розряд) x_2	Денний виробіток робітника, деталі y
1	5	3	86
2	8	4	88
3	15	5	94
…	…	…	…
20	14	4	92
Усього	**220**	**80**	**1800**
Середні значення	$\bar{x}_1 = 11$	$\bar{x}_2 = 4$	$\bar{y} = 90$

Регресійну двохфакторну модель побудуємо у лінійній формі: $\hat{y}_{x_1x_2} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2$ і перевіримо її адекватність.

Для знаходження параметрів цього рівняння виконаємо розрахунки допоміжних величин у таблиці 12.1.5, використовуючи систему нормальних рівнянь 12.1.19.

Таблиця 12.1.5 – До розрахунку параметрів і оцінювання лінійної двохфакторної регресійної моделі

y^2	x_1^2	x_2^2	yx_1	yx_2	x_1x_2	$\hat{y}_{x_1x_2}$	$y-\hat{y}_{x_1x_2}$	$(y-\hat{y}_{x_1x_2})^2$
7396	25	9	430	258	15	89,0	–3,0	9,00
7744	64	16	704	352	32	91,2	–3,2	10,24
8836	225	25	1410	470	45	91,7	2,3	5,29
…	…	…	…	…	…	…	…	…
8464	196	16	1288	368	56	88,7	3,3	10,89
162640	**2830**	**342**	**19436**	**7298**	**822**	**1800**	-	**177,20**

$\overline{y^2}$ =8132; $\overline{x_1^{\,2}}$ =141,5; $\overline{yx_1}$ =971,8; $\overline{yx_2}$ =364,9; $\overline{x_1x_2}$ =41,1; $\overline{x_2^{\,2}}$ =17,1;

$\sigma^2_{зал} = \dfrac{\sum(y-\hat{y}_{x_1x_2})^2}{n} = \dfrac{177{,}2}{20} = 8{,}86$ – залишкова дисперсія.

Складемо систему нормальних рівнянь:

$$\begin{cases} 20a_0 + 220a_1 + 80a_2 = 1800 \\ 220a_0 + 2830a_1 + 822a_2 = 19436 \\ 80a_0 + 822a_1 + 342a_2 = 7298. \end{cases}$$

Розв'язуючи цю систему методом Гаусса, отримаємо:
a_0=81,03, a_1=–0,41, a_2=3,37.

Таким чином, рівняння множинної регресії, що відображає залежність продуктивності праці $\hat{y}$ від простоїв усередині зміни x_1 і кваліфікації робітників x_2, набуде вигляду:

$$\hat{y}_{x_1x_2} = 81{,}03 - 0{,}41\,x_1 + 3{,}37x_2$$

Це означає, що зі збільшенням тривалості простоїв усередни зміни на 1 хв. слід очікувати зниження продуктивності праці (денного виробітку деталей одним робітником) на 0,41 деталі (обернений зв'язок); а підвищення кваліфікації робітника на 1 розряд може привести до збільшення виробітку на 3,37 деталі.

Однак на підставі коефіцієнтів регресії не можна сказати, яка з факторних ознак має найбільший вплив на результативну ознаку, тобто коефіцієнти регресії незіставні один з одним, оскільки вони виміряні у різних одиницях.

Аби мати можливість судити про порівняльну силу впливу окремих факторів і про ті три резерви, які в них закладені, треба обчислити *частинні коефіцієнти еластичності* E_i:

$$E_i = a_i\,\frac{\bar{x}_i}{\bar{y}}, \qquad \textbf{(12.1.20)}$$

де a_i – коефіцієнт регресії за i-го фактора;

$\bar{x}_i$ – середнє значення i-го фактора;

$\bar{y}$ – середнє значення досліджуваного показника.

Для прикладу 12.1.3: $E_1 = -0{,}41\dfrac{11}{90} \approx -0{,}05$,

$$E_2 = 3{,}37\,\frac{4}{90} \approx 0{,}15.$$

Аналіз частинних коефіцієнтів еластичності показує, що найбільший вплив на продуктивність праці має фактор x_2 – підвищення кваліфікації робітників на 1 % приводить до зростання продуктивності праці на 0,15 %. Зростання тривалості простоїв усередині зміни на 1 % знижує продуктивність праці на 0,05 %.

Для вимірювання щільності зв'язку між двома аналізованими змінними (без урахування їхньої взаємодії з іншими змінними) застосовують *парні коефіцієнти кореляції.* Методика розрахунку таких коефіцієнтів і їхня інтерпретація аналогічні методиці розрахунку лінійного коефіцієнта кореляції для однофакторного зв'язку.

Спрощені формули розрахунку парних коефіцієнтів кореляції:

$$r_{yx_1} = \frac{\overline{x_1 y} - \bar{x}_1 \bar{y}}{\sigma_{x_1} \sigma_y}, \qquad \textbf{(12.1.21)}$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \bar{x}_2 \bar{y}}{\sigma_{x_2} \sigma_y}, \qquad \textbf{(12.1.22)}$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{\sigma_{x_1} \sigma_{x_2}}. \qquad \textbf{(12.1.23)}$$

Розрахуємо середні квадратичні відхилення:

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{8132 - 8100} = 5{,}66,$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^{\ 2}} - \bar{x}_1^{\ 2}} = \sqrt{141{,}5 - 121} = 4{,}53,$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^{\ 2}} - \bar{x}_2^{\ 2}} = \sqrt{17{,}1 - 16} = 1{,}05.$$

Підставивши значення середніх квадратичних відхилень у формули 12.1.21 – 12.1.23 для парних коефіцієнтів кореляції, отримаємо:

$$r_{yx_1} = -0{,}710,$$

$$r_{yx_2} = 0{,}825,$$

$$r_{x_1x_2} = -0{,}609.$$

Отримані результати є допустимими, і всі фактори залишаються у моделі. Якщо ж, наприклад, отримано значення $r_{yx_1} = -0{,}80$; $r_{yx_2} = 0{,}65$; $r_{x_1x_2} = -0{,}88$, то в регресійне рівняння слід включити фактор x_1, а x_2 не включати, оскільки він щільно пов'язаний (колінеарний) з x_1, а його кореляція з y слабше, ніж кореляція x_1 з y.

У реальних умовах усі змінні зазвичай є взаємопов'язаними. Щільність цього зв'язку визначають *частинними коефіцієнтами кореляції*, які характеризують ступінь і вплив одного з аргументів на функцію за умови, що решта незалежних змінних закріплені на постійному рівні:

$$r_{yx_1(x_2)} = \frac{r_{yx1} - r_{yx2} r_{x1x2}}{\sqrt{(1 - r^2_{yx2})(1 - r^2_{x1x2})}}$$ – за вилучення впливу x_2, **(12.1.24)**

$$r_{yx_2(x1)} = \frac{r_{yx2} - r_{yx1} r_{x1x2}}{\sqrt{(1 - r^2_{yx1})(1 - r^2_{x1x2})}}$$ – за вилучення впливу x_1, **(12.1.25)**

$$r_{x_1x2(y)} = \frac{r_{x1x2} - r_{yx1} r_{yx2}}{\sqrt{(1 - r^2_{yx1})(1 - r^2_{yx2})}}$$ – за вилучення впливу y. **(12.1.26)**

Для прикладу 12.1.3:

$r_{yx_1(x_2)} \approx -0{,}465$; $r_{yx_2(x1)} \approx 0{,}696$; $r_{x_1x_2(y)} \approx -0{,}058$.

Отже, зв'язок кожного фактора з досліджуваним показником за умови комплексного впливу факторів є слабшим. Практично відсутній зв'язок між факторними ознаками за елімінування (вилучення) результативного показника. Це цілком зрозуміло – адже простої всередині зміни і кваліфікація робітників ніяк не пов’язані одне з одним (якщо не брати до уваги необхідність виконання завдання).

Показником щільності зв’язку між результативною і двома або більше факторними ознаками є *сукупний коефіцієнт множинної кореляції* $R_{yx_1,x_2,...,x_n}$. У разі лінійного двохфакторного зв’язку сукупний коефіцієнт множинної кореляції можна розрахувати за формулою:

$$R_{yx_1x_2.} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2 r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r^2_{x_1x_2}}}, \quad \textbf{(12.1.27)}$$

де r – парні лінійні коефіцієнти кореляції.

Його значення знаходяться в межах від –1 до +1. Чим ближче значення R до ± 1, тим більш інтенсивним буде кореляційний зв'язок.

Для прикладу 12.1.3: $R_{yx_1x_2.} = \sqrt{0{,}749} = 0{,}865$.

Сукупний коефіцієнт множинної детермінації R^2 – величина, що показує, яка частка варіації досліджуваного показника пояснюється впливом факторів, включених до рівняння множинної регресії. $0 \le R^2 \le 1$. Чим ближче R^2 до одиниці, тим більшою мірою варіація досліджуваного показника характеризується впливом відібраних для дослідження факторів. У прикладі 12.1.3: R^2=0,749 показує, що варіація продуктивності праці на 74,9 % обумовлена аналізованими факторами, а на 25,1 % – впливом неврахованих факторів. Отже, відібрані для дослідження фактори суттєво впливають на показник продуктивності праці.

Перевірку значимості (адекватності) рівняння регресії виконують на основі розрахунку F-критерію Фішера:

$$F = \frac{\sigma_y^2}{\sigma_{зал}^2} \cdot \frac{n-m}{m-1}, \qquad \textbf{(12.1.28)}$$

де m – кількість параметрів у рівнянні регресії (для рівнянь множинної регресії кількість параметрів відповідає кількості незалежних змінних).

Отримане значення критерію $F_{розр.}$ порівнюють з критичним (табличним) для прийнятого рівня значимості 0,05 або 0,01 і чисел ступенів свободи $\nu_1 = m - 1$ і $\nu_2 = n - m$. Якщо $F_{розр.}$ виявиться більшим, ніж відповідне табличне значення, то таке рівняння регресії є статистично значимим (тобто частка варіації, обумовлена регресією, набагато перевищує випадкову помилку).

Зазвичай вважають, що рівняння регресії є придатним для практичного використання, якщо $F_{розр.} > F_{табл}$ не менше, ніж у 4 рази.

У прикладі 12.1.3:

$$F_{розр.} = \frac{32}{8{,}86} \cdot \frac{20-2}{2-1} = 65.$$

$F_{табл}$=4,41 (знаходимо за таблицею “Значення F-критерію Фішера за рівня значимості 0,05”, яка знаходиться у будь-якому статистичному підручнику або довіднику).

$F_{розр.} > F_{табл}$, отже, рівняння регресії:

$\hat{y}_{x_1x_2}$ =81,03–0,41 $x_1 + 3{,}37x_2$ можна визнати адекватним.

Оцінювання значимості коефіцієнтів регресії і суттєвості коефіцієнта кореляції за лінійної залежності y від x_1 і x_2 (двох факторів) здійснюють за допомогою t-критерію Ст’юдента за $\nu = n$-m–1 ступенів свободи (де n=20, m=2):

$$t_{a_1} = \frac{a_1\sigma_{x_1}\sqrt{1 - r^2_{x_1x_2}} \cdot \sqrt{n-m-1}}{\sigma_y\sqrt{1 - R^2_{yx_1x_2}}}, \qquad \textbf{(12.1.29)}$$

$$t_{a_2} = \frac{a_2\sigma_{x_2}\sqrt{1 - r^2_{x_1x_2}} \cdot \sqrt{n-m-1}}{\sigma_y\sqrt{1 - R^2_{yx_1x_2}}}, \qquad \textbf{(12.1.30)}$$

$$t_{R_{yx_1x_2}} = \frac{R_{yx_1x_2} \cdot \sqrt{n-m-1}}{1 - R^2_{yx_1x_2}}. \qquad \textbf{(12.1.31)}$$

У прикладі 12.1.3: t_{a_1} = =2,14; t_{a_2} = =4,07; $t_{R_{yx_1x_2}}$ = 14,20.

Табличне значення t-критерію за 5-процентного рівня значимості (тобто $\alpha = 0{,}05$) і $\nu = 17$ складає 2,11.

Відповідні $t_{розр.} > t_{табл}$, отже, обидва фактори a_1 і a_2 та сукупний коефіцієнт кореляції $R_{yx_1x_2.}$ є значимими (суттєвими).

Таким чином, побудована регресійна модель:

$\hat{y}_{x_1x_2}$ =81,03–0,41 x_1 + 3,37x_2 придатна для практичного застосування. Її можна використовувати для виявлення резервів продуктивності праці.

Непараметричні методи. Поряд з використанням параметричних методів, коли всі ознаки кількісно виміряні (ці методи розглянуто вище), у статистиці також застосовуються непараметричні методи, за допомогою яких встановлюється зв'язок між якісними (атрибутивними) ознаками. Тут не ставиться задача представлення зв'язку через рівняння, а тільки йдеться про його наявність і щільність. Аналіз виконується за допомогою *таблиць взаємного сполучення.*

Якщо необхідно провести дослідження зв'язку між альтернативними ознаками, представленими тільки групами з протилежними (взаємовиключними) характеристиками, то щільність зв'язку можна оцінити, обчисливши коефіцієнт асоціації (k_{ac}) або коефіцієнт контингенції ($\boldsymbol{k}_{кон.}$).

Для цього будують чотирьохклітинну таблицю (таблицю "чотирьох полів") з частотами (*a*, *b*, *c*, *d*) описових ознак (табл. 12.1.6):

Таблиця 12.1.6 – Таблиця "чотирьох полів"

Зріст сина	Зріст батька		Усього
	Нижче середнього, чол.	Вище середнього, чол.	
Нижче середнього, чол.	*a*	*B*	*a*+*b*
Вище середнього, чол.	*c*	*D*	*c*+*d*
Усього	*a*+*c*	*b*+*d*	*A*+*b*+ *c*+*d*

$$k_{ac} = \frac{ad - bc}{ad + bc}, \qquad \textbf{(12.1.32)}$$

$$k_{кон.} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \qquad \textbf{(12.1.33)}$$

$$-1 \le k_{ac.} \le 1,$$

$$-1 \le k_{кон.} \le 1.$$

Властивості у цих коефіцієнтів такі ж самі, як і у коефіцієнта кореляції.

kас завжди більше $k_{кон.}$. Зв'язок вважається підтвердженим, якщо $|\,kас| > 0{,}5$; $|\,k_{кон.}\,| > 0{,}3$.

Якщо для кожної з взаємопов’язаних ознак виділяють кількість груп, більшу двох, то з використанням подібних таблиць щільність зв’язку між якісними ознаками можна виміряти за допомогою коефіцієнтів взаємного сполучення Пірсона ($К_П$) і Чупрова ($Кч$).

$$К_П = \sqrt{\frac{\varphi^2}{1+\varphi^2}}, \qquad \textbf{(12.1.34)}$$

$$К_Ч = \sqrt{\frac{\varphi^2}{\sqrt{(К_1 - 1)(К_2 - 1)}}}, \qquad \textbf{(12.1.35)}$$

де $К_1$, $К_2$ – кількість значень 1-ї і 2-ї ознак, відповідно,

φ^2 – показник взаємного сполучення, визначається як сума відношень квадратів частот кожної клітині таблиці до добутку підсумкових частот відповідного стовпчика і строки мінус 1 (табл. 12.1.7).

Таблиця 12.1.7 – Допоміжна таблиця для розрахунку коефіцієнта взаємного сполучення φ^2:

x \ y	I	II	III	Усього
I	…	…	$\boldsymbol{n_{xy}}$	$\boldsymbol{n_x}$
II	…	…	…	$\boldsymbol{n_x}$
III	…	…	…	$\boldsymbol{n_x}$
Усього	$\boldsymbol{n_y}$	$\boldsymbol{n_y}$	$\boldsymbol{n_y}$	$\boldsymbol{n}$

$$\varphi^2 = \sum \frac{n^2{}_{xy}}{n_x n_y} - 1. \qquad \textbf{(12.1.36)}$$

Чим ближче величини $К_П$ і $Кч$ до 1, тим щільнішим буде зв'язок. Казати про щільний зв'язок між варіацією досліджуваних ознак можна вже тоді, коли вони досягають значення 0,3.

Приклад 12.1.4. Для одного з факультетів є дані про розподіл 600 студентів-заочників за двома ознаками: характером роботи і результатами здачі екзаменів зі спеціальних предметів (табл. 12.1.8).

Визначити щільність зв’язку між ознаками, використовуючи:

- $k_{ас}$;
- $k_{кон.}$;
- $К_П$;
- $Кч$.

Таблиця 12.1.8 – Вихідні дані

Характер роботи	Студенти, чол.		Усього студентів
	Ті, хто здав сесію без незадовільних оцінок	Ті, хто отримав незадовільні оцінки	
Ті, хто працює за профілем факультету, чол.	**270** (a)	**50** (b)	**320** (a+b)
Ті, хто працює не за профілем факультету, чол.	**150** (c)	**130** (d)	**280** (c+d)
Усього студентів	**420** (a+c)	**180** (b+d)	**600** (a+b+ c+d)

Розв'язок

За формулою 12.1.32:

$$k_{ac} = \frac{270 \times 130 - 150 \times 50}{270 \times 130 + 150 \times 50} = 0{,}648.$$

За формулою 12.1.33:

$$k_{кон.} = \frac{270 \times 130 - 150 \times 50}{\sqrt{320 \times 280 \times 420 \times 180}} = 0{,}335.$$

$|\, k_{ac} \,| > 0{,}5;\ |\, k_{кон.} \,| > 0{,}3 \quad \Rightarrow$ зв'язок вважається підтвердженим.

За формулами 12.1.34 и 12.1.36:

$$\varphi^2 = \left(\frac{270^2}{420 \times 320} + \frac{50^2}{180 \times 320} + \frac{150^2}{420 \times 280} + \frac{130^2}{180 \times 280}\right) - 1 = 0{,}113,$$

тоді $K_П = \sqrt{\dfrac{0{,}113}{1 + 0{,}113}} = 0{,}319$.

За формулою 12.1.35:

$$K_Ч = \sqrt{\frac{0{,}113}{\sqrt{(2-1)(2-1)}}} = 0{,}336,$$

де K_1=2, K_2=2.

Висновок: успішність студентів у навчанні сильно залежить від характеру їхньої роботи.

12.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. У чому полягає відмінність між функціональним і стохастичним зв’язком?
2. Що таке кореляційний зв'язок?
3. Якими статистичними методами досліджуються функціональні та кореляційні зв’язки?
4. У чому полягає сутність паралельних рядів і аналітичних групувань?
5. Які основні задачі розв’язують за допомогою кореляційного і регресійного аналізу?
6. Охарактеризуйте основні проблеми і правила побудови однофакторної лінійної регресійної моделі.
7. Що характеризують коефіцієнти регресії?
8. Метод визначення параметрів рівняння регресії.
9. Навіщо необхідна перевірка адекватності регресійної моделі?
10. Як здійснюється перевірка значимості коефіцієнтів регресії?
11. Якими показниками вимірюється щільність кореляційного зв’язку?
12. Яке значення має розрахунок коефіцієнта детермінації?
13. Лінійні коефіцієнти кореляції та детермінації, їхній смисл і призначення.
14. Як охарактеризувати однофакторну регресійну модель з погляду на її економічну сутність?
15. Який економічний смисл мають коефіцієнти еластичності?
16. Основні проблеми і правила побудови багатофакторної кореляційної моделі.
17. Сутність і призначення парних і непарних коефіцієнтів кореляції.
18. Сутність і призначення сукупного коефіцієнта множинної кореляції і сукупного коефіцієнта детермінації.
19. Як перевірити адекватність рівняння в цілому? Як перевірити значимість коефіцієнта регресії і які критерії можна для цього використати?
20. Як інтерпретувати багатофакторну регресійну модель з погляду на її економічну сутність?
21. Які непараметричні методи застосовують для моделювання зв’язку?
22. Для 10 однотипних підприємств є дані про випуск продукції (Х) у тис. одиниць і про витрати умовного палива (У) у тонах. Потрібно знайти рівняння залежності витрат палива від випуску продукції (або рівняння регресії У по Х) і виміряти щільність залежності між ними.

Х	5	6	8	8	10	10	14	20	20	24
У	4	4	6	5	7	8	8	10	12	16

23. За даними від 10 підприємств потрібно знайти рівняння залежності між обсягом випуску продукції (У) млн. грн. і вартістю основних виробничих фондів (Х) млн. грн.

Х	1,5	1,8	2,0	2,2	2,3	2,6	3,0	3,1	3,5	3,8
У	3,9	4,4	3,8	3,5	4,8	4,3	7,0	6,5	6,1	8,2

24. За даними від 10 господарств виміряти щільність залежності між врожайністю картоплі і кількістю внесених мінеральних добрив.

Добрива, кг/га	140	148	150	150	185	190	202	220	220	240
Врожайність картоплі, ц/га	135	135	182	175	200	200	200	210	265	250

25. За даними про суму активів і кредитних вкладень комерційних банків необхідно визначити щільність зв'язку між ознаками:

Кредитні вкладення, грош. од., Х	311	658	783	1142	1319	1962	2496
Сума активів, грош. од., У	518	1194	2941	1865	1997	3066	3176

26. Компанія "GG" включає велику розгалужену систему збуту з великою кількістю товарних вагонів, і кожному постачальнику відповідає певна група споживачів. Постачальник постачає зі складу картон кожному споживачу, і цей картон (у вагоні) слідує певним маршрутом. За наведеними даними виміряти щільність зв'язку між кількістю часу на доставку вантажу і загальною вагою вантажу.

Група постачальників	1	2	3	4	5	6
Час на доставку вантажу, дні	6,3	4,7	5,0	8,5	3,8	4,1
Загальна вага вантажу, т	24	15	17	30	14	20

27. Було обстежено 1000 жінок з метою дослідження залежності між кольором очей матерів і дочок, і отримано такі дані.

Визначити щільність зв'язку між ознаками.

Колір очей матері	Колір очей дочки		
	1	2	Усього
1	471	148	619
2	151	230	381
Усього	622	378	1000

28. Є дані від 10 заводів галузі. Визначити, чи існує кореляційний зв'язок між вартістю основних фондів (факторна ознака) і обсягом випуску продукції (результативна ознака).

№ заводу	Вартість основних фондів (млн. грн.)	Обсяг випуску продукції (млн. грн.)
1	2	2,0
2	1	1,2
3	3	3,6
4	5	6,8
5	4	4,4
6	3	3,8
7	1	0,8
8	2	2,2
9	4	5,0
10	5	4,6

29. Компанія "GG" володіє 12 магазинами. Розглядається можливість злиття магазинів для збільшення прибутку. За наведеними даними виміряти щільність залежності між прибутком і товарооборотом.

Магазин	1	2	3	4	5	6	7	8	9	10	11	12
Прибуток ($тис.)	2	4	11	17	18	28	34	36	48	55	71	85
Оборот ($тис.)	50	60	85	85	100	120	140	155	180	210	250	300

30. Припустимо, що є розподіл 100 дослідних ділянок (під овочевою культурою) за двома ознаками: інтенсивністю поливу (Х) і рівнем врожайності (У).

1. Визначити, чи є випадковим такий розподіл, тобто чи є залежність між поливом і врожайністю.
2. Виміряти щільність залежності між інтенсивністю поливу і рівнем врожайності.

Х\У	Висока	Середня	Низька	Усього
Інтенсивний	40	10	5	
Середній	20	7	3	
Слабий	-	5	10	
Усього				

31. З метою дослідження впливу професійної підготовки на виконання норм виробітку провели статистичне дослідження, результати якого наведено у таблиці:

Групи робітників	Кількість робітників у групі	
	Ті, хто виконав норму виробітку	Ті, хто не виконав норму виробітку
Ті, хто пройшов професійну підготовку	115	20
Ті, хто не пройшов професійну підготовку	15	50

Оцініть ступінь зв'язку між ознаками, зробіть висновки.

32. Торговець продає пучки салату у роздріб щоденно. Визначте щільність залежності між кількістю проданих пучків салату і їхньою ціною.

Кількість проданих за день пучків	28	29	34	35	37	37	41	46
Ціна за пучок, грош.од.	30	31	25	26	22	24	16	12

33. Є дані про зріст батьків і синів, які представлено у таблиці. Розрахуйте коефіцієнт асоціації і коефіцієнт взаємного сполучення Чупрова. Зробіть висновки.

Зріст сина	Зріст батька		Усього
	Нижче середнього	Вище середнього	
Нижче середнього	70	20	90
Вище середнього	30	80	110
Усього	100	100	200

34. Дані про рівень освіти членів 100 сімей наведено у таблиці:

Освіта чоловіка	Освіта дружини			Усього
	Неповна середня	Середня і середня спеціальна	Вища	A
Неповна середня	15	11	2	28
Середня і середня спеціальна	8	32	8	48
Вища	1	8	15	24
Усього	24	51	25	100

Визначити коефіцієнт взаємного сполучення Чупрова.

35. Дослідити залежність між оцінками рівня життя респондентів з Києва і типом підприємства, на якому вони працюють. Дані наведено у таблиці:

Тип підприємства	Оцінки рівня життя респондентами				Усього
	Цілком задовільне	Скоріше задовільне	Скоріше незадовільне	Цілком незадовільне	
Державне підприємство	31	35	35	35	136
Акціонерне підприємство	17	13	14	9	53
Орендне підприємство	4	2	1	1	8
Приватне підприємство	8	5	4	3	20
Усього	60	55	54	48	217

Визначити коефіцієнти сполучення Пірсона і Чупрова.

36. Є дані від дослідження соціально-демографічної характеристики випадкових споживачів тютюнових виробів у одному з регіонів, тис.чол.

Групи споживачів тютюнових виробів	Сімейний стан		Усього
	Заміжня (жонатий)	Незаміжня (не жонатий)	
Споживав	10,0	14,5	24,5
Не споживав	2,5	4,5	7,0
Усього	12,5	19,0	31,5

Розрахувати коефіцієнти асоціації і контингенції.

37. Припустимо, що є розподіл 100 дослідних ділянок (під овочевою культурою) за двома ознаками: інтенсивністю поливу (х) і рівнем врожайності (у).

Врожайність (у) / Полив (х)	Висока	Середня	Низька	Усього
Інтенсивний	40	10	5	55
Середній	20	7	3	30
Слабкий	-	5	10	15
Усього	60	22	18	100

1. Визначити, випадковим чи невипадковим є розподіл у таблиці, тобто зробити висновок про наявність або відсутність залежності врожайності від поливу.
2. Виміряти щільність цієї залежності, якщо вона є.

38. Розрахувати щільність зв'язку між ставленням робітників до навчання і їхньою участю у винахідництві за вихідними даними, представленими у таблиці:

Ставлення до навчання	Участь у винахідництві		Усього
	Ті, хто не давав пропозицій	Ті, хто дав одну пропозицію	
Навчаються в інституті ІІІ ступеня акредитації	61	6	67
Навчаються в університеті IV ступеня акредитації	32	3	35
Усього	93	9	102

ГЛАВА 13

БАГАТОМІРНІ ГРУПУВАННЯ

13.1
МЕТОДИЧНІ ВКАЗІВКИ

У главі 3 було розглянуто групування об’єктів за однією ознакою. Однак на практиці об’єкти характеризуються багатьма ознаками, які часто (для потреб більш точних досліджень) необхідно враховувати у комплексі. У розділі “Багатомірні групування” глави 3 було коротко сформульовано їхню сутність. Необхідно також зазначити, що багатомірні групування дозволяють розв’язувати такі важливі задачі економіко-статистичного дослідження, як формування однорідних сукупностей, вибір суттєвих ознак, виділення типових груп об’єктів тощо.

У цій главі розглядаються:

1) групування об’єктів на основі багатомірних середніх;
2) кластерний аналіз:
   а) на основі міри схожості – евклідова відстань,
   б) векторна модифікація кластерного аналізу.

Групування об’єктів на основі багатомірних середніх – найпростіший варіант багатомірної класифікації. *Багатомірна середня* – це середня величина декількох ознак для однієї одиниці сукупності. Оскільки не можна розрахувати середню величину абсолютних значень різних ознак, виражених у різних одиницях вимірювання, то багатомірну середню обчислюють з відносних величин, зазвичай з відношень значень ознак для одиниці сукупності до середніх значень цих ознак:

$$\overline{P_i} = \frac{\sum_{j=1}^{k} P_{ij}}{k} = \sum_{j=1}^{k} \left(\frac{x_{ij}}{\overline{x_j}}\right) : k, \qquad \textbf{(13.1.1)}$$

де $\overline{P_i}$ – багатомірна середня для i-ї одиниці;

x_{ij} – значення ознаки x_j для i –ї одиниці;

$\overline{x_j}$ – середнє значення ознаки x_j;

k – кількість ознак;

j – номер ознаки;

i – номер одиниці сукупності.

У разі великого обсягу сукупності для виділення груп на основі багатомірної середньої встановлюють інтервали значень багатомірної середньої:

$$i_{\overline{P}} = \frac{\overline{P}_{max} - \overline{P}_{min}}{кількість\ груп}. \qquad \textbf{(13.1.2)}$$

Потім слід провести групування одиниць: визначити їхню кількість у кожній групі і спробувати показати, у чому полягають якісні відмінності між групами.

Приклад 13.1.1. Розглянемо використання багатомірних середніх на прикладі сільськогосподарських підприємств регіону, якщо для кожного підприємства наведено чотири ознаки (табл. 13.1.1, графи: А, 1, 2, 3, 4):

- середньомісячна оплата праці робітника, грош.од., x_1;
- валовий доход на 1 га сільськогосподарських угідь, грош.од./га, x_2;
- середньорічна вартість основних виробничих фондів на 1 га сільськогосподарських угідь, грош.од./га, x_3;
- відношення дебіторської заборгованості до кредиторської заборгованості, %, x_4.

Таблиця 13.1.1 – Розрахункова

Підприємства	Значення ознак				В % до середньої				Багатомірна середня, %
	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	
А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
А	597	390	20,6	72	148	199	106	107	140
Б	353	96	12,1	30	88	49	62	45	61
В	403	84	20,6	26	100	43	106	39	72
Г	231	71	15,1	74	57	36	78	110	70
Д	330	114	14,8	159	82	58	76	237	113
Е	540	235	24,0	26	134	120	184	39	104
Ж	372	461	33,2	85	93	235	171	127	156
З	393	113	15,0	62	98	58	77	92	81
Середні величини ($\bar{x}$)	402	196	19,4	67	100	100	100	100	100
Середні квадратичні відхилення (σ)	109	142	6,4	41	-	-	-	-	-

Багатомірні середні, наведені у графі 9 табл. 13.1.1, узагальнюють чотири ознаки (при цьому значимість ознак для оцінювання підприємства припускається однаковою, що, звісно, є суперечним). Можна ускладнити методику, приписавши ознакам, на підставі експертної оцінки або кореляційно-регресійного аналізу, різні ваги і обчислити багатомірні середні.

Судячи з оцінок, отриманих для прикладу 13.1.1, підприємства поділяються на групи з багатомірними середніми нижче 100% (чотири підприємства), дещо вище 100 % (два підприємства) і набагато вище 100 % (два підприємства).

Кластерний аналіз є більш обґрунтованим методом багатомірної класифікації. Такий аналіз дозволяє виконувати розбиття об'єктів на групи за набором ознак, де кожна одиниця сукупності розглядається як точка в заданому ознаковому просторі. Потрапляння до одного або різних кластерів j-ї і k-ї одиниць сукупності визначається поняттям відстані між ними, яке характеризує ступінь схожості або однорідності сукупності.

Методологічний принцип класифікації об'єктів, оснований на поняттях схожості і відмінності, містить два основних положення:

- в один клас об'єднуються подібні, схожі між собою одиниці сукупності;
- ступінь схожості, подібності одиниць, які належать до одного класу, є вищою, ніж ступінь схожості одиниць, включених до різних класів.

Кластерний аналіз на основі міри схожості. Для подальших пояснень скористаємося поняттям *кортежу*. У теорії множин (множина – це сукупність деяких різних об'єктів, що розглядаються як одне ціле; природа об'єктів тут не має значення) одним з об'єктів вивчення є так звані впорядковані ($x_1, x_2, \ldots, xn$) множини, де величини $x_1, x_2, \ldots, xn$ є скалярами. Такі множини називаються впорядкованими тому, що порядок елементів у них чітко закріплений і два чи більше елементів можуть бути однаковими. Кількість елементів n у кортежі називається його довжиною. Зазвичай кортежі з двох елементів називаються впорядкованими двійками, з трьох елементів – впорядкованими трійками, з n елементів – впорядкованими n-ками. Елементами кортежу можуть бути об'єкти будь-якої природи. Однак на практиці найбільшу роль відіграють кортежі, елементами яких є неіменовані (тобто без розмірності) числа. Якщо розглядаються декілька таких множин, то вони повинні мати однакову довжину і однакове впорядкування. Тоді такий кортеж можна буде зобразити точкою у n-мірному декартовому просторі. Найпростіший випадок двох двомірних кортежів зображено на рис. 13.1.1.

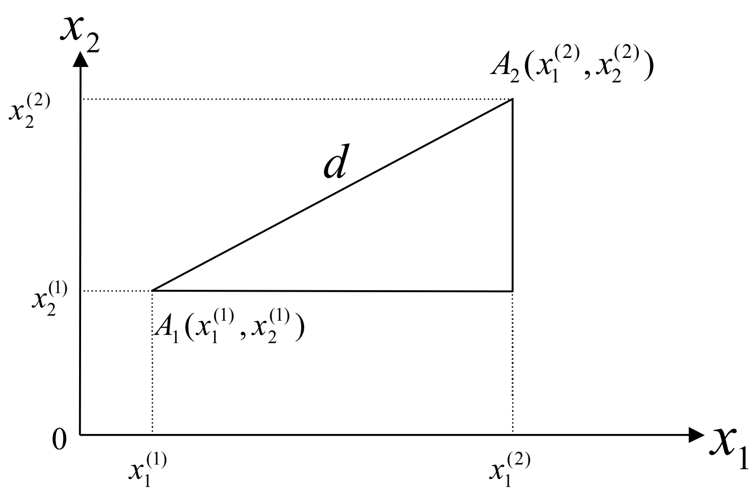


Рисунок 13.1.1 – Геометричне скалярне представлення двох двомірних кортежів

На рис. 13.1.1 точка A_1 відповідає кортежу $(x_1^{(1)}, x_2^{(1)})$, а точка A_2 – кортежу $(x_1^{(2)}, x_2^{(2)})$; d – це відстань між кортежами.

Очевидно, що це представлення можна поширити на будь-яку кількість вимірів n (тобто отримуємо декартовий гіперпростір) і визначити відстань d між точками A_1 і A_2. Для двомірного випадку (за теоремою Піфагора):

$$d^{(1,2)} = \sqrt{(x_1^{(2)} - x_1^{(1)})^2 + (x_2^{(2)} - x_2^{(1)})^2}\,. \qquad \textbf{(13.1.3)}$$

У разі поширення на n вимірів цей результат відстані між двома кортежами називається евклідовою відстанню (узагальнена формула 13.1.4).

$$d = \sqrt{\sum_{n=1}^{\kappa} (x_n^{(i)} - x_n^{(j)})^2}\,, \qquad \textbf{(13.1.4)}$$

де $x_n^{(i)}$ і $x_n^{(j)}$ – значення n-ї ознаки для i-го і j-го об’єктів.
На цій відстані основане поняття розподілу на кластери.

У нашому випадку (у прикладі 13.1.1) об’єктами дослідження (кортежів) є підприємства з впорядкованими характеристиками (x_1, x_2, x_3, x_4). Цілком очевидно, що не можна сумувати квадрати відхилень однієї точки від іншої в абсолютних значеннях різноякісних ознак. Необхідно спочатку виразити відмінності між одиницями сукупності для кожної ознаки у якомусь відносному безвимірному показнику. У якості такого показника часто застосовують нормовану різницю:

$$d_{n_{i,j}} = \frac{x_n^{(i)} - x_n^{(j)}}{\sigma_{x_n}}\,, \qquad \textbf{(13.1.5)}$$

де σ_{x_n} – середнє квадратичне відхилення ознаки x_n.

Знаки нормованих різностей не мають значення, оскільки “відстань” в ознаковому просторі є скалярною, а не векторною величиною.

За даними табл. 13.1.1 (приклад 13.1.1), середнє квадратичне відхилення ознаки x_1 дорівнює 109. Розділивши всі попарні різниці значень цієї ознаки на 109, отримаємо матрицю нормованих різниць D_1 (табл. 13.1.2). Очевидно, що ця матриця є симетричною.

Таблиця 13.1.2 – Матриця нормованих різниць між підприємствами (d_1) за середньомісячною оплатою праці (x_1)

Підприємства	А	Б	В	Г	Д	Е	Ж	З
А	0							
Б	2,239	0						
В	1,780	0,459	0					
Г	3,358	1,119	1,578	0				
Д	2,450	0,211	0,670	0,908	0			
Е	0,523	1,716	1,257	2,835	1,927	0		
Ж	2,064	0,174	0,284	1,294	0,385	1,541	0	
З	1,872	0,367	0,092	1,486	0,518	1,349	0,193	0

Середня нормована різниця $\bar{d}_1$ за даними табл. 13.1.2 склала 0,962.

У нормально розподіленій сукупності нормована різниця у середньому дорівнює одиниці. Це дуже важливо для встановлення граничної (критичної) відстані в ознаковому просторі, при досягненні якої припиняють об’єднання кластерів.

Аналогічно обчислюють матриці нормованих різниць за ознаками x_2, x_3, x_4 (див. табл.13.1.3 –13.1.5).

Таблиця 13.1.3 – Матриця нормованих різниць між підприємствами (d_2) за валовим доходом на 1 га сільськогосподарських угідь (x_2)

Підприємства	А	Б	В	Г	Д	Е	Ж	З
А	0							
Б	2,070	0						
В	2,155	0,085	0					
Г	2,246	0,176	0,092	0				
Д	1,944	0,127	0,211	0,303	0			
Е	1,092	0,972	1,063	1,155	0,852	0		
Ж	0,500	2,552	2,636	2,746	2,444	1,592	0	
З	1,951	0,119	0,203	0,296	0,007	0,859	2,451	0

$\bar{d}_2$ =0,914.

Таблиця 13.1.4 – Матриця нормованих різниць між підприємствами (d_3) за середньорічною вартістю основних виробничих фондів на 1 га сільськогосподарських угідь (x_3)

Підприємства	А	Б	В	Г	Д	Е	Ж	З
А	0							
Б	1,328	0						
В	0	1,328	0					
Г	0,859	0,469	0,859	0				
Д	0,906	0,422	0,906	0,047	0			
Е	0,531	1,859	0,531	1,391	1,438	0		
Ж	1,969	3,297	1,906	2,828	2,875	1,438	0	
З	0,875	0,453	0,875	0,016	0,031	1,406	2,844	0

$\bar{d}_3 = 0{,}936$.

На підставі даних з таблиць 13.1.2 – 13.1.5 формується матриця евклідових відстаней (d) (табл. 13.1.6).

Наприклад, відстань між підприємствами А і Б (з використанням формули 13.1.4) складає:

$$d = \sqrt{\sum_{n=1}^{4} (x_n^{(A)} - x_n^{(Б)})^2} = \sqrt{2{,}239^2 + 2{,}070^2 + 1{,}328^2 + 1{,}024^2} = 3{,}480 \ .$$

І таке інше.

Таблиця 13.1.5 – Матриця нормованих різниць між підприємствами (d_4) щодо відношення дебіторської заборгованості до кредиторської (x_4)

Підприємства	А	Б	В	Г	Д	Е	Ж	З
А	0							
Б	1,024	0						
В	1,122	0,098	0					
Г	0,049	1,073	1,171	0				
Д	2,122	3,146	3,244	2,073	0			
Е	1,122	0,098	0	1,171	3,244	0		
Ж	0,317	1,341	1,439	0,268	1,805	1,439	0	
З	0,244	0,780	0,878	0,293	2,366	0,878	0,561	0

$\bar{d}_4$ =0,927

Таблиця 13.1.6 – Матриця нормованих евклідових відстаней (d)

Підприємства	А	Б	В	Г	Д	Е	Ж	З
А	0							
Б	3,480	0						
В	3,012	1,411	0					
Г	4,130	1,629	2,885	0				
Д	3,887	3,184	3,441	2,284	0			
Е	1,734	2,712	1,373	3,559	4,127	0		
Ж	2,913	4,383	3,568	4,157	4,188	3,008	0	
З	2,852	0,981	1,130	1,541	2,422	2,281	3,775	0

Матриця евклідових відстаней d (табл. 13.1.6) є основою *агломеративно-ієрархічного методу класифікації,* який полягає у послідовному об’єднанні об’єктів, що групуються, – спочатку найближчих один до одного, а потім дедалі більш віддалених один від одного. Процедура класифікації складається з послідовних кроків, на кожному з яких виконується об’єднання двох найближчих груп об’єктів (кластерів). На нульовому кроці кожний об’єкт розглядається як окремий кластер. На першому кроці об’єднаємо у кластер підприємства з найменшою евклідовою відстанню (Б і З). Знайдемо середні всіх ознак для цього кластеру і евклідові відстані від кластеру до інших підприємств (табл. 13.1.7).

Таблиця 13.1.7 – Нормовані відстані і евклідові відстані для кластеру “Б+З”

Підприємства	Значення ознак				Евклідова відстань
	x_1	x_2	x_3	x_4	
Середні величини для кластеру	373	104,5	13,55	46	0
А	2,055	2,011	1,094	0,634	3,141
В	0,275	0,144	1,094	0,488	1,237
Г	1,303	0,236	0,242	0,683	1,509
Д	0,394	0,667	0,195	2,756	2,869
Е	1,532	0,919	1,633	0,488	2,469
Ж	0,003	2,511	3,070	0,951	4,079

Замінивши в матриці евклідових відстаней (табл. 13.1.6) відстані між підприємствами, що увійшли до першого кластеру, на числа останньої графи табл. 13.1.7, бачимо, що зараз мінімальною є відстань між підприємством В і першим кластером: *d*=1,237 (табл. 13.1.8).

Таблиця 13.1.8 – Матриця евклідових відстаней після утворення кластеру “Б+З”

Підприємства	“Б+З”	А	В	Г	Д	Е	Ж
«Б+З»	0						
А	3,141	0					
В	1,237	3,012	0				
Г	1,509	4,130	2,885	0			
Д	2,869	3,887	3,441	2,284	0		
Е	2,469	1,734	1,373	3,559	4,127	0	
Ж	4,079	2,913	3,568	4,157	4,188	3,008	0

Отже, на другому кроці до першого кластеру приєднується підприємство В. Обчислюємо середні величини, нормовані різниці для кожної ознаки і евклідові відстані від кластеру, що включає три підприємства “Б+З+В”, до кожного з решти підприємств. Результати представлено у табл. 13.1.9.

Замінивши евклідові відстані між підприємствами, що увійшли до кластеру, даними останньої графи табл. 13.1.9, отримаємо нову матрицю евклідових відстаней (табл. 13.1.10).

Таблиця 13.1.9 – Нормовані різниці і евклідові відстані для кластеру“Б+З+В”

Підприємства	Значення ознак				Евклідова відстань
	x_1	x_2	x_3	x_4	
Середні величини для кластеру	383	98	15,9	39	0
А	1,963	2,056	0,734	0,805	3,044
Г	1,394	0,190	0,125	0,854	1,651
Д	0,486	0,113	0,172	2,927	2,974
Е	1,440	0,965	1,266	0,317	2,170
Ж	0,101	2,556	2,703	1,122	3,887

Таблица 13.1.10 – Матриця евклідових відстаней після утворення кластеру “Б+З+В”

Підприємства	“Б+З+В”	А	Г	Д	Е	Ж
“Б+З+В”	0					
А	3,044	0				
Г	1,651	4,130	0			
Д	2,974	3,887	2,284	0		
Е	2,170	1,734	3,559	4,127	0	
Ж	3,887	2,913	4,157	4,188	3,008	0

Мінімальною є евклідова відстань від кластеру підприємства Г. На третьому кроці утворюємо кластер “Б+З+В+Г”. Отримані середні величини для кластеру, нормовані різниці і евклідові відстані представлено у табл. 13.1.11 и 13.1.12.

Таблиця 13.1.11 – Нормовані різниці і евклідові різниці для кластеру “Б+З+В+Г”

Підприємства	Значення ознак				Евклідова відстань
	x_1	x_2	x_3	x_4	
Середні величини для кластеру	345	91	15,7	48	0
А	2,312	2,106	0,766	0,585	3,273
Д	0,138	0,162	0,141	2,707	2,719
Е	1,789	1,014	1,297	0,537	2,490
Ж	0,248	2,606	2,734	0,902	3,891

Таблиця 13.1.12 – Матриця евклідових відстаней після утворення кластеру “Б+З+В+Г”

Підприємства	“Б+З+В+Г”	А	Д	Е	Ж
“Б+З+В+Г”	0				
А	3,273	0			
Д	2,719	3,887	0		
Е	2,490	1,734	4,127	0	
Ж	3,891	2,913	4,188	3,008	0

Мінімальною евклідова відстань є між підприємствами Е і А (вона менше 2), отже, ці підприємства об’єднуються в кластер 2 (табл. 13.1.13). Кластер “Б+З+В+Г” називатимемо кластером 1.

Таблиця 13.1.13 – Нормовані різниці і евклідові відстані для кластерів 1 і 2

Підприємства	Значення ознак				Евклідова відстань
	x_1	x_2	x_3	x_4	
Середні кластеру 2	568	312	22,3	49	0
Кластер 1	2,046	1,556	1,031	0,024	2,770
Д	2,183	1,394	1,172	2,683	3,904
Ж	1,798	1,049	1,703	0,878	2,829

Після четвертого кроку отримуємо нову матрицю евклідових відстаней (табл. 13.1.14).

Таблиця 13.1.14 – Матриця евклідових відстаней після утворення кластеру "Б+З+В+Г"

Підприємства	Кластер 1	Кластер 2	Кластер 3 (Д)	Кластер 4 (Ж)
Кластер 1	0			
Кластер 2	2,770	0		
Д	2,719	3,904	0	
Ж	3,891	2,829	4,188	0

Згідно з табл. 13.1.4, усі відстані більше 2. Залишаємо 4 типи підприємств: підприємства, що увійшли в кластер 1; кластер 2; кластер 3 (підприємство Д); кластер 4 (підприємство Ж).

Порівнюючи результати кластерного аналізу з багатомірними середніми (табл. 13.1.1), бачимо, що склад кластеру 1 точно відповідає тим підприємствам, чиї багатомірні середні нижче 100 %. Виділення у самостійний кластер підприємства Ж також відповідає найвищому значенню його багатомірної середньої. А ось об'єднання у кластер 2 підприємств А і Е не відповідає багатомірним середнім, за якими ближчим до підприємства А було б підприємство Д. Внаслідок різкої відмінності за ознакою x_4 підприємство Д виділилося в окремий кластер 3.

Отже, ***послідовність проведення кластерного аналізу є такою***

1. Обчислення середніх величин кожної з класифікованих ознак $\bar{x}_n$ загалом для сукупності.
2. Обчислення середніх квадратичних відхилень кожної ознаки для сукупності σ_{x_n} .

3. Обчислення матриць нормованих різниць для кожної групувальної ознаки $d_{n_{i,j}}$.
4. Обчислення евклідових відстаней між кожною парою сполучень одиниць сукупності d.
5. Вибір найменшої з евклідових відстаней.
6. Об'єднання одиниць сукупності з найменшою евклідовою відстанню між ними в один кластер.
7. Обчислення середніх значень усіх ознак для одиниць, об'єднаних в один кластер.
8. Обчислення нових нормованих відстаней між об'єднаним кластером і рештою одиниць (або кластерів).
9. Обчислення нових евклідових відстаней між об'єднаним кластером і рештою одиниць (або кластерів).
10. Вибір найменшої з евклідових відстаней.
11. Повторення операцій 6 – 10 і таке інше.

Об'єднання в кластери припиняють, коли всі евклідові відстані перевищать задану критичну величину $d_{крит}$.

Розглянута вище методика обчислення евклідової відстані будується на припущенні, що всі ознаки вважаються рівноправними. Але насправді при виділенні типів соціально-економічних явищ групувальні ознаки не є рівноправними (зазвичай деякі ознаки мають більше, а інші – менше значення). Отже, більш досконала методика кластерного аналізу має враховувати різну значимість, різну вагу групувальних ознак за формулою:

$$d = \sqrt{\sum_{n=1}^{\kappa} w_n (x_n^{(i)} - x_n^{(j)})^2}\,, \qquad \textbf{(13.1.6)}$$

де w_n – вага n-ї ознаки.

Проблема визначення w_n є дуже складною і не розглядається у цьому курсі.

Таким чином, сукупність об'єктів – кортежів може бути розбита на кластери, що являють собою вибіркові сукупності кортежів, які знаходяться всередині n-мірних гіперсфер з евклідовими радіусами d, які визначаються у процесі розрахунку.

Векторна модифікація кластерного аналізу. Класичний спосіб кластеризації ґрунтується на тому, що величини ($x_1, x_2, ..., x_n$, d), які нами розглядаються, є скалярами (див. вище).

В аналітичній геометрії показано, що кортеж виду ($x_1, x_2, ...,$, xn) може розглядатися як вектор $\overrightarrow{x^{(1)}}$ у n-мірному просторі. Починається цей вектор на початку координат – точка (0, 0, …, 0), а закінчується в точці з координатами ($x_1, x_2, ...,$, xn). Цих векторів може бути декілька. Припустимо, що ми розглядаємо два кортежі $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$. Найпростіший випадок векторного представлення двомірних кортежів зображено на рис. 13.1.2.

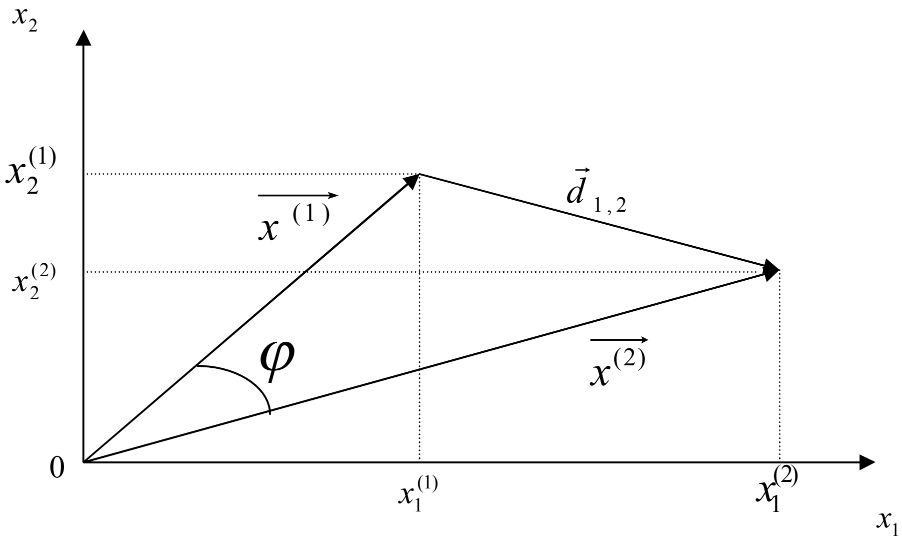


Рисунок 13.1.2 – Геометричне векторне представлення двомірних кортежів

За такої інтерпретації різниця між векторами $\overrightarrow{x^{(2)}}$ і $\overrightarrow{x^{(1)}}$ є вектором $\vec{d}_{1,2}$:

$$\vec{d}_{1,2} = \overrightarrow{x^{(2)}} - \overrightarrow{x^{(1)}} \qquad \textbf{(13.1.7)}$$

Тут з'являється нова якість: якщо при скалярному представленні кортежів різниця між ними характеризується тільки скаляром d, то при векторному представленні кортежів з'являються нові характеристики, а саме, кути нахилів векторів $\vec{d}$, $\overrightarrow{x^{(2)}}$ і $\overrightarrow{x^{(1)}}$ до осей координат і один відносно одного. Кут φ між кортежами $\overrightarrow{x^{(2)}}$ і $\overrightarrow{x^{(1)}}$ дозволяє певною мірою прояснити схожість об'єктів [25].

Наприклад, одне підприємство має кортеж $\overrightarrow{x^{(1)}}=(x_1, x_2, \ldots, x_n)$, а інше підприємство має кортеж $\overrightarrow{x^{(2)}}=(y_1, y_2, \ldots, y_n)$.

Якщо

$$y_i = kx_i, \qquad \textbf{(13.1.8)}$$

де k – коефіцієнт,

то за $k>0$ вектори $\overrightarrow{x^{(2)}}$ і $\overrightarrow{x^{(1)}}$ є односпрямованими ($\overrightarrow{x^{(2)}} \uparrow\uparrow \overrightarrow{x^{(1)}}$, $\varphi = 0^0$), тобто політики підприємств є однаковими;

за $k<0$ вектори $\overrightarrow{x^{(2)}}$ і $\overrightarrow{x^{(1)}}$ є протилежно спрямованими ($\overrightarrow{x^{(2)}} \uparrow\downarrow \overrightarrow{x^{(1)}}$, $\varphi = 180^0$), тобто політики підприємств є протилежними.

Якщо $x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = 0$, то вектори $\overrightarrow{x^{(2)}}$ і $\overrightarrow{x^{(1)}}$ є перпендикулярними ($\overrightarrow{x^{(2)}} \perp \overrightarrow{x^{(1)}}$, $\varphi = \pm 90^0$), тобто політики підприємств є різними, не пов'язаними одна з одною.

Решта співвідношень дає інші кути, більшою або меншою мірою близькі до вказаних.

Визначимо ступінь статистичного зв'язку між кортежами $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$. Для цього застосуємо операцію скалярного добутку між векторами і введемо поняття початкового змішаного моменту для двох кортежів.

Результатом скалярного добутку двох векторів є сума всіх попарних добутків однойменних компонентів кортежів:

$$\| \overrightarrow{x^{(1)}}, \overrightarrow{x^{(2)}} \| = x_1^{(1)} x_1^{(2)} + x_2^{(1)} x_2^{(2)} + x_3^{(1)} x_3^{(2)} + \ldots + x_n^{(1)} x_n^{(2)} . \qquad \textbf{(13.1.9)}$$

Якщо вектори $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$ є однаковими, тобто $\overrightarrow{x^{(1)}} = \overrightarrow{x^{(2)}}$, то вираз 13.1.9 матиме вигляд:

$$\| \overrightarrow{x^{(1)}}, \overrightarrow{x^{(1)}} \| = (x_1^{(1)})^2 + (x_2^{(1)})^2 + (x_3^{(1)})^2 + \ldots + (x_n^{(1)})^2 . \qquad \textbf{(13.1.10)}$$

Формула 13.1.10 є квадратом довжини вектора $\overrightarrow{x^{(1)}}$.

Співвідношення 13.1.9 і 13.1.10 дозволяють визначити косинус кута φ між векторами $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$:

$$\cos\varphi = \frac{\left\| \overrightarrow{x^{(1)}}, \overrightarrow{x^{(2)}} \right\|}{\sqrt{\left\| \overrightarrow{x^{(1)}}, \overrightarrow{x^{(1)}} \right\| \times \left\| \overrightarrow{x^{(2)}}, \overrightarrow{x^{(2)}} \right\|}} . \qquad \textbf{(13.1.11)}$$

Для подальшого викладення скористуємось поняттям моменту. Припустимо, що є деякий кортеж $\vec{x} = (x_1, x_2, \ldots, x_n)$. Початковим моментом k –го порядку для цього кортежу є вираз:

$$M_k(\vec{x}) = \frac{1}{n} \sum_{i=1}^{n} x_i^k , \qquad \textbf{(13.1.12)}$$

де $k = 0, 1, 2, \ldots$.

Для двох кортежів x і y однакової довжини n вводиться поняття початкового змішаного моменту порядку ($k+l$):

$$M_{kl}(\vec{x}, \vec{y}) = \frac{1}{n} \sum_{i=1}^{n} x_i^k y_i^l . \qquad \textbf{(13.1.13)}$$

З порівняння формул 13.1.9 и 13.1.13 слідує, що

$$\| \overrightarrow{x^{(1)}}, \overrightarrow{x^{(2)}} \| = n M_{11}(\overrightarrow{x^{(1)}}, \overrightarrow{x^{(2)}}) . \qquad \textbf{(13.1.14)}$$

Аналогічно, з порівняння формул 13.1.10 і 13.1.13 слідує, що

$$\| \overrightarrow{x^{(1)}}, \overrightarrow{x^{(1)}} \| = n M_2(\overrightarrow{x^{(1)}}) . \qquad \textbf{(13.1.15)}$$

Змішані початкові моменти другого порядку M_{11} і M_2 у вітчизняній літературі називаються коваріаційними моментами.

Тоді формула 13.1.11 набуває вигляду:

$$\cos\varphi = \frac{nM_{11}(\overrightarrow{x^{(1)}}, \overrightarrow{x^{(2)}})}{\sqrt{nM_2(\overrightarrow{x^{(1)}}) \cdot nM_2(\overrightarrow{x^{(2)}})}} \qquad \mathbf{(13.1.16)}$$

або

$$\cos\varphi = \frac{M_{11}(\overrightarrow{x^{(1)}}, \overrightarrow{x^{(2)}})}{\sqrt{M_2(\overrightarrow{x^{(1)}}) \cdot M_2(\overrightarrow{x^{(2)}})}} \quad . \qquad \mathbf{(13.1.17)}$$

Права частина отриманої формули у статистиці має назву коефіцієнта коваріації (в англомовній літературі вона називається коефіцієнтом кореляції). Доведено, що вона показує ступінь статистичного зв’язку між кортежами (або векторами) $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$. Тобто якщо коефіцієнт коваріації дорівнює нулю, то зв'язок між кортежами відсутній; якщо він дорівнює одиниці, то зв'язок є функціональним; якщо він близький до одиниці, то зв'язок є сильним; якщо він близький до нуля, то зв'язок є слабким; а якщо він від’ємний, то зв'язок є оберненим. Отже, кут φ так само дозволяє оцінити ступінь статистичного зв’язку.

Таким чином, сукупність об’єктів – кортежів може бути розбита на кластери, у яких враховуються напрямки кортежів один відносно одного.

На рис. 13.1.3 на колі показано умовні області, що характеризують міру схожості політик підприємств (об’єктів) залежно від змін кута φ між підприємствами $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$.

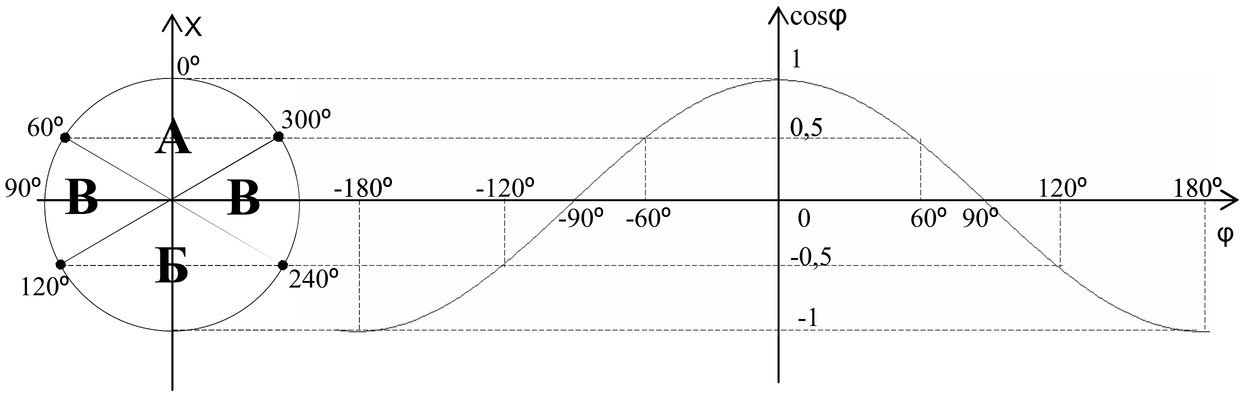


Рисунок 13.1.3 – Значення функції $\cos\varphi$ для кутів φ

На зображеному графіку кола:

- область А відповідає кутам $|\varphi| < 60^0$ і характеризує підприємства як установи з близькою політикою;
- область Б відповідає кутам $120^0 < \varphi < 240^0$ і характеризує підприємства як установи з протилежною політикою;

- область В відповідає кутам $60^0 \le \varphi \le 120^0$, $240^0 \le \varphi \le 300^0$ і характеризує підприємства як установи, політики яких індиферентними (і майже не пов'язаними одна з одною).

Характер зв'язку між кортежами $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$ залежно від коефіцієнта коваріації ($\cos\varphi$) представлено у табл. 13.1.15.

Таблиця 13.1.15 – Ступінь статистичного зв'язку між кортежами $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$

φ	$\cos\varphi$	Характер зв'язку	Політики об'єктів
$\lvert\varphi\rvert < 60^0$	$0{,}5 < \cos\varphi \le 1$	сильна, пряма	близькі
$120^0 < \varphi < 240^0$	$-1 \le \cos\varphi < -0{,}5$	сильна, обернена	протилежні
$60^0 \le \varphi \le 120^0$ $240^0 \le \varphi \le 300^0$	$-0{,}5 \le \cos\varphi \le 0{,}5$	слабка	індиферентні

Розрахунок кутів між кортежами $\overrightarrow{x^{(1)}}$ і $\overrightarrow{x^{(2)}}$ (де кортеж $\overrightarrow{x^{(1)}}$ – це об'єкт, взятий за базу порівняння, а кортеж $\overrightarrow{x^{(2)}}$ – це інші об'єкти, які почергово порівнюються з базисним об'єктом) реалізується за перетвореною формулою 13.1.18:

$$\cos\varphi = cos(\overrightarrow{x^{(1)}}, \overrightarrow{x^{(2)}}) = \frac{x_1^{(1)}x_1^{(2)} + x_2^{(1)}x_2^{(2)} + \ldots + x_n^{(1)}x_n^{(2)}}{\sqrt{((x_1^{(1)})^2 + (x_2^{(1)})^2 + \ldots + (x_{10}^{(1)})^2) \times ((x_1^{(2)})^2 + (x_2^{(2)})^2 + \ldots + (x_n^{(2)})^2)}}. \quad \textbf{(13.1.18)}$$

Приклад 13.1.2: За даними прикладу 13.1.1 (табл. 13.1.1, графи А, 1, 2, 3, 4) провести багатомірне групування підприємств з використанням векторної модифікації кластерного аналізу.

<u>Розв'язок.</u>

Виконаємо нормування значень ознак (x_1, x_2, x_3, x_4) з використанням формули 13.1.19:

$$x_{норм} = \frac{x_n - \bar{x}}{\sigma}. \quad \textbf{(13.1.19)}$$

Це дозволить нам розглядати відносні показники, з якими можна виконувати будь-які математичні розрахунки (табл. 13.1.16).

У табл. 13.1.16, наприклад (за даними таблиці 13.1.1):

$1,789=\dfrac{597-402}{109}$; $1,366=\dfrac{390-196}{142}$ і таке інше.

Ці дані є безрозмірними, що дозволяє досліджувати їх за єдиною схемою представлення підприємств як векторних кортежів з впорядкованими елементами (x_1, x_2, x_3, x_4).

Таблиця 13.1.16 – Розрахункова

Підприємства	Нормовані значення ознак			
	x_1	x_2	x_3	x_4
А	*1*	*2*	*3*	*4*
А	1,789	1,366	0,188	0,122
Б	– 0,450	–0,704	– 1,141	– 0,902
В	0,009	–0,789	0,188	–1,000
Г	–1,569	–0,880	– 0,672	0,171
Д	– 0,661	–0,577	– 0,719	2,244
Е	1,266	0,275	0,719	–1,000
Ж	– 0,275	1,866	2,156	0,439
З	– 0,083	– 0,585	– 0,688	– 0,122

Оберемо базисним підприємство А і розрахуємо відносно нього $cos\varphi$ для решти підприємств з використанням формули 13.1.18 (табл. 13.1.17).

Таблиця 13.1.17 – $cos\varphi$ між підприємствами – кортежами

Підприємства	А (базове підприємство)
А (базове підприємство)	1,000
Б	– 0,551
В	– 0,394
Г	– 0,943
Д	– 0,322
Е	0,656
Ж	0,384
З	– 0,527

Згрупуємо підприємства згідно з табл. 13.1.15 і представимо результати у табл. 13.1.18.

Порівнявши дані табл. 13.1.18 і 13.1.14, можна зробити висновок про те, що отримані результати дещо розходяться одне з одним. Це, очевидно, пов’язано з тим, що і та, і інша кластеризація об’єктів має як переваги, так і недоліки, пов’язані з певними математичними обмеженнями. У будь-якому разі, групування об’єктів на основі багатьох ознак є досить складним процесом, який потребує комплексного і всебічного підходу для отримання більш точних результатів.

Таблиця 13.1.18 – Кластеризація підприємств на підставі міри схожості – кута нахилу між векторними кортежами

<table>
<tr><th>Підприємства</th><th>Політики підприємств</th></tr>
<tr><td>А</td><td rowspan="2">близькі
(кластер 1)</td></tr>
<tr><td>Е</td></tr>
<tr><td>Б</td><td rowspan="3">протилежні
(кластер 2)</td></tr>
<tr><td>Г</td></tr>
<tr><td>З</td></tr>
<tr><td>В</td><td rowspan="3">індиферентні
(кластер 3)</td></tr>
<tr><td>Д</td></tr>
<tr><td>Ж</td></tr>
</table>

Оброблення інформації з використанням комп’ютерних технологій.

Наведений нижче матеріал пропонується як приклад оброблення великих обсягів інформації, які неможливо розрахувати без певного програмного забезпечення.

Пропонується дослідити сукупність з 70 банків України, кожен з яких характеризується десятьма показниками–скалярами (активами (1), зобов’язаннями (2), кредитно-інвестиційним портфелем (3), кредитами юридичним (4) і фізичним (5) особам, власним капіталом (6), статутним капіталом (7), депозитами фізичних (8) і юридичних (9) осіб, фінансовим результатом (10)), які повно і об’єктивно характеризують політику банків станом на 01.12.05 р. Це реальні дані, взяті з офіційного сайту Асоціації українських банків і представлені у додатку 3 (табл. 3.1) [26]. При вивченні цих даних видно, що кожний банк має свої характеристики, відмінні від інших банків. Очевидно, що розкид цифр тут є величезним. Крім цього, для характеристики параметрів діяльності банків використовують різні одиниці вимірювання (млн. грн, млн. євро). Аби запобігти великому розкиду і втраті даних, було застосовано метод нормування з використанням формули 13.1.19. У результаті отримано величини, представлені у додатку 3 (табл. 3. 2). Ці дані є безрозмірними, що дозволяє дослідити їх за єдиною схемою представлення банків як кортежів з впорядкованими елементами ($x_1, x_2, \ldots, x_{10}$), де:

x_1 – нормовані активи;

x_2 – нормовані зобов’язання;

x_3 – нормований кредитно-інвестиційний портфель;

x_4 – нормовані кредити юридичним особам;

x_5 – нормовані кредити фізичним особам;
x_6 – нормований капітал;
x_7 – нормований статутний капітал;
x_8 – нормовані депозити фізичних осіб;
x_9 – нормовані депозити юридичних осіб;
x_{10} – нормований фінансовий результат.

Фрагмент округлених значень, що містяться у додатку 3 (табл. 3.2) представлено у таблиці 13.1.19.

Таблиця 13.1.19 – Нормовані показники діяльності банків станом на 01.12.05 р. (фрагмент)

Банк	x_1	x_2	x_3	І т.д.	x_8	x_9	x_{10}
АВАЛЬ (01.01.04)	2,2	2,7	2,1	-	2,3	1,9	0,0
УКРСИББАНК	2,2	2,1	2,2	-	1,2	1,7	0,4
РАЙФФАЙЗЕН-БАНК Україна	1,3	1,3	1,4	-	0,4	0,9	0,9
І т.д.	-	-	-	-	-	-	-
ПРИВАТБАНК	5,1	5,0	5,3	-	5,2	3,0	8,1

Показники банку АВАЛЬ (на 01.01.04 р.) представлено у табл. 13.1.19, аби у подальшому розглядати не лише міру схожості банків один відносно одного, а й відносно базисного банку, яким обрано АВАЛЬ зі значеннями станом на 01.01.04 р. Тобто до сукупності з 70 банків додано й базисний АВАЛЬ (у результаті розглядається сукупність з 71 банку).

Класифікацію досліджуваної сукупності банків можна здійснити з використанням процедур Кластерний аналіз (алгоритм К-means) і Дискримінантний аналіз програми SPSS 12.0 для Windows.

Алгоритм К-means – це автоматичний метод розбиття на кластери, де в якості початкового розбиття обирається гіпотеза про існування конкретної кількості кластерів, які необхідно виділити. На першому кроці вказується ця кінцева кількість кластерів (гіпотеза про К-кластери). У нашому випадку кількість кластерів дорівнює 4 (саме на 4 групи за активами найчастіше групують комерційні банки економісти: найбільші банки, великі банки, середні банки і дрібні банки, припускаючи, що принципи роботи цих банків всередині груп є схожими).

На другому кроці для кожного з кластерів обчислюють центр кластеру. Наприклад, у якості центру може бути обрано довільний об’єкт. На практиці центр кластеру обчислюють на основі випадкових вибірок, які виділяють з усієї сукупності аналізованих банків. Після визначення центрів кластерів проглядають об'єкти з аналізованої сукупності банків. Відповідно до встановленої міри схожості обирають той об'єкт, який має кращу міру схожості відносно інших об'єктів і кластерів, що формуються. На підставі обраної міри об'єкт включають до одного з кластерів. Далі процес розрахунків триває, поки всі об'єкти не будуть рознесені по кластерах. Близькість між банками (кортежами) як об'єктами, що характеризуються десятьма ознаками у декартовому гіперпросторі, який має евклідову метрику, оцінювалась за допомогою евклідової відстані між десятимірними кортежами за узагальненою формулою 13.1.4.

За допомогою дискримінантного аналізу отриманої класифікації роблять імовірнісне оцінювання, тобто передбачать кластери, до яких належатимуть банки, та імовірність включення банків до певного кластеру (додаток 3, табл. 3.3).

Результати кластеризації кортежів станом на 01.12.05 р. за десятьма ознаками представлено у табл. 13.1.20.

Таблиця 13.1.20 – Кластеризація банків України за 10 ознаками станом на 01.12.05 р. (міра схожості – евклідова відстань)

Кластер 1	**Кластер 2**	**Кластер 3**	**Кластер 4**
АВАЛЬ	АВАЛЬ (1.01.04); УКРСИББАНК; РАЙФФАЙЗЕНБАНК УКРАЇНА; УКРСОЦБАНК; ФОРУМ	ІНШІ	ПРИВАТБАНК
Усього – 1 банк	Усього – 5 банків	Усього – 64 банки	Усього –1 банк

Якщо порівнювати групування банків, представлене в табл. 13.1.20, з групуванням банків, здійсненим Національним банком України (табл. 13.1.21), то стає очевидним, що при характеристиці принципів діяльності об'єктів необхідно враховувати комплекс показників. Тоді результат буде точнішим.

Таблиця 13.1.21 – Класифікація банків за активами, здійснена НБУ станом на 01.12.05 р.

Вище 2500 млн.грн.	Від 1300 до 2500 млн.грн.	Від 400 до 1300 млн.грн.	Нижче 400 млн.грн.
АВАЛЬ (1.01.04); АВАЛЬ; ПРИВАТБАНК; УКРСИББАНК; РАЙФФАЙЗЕН-БАНК УКРАЇНА; УКРСОЦБАНК; ОЩАДБАНК; НАДРА; ФІНАНСИ І КРЕДИТ; УКРПРОМБАНК; БРОКБІЗНЕСБАНК; ФОРУМ; ПУМБ; КРЕДИТПРОМ- БАНК; ХРЕЩАТИК	УКРГАЗБАНК; ДОНГОРБАНК; ПІВДЕННИЙ; ВАБАНК; КРЕДИТ БАНК (УКРАЇНА); АЛЬФА-БАНК; МРІЯ; ТАС-КОМЕРЦ- БАНК; ПРАВЕКС-БАНК; СІТІБАНК УКРАЇНА; ЕКСПРЕС-БАНК; ХФБ УКРАЇНА; КИЇВ; ІМЕКС-БАНК	ІНДУСТРІАЛ- БАНК; КРЕДИТ-ДНІПРО; БІГ ЕНЕРГІЯ; МОРСЬКИЙ ТРАНСПОРТНИЙ БАНК; МЕГАБАНК; ТАВРІКА; ЗАХІДІНКОМ- БАНК; ЕКСПОБАНК; УКРІНБАНК; ДІАМАНТ; АЖІО; ФАКТОРІАЛ-БАНК; АРКАДА; ЕЛЕКТРОН- БАНК; БАЗИС; ЗОЛОТІ ВОРОТА; УКРАЇНСЬКИЙ ПРОФЕСІЙНИЙ БАНК; ТРАНСБАНК; ПОЛТАВА-БАНК; МІЖНАРОДНИЙ КОМЕРЦІЙНИЙ БАНК; НАЦІОНАЛЬНІ ІНВЕСТИЦІЇ	ІНШІ
Усього – 15 банків	Усього – 14 банків	Усього – 21 банк	Усього –21 банк

13.2
КОНТРОЛЬНІ ПИТАННЯ І ВПРАВИ

1. У чому полягає сутність групування?

2. Принципи групування об’єктів на основі багатомірних середніх.

3. Кластерний аналіз, способи проведення.
4. Кластерний аналіз на основі міри схожості – евклідової відстані.
5. Необхідність векторної модифікації кластерного аналізу.
6. Основні прийоми нормування даних, їхній смисл.
7. За даними таблиці 3.1 додатку 3 провести кластеризацію об’єктів за будь-якими 4 ознаками.
8. За даними таблиці 3.1 додатку 3 провести кластеризацію об’єктів за будь-якими 7 ознаками.
9. За даними таблиці 3.1 додатку 3 провести кластеризацію банків за активами. Порівняти отримані результати з групуванням банків НБУ (табл. 13.1.21).
10. За даними таблиці 3.1 додатку 3 провести класифікацію об’єктів методом багатомірних середніх. Порівняти отримані результати з даними таблиці 13.1.20.

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Андерсон. Т. Статистический анализ временных рядов / Т. Андерсон. – М. : Мир, 1976. – 760 с.
2. Андерсон. Т. Введение в многомерный статистический анализ / Т. Андерсон. – М. : Физматгиз, 1963. – 500 с.
3. Афифи А. Статистический анализ: подход с использованием ЭВМ: [пер. с англ.]. / А. Афифи, С. Эйзен. – М. : Мир, 1982. – 488 с.
4. Боровиков В.П. STATISTICA: Статистический анализ и обработка данных в среде Windows / В.П. Боровиков, И.П. Боровиков. – М. : Информационно - издательский дом “Филинъ”, 1997. – 608 с.
5. Боровиков В.П. Искусство анализа данных на компьютере. Для профессионалов / В.П. Боровиков. – СПб. : Питер, 2001. – 656 с.
6. Боровиков В.П. Программа STATISTICA для студентов и инженеров / В.П. Боровиков. – 2-е изд. – М. : КомпьютерПресс, 2001. – 301 с.
7. Бююль А. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей / А. Бююль, П. Цефель. – СПб. : ДиаСофт, 2005. – 608 с.
8. Вентцель Е.С. Теория вероятностей: учеб. для вузов / Е.С. Вентцель. – 7-е изд. стер. – М. : Высш. шк., 2001. – 575 с.
9. Гусаров В.М. Статистика: учеб. пособие для вузов / В.М. Гусаров. – М. : ЮНИТИ-ДАНА, 2002. – 463 с.
10. Дюран Б. Кластерный анализ: [пер. с англ.]. / Б. Дюран, П. Одел. – М. : Статистика, 1977. – 128 с.
11. Дюран Б. Факторный, дискриминантный, кластерный анализ: [пер. с англ.]. / Б. Дюран, П. Одел. – М. : Финансы и статистика, – 1989. – 215 с.
12. Елисеева И.И. Группировка, корреляция, распознавание образов: статистические методы и измерения связей / И.И Елисеева. – М. : Статистика, 1977. – 143 с.
13. Елисеева И.И. Статистические методы измерения связей / И.И Елисеева; под ред. А.Н. Жигарева. – Л. : Изд-во Ленинградского ун-та, 1982. – 134 с.
14. Елисеева И.И. Общая теория статистики: учебник / И.И Елисеева, М.М. Юзбашев; под ред. чл.-корр. РАН И.И. Елисеевой. – 4-е изд. – М. : Финансы и статистика, 2002. – 480 с.
15. Ерина А.М. Теория статистики: Практикум / А.М. Ерина, З.О. Пальян. – К. : Общество «Знання», 2001. – 267 с.
16. Общая теория статистики: учебник / М.Р. Ефимова [и др.]. – 2-е изд. – М. : ИНФРА-М, 2002. – 416 с.

17.Колмогоров А.Н. Основные понятия теории вероятностей / А.Н. Колмогоров. – М. : Наука, 1974. – 119 с.
18.Кулинич О.І. Теорія статистики: підручник / О.І .Кулинич, Р.О. Кулинич. – 3-тє вид., переробл. і доповн. – К. : Знання, 2006. – 294 с.
19.Марюта О.М. Статистические методы и модели в экономике: монография / О.М. Марюта, Н.Е. Бойцун. – Днепропетровск: Пороги, 2002. – 384 с.
20.Моторин Р.М. Статистика для економістів: навч. посіб. / Р.М. Моторин, Е.В. Чекотовський. – К. : Знання, 2009. – 430 с.
21.Практикум по теории статистики: учеб. пособие / под ред проф. Р.А. Шмойловой. – М. : Финансы и статистика, 2000. – 416 с.
22.Салин В.Н. Практикум по курсу «Статистика»: учеб. пособие / В.Н. Салин, Э.Ю. Чурилова. – М. : Перспектива, 2002. – 188 с.
23.Теория статистики: учебник / под ред. проф. Р.А. Шмойловой. – М. : Финансы и статистика, 2001. – 240 с.
24.Ферстер Э. Методы корреляционного и регрессионного анализа / Э. Ферстер, Б. Ренц; пер. В.М. Ивановой. – М. : Финансы и статистика, 1983. – 302 с.
25. Харкевич А.А. Борьба с помехами / А.А. Харкевич. – М. : Изд-во Наука, 1965. – 274 с.
26.http://www.aub.com.ua

ДОДАТОК 1

Теорема П.Л.Чебишева (з уточненнями А.М. Ляпунова) [9, 16]:

з імовірністю, скільки завгодно близької до одиниці, можна стверджувати, що за достатньо великого обсягу вибірки і обмеженої генеральної дисперсії вибіркові узагальнені показники (середня, частка) будуть скільки завгодно мало відрізнятися від відповідних генеральних показників.

Для находження *середнього* значення ознаки ця теорема може бути записана так:

$$\mathrm{P}\left[\left|\tilde{X} - \overline{X}\right| \le \Delta_{\tilde{x}}\right] = \Phi(t),$$

а для находження *частки* ознаки:

$$\mathrm{P}\left[\left|w - p\right| \le \Delta_{p}\right] = \Phi(t),$$

де $\Phi(t) = \dfrac{1}{2\pi}\displaystyle\int_{-t}^{+t} e^{-\frac{x^2}{2}} dx$ – нормована функція Лапласа

Значення функції $\Phi(t)$ за різних значень t як коефіцієнта кратності середньої помилки вибірки визначаються на основі спеціально складених таблиць. Наведемо деякі значення, найчастіше використовувані для вибірок достатньо великого обсягу ($n \ge 30$).

Таблица 1.1

t	1,000	1,960	2,000	2,580	3,000
$\Phi(t)$	0,683	0,950	0,954	0,990	0,997

Метод найменших квадратів (МНК) – розроблений К.Ф. Гауссом [8, 14]:

виходячи з нормального закону помилок вимірювання і вимоги щодо максимальної імовірності певної сукупності помилок, вимога щодо найкращого узгодження кривої $y = \varphi(x)$ і експериментальних точок зводиться до того, **аби сума квадратів відхилень експериментальних точок від згладжувальної кривої перетворювалась на мінімум:**

$$\sum_{i=1}^{n}[y_i - \varphi(x_i)]^2 = min$$ – “принцип максимальної правдоподібності”.

ДОДАТОК 2

Нормальний закон розподілу (закон Гаусса) [8, 16]**:**

головна особливість, що виділяє нормальний закон з-серед інших законів, полягає у тому, що він є *граничним законом*, до якого наближуються інші закони розподілу за типових умов, які доволі часто мають місце. **Нормальний розподіл є можливим тоді, коли на величину ознаки впливає велика кількість випадкових причин. Дія цих причин є незалежною, і жодна з причин не має переважного впливу порівняно з іншими.**

Якщо неперервна випадкова величина має щільність розподілу:

$$f(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} ,$$

то вона підпорядковується закону нормального розподілу.

Особливості кривої нормального розподілу

1. Крива є симетричною відносно максимальної ординати. Максимальна ордината відповідає значенню $x=M_0=Me$, її величина дорівнює $\frac{1}{\sqrt{2\pi}\,\sigma}$.

2. Крива асимптотично наближується до вісі абсцис, продовжуючись в обидві сторони до безкінечності. Отже, чим більше значення відхиляються від $\bar{x}$, тим рідше вони зустрічаються. Однакові за абсолютним значенням, але протилежні за знаком відхилення значень змінної x от $\bar{x}$ є рівноімовірними

3. Крива має дві точки перегину, що знаходяться на відстані $\pm\sigma$ від $\bar{x}$.

4. За $\bar{x} = const$, зі зростанням σ крива стає більш пологою. За $\sigma = const$, зі зміною $\bar{x}$ крива не змінює свою форму, а лише зрушується вправо або вліво по вісі абсцис.

5. У проміжку $\bar{x} \pm \sigma$ знаходяться 68,3 % усіх значень ознаки. У проміжку $\bar{x} \pm 2\sigma$ знаходяться 95,4 % усіх значень ознаки. У проміжку $\bar{x} \pm 3\sigma$ знаходяться 99,7 % усіх значень ознаки.

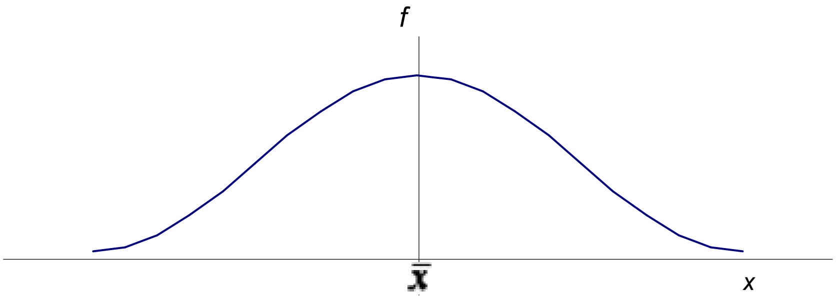


Рисунок 2.1 – Крива нормального розподілу

ДОДАТОК 3

Таблиця 3.1 – Показники діяльності банків на 01.12.05 р.

Банк	Активи (млн.грн.)	Зобов'я-зання (млн. грн.)	КІП (млн. грн.)	Кредити юридич-ним особам (млн. грн.)	Кредити фізичним особам (млн.грн.)	Капітал (млн. грн.)	Статутний капітал (млн. ЄВРО)	Депозити фізичних осіб (млн.грн.)	Депозити юридич-них осіб (млн.грн.)	Фінансовий результат (млн.грн.)
А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
АВАЛЬ (01.01.04)	9923,5	11293,1	7433,2	5372,4	1398,6	877,7	120,1	4291,0	2929,5	11,7
УКРСИББАНК	9941,3	9020,7	7710,2	4205,5	2739,0	1104,7	125,9	2410,4	2665,1	28,9
РАЙФФАЙЗЕН-БАНК УКРАЇНА	6452,9	5852,0	5274,7	3682,4	1389,4	769,1	87,3	1021,2	1575,2	58,1
ОЩАДБАНК	8347,9	7960,0	4947,1	330,9	1207,2	373,0	51,0	5257,0	2221,3	32,4
УКРПРОМБАНК	3872,7	3243,0	3352,8	3095,2	256,9	644,0	99,1	1848,0	1053,4	20,9
ФІНАНСИ І КРЕДИТ	4105,9	3620,4	3388,6	2339,0	610,9	570,6	70,4	1653,9	1265,3	11,6
НАДРА	5232,4	4844,1	4577,9	2409,5	1441,1	484,3	13,6	2302,4	817,3	22,4
БРОКБІЗНЕСБАНК	4086,3	3614,4	3314,2	1705,6	765,0	479,4	58,8	972,6	1928,3	15,4
ФОРУМ	3308,5	2961,0	2690,6	1915,1	287,5	402,4	45,3	928,2	1118,7	27,8
УКРСОЦБАНК	14141,7	13137,5	7608,0	4133,5	2714,8	996,5	11,6	3279,9	8137,9	95,8
КРЕДИТПРОМ-БАНК	2684,4	2359,2	2136,9	1208,0	428,6	329,7	41,8	641,2	774,9	13,4
ХРЕЩАТИК	3043,6	2820,0	1585,3	1127,1	106,1	223,7	28,9	476,8	2041,8	18,7
ПУМБ	2666,1	2190,4	2213,1	1681,9	57,3	462,0	3,9	603,3	1161,3	38,6
УКРГАЗБАНК	2278,6	2068,6	1866,4	1165,8	158,5	260,2	28,5	959,8	561,7	9,7
ПІВДЕННИЙ	2015,5	1815,6	1647,0	1477,0	129,6	207,0	28,0	725,7	757,1	24,1
КРЕДИТ БАНК (УКРАЇНА)	1936,6	1781,7	1566,9	1200,4	260,5	191,2	24,1	854,2	520,6	6,7
ДОНГОРБАНК	2121,4	1843,8	1568,0	932,9	59,9	280,2	32,6	281,0	1153,9	63,8

Продовження табл. 3.1

А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
МРІЯ	1801,6	1588,9	1459,7	1224,7	158,8	214,2	21,6	666,4	663,0	30,8
ТАС-КОМЕРЦБАНК	1815,5	1608,4	1455,5	890,0	282,2	205,5	28,1	516,2	700,4	15,0
ВАБАНК	2026,6	1858,5	1615,7	1084,8	119,2	218,5	13,4	524,3	631,1	14,4
АЛЬФА-БАНК	1823,3	1625,7	1453,5	1208,7	30,1	256,7	21,3	75,8	599,7	2,6
ПРАВЕКС-БАНК	1945,6	1793,0	1347,3	225,8	1068,5	189,0	15,9	1240,4	405,7	11,3
ІМЕКС-БАНК	1432,8	1294,1	1016,1	779,7	70,5	156,2	17,6	577,5	546,0	9,0
КИЇВ	1373,3	1227,4	1081,1	934,4	102,9	146,3	10,1	656,3	435,2	19,9
ЕКСПРЕС-БАНК	1452,2	1270,7	1083,8	620,7	244,1	180,6	8,3	401,6	653,9	18,2
СІТІБАНК УКРАЇНА	1706,1	1509,6	1347,6	821,2	7,0	197,9	8,4	2,9	701,6	45,3
ХФБ УКРАЇНА	1366,6	1223,5	1123,1	978,2	35,5	187,4	13,3	19,4	487,2	38,3
КРЕДИТ-ДНІПРО	1148,2	1016,0	886,0	776,3	87,7	123,6	12,4	278,3	654,8	16,8
ІНДУСТРІАЛБАНК	1089,3	897,5	775,0	431,1	18,3	192,7	16,9	105,3	482,9	41,9
МОРСЬКИЙ ТРАНСПОРТНИЙ БАНК	863,6	724,5	668,6	463,1	77,1	145,7	8,4	197,2	372,6	20,5
МЕГАБАНК	792,7	681,8	588,4	424,5	62,3	128,1	7,6	339,4	269,6	5,0
УКРІНБАНК	663,3	560,1	469,5	384,2	66,4	98,9	8,4	299,2	237,9	5,0
ТАВРІКА	625,3	543,3	585,0	433,6	71,0	82,1	8,6	269,1	190,4	7,6
ЕКСПОБАНК*	638,8	544,6	501,7	464,9	32,9	102,2	7,9	198,7	155,7	7,2
ЗАХІДІНКОМБАН К	696,3	623,6	553,1	386,6	124,5	72,6	4,6	272,4	242,4	13,8
ДІАМАНТ	565,2	445,0	467,9	270,2	23,7	120,9	16,8	101,7	156,4	8,8
ФАКТОРІАЛ-БАНК	612,7	523,1	449,7	247,5	73,0	88,8	9,6	248,2	167,5	1,3
АРКАДА	467,5	353,5	332,4	65,1	246,1	104,8	12,8	83,3	249,5	14,3
ЗОЛОТІ ВОРОТА	465,5	414,1	369,2	269,0	95,4	50,9	8,2	185,1	208,2	1,4
АЖІО	627,0	499,7	439,4	296,2	27,1	105,7	2,1	157,0	124,7	1,7
УКРАЇНСЬКИЙ ПРОФЕСІЙНИЙ БАНК	422,4	334,4	300,8	249,4	3,9	87,6	11,8	123,5	113,5	3,6

Продовження табл. 3.1

A	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
БАЗИС	487,9	441,9	351,5	279,8	49,0	60,5	5,1	192,0	223,3	4,6
ЕЛЕКТРОНБАНК	471,4	384,3	344,6	194,0	119,6	86,9	6,7	254,6	75,2	1,7
МІЖНАРОДНИЙ КОМЕРЦІЙНИЙ БАНК	451,4	399,5	347,5	268,1	58,7	48,1	5,9	233,7	121,2	2,4
ПОЛТАВА-БАНК	435,8	372,0	339,6	287,4	34,8	63,8	3,4	234,5	115,5	7,2
ТРАНСБАНК	432,4	372,1	359,2	245,6	43,7	59,8	6,1	160,5	121,1	3,4
МЕРКУРІЙ	340,5	280,8	245,2	173,4	62,5	60,7	7,5	126,2	135,4	4,1
АКТИВ-БАНК	377,3	329,8	283,2	198,3	41,0	54,2	6,2	135,5	110,3	4,7
ЕНЕРГОБАНК	407,4	362,5	232,4	161,2	9,8	55,9	5,8	133,7	200,7	2,7
СИНТЕЗ	265,5	194,5	213,8	96,5	114,2	75,5	10,7	114,6	69,7	2,0
ФОРТУНА-БАНК	301,3	246,6	252,4	202,1	23,2	54,7	8,4	104,7	96,2	3,9
НАЦІОНАЛЬНІ ІНВЕСТИЦІЇ	413,4	375,9	326,4	202,8	61,2	45,8	3,1	82,5	178,8	5,8
СХІДНО-ЄВРОПЕЙСЬКИЙ	309,6	257,3	245,6	169,0	23,4	57,2	7,6	110,3	54,0	2,5
КАПІТАЛ	305,0	266,3	247,8	134,1	62,3	43,7	4,9	110,2	94,4	1,4
ІНТЕГРАЛ	303,4	252,4	275,0	104,2	38,3	51,0	6,6	69,1	45,0	3,1
МЕТАЛУРГ	268,5	177,6	202,5	139,9	42,2	90,2	2,2	117,8	48,9	12,5
РЕАЛ-БАНК	197,8	127,3	132,1	78,6	1,4	69,9	9,6	19,6	43,7	1,0
ІКАР-БАНК	167,2	118,8	113,3	40,3	73,0	49,5	5,9	72,0	41,0	3,4
ЗЕМЕЛЬНИЙ БАНК	156,7	108,5	140,8	89,4	1,8	48,4	6,2	22,9	71,3	1,8
ПРИКАРПАТТЯ	188,3	152,2	129,6	97,2	30,8	39,0	3,4	90,8	46,1	1,4
УКРАЇНСЬКИЙ КРЕДИТНО-ТОРГОВИЙ БАНК	147,8	108,4	107,2	71,1	11,2	46,2	5,7	34,2	30,6	1,3
ПОЛІКОМБАНК	147,3	107,0	118,0	106,0	8,6	39,7	3,8	62,9	28,7	0,1
УКООПСПІЛКА	116,7	67,3	73,0	56,2	9,3	49,8	4,3	13,4	43,2	1,4
РЕГІОН-БАНК	103,8	67,1	75,5	60,6	10,9	36,7	4,0	36,0	27,5	1,3

Продовження табл. 3.1

А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
СХІДНО-ПРОМИСЛОВИЙ*	86,5	58,1	68,3	53,7	9,7	29,7	4,3	34,3	15,2	1,7
СЛАВУТИЧ	108,3	78,1	69,2	38,7	13,4	30,2	3,7	22,6	7,4	0,6
МОРСЬКИЙ	96,4	64,0	70,9	32,1	4,3	32,3	4,0	17,1	20,9	0,3
ТММ-БАНК	67,4	36,2	38,0	24,6	13,4	31,5	4,4	1,2	22,7	2,6
АВАЛЬ	19102,1	17292,9	16057,7	7754,6	3919,8	2166,7	251,0	8672,7	6290,3	1,0
ПРИВАТБАНК	21719,2	19606,1	17437,7	8175,2	6149,9	2165,4	189,7	9210,6	4416,5	437,1

Таблиця 3.2 – Нормовані показники діяльності банків на 01.12.05 р.

Банк	Активи	Зобов'язання	КІП	Кредити юридичним особам	Кредити фізичним особам	Капітал	Статутний капітал)	Депозити фізичних осіб	Депозити юридичних осіб	Фінансовий результат
А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
АВАЛЬ (01.01.04)	2,2	2,7	2,1	3,0	1,2	1,9	2,5	2,3	1,9	0,0
УКРСИББАНК	2,2	2,1	2,2	2,3	2,6	2,4	2,7	1,2	1,7	0,4
РАЙФФАЙЗЕНБАНК УКРАЇНА	1,3	1,3	1,4	2,0	1,2	1,6	1,8	0,4	0,9	0,9
ОЩАДБАНК	1,8	1,8	1,3	-0,1	1,0	0,6	0,9	2,9	1,3	0,4
УКРПРОМБАНК	0,7	0,6	0,8	1,6	0,0	1,3	2,0	0,9	0,5	0,2
ФІНАНСИ І КРЕДИТ	0,7	0,7	0,8	1,1	0,4	1,1	1,4	0,7	0,6	0,0
НАДРА	1,0	1,0	1,2	1,2	1,3	0,9	0,0	1,1	0,3	0,2
БРОКБІЗНЕСБАНК	0,7	0,7	0,8	0,7	0,6	0,9	1,1	0,3	1,1	0,1
ФОРУМ	0,5	0,5	0,6	0,9	0,1	0,7	0,8	0,3	0,5	0,3
УКРСОЦБАНК	3,2	3,2	2,2	2,2	2,6	2,2	0,0	1,7	5,7	1,6
КРЕДИТПРОМБАНК	0,4	0,3	0,4	0,4	0,2	0,5	0,7	0,1	0,3	0,1
ХРЕЩАТИК	0,5	0,5	0,2	0,4	-0,1	0,2	0,4	0,0	1,2	0,2
ПУМБ	0,4	0,3	0,4	0,7	-0,2	0,8	-0,2	0,1	0,6	0,5

Продовження табл. 3.2

A	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
УКРГАЗБАНК	0,3	0,3	0,3	0,4	-0,1	0,3	0,4	0,3	0,1	0,0
ПІВДЕННИЙ	0,2	0,2	0,2	0,6	-0,1	0,2	0,4	0,2	0,3	0,3
КРЕДИТ БАНК (УКРАЇНА)	0,2	0,2	0,2	0,4	0,1	0,1	0,3	0,3	0,1	-0,1
ДОНГОРБАНК	0,2	0,2	0,2	0,2	-0,2	0,4	0,5	-0,1	0,6	1,0
МРІЯ	0,2	0,1	0,2	0,4	-0,1	0,2	0,2	0,1	0,2	0,4
ТАС-КОМЕРЦБАНК	0,2	0,1	0,2	0,2	0,1	0,2	0,4	0,1	0,2	0,1
ВАБАНК	0,2	0,2	0,2	0,3	-0,1	0,2	0,0	0,1	0,2	0,1
АЛЬФА-БАНК	0,2	0,1	0,2	0,4	-0,2	0,3	0,2	-0,2	0,2	-0,1
ПРАВЕКС-БАНК	0,2	0,2	0,1	-0,2	0,9	0,1	0,1	0,5	0,0	0,0
ІМЕКС-БАНК	0,1	0,1	0,0	0,2	-0,1	0,1	0,1	0,1	0,1	0,0
КИЇВ	0,0	0,0	0,1	0,3	-0,1	0,0	-0,1	0,1	0,0	0,2
ЕКСПРЕС-БАНК	0,1	0,0	0,1	0,1	0,0	0,1	-0,1	0,0	0,2	0,2
СІТІБАНК УКРАЇНА	0,1	0,1	0,1	0,2	-0,2	0,2	-0,1	-0,2	0,2	0,7
ХФБ УКРАЇНА	0,0	0,0	0,1	0,3	-0,2	0,1	0,0	-0,2	0,1	0,5
КРЕДИТ-ДНІПРО	0,0	0,0	0,0	0,2	-0,1	0,0	0,0	-0,1	0,2	0,1
ІНДУСТРІАЛБАНК	0,0	-0,1	0,0	-0,1	-0,2	0,1	0,1	-0,2	0,1	0,6
БІГ ЕНЕРГІЯ	-0,1	-0,1	0,0	0,0	-0,1	0,0	0,1	0,0	-0,2	-0,1
МОРСЬКИЙ ТРАНСПОРТНИЙ БАНК	-0,1	-0,1	-0,1	0,0	-0,1	0,0	-0,1	-0,1	0,0	0,2
МЕГАБАНК	-0,1	-0,1	-0,1	-0,1	-0,2	0,0	-0,1	0,0	-0,1	-0,1
УКРІНБАНК	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1
ТАВРІКА	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	0,0
ЕКСПОБАНК	-0,1	-0,1	-0,1	0,0	-0,2	-0,1	-0,1	-0,1	-0,2	-0,1
ЗАХІДІНКОМБАНК	-0,1	-0,1	-0,1	-0,1	-0,1	-0,2	-0,2	-0,1	-0,1	0,1
ДІАМАНТ	-0,2	-0,2	-0,1	-0,2	-0,2	0,0	0,1	-0,2	-0,2	0,0
ФАКТОРІАЛ-БАНК	-0,1	-0,2	-0,1	-0,2	-0,1	-0,1	-0,1	-0,1	-0,2	-0,2
АРКАДА	-0,2	-0,2	-0,2	-0,3	0,0	-0,1	0,0	-0,2	-0,1	0,1
ЗОЛОТІ ВОРОТА	-0,2	-0,2	-0,2	-0,2	-0,1	-0,2	-0,1	-0,1	-0,1	-0,2
АЖІО	-0,1	-0,2	-0,2	-0,1	-0,2	-0,1	-0,3	-0,2	-0,2	-0,2

Продовження табл. 3.2

A	1	2	3	4	5	6	7	8	9	10
УКРАЇНСЬКИЙ ПРОФЕСІЙНИЙ БАНК	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	0,0	-0,2	-0,2	-0,1
БАЗИС	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,1	-0,1
ЕЛЕКТРОНБАНК	-0,2	-0,2	-0,2	-0,2	-0,1	-0,1	-0,2	-0,1	-0,2	-0,2
МІЖНАРОДНИЙ КОМЕРЦІЙНИЙ БАНК	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,2	-0,1
ПОЛТАВА-БАНК	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,2	-0,1
ТРАНСБАНК	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1
МЕРКУРІЙ	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,2	-0,2	-0,1
АКТИВ-БАНК	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1
ЕНЕРГОБАНК	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,1
СИНТЕЗ	-0,2	-0,2	-0,2	-0,3	-0,1	-0,1	-0,1	-0,2	-0,2	-0,2
ФОРТУНА-БАНК*	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,2	-0,2	-0,1
НАЦІОНАЛЬНІ ІНВЕСТИЦІЇ	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1
СХІДНО-ЄВРОПЕЙСЬКИЙ	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,2	-0,3	-0,1
КАПІТАЛ	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2
ІНТЕГРАЛ	-0,2	-0,2	-0,2	-0,3	-0,2	-0,2	-0,2	-0,2	-0,3	-0,1
МЕТАЛУРГ	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,3	-0,2	-0,3	0,0
РЕАЛ-БАНК	-0,2	-0,3	-0,3	-0,3	-0,2	-0,2	-0,1	-0,2	-0,3	-0,2
ІКАР-БАНК	-0,3	-0,3	-0,3	-0,3	-0,1	-0,2	-0,2	-0,2	-0,3	-0,1
ЗЕМЕЛЬНИЙ БАНК	-0,3	-0,3	-0,2	-0,3	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2
ПРИКАРПАТТЯ	-0,3	-0,3	-0,3	-0,3	-0,2	-0,2	-0,2	-0,2	-0,3	-0,2
УКРАЇНСЬКИЙ КРЕДИТНО-ТОРГОВИЙ БАНК	-0,3	-0,3	-0,3	-0,3	-0,2	-0,2	-0,2	-0,2	-0,3	-0,2
ПОЛІКОМБАНК	-0,3	-0,3	-0,3	-0,3	-0,2	-0,2	-0,2	-0,2	-0,3	-0,2
УКООПСПІЛКА	-0,3	-0,3	-0,3	-0,3	-0,2	-0,2	-0,2	-0,2	-0,3	-0,2

Продовження табл. 3.2

А	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
РЕГІОН-БАНК	-0,3	-0,3	-0,3	-0,3	-0,2	-0,2	-0,2	-0,2	-0,3	-0,2
СХІДНО-ПРОМИСЛОВИЙ	-0,3	-0,3	-0,3	-0,3	-0,2	-0,3	-0,2	-0,2	-0,3	-0,2
СЛАВУТИЧ	-0,3	-0,3	-0,3	-0,3	-0,2	-0,3	-0,2	-0,2	-0,3	-0,2
МОРСЬКИЙ	-0,3	-0,3	-0,3	-0,3	-0,2	-0,3	-0,2	-0,2	-0,3	-0,2
ТММ-БАНК	-0,3	-0,3	-0,3	-0,3	-0,2	-0,3	-0,2	-0,3	-0,3	-0,1
АВАЛЬ	4,5	4,3	4,9	4,5	3,8	5,1	5,6	4,9	4,3	-0,2
ПРИВАТБАНК	5,1	5,0	5,3	4,8	6,1	5,1	4,2	5,2	3,0	8,1

Таблиця 3.3 – Кластерний і дискримінантний аналіз за 10 ознаками (для 4 кластерів) на 01.12.05 р.

Банк	Умовний номер кластеру	Група за передбаченням у результаті дискримінантного аналізу	Імовірність належності банку до групи 1	Імовірність належності банку до групи 2	Імовірність належності банку до групи 3	Імовірність належності банку до групи 4
А	*1*	*2*	*3*	*4*	*5*	*6*
АВАЛЬ (01.01.04)	2	2	,00000	1,00000	,00000	,00000
УКРСИББАНК	2	2	,00000	1,00000	,00000	,00000
РАЙФФАЙЗЕНБАНК УКРАЇНА	2	2	,00000	1,00000	,00000	,00000
УКРСОЦБАНК	2	2	,00000	1,00000	,00000	,00000
ОЩАДБАНК	3	3	,00000	,00000	1,00000	,00000
НАДРА	3	3	,00000	,00000	1,00000	,00000
ФІНАНСИ І КРЕДИТ	3	3	,00000	,00000	1,00000	,00000
УКРПРОМБАНК	3	3	,00000	,00000	1,00000	,00000
БРОКБІЗНЕСБАНК	3	3	,00000	,00000	1,00000	,00000
ФОРУМ	2	2	,00000	1,00000	,00000	,00000

Продовження табл. 3.3

A	*1*	*2*	*3*	*4*	*5*	*6*
ПУМБ	3	3	,00000	,00000	1,00000	,00000
КРЕДИТПРОМБАНК	3	3	,00000	,00000	1,00000	,00000
ХРЕЩАТИК	3	3	,00000	,00000	1,00000	,00000
УКРГАЗБАНК	3	3	,00000	,00000	1,00000	,00000
ДОНГОРБАНК	3	3	,00000	,00000	1,00000	,00000
ПІВДЕННИЙ	3	3	,00000	,00000	1,00000	,00000
ВАБАНК	3	3	,00000	,00000	1,00000	,00000
КРЕДИТ БАНК (УКРАЇНА)	3	3	,00000	,00000	1,00000	,00000
АЛЬФА-БАНК	3	3	,00000	,00000	1,00000	,00000
МРІЯ	3	3	,00000	,00000	1,00000	,00000
ТАС-КОМЕРЦБАНК	3	3	,00000	,00000	1,00000	,00000
ПРАВЕКС-БАНК	3	3	,00000	,00000	1,00000	,00000
СІТІБАНК УКРАЇНА	3	3	,00000	,00000	1,00000	,00000
ЕКСПРЕС-БАНК	3	3	,00000	,00000	1,00000	,00000
ХФБ УКРАЇНА	3	3	,00000	,00000	1,00000	,00000
КИЇВ	3	3	,00000	,00000	1,00000	,00000
ІМЕКС-БАНК	3	3	,00000	,00000	1,00000	,00000
ІНДУСТРІАЛБАНК	3	3	,00000	,00000	1,00000	,00000
КРЕДИТ-ДНІПРО	3	3	,00000	,00000	1,00000	,00000
БІГ ЕНЕРГІЯ	3	3	,00000	,00000	1,00000	,00000
МОРСЬКИЙ ТРАНСПОРТНИЙ БАНК	3	3	,00000	,00000	1,00000	,00000
МЕГАБАНК	3	3	,00000	,00000	1,00000	,00000
ТАВРІКА	3	3	,00000	,00000	1,00000	,00000
ЗАХІДІНКОМБАНК	3	3	,00000	,00000	1,00000	,00000
ЕКСПОБАНК*	3	3	,00000	,00000	1,00000	,00000
УКРІНБАНК	3	3	,00000	,00000	1,00000	,00000
ДІАМАНТ	3	3	,00000	,00000	1,00000	,00000
АЖІО	3	3	,00000	,00000	1,00000	,00000
ФАКТОРІАЛ-БАНК	3	3	,00000	,00000	1,00000	,00000

Продовження табл. 3.3

A	*1*	*2*	*3*	*4*	*5*	*6*
АРКАДА	3	3	,00000	,00000	1,00000	,00000
ЕЛЕКТРОНБАНК	3	3	,00000	,00000	1,00000	,00000
БАЗИС	3	3	,00000	,00000	1,00000	,00000
ЗОЛОТІ ВОРОТА	3	3	,00000	,00000	1,00000	,00000
УКРАЇНСЬКИЙ ПРОФЕСІЙНИЙ БАНК	3	3	,00000	,00000	1,00000	,00000
ТРАНСБАНК	3	3	,00000	,00000	1,00000	,00000
ПОЛТАВА-БАНК	3	3	,00000	,00000	1,00000	,00000
МІЖНАРОДНИЙ КОМЕРЦІЙНИЙ БАНК	3	3	,00000	,00000	1,00000	,00000
НАЦІОНАЛЬНІ ІНВЕСТИЦІЇ	3	3	,00000	,00000	1,00000	,00000
АКТИВ-БАНК	3	3	,00000	,00000	1,00000	,00000
ЕНЕРГОБАНК	3	3	,00000	,00000	1,00000	,00000
МЕРКУРІЙ	3	3	,00000	,00000	1,00000	,00000
МЕТАЛУРГ	3	3	,00000	,00000	1,00000	,00000
ІНТЕГРАЛ	3	3	,00000	,00000	1,00000	,00000
СХІДНО-ЄВРОПЕЙСЬКИЙ	3	3	,00000	,00000	1,00000	,00000
ФОРТУНА-БАНК	3	3	,00000	,00000	1,00000	,00000
СИНТЕЗ	3	3	,00000	,00000	1,00000	,00000
КАПІТАЛ	3	3	,00000	,00000	1,00000	,00000
РЕАЛ-БАНК	3	3	,00000	,00000	1,00000	,00000
ЗЕМЕЛЬНИЙ БАНК	3	3	,00000	,00000	1,00000	,00000
ПРИКАРПАТТЯ	3	3	,00000	,00000	1,00000	,00000
ІКАР-БАНК	3	3	,00000	,00000	1,00000	,00000
УКРАЇНСЬКИЙ КРЕДИТНО-ТОРГОВИЙ БАНК	3	3	,00000	,00000	1,00000	,00000
ПОЛІКОМБАНК	3	3	,00000	,00000	1,00000	,00000
УКООПСПІЛКА	3	3	,00000	,00000	1,00000	,00000
РЕГІОН-БАНК	3	3	,00000	,00000	1,00000	,00000
МОРСЬКИЙ	3	3	,00000	,00000	1,00000	,00000
СЛАВУТИЧ	3	3	,00000	,00000	1,00000	,00000

Продовження табл. 3.3

A	*1*	*2*	*3*	*4*	*5*	*6*
СХІДНО-ПРОМИСЛОВИЙ	3	3	,00000	,00000	1,00000	,00000
ТММ-БАНК	3	3	,00000	,00000	1,00000	,00000
АВАЛЬ	1	1	1,00000	,00000	,00000	,00000
ПРИВАТБАНК	4	4	,00000	,00000	,00000	1,00000

Навчальний посібник

Доценко Оксана Станіславівна

ПРАКТИКУМ
З ЗАГАЛЬНОЇ ТЕОРІЇ СТАТИСТИКИ

Учебное пособие

Доценко Оксана Станиславовна

ПРАКТИКУМ
ПО ОБЩЕЙ ТЕОРИИ СТАТИСТИКИ
(українською мовою)

Коректор Л.П.Светлих
Нормоконтролер І.О.Черевкова
Комп'ютерний набір і верстка А.В.Іванова

Здано в набір 20.02.2010. Підп. до друку 09.04.2010 ДК № 1272 від 17.03.03. Формат 89 х 124/16. Пап. тип. № 1. Офсет. друк. Ум. друк. арк. 26,01.
Тираж 300 прим. Зам. № 252