

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

**Л. С. СТРИГУЛЬ**  
**В. О. АЛЕКСАНДРОВА**

**СТАТИСТИКА**

ТЕКСТ ЛЕКЦІЙ

Харків 2015

УДК 311.1  
ББК 60.6я73  
С-8

*Рецензенти:* П. Г. Перерва, д-р економ. наук, проф., зав. каф. організації виробництва та управління персоналом, декан економічного факультету Національного технічного університету «Харківський політехнічний інститут»;

А. А. Пилипенко, д-р економ. наук, проф., зав. каф. бухгалтерського обліку Харківського національного економічного університету імені Семена Кузнеця.

**С 8 Стригуль Л. С., Александрова В. О.**

Статистика : текст лекцій для студ. спец. 6.030509 «Облік і аудит» всіх форм навч. / Л. С. Стригуль, В. О. Александрова – Х. : НТУ «ХПІ», 2014. – 220 с.

ISBN

Навчальне видання подано у вигляді тексту лекцій зі статистики. У ньому детально розглянуто різноманітні методики обчислення статистичних показників, необхідні для проведення комплексного аналізу явищ та процесів, діяльності підприємств, організацій і установ. Виклад матеріалу супроводжується докладними прикладами, наочними таблицями і графіками.

Призначено для студентів економічних спеціальностей.

Рис. 35. Табл. 36. Бібліогр. 28 назв.

УДК 311.1  
ББК 60.6я73

ISBN

©Л.С.Стригуль,  
В.О.Александрова, 2014  
© НТУ «ХПІ», 2014

## ПЕРЕДМОВА

Статистика є базовою загальнопрофесійною дисципліною для студентів, що навчаються за напрямом «Облік і аудит», та студентів інших економічних спеціальностей, оскільки в даний час у всьому світі вона є універсальною мовою для спілкування економістів усіх країн і всіх напрямів, державних діячів, політиків, соціологів, бізнесменів, фінансистів і фахівців в області держуправління. Тому важливо навчити студентів цієї універсальної мови спілкування фахівців у всьому світі.

Кількісні вимірювання соціально-економічних процесів і явищ ґрунтуються на положеннях економічної теорії, результатах вивчення якісних аспектів економічних процесів, отриманих в рамках загальноекономічної теорії і різних прикладних розділів економічної науки. У свою чергу, розвиток економічної теорії неможливий без використання результатів статистичного опису і аналізу економічних процесів для перевірки і уточнення окремих постулатів, концепцій, положень і висновків. Статистика за допомогою своїх методів не тільки перевіряє коректність запропонованої теоретичної гіпотези, але і підтверджує конкретними числовими характеристиками розвиток встановлених тенденцій у часі, а також встановлює роль різних чинників та визначає їх вплив на загальний процес розвитку досліджуваного явища.

Навчальна програма нормативної дисципліни циклу професійної та практичної підготовки «Статистика» є складовою системи стандартів вищої освіти і враховує нормативні документи МОН України, Освітньо-професійної програми підготовки бакалавра за напрямом підготовки 6.030509 «Облік і аудит» галузі знань 0305 «Економіка та підприємництво», Освітньо-кваліфікаційних характеристик (ОКХ) даного напрямку.

Нормативна навчальна дисципліна циклу професійної та практичної підготовки «Статистика» є актуальною для підготовки фахівців у формуванні професійних знань бакалаврів за напрямом підготовки 6.030509.

Метою навчальної дисципліни є системний виклад статистичних методів, які застосовуються на основних стадіях економіко-статистичного дослідження. При цьому розгляд статистичної методології вивчення закономірностей економічної діяльності ґрунтується на широкому використанні результатів теорії вірогідності і математичної статистики із застосуванням засобів сучасної обчислювальної техніки.

Текст лекцій містить інформацію, яка дає змогу підвищити рівень коректності сприйняття студентами статистичної інформації з метою її кваліфікованої обробки, для подальшого обґрунтованого прогнозування і ефективного використання в економічних дослідженнях отриманих результатів. Також метою курсу є формування у студентів уявлення про статистику, не тільки як про наукову дисципліну, але і як про сферу практичної діяльності, яка дозволяє збирати і обробляти реальну інформацію, на засаді ознайомлення з основними статистичними показниками, отримання первинних навиків збору, систематизації, обробки і аналізу статистичних даних. У процесі вивчення дисципліни у студентів формується адекватне уявлення про економічний зміст і інформаційні можливості статистичних методів.

Завдання вивчення дисципліни «Статистика» визначаються вимогами до підготовки кадрів, встановленими в кваліфікаційній характеристиці підготовки фахівців (бакалаврів) з напрямку «Облік і аудит», і передбачають:

- поглиблення знань з теорії статистичного аналізу економічних процесів, освоєння статистичної методики збору і обробки даних при соціально-економічних дослідженнях;

- вивчення апарату, техніки і засобів статистичного аналізу соціально-економічних процесів;

- формування навиків проведення складних статистичних розрахунків при обчисленні статистичних показників і інтерпретації отриманих результатів;

- підготовку фахівців, які володіють дослідницьким потенціалом.

У результаті вивчення дисципліни «Статистика» студент повинен знати: зміст основних етапів статистичного дослідження, основні способи і види статистичного спостереження, основні вимоги до первинної інформації, методи збору і обробки статистичних даних для проведення якісного економічного аналізу, методи статистичних угруповань, основні правила складання таблиць та умови застосування методів графічного відображення статистичних даних, основні види узагальнюючих статистичних показників і їх роль у вивченні економічної діяльності, основні формули і залежності, які використовуються при аналізі статистичних даних, методи обчислення основних статистичних показників, способи аналізу рядів динаміки виявлення основних тенденцій і сезонних коливань, специфіку індексного методу в дослідженнях комерційної діяльності, зв'язок між показниками

комерційної діяльності.

У результаті вивчення тексту лекцій з дисципліни «Статистика» студенти повинні навчитись проводити статистичний аналіз із застосуванням отриманих знань і методів і інтерпретувати статистичні показники, аналізувати і знаходити залежності в явищах, які підлягають вивченню, організовувати статистичні дослідження і проводити статистичний аналіз реальних економічних ситуацій, використовувати статистичні дані як інформаційну основу статистичного аналізу, перевіряти достовірність первинних даних. Також студенти повинні мати уявлення про можливості і межі застосування вивчених в дисципліні статистичних методів, про основні джерела статистичної інформації, про прийоми статистичної оцінки значущості отриманих результатів. Студенти у результаті вивчення тексту лекцій з дисципліни «Статистика» повинні володіти навиками самостійного вибору і застосування статистичних методів для обробки наявної інформації.

У пропонованому тексті лекцій міститься системний виклад основних розділів дисципліни «Статистика», вивчення яких багато в чому сприяє формуванню ділових якостей фахівця в різних областях професійної діяльності.

Для успішного вивчення курсу студент повинен володіти основами математичного аналізу, лінійної алгебри, апаратом теорії вірогідності, мати базові уявлення в області загальної економічної теорії.

Особливістю подання даного тексту лекцій є те, що питання статистичної методології розглядаються безпосередньо на прикладах вирішень типових економічних завдань.

Основну увагу при вивченні дисципліни «Статистика» слід приділяти системам показників і можливостям їх інтерпретації, методам статистичного аналізу соціально-економічних явищ і процесів. При вивченні кожної теми спочатку рекомендується уважно розібратись у викладеному теоретичному матеріалі. Після осмислення теоретичного матеріалу слід розібрати приклади вирішення завдань. Аналіз пропонованих у тексті лекцій прикладів дозволить студентам краще з'ясувати сферу застосування статистичних показників і методів їх обчислень. Для глибшого розуміння і якіснішого сприйняття теоретичного матеріалу тексту лекцій з дисципліни «Статистика» запропоновано теоретичні контрольні питання для самостійного опрацювання.

# **Тема 1. ПРЕДМЕТ, МЕТОД І ОСНОВНІ ЗАВДАННЯ СТАТИСТИКИ. БАЗОВІ ПОНЯТТЯ, ТЕРМІНИ, ФОРМИ, СПОСОБИ І ВИДИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ**

- 1.1. Статистика як суспільна наука.
- 1.2. Предмет, метод і основні завдання статистики.
- 1.3. Поняття, категорії та показники статистики.
- 1.4. Базові поняття, терміни, форми, способи і види статистичного спостереження.

## ***1.1. Статистика як суспільна наука***

В умовах ринкових відносин робота економістів, комерсантів, менеджерів та підприємців наповнюється новим змістом, при цьому роль статистичної науки зростає. Тому володіння методологією статистики є неодмінною умовою вивчення тенденцій попиту та пропозицій, результатів підприємницької діяльності, прийняття оптимальних управлінських рішень.

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

Термін «Статистика» був введений у науковий світ в 1749 році німецьким вченим Готфридом Ахенвалем.

Вчені, які здійснили вагомий внесок у розвиток статистики, це: *Вільям Петті* – загальновизнаний засновник статистики, заслуга якого полягає в тому, що він вперше застосував числовий метод для аналізу закономірностей громадського життя, йому належить наукова робота – «Політична арифметика»; *Адольф Кетле* – бельгійський статистик, який довів, що навіть ймовірнісні випадковості суспільного життя мають внутрішню закономірність; *К.Ф. Герман* – російський статистик, якому належить наукова робота «Загальна теорія статистики»; *В. І. Ленін* – який сформулював теорію угруповань і теорію статистичного спостереження.

Статистика – це багатогалузева наука. Основними напрямками її досліджень є:

1. Теорія статистики – містить статистичні категорії, методи й способи аналізу масових явищ і процесів.

2. Соціальна і демографічна статистика – вивчає природний механічний рух населення, соціальні умови й характер праці, рівень життя населення.

3. Економічна статистика – містить системи економічних показників і методи вивчення економіки країни або регіону як єдиного цілого;

4. Галузева статистика – вивчає особливості окремих видів економічної діяльності (статистика інвестицій, фінансів, торгівлі і т. ін.)

Термін «статистика» (від лат. status – стан речей) означає кількісний облік та аналіз масових соціально-економічних явищ та процесів. Статистика – це наука, яка об'єднує принципи та методи роботи з масовими числовими даними – кількісними характеристиками визначених явищ та процесів.

Термін «статистика» в практичній та науковій сферах застосовують у різних формулюваннях:

– *статистика* – це галузь практичної діяльності, яка спрямована на збирання, обробку та аналіз масових суспільно-економічних явищ і процесів;

– *статистика* – це сукупність зведених підсумкових цифрових показників, зібраних для кількісної характеристики будь-якої сфери суспільних явищ, процесів або окремого практичного завдання;

– *статистика* – це галузь знань, тобто спеціальна наукова дисципліна (статистична наука) і відповідно навчальна дисципліна, що викладається у ВНЗ. Статистика є важливою частиною навчального плану підготовки економістів, менеджерів, бухгалтерів, маркетологів, комерсантів, підприємців найвищої кваліфікації.

*Статистична сукупність* характеризується системою показників, які пояснюють явище з різних сторін, тобто статистична сукупність – це сукупність соціально-економічних об'єктів або явищ суспільного життя, об'єднана якісною основою.

*Одиниця сукупності* – це первинний елемент статистичної сукупності який є носієм ознак, що підлягають реєстрації. *Ознака* – це якісна особливість одиниці сукупності. Ознаки бувають кількісними або якісними (атрибутивними). Одиниці сукупності, поряд із загальними для всіх одиниць ознаками, які обумовлюють якісну визначеність сукупності, мають індивідуальні особливості і відмінності, що відрізняють їх одна від одної, тобто існує варіація ознак. Необхідність статистики визначається наявністю варіації. *Варіація* – це зміна значень

ознаки.

Кількісна визначеність – це об'єктивна властивість предмета пізнання статистики. Кількісну характеристику статистика виражає через певний ряд чисел, які називають статистичними показниками. *Статистичний показник* – це кількісно-якісна характеристика статистичної сукупності або соціально-економічного явища. Статистичний показник має три обов'язкові атрибути:

- кількісну визначеність;
- місце;
- час (момент або період часу).

Конкретні значення статистичних показників називають статистичними даними. У статистиці всі числа завжди іменовані, відносяться до певного місця і часу.

Статистика також характеризує структуру суспільних явищ. *Структура* – це внутрішня побудова масових явищ, тобто внутрішня побудова статистичної безлічі. Кожному суспільному явищу властиві зміни в просторі і часі. Зміни в просторі виявляються за допомогою аналізу структури суспільного явища, а зміни рівня явища досліджуються в часі тобто в динаміці. Явища суспільного життя взаємозалежні і взаємообумовлені. Тому одним з завдань статистики є виявлення причинно-наслідкових зв'язків з метою впливу на суспільні явища в інтересах суспільства.

*Завдання статистичного дослідження* – отримання узагальнювальних показників і виявлення закономірностей соціально-економічних явищ і процесів в конкретних умовах місця і часу.

*Статистична закономірність* – це форма прояву причинно-наслідкового зв'язку, що виражається в повторюваності подій з достатньо високою вірогідністю.

Сьогодні статистику використовують, вивчаючи життєвий рівень населення та громадську думку, оцінюючи підприємницькі та фінансові ризики, у маркетингових дослідженнях, страхуванні і т. ін. Основними розділами статистичної науки є:

- 1) *теорія статистики* – розглядає загальні принципи і методи вивчення економічних явищ і процесів;
- 2) *соціально-економічна статистика* – вивчає методологію побудови макроекономічних показників та їх аналіз на рівні народного господарства країни чи регіону як єдиного цілого, а також соціальних умов життя і праці населення, споживання ним матеріальних благ і послуг;



3) демографічна, промислова, сільськогосподарська та інші галузеві статистики – вивчають окремі галузі народного господарства чи суспільного життя.

### **1.2. Предмет, метод і основні завдання статистики**

Статистика є наукою, яка вивчає розміри та кількісні співвідношення масових суспільно-економічних явищ і процесів у нерозривному зв'язку з їх змістом, тобто вона кількісно досліджує закономірності розвитку суспільних явищ за конкретних умов, місця і часу.

*Предмет статистики* – кількісна сторона масових суспільних явищ у нерозривному зв'язку з їх якісною стороною, а також кількісне вираження закономірностей суспільного розвитку в конкретних умовах місця і часу.

Виділяють *три основні елементи предмета* статистики:

– вивчення суспільних явищ, зумовленого їх особливостями, законами розвитку і методами пізнання. Щоб з'ясувати закони суспільного розвитку, треба зібрати та узагальнити чисельні факти суспільного життя та науково виявити і дослідити їх сутність. Статистика збирає дані про масові явища і процеси, відповідно опрацьовує їх та робить висновки. Без науково опрацьованих статистичних даних не можна глибоко вивчити стан та розвиток суспільства;

– вивчення суспільних явищ за допомогою кількісних характеристик, якими можуть бути обсяги, рівні, кількісні співвідношення і пропорції, темпи розвитку, що виражені в певних числах-показниках. Статистика спрямована на пізнання масових суспільних явищ у єдності їх якісних та кількісних сторін. Саме тому в визначенні статистики вказано, що ця наука вивчає кількісні характеристики явищ та процесів, для подальшого пізнання на цій основі їх якісного змісту. Кількісні характеристики змінюються у часі та просторі. Статистичним даним притаманна конкретність з огляду на час, місце та обсяг сукупності даних досліджуваних масових явищ;

– вивчення масових суспільних явищ, тобто тих, що складаються з достатньо великої сукупності одиниць чи фактів. Статистика вивчає закономірності зміни кількісних характеристик на основі масового узагальнення фактів.

*Мова статистики* – це мова цифр.

Умовно можна поділити статистику на дві складові частини:

– описову статистику, призначену для застосування методів збору і обробки числових даних;

– статистику висновків, сутність якої – здатність робити висновки про сукупність на основі даних їх частки (вибірки).

*Метод статистики* – це ціла сукупність прийомів, користуючись якими статистика досліджує свій предмет. Виявити та охарактеризувати розміри, зміни і кількісні співвідношення певних масових явищ можна, здійснюючи послідовно *три основні стадії економіко-статистичного дослідження*:

– перша стадія – це застосування методу масових спостережень, який полягає у збиранні первинного статистичного матеріалу і у науково-організованій реєстрації всіх існуючих фактів, які відносяться до розглянутого об'єкта.

– друга стадія – застосування методу групувань, який дає можливість усі зібрані в результаті масового статистичного спостереження факти піддавати систематизації і класифікації.

– третя стадія – застосування методу узагальнюючих показників, який дозволяє характеризувати досліджувані явища і процеси за допомогою статистичних величин – абсолютних, відносних і середніх. На цьому етапі статистичного дослідження виявляються взаємозв'язки і масштаби явищ, визначаються закономірності їх розвитку, даються прогнозні оцінки.

На кожному етапі застосовують комплекс специфічних, властивих лише статистиці методів, який утворює статистичну методологію і який обумовлений специфікою предмета статистики.

Об'єкт статистичного дослідження (у кожному конкретному випадку) називають *статистичною сукупністю*. Об'єктами статистики є різноманітні явища і процеси суспільного життя.

*Предметом статистичного дослідження* є розміри і кількісні співвідношення між масовими суспільними явищами, закономірності їх формування, розвитку, взаємозв'язку.

*Завдання статистики як науки:*

1. Вивчення рівня і структури масових соціально-економічних явищ.

2. Вивчення взаємозв'язків масових соціально-економічних явищ.

3. Вивчення динаміки масових соціально-економічних явищ.

*Загальним принципом*, що лежить в основі дослідження статистичних закономірностей, виступає так званий «закон великих

*чисел*», що виражає діалектику випадкового і необхідного.

*Сутність закону великих чисел* полягає в тому, що при підсумовуванні даних щодо достатньо великої кількості випадків, взаємопогашаються і у загальних середніх числах виступають істотні, характерні риси і взаємозв'язки явища в цілому; тобто сукупні дії великої кількості випадкових факторів приводять до результату, який майже не залежить від випадку. Закон великих чисел у найбільш простій формі говорить, що кількісні закономірності масових явищ чітко виявляються лише при достатньо великій їх кількості.

Статистичні закономірності вивчають розподіл одиниць статистичної множини за окремими ознаками під впливом усієї сукупності факторів. Статистична закономірність виступає як об'єктивна закономірність складного масового процесу і є формою причинного зв'язку, яка виявляється в підсумку масового статистичного спостереження. Цим обумовлюється її зв'язок із законом великих чисел. Статистична закономірність, з певною вірогідністю, гарантує стабільність середніх величин при збереженні постійного комплексу умов, що породжують дане явище.

Основним завданням *статистичного спостереження* є отримання певних значень досліджуваних ознак від кожної одиниці статистичної сукупності шляхом реєстрації (обліку) їх на основі ретельно розробленої програми спостереження. Масове статистичне спостереження дає інформаційну базу для статистичних узагальнень і характеристики об'єктивних закономірностей. Спостереження має набувати масового характеру. Вимога масовості одиниць спостереження зумовлена тим, що вивчення статистичної закономірності виявляється у достатньо великому масиві даних на основі дії закону великих чисел.

*Зведення та групування первинних даних* має завдання всебічно систематизувати матеріали статистичного спостереження. Сутність другої стадії статистичного дослідження зводиться до перевірки даних, їх групування за певними ознаками, підбиття групових і загальних підсумків, розрахунку різних показників, проектування таблиць і внесення в них даних. Результати зведеної обробки статистичних матеріалів, що становлять впорядковану сукупність взаємопов'язаних цифрових характеристик суспільних явищ і процесів, зображують у вигляді системи таблиць та графічних ілюстрацій. Метод групувань дає змогу виділити в досліджуваній сукупності соціально-економічні типи явищ, охарактеризувати їх структуру, виявити взаємозв'язки і

взаємозалежності між показниками.

Впорядковану сукупність називають *статистичним рядом*. У статистиці розглядають такі типи рядів:

– *ряди розподілу* – це результат класифікації, групування та ранжування елементів сукупності у статистиці;

– *ряди динаміки* – описують динаміку розвитку масового процесу, класифікують значення статистичних показників у часі.

*Аналіз статистичної інформації* передбачає проведення аналізу даних на основі обчислення узагальнюючих показників: абсолютних, відносних і середніх величин, статистичних коефіцієнтів, показників варіації ознак і динаміки явищ, індексів та показників, які характеризують щільність зв'язку між явищами тощо. Це дає змогу розкрити причинні зв'язки досліджуваних явищ, визначити вплив і взаємодію різних чинників, оцінити ефективність прийнятих управлінських рішень та можливі економічні і соціальні наслідки ситуації, яка склалася на підприємстві. Порівнюючи статистичні показники, визначають кількісні оцінки їх поширення в просторі та розвитку у часі, виявляють характеристики зв'язку та взаємозалежності, формулюють наукові і практичні висновки. При аналізі статистичної інформації застосовують методи вивчення варіації, диференціації та усталеності, швидкості та інтенсивності розвитку, узагальнюючі індекси, регресійні моделі. Аналітичні можливості статистичних методів поглиблюються завдяки використанню компактної та раціональної форми подання результатів узагальнення інформації, а також аналізу виявлених закономірностей. Такими формами є *статистичні таблиці та графіки*.

Статистичні методи пов'язані з математикою. У них є спільні методи обробки і оцінювання даних, але різні предмети пізнання. Математична статистика вивчає закономірності масових явищ в абстрактній формі, статистика є суспільною наукою і характеризує розміри і співвідношення суспільних явищ в конкретних умовах їх існування та розвитку. Статистичний аналіз є необхідною ланкою в системі управління економікою та державою в цілому.

### ***1.3. Поняття, категорії та показники статистики***

Статистика використовує при вивченні явищ та процесів цілу низку понять та категорій: *ознака; варіація; закономірність; статистична сукупність; показник; система показників*.

*Ознака* – це відмітна риса, властивість, якість, що є характерною для окремих одиниць, об'єктів (явищ). Кожний елемент сукупності характеризується багатьма ознаками, значення яких змінюється від елемента до елемента або від одного періоду до іншого.

У статистиці *ознаки поділяють* на:

– *якісні (атрибутивні)* – ознаки, варіанти яких, характеризуючи особливості окремих одиниць, не мають кількісного вираження. Серед атрибутивних ознак деякі чітко окреслені (стать, професія, галузь), інші невизначені (суб'єктивні оцінки, твердження, думки);

– *кількісні* – ознаки, які мають числове вираження (наприклад, стаж роботи працівників, врожайність певної культури, вага виробів);

– *основні* – ознаки, що розкривають головну сутність досліджуваних явищ і процесів;

– *другорядні* – ознаки, які не пов'язані безпосередньо зі змістом явищ, але які здійснюють опосередкований вплив на них;

– *варіаційні* – ознаки, які набувають різних значень окремих одиниць досліджуваного явища (наприклад, обсяг реалізованої продукції є ознакою варіаційною, тому що у окремих підприємств він різний);

– *постійні* – ознаки, які мають незмінні значення усіх одиниць досліджуваного явища;

– *дискретні* – ознаки, які мають лише цілочислові значення (кількість укладених на біржі угод, кількість дітей в сім'ї);

– *безперервні* – мають будь-які значення в певних межах варіації (наприклад, вік людини в межах від 0 до 100 і більше років, народжуваність, урожайність, балансова ліквідність).

*Варіація* – це відмінність, коливання значень ознаки. Ознаки мають різний рівень вимірювання, що відображується у відповідних *типах шкал вимірювання*:

1) *номінальна шкала* – шкала найменувань, суть якої полягає в тому, що подібним елементам відповідає те саме число, а неподібним – різні числа;

2) *порядкова (рангова) шкала* – встановлює не лише відношення подібності елементів, а і відношення послідовності – порядку;

3) *метрична шкала* – це звичайна шкала дійсних чисел, за допомогою якої вимірюються натурально-речові явища, ресурси та результати фінансово-господарської діяльності.

*Закономірність* – це повторюваність, послідовність і порядок у масових процесах. Виявити і виміряти статистичну закономірність

можна лише з урахуванням дії закону великих чисел, основними принципами якого є масовість і причинна зумовленість явищ. Об'єктивною основою існування закономірностей є складне переплетення причин, які формують масовий процес. Статистичні закономірності притаманні лише сукупностям.

*Статистична сукупність* – це велика кількість одиниць, об'єктів, явищ, об'єднаних будь-якими загальними властивостями (ознаками), які підлягають статистичному вивченню, тобто це певна множина елементів, поєднаних умовами існування і розвитку. Окремі об'єкти, явища, що складають статистичну сукупність, називають *одиницями сукупності*. *Статистичний показник* – узагальнююча кількісна характеристика соціально-економічних явищ і процесів у їх якісній визначеності щодо конкретних умов місця і часу.

У статистиці виділяють такі показники:

– *натуральні* – виражені у фізичних одиницях (тонни, кілограми, штуки);

– *вартісні* – відповідають грошовій оцінці економічних об'єктів;

– *об'ємні* – вимірюють обсяг сукупності об'єктів (елементів), (наприклад, обсяг виробництва, чисельність працівників);

– *якісні* – характеризують рівень розвитку явища (рівень рентабельності, рівень продуктивності).

*Система показників* – сукупність показників, які всебічно характеризують розвиток суспільства.

У статистиці важливо правильно визначати зміст статистичних показників та методи їх побудови, з'ясування цих питань здійснюється на основі економічної теорії, яка розкриває економічну сутність явищ.

Приклад формування статистичної сукупності з одиниць сукупності з притаманними їм ознаками та відповідною шкалою вимірювання наведено в табл. 1.1.

#### ***1.4. Базові поняття, терміни, форми, способи і види статистичного спостереження***

Для дослідження соціально-економічних явищ і процесів суспільного життя необхідно, перш за все, зібрати про них необхідні відомості – статистичні дані. Під *статистичними даними* (інформацією) розуміють сукупність кількісних характеристик соціально-економічних явищ і процесів, отриманих в результаті статистичного спостереження,

їх обробки або відповідних розрахунків.

Таблиця 1.1 – Складові статистичної сукупності

Статистична сукупність	Одиниця сукупності	Ознака
Сукупність комерційних банків	Банк	Кредитний портфель
Сукупність фермерських господарств	Фермерське господарство	Розмір посівних площ Кількість поголів'я худоби
Види ознак		Типи шкали виміру
<i>Кількісна</i> (виражається числом)		<i>Метрична шкала</i> – ряд дійсних чисел
Дискретна (кількість дітей у родині)		
Безперервна (вік, ціна)		
<i>Атрибутивна</i> – словесна, описова. Її не можна виміряти, можна тільки зафіксувати (професія, стать)		<i>Номінальна шкала</i> – шкала найменувань. Порядок її розташування не має значення (перелік фірм, видів економічної діяльності). Для ідентифікації найменувань використовуються певні числові коди або натуральні числа
<i>Альтернативна</i> – тільки два значення (так/ні, задоволений/не задоволений)		
<i>Рангова</i> – характеризує послідовність інтенсивності прояву ознаки (задоволений / не дуже задоволений / не задоволений)		<i>Порядкова шкала</i> – встановлює співвідношення послідовності: порядок – ранги

Важливою частиною будь-якого статистичного дослідження є статистичне спостереження. *Статистичне спостереження* – це масове, планомірне, науково організоване спостереження за явищами соціального і економічного життя, яке полягає в реєстрації відібраних ознак кожної одиниці сукупності.

*Метою статистичного спостереження* є отримання достовірної інформації для виявлення закономірностей розвитку явищ і процесів. Метою *визначається* програма і форми організації. Некоректно поставлена мета може призвести до того, що в процесі спостереження будуть зібрані непотрібні дані або не будуть отримані відомості, необхідні для аналізу.

*Об'єктом спостереження* називається конкретна сукупність одиниць, явищ або процесів, що підлягають обстеженню. Для визначення об'єкта статистичного спостереження необхідно встановити границі досліджуваної сукупності. Об'єкт спостереження *складається* з окремих елементів – одиниць спостереження. *Одиницею спостереження* (у закордонній літературі використовується термін «елементарна одиниця») називають складний елемент об'єкта, що є носієм ознак, які підлягають реєстрації. (Приклад: при демографічних дослідженнях одиницею може

бути як людина, так і родина; при бюджетних – родина або домогосподарство і т. д.). *Звітною одиницею* виступає суб'єкт, від якого надходять дані про одиницю спостереження. Одиниці сукупності і одиниці спостереження можуть іноді розрізнятися. *Одиниця сукупності* – первинний осередок, від якого повинні бути отримані необхідні статистичні відомості. Наприклад, при перепису верстатного парку заводів одиниця спостереження – завод, а одиниці сукупності – верстати. Перелік ознак (питань), що підлягають реєстрації в процесі спостереження, складають зміст *програми спостереження*.

Статистичне спостереження є першим кроком будь-якого статистичного дослідження і в ньому можна виділити три етапи:

1. Підготовка спостереження.
2. Збір матеріалів.
3. Контроль отриманого матеріалу.

При підготовці статистичного спостереження визначають цілі і об'єкт спостереження, склад ознак, що підлягають реєстрації; розробляють документи для збору даних; також визначають методи і засоби отримання даних. Окрім цього вирішують проблеми організаційного характеру.

*Форми, способи і види статистичного спостереження.* У вітчизняній статистиці використовуються три організаційні форми (типи) статистичного спостереження:

1. Звітність (підприємств, організацій, установ і т. ін.).
2. Спеціально організоване статистичне спостереження (переписи, одноразовий облік, обстеження).
3. Регістрове спостереження.

*Звітність* – це основна форма статистичного спостереження, за допомогою якої статистичні органи в певні терміни отримують від підприємств, установ і організацій необхідні дані у вигляді встановлених в законному порядку звітних документів. Звітність як форма статистичного спостереження заснована в первинному обліку і є реєстрацією різних фактів та подій, які обліковуються у міру їх здійснення.

*Спеціально організоване спостереження* проводиться з метою отримання відомостей відсутніх в звітності або для перевірки її даних.

*Регістрове спостереження* – це форма безперервного статистичного спостереження за довготривалими процесами, які мають фіксований початок, стадію розвитку і фіксоване закінчення. Воно засноване на веденні статистичного *реєстру*. У реєстрі кожна одиниця спосте-



реження характеризується сукупністю показників, і всі показники зберігаються до повного завершення спостереження за одиницею обстежуваної сукупності. Проведення спостереження здійснюють спеціально навчені кваліфіковані працівники, які в програмі спостереження мають назву - «лічильники».

Існують різні *способи* отримання статистичної інформації:

1. *Безпосереднє спостереження*, при якому факт, що підлягає реєстрації, фіксується на місці;

2. *Документальний облік фактів*, заснований на використанні в якості джерел статистичної інформації різного роду документів, як правило, облікового характеру;

3. *Опит* – спосіб спостереження, при якому необхідні відомості отримують зі слів респондента.

У статистиці застосовуються такі *види опитів*:

– усний – «лічильники» отримують необхідну інформацію на основі опиту відповідних осіб і самі фіксують відповіді;

– самореєстрація – роздані «лічильниками» формуляри заповнюються самими респондентами;

– кореспондентський – інформація повідомляється добровільними кореспондентами;

– анкетний – збір інформації у вигляді анкет;

– явочний – інформація надається в явочному порядку.

Класифікація *видів статистичного спостереження* наведена на рис. 1.1.

У сучасному статистичному спостереженні також використовують метод *ведення щоденників* для аналізу явищ, які важко відновити по пам'яті, крім того, набув популярності метод *бізнес-обстеження* для аналізу діяльності підприємства. Точність статистичного спостереження є важливою і основною вимогою органів державної статистики. Однак, хоч як би старанно не було підготовлене статистичне спостереження, в процесі його проведення трапляються помилки, які призводять до зниження його точності.

*Точністю статистичного спостереження* називають ступінь відповідності значення будь-якої ознаки визначеної за допомогою статистичного спостереження, до її дійсного значення. Чим ближчі значення ознак, отриманих в результаті статистичного спостереження до їх фактичних значень, тим точніше проведено спостереження. Точність статистичного спостереження визначається як відношення даних спостере-

ження до дійсних значень досліджуваних величин, або як різниця між ними.

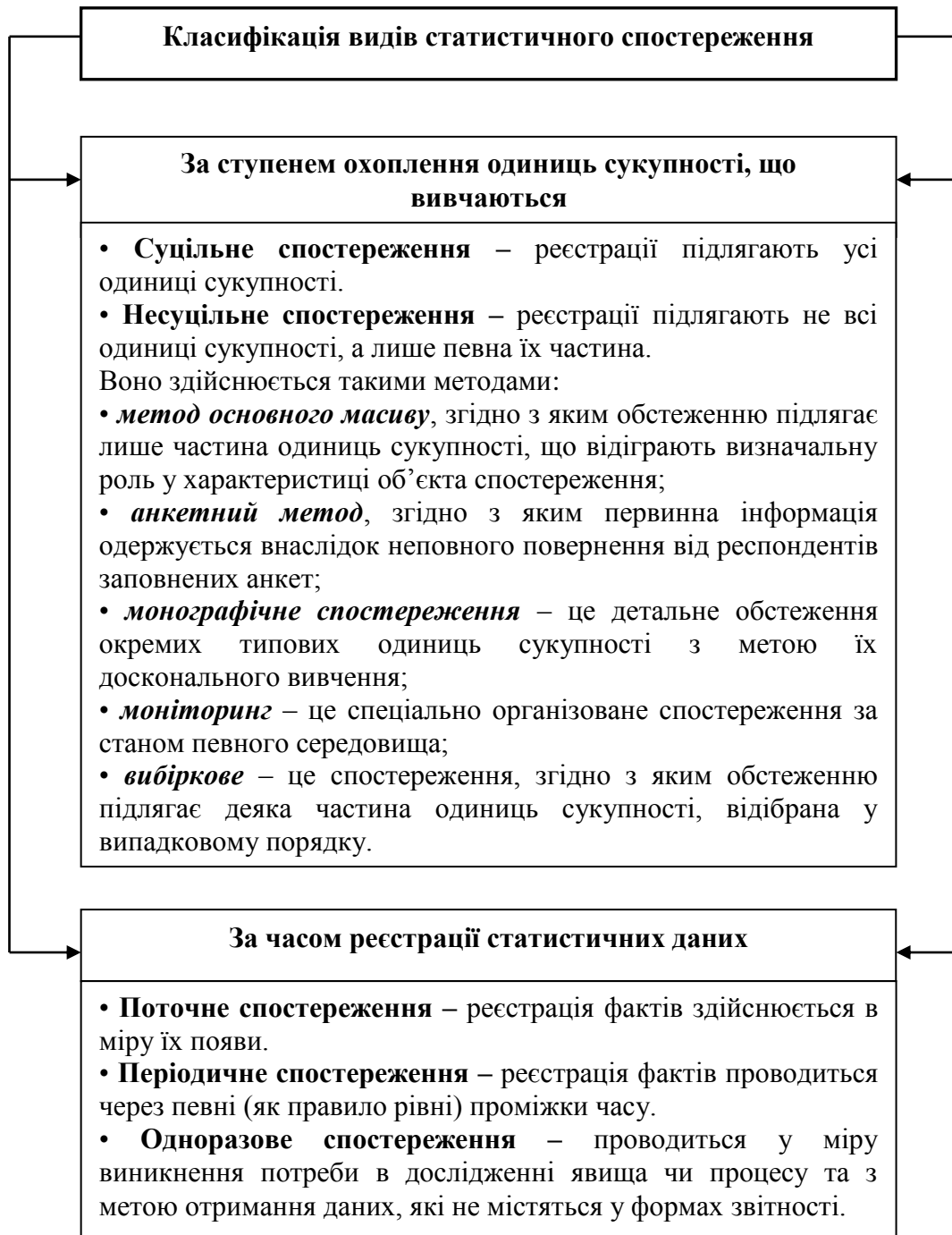


Рисунок 1.1 – Класифікація видів статистичного спостереження

*Помилками спостереження* називаються розходження між встановленими даними статистичних спостережень і дійсними значеннями досліджуваних величин. Помилки спостереження виникають внаслідок

неточностей при збиранні і реєстрації значень досліджуваних ознак. Недопущення і попередження помилок є одним з важливих завдань організації і проведення статистичного спостереження.

Невірні статистичні дані можуть призвести до прорахунків в державному управлінні економікою, серйозних помилок в науковому плануванні і прогнозуванні та інших негативних наслідків. Тому в Україні встановлена сувора відповідальність посадових осіб за навмисні викривлення статистичних даних.

Залежно від характеру, ступеня впливу на кінцеві результати, джерел і причин виникнення неточностей розрізняють такі типи помилок статистичного спостереження:

- а) помилки реєстрації;
- б) помилки репрезентативності.

Кожний з цих типів помилок ділиться на випадкові (ненавмисні) і систематичні (навмисні і ненавмисні). *Систематичні помилки* – це постійна зміна статистичних даних.

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

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

*Систематичні* помилки реєстрації виникають внаслідок певних причин, діють в тому ж самому напрямку і спричиняють серйозні викривлення загальних результатів статистичного спостереження. Наприклад, під час перепису населення опитувані особи часто округлюють свій вік, як правило, на цифрах, які закінчуються на «5» і «0». Замість 24–26 років говорять 25, замість 39–41 говорять, що їм 40 років і т.д. Внаслідок цього виходить, що 25-, 30-, 35-, 40-річних громадян значено більше, ніж 24-, 26-, 31-, 36-, 39-, 41-річних.

Статистичні помилки реєстрації можуть бути отримані внаслідок свідомого викривлення фактів. Це навмисні приписки або приховування у звітах фактичних даних. Помилки реєстрації виникають як при суцільному, так і при несуцільному спостереженні. На відміну від помилок реєстрації, помилки репрезентативності властиві тільки

несуцільному спостереженню. Для запобігання помилкам, перш за все здійснюється зовнішній контроль статистичного формуляра – правильність заповнення, відповідність даних до запитань тощо. Потім здійснюється логічний контроль – зіставлення відповідей на взаємозв'язані питання. Потім проводиться арифметичний контроль – він дозволяє перевірити правильність кількості взаємозв'язаних одиниць сукупності.

*Помилками репрезентативності* називаються відхилення значень ознак відібраної і обстеженої частини сукупності від значень ознак всієї досліджуваної сукупності. *Помилки репрезентативності* – це помилки, пов'язані з неправильним вибором типових представників сукупності, тобто коли сукупність представлена нетиповими представниками.

*Випадкові помилки репрезентативності* виникають внаслідок того, що відібрана випадковим, неупередженим способом частина досліджуваної сукупності недостатньо повно відтворює всю сукупність в цілому. *Систематичні помилки репрезентативності* виникають внаслідок порушення принципів неупередженого, випадкового відбору одиниць для обстеження.

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

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

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

*Арифметичний контроль* полягає в лічильній перевірці підсумкових даних звітів або формулярів і погодженні тих показників, які взаємопов'язані між собою і можуть бути виведені один з одного. Наприклад, в шаховій таблиці будь-якого значення підсумки рядків і колонок повинні збігатись, а якщо такого збігу немає, тоді шукають помилку в рядках чи колонках.

*Логічний контроль* полягає у зіставленні пов'язаних між собою відповідей на питання формуляра статистичного спостереження і у з'ясуванні їх логічної сумісності. Якщо виявляють логічно несумісні відповіді, шляхом їх подальшого зіставлення з відповідями на інші

питання, встановлюють, яка з відповідей записана невірно. Наприклад, якщо у формулярі переписного листа перепису населення записано, що опитувана особа у віці 7 років має сім'ю, вищу освіту, працює лікарем, то зрозуміло, що неправильно записаний вік.

Основною умовою успішного проведення будь-якого статистичного дослідження на всіх його стадіях, у тому числі і на стадії спостереження, є висока якість зібраного матеріалу.

*Організаційні питання статистичного спостереження.* Під час статистичного спостереження вирішується ціла низка організаційних питань. Послідовність їх вирішення відображається в організаційному плані статистичного спостереження, який є складовою плану спостереження. Організаційні питання стосуються місця та часу проведення обстеження, його матеріально-технічного забезпечення, а також системи гарантування точності результатів.

*План спостереження* – це основний організаційний документ, в якому зафіксовано всі важливі організаційні заходи, здійснення яких потрібне для успішного проведення статистичного спостереження. В організаційному плані спостереження вказують на способи збирання статистичних даних, терміни підготовки і забезпечення спостереження необхідним інструментарієм.

Статистичне спостереження може проводитися:

- органами державної статистики;
- науково-дослідними інститутами;
- економічними службами банків, бірж, фірм, підприємств.

*Орган спостереження* – це організатор і виконавець статистичного спостереження, в Україні ці функції переважно виконує Державний Комітет статистики з його широкою мережею управлінь та інспекцій. Саме тут вирішують питання про час проведення спостережень, в тому числі вибирають сезон спостереження, встановлюють термін і критичний момент спостереження.

Вибір *місця* проведення обстеження залежить від його мети.

Вибір *часу* спостереження полягає у вирішенні двох питань:

- 1) встановлення критичного моменту або інтервалу часу;
- 2) визначення строку (періоду) спостереження.

*Критичний момент* – конкретний день року і година доби за станом на який повинна бути проведена реєстрація ознак за кожною одиницею досліджуваної сукупності.

*Строк (період) спостереження* – час, протягом якого відбувається

заповнення статистичних формулярів, тобто час для проведення масового збору даних. Цей строк визначається, виходячи з обсягу робіт. Але потрібно враховувати, що його віддалення від критичного моменту призводить до зниження вірогідності отриманих даних.

*Термін або час спостереження* – це час, протягом якого збираються дані про об'єкт спостереження. Коли об'єктом спостереження є процес, то вибирають інтервал часу, протягом якого нагромаджуються дані. Якщо об'єктом спостереження є певний стан, то вибирають критичний момент – момент часу, станом на який реєструються дані.

*Сезон спостереження* треба добирати в такий період року, коли об'єкт перебуває у звичайному для нього стані. Термін спостереження має максимально наближатися до критичного моменту.

Під *періодом проведення спостереження* розуміють суб'єктивний час чи інтервал, в якому вказано початок та закінчення збирання фактів, статистичних даних.

Серед розділів організаційного плану спостереження великого значення набувають підготовчі заходи щодо його проведення, такі як складання різних списків одиниць спостереження, підготовка картографічного матеріалу, розбиття території, підготовка та інструктаж кадрів щодо проведення та заповнення формулярів тощо.


Контроль даних спостереження є методологічним та організаційним питанням. Контроль – це засіб попередження помилок та виявлення і виправлення помилок. Він означає перевірку даних обстежень щодо їх повноти та вірогідності. Повнота даних перевіряється візуально, тобто перевіряють наявність даних за всіма одиницями та позиціями.

Вимоги до даних статистичного спостереження: *повнота; вірогідність; своєчасність; порівнянність даних у часі і просторі; доступність.*

*Способами* перевірки даних на вірогідність є:

– *логічний контроль* – перевірка сумісності даних, яка полягає в порівнянні взаємозалежних ознак;

– *арифметичний контроль* – прямий чи побічний перерахунок зареєстрованих даних.

 Література до теми 1: [4; 5; 7; 8; 15; 16; 17; 18; 21–28].

### ***Контрольні запитання до теми 1***

1. Дайте визначення терміну «статистика».
2. Що є предметом статистики?
3. Що таке статистична методологія?
4. Що називається статистичною сукупністю?
5. Що є одиницею сукупності?
6. Що називається ознакою?
7. Що таке варіація?
8. Що називається статистичним показником?
9. Що є завданням статистичного дослідження?
10. Що таке статистична закономірність?
11. Що називається статистичним спостереженням?
12. Що є метою статистичного спостереження?
13. Що є об'єктом спостереження?
14. Які існують форми, способи і види статистичного спостереження?
15. За якими ознаками можливо класифікувати статистичне спостереження?
16. Що ви розумієте під поточним спостереженням?
17. Що ви розумієте під періодичним спостереженням?
18. Що ви розумієте під одноразовим спостереженням?
19. Які види помилок бувають залежно від причини виникнення?
20. Арифметичний та логічний контроль – це методи чи способи контролю?

## **Тема 2. СТАТИСТИЧНЕ ЗВЕДЕННЯ, ГРУПУВАННЯ. СТАТИСТИЧНІ ТАБЛИЦІ І ГРАФІКИ**

- 2.1. Сутність, організація і техніка статистичного зведення.
- 2.2. Основні види та завдання статистичних групувань.
- 2.3. Принципи вибору групувальної ознаки та утворення груп.
- 2.4. Табличне відображення статистичних даних.
- 2.5. Графічне відображення статистичних даних.

### ***2.1. Сутність, організація і техніка статистичного зведення***

Спостереження дає відомості по кожній одиниці досліджуваного об'єкта, але отримані дані не є узагальнюючими показниками. З їх допомогою не можна зробити висновки про об'єкт без попередньої обробки даних. Зареєстровані у процесі масового статистичного спостереження значення ознак відображають весь діапазон об'єктивно існуючої в сукупності варіації. Перехід від одиничного до загального відбувається завдяки зведенню.

Тому *мета* наступного етапу статистичного дослідження полягає в систематизації первинних даних і одержанні на цій основі зведеної характеристики всього об'єкта за допомогою узагальнюючих статистичних показників. *Зведення* – комплекс послідовних операцій з узагальнення конкретних поодиноких фактів, що утворюють сукупність, для виявлення типових рис і закономірностей, властивих досліджуваному явищу в цілому. *Результат зведення* – докладні дані, що відображають у цілому всю сукупність.

*Статистичне зведення* є другим етапом дослідження масових суспільних явищ, на якому здійснюється систематизація і підсумовування даних з метою отримання узагальнюючої характеристики досліджуваного явища за деякими істотними ознаками. Суть зведення полягає в тому, що матеріали спостереження класифікують та агрегують.

Перехід до ринкових відносин об'єктивно змінює принципові підходи до статистичного зведення та групування. *Основним завданням* зведення є виявлення типових рис та закономірностей масових явищ чи процесів. Зведення є основою подальшого аналізу статистичної інформації. За зведеними даними обчислюють узагальнюючі показники, виконують порівняльний аналіз, а також аналіз причин групових відмінностей, вивчають взаємозв'язки між ознаками.

*Складовими статистичного зведення є:*



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

Зведення статистичних матеріалів здійснюється на основі науково-розробленої програми. Вся багатогранна та складна робота, пов'язана зі статистичним зведенням вихідної інформації, поділяється на такі етапи:

- визначення завдання статистичного зведення;
- вибір груповальної ознаки, кількості груп та інтервалу, формування груп та підгруп;
- перевірка повноти і якості матеріалів спостереження, підбиття підсумків, обчислення необхідних показників для характеристики всієї сукупності та їх окремих частин.

На практиці використовуються різні технологічні схеми комп'ютерної обробки первинних даних, спільними для яких є дві операції:

- 1) кодування даних;
- 2) перенесення первинних даних з документів на технічні носії інформації.

*Коди* – це умовні ідентифікатори ознак. Централізовано розроблено єдині класифікатори, оформлені у вигляді словників, які забезпечують однозначність кодів. Сформовані автоматизовані банки даних уможливають багаторазове використання інформації, при цьому поглиблюється аналіз, зменшується кількість помилок.

За складністю побудови зведення поділяють на прості та групові.

*Просте підсумкове зведення* не передбачає попереднього розподілу на групи одержаних даних. У цьому разі лише визначають загальний підсумок усіх одиниць сукупності або загальний обсяг досліджуваного показника.

*Групове зведення* передбачає попередній розподіл одиниць сукупності на групи. Це дає змогу підрахувати кількість одиниць сукупності та обсяг досліджуваної ознаки в кожній групі.

За способом здійснення статистичне зведення поділяється:

– на *централізоване*, коли первинний матеріал спостереження зосереджується, систематизується та узагальнюється в центральному органі державної статистики – комітеті статистики України;

– на *децентралізоване*, що передбачає узагальнення матеріалів знизу догори за ієрархічними сходинками управління з відповідною обробкою на кожній із них. Під час децентралізованого зведення матеріалів послідовно розробляють відповідні ланки по інстанції за єдиною програмою.

Таким чином, *статистичне зведення* у широкому розумінні – це складна операція наукової обробки первинних статистичних даних, яка охоплює групування матеріалів, розробку системи показників для характеристики типових груп і підгруп, підбиття підсумків у розрізі груп і в цілому за всією сукупністю та зображення згрупованих матеріалів у вигляді таблиць.

## **2.2. Основні види та завдання статистичних групувань**

Окремі одиниці статистичної сукупності поєднуються в групи за допомогою методу групування, що дозволяє «стиснути» інформацію, отриману в ході спостереження і на цій основі *виділити* закономірності, властиві досліджуваному явищу.

*Групування* – це поділ безлічі досліджуваної сукупності на групи за певними істотними для них ознаками.

Завдання, які вирішуються за допомогою методу групувань:

– поділ всієї сукупності на якісно однорідні групи, тобто виділення соціально-економічних типів явищ;

– вивчення складу досліджуваних явищ і структурних змін;

– дослідження взаємозв'язку і залежності між ознаками суспільних явищ.

Групування є одним з найефективніших методів, які перетворюють статистику на одне з наймогутніших знарядь соціального пізнання.

Методологічна сутність групування полягає в тому, що *групування* – це процес утворення однорідних груп на основі розподілу всієї сукупності досліджуваного явища на окремі групи (частини) за найістотнішими ознаками.

Поділ сукупностей на групи пов'язаний з систематизацією, типологією та групуванням. Залежно від змісту і форм досліджуваних ознак статистичне групування здійснюється за допомогою розподілу сукуп-

ності на окремі частини, які характеризуються внутрішньою однорідністю і відрізняються низкою ознак або завдяки об'єднанню окремих одиниць сукупності в групи за типовими ознаками.

Особливим видом групувань є *класифікації*, які широко використовуються у статистиці. За допомогою класифікацій суспільних явищ варіація їхніх ознак фіксується в певному системному вигляді. *Класифікація* – систематизований розподіл явищ і об'єктів на певні групи, класи, розряди на основі їх подібності і відмінності. У класифікації точно визначені різноманітні групи і є докладні показники, які допомагають віднести кожен одиницю до тієї або іншої групи. Класифікація є *основою* групування. Відмітні риси класифікації:

1. В основу класифікації завжди покладена якісна ознака.
2. Класифікації стандартні.
3. Класифікації стійкі. Залишаються незмінними протягом тривалого часу.

Значення статистичних групувань полягає в тому, що вони дають змогу виявити об'єктивний стан речей, властивості досліджуваних явищ, здобути інформацію про розміри окремих груп, їх співвідношення в загальній сукупності та про зв'язки між досліджуваними показниками.

Головний принцип будь-якого поділу ґрунтується на двох положеннях:

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

У кожному дослідженні вирішуються три питання:

- що взяти за основу групування;
- скільки груп, позицій необхідно відокремити;
- як розмежувати групи.

Виділяють такі *види групування*:

1) *Типологічне групування* – поділ якісно неоднорідної сукупності на класи, соціально-економічні типи, однорідні групи, тобто визначення істотних відмінностей між ними та ознаки, які є спільними для всіх груп. Їхнє завдання – виявлення соціально-економічних типів або однорідних в істотному відношенні груп. Необхідність даного групування зумовлена потребою теоретичного узагальнення первинної статистичної інформації та подальшого одержання узагальнюючих показників. Застосовуються при вивченні розподілу підприємств за формами власності та

суспільного виробництва за економічним призначенням продукції, групування населення за суспільними групами (наприклад, поділ держав за рівнем затрат праці на виробництво 1 ц сільськогосподарської продукції).

2) *Структурне групування* – характеризує розподіл якісно однорідної сукупності на групи за певною ознакою. Їх завдання – вивчення складу окремих типових груп за допомогою об'єднання одиниць сукупності, близьких одна до одної за величиною групувальної ознаки. Використовують для пізнання явищ суспільного життя, виявлення закономірностей розподілу одиниць сукупності за варіюючими значеннями досліджуваної ознаки, для вивчення складу сукупності та структурних зрушень, а також в разі вивчення підприємств за галузями виробництва, кількістю працівників, обсягом продукції, для дослідження складу населення – за статтю, віком, національністю, освітою тощо.

Структурні групування, як і типологічні, можна здійснювати:

– за *атрибутивною ознакою*, яка передбачає, що групи відрізняються між собою не розміром, а характером ознаки. Кількість груп визначається кількістю різновидів атрибутивної ознаки (наприклад, групування працівників за статтю припускає дві групи, а групування за професіями – стільки груп, скільки професій);

– за *кількісною ознакою* потрібно визначити оптимальну кількість груп та простежити, щоб не зникли особливості досліджуваного явища.

3) *Аналітичне групування* – за його допомогою виявляють та вивчають зв'язок між показниками, а також пояснюють причини закономірностей та їх зміни в часі і просторі. Їх завдання – виявлення впливу одних ознак на інші (виявлення зв'язку між соціально-економічними явищами). *Специфічною особливістю аналітичних групувань* є те, що кожна група факторної ознаки характеризується середніми значеннями результативної ознаки. *Перевага методу аналітичних групувань* перед іншими методами аналізу зв'язку (наприклад, кореляційно-регресійним) полягає в тому, що він не потребує дотримання яких-небудь умов для свого застосування, крім однієї – якісної однорідності сукупності.

4) *Комбінаційні групування* – в них проводиться поділ сукупності на групи за двома або більше ознаками. При цьому групи, утворені за однією ознакою, розбиваються на підгрупи за іншою ознакою. Такі групування дають можливість вивчити структуру сукупності за декількома ознаками одночасно.

Соціально-економічний аналіз припускає використання системи

простих і комбінаційних групувань. Також дуже часто вдаються до вторинного групування – перегрупування вже згрупованих даних. Вторинне групування може бути проведене методом простого укрупнення інтервалів, крім того, часто використовується процентне перегрупування.

Всі явища суспільного життя та їх ознаки щільно пов'язані між собою і залежать одне від одного. У підприємницькій діяльності трапляються різні взаємозв'язки між ознаками, які можуть впливати на причину або результат явища. З них можна виділити такі:

1) фактор – кількісна ознака, а результат – якісна (наприклад, взаємозв'язок між стажем роботи і кваліфікацією робітника, тривалістю договірних зобов'язань з постачальниками матеріальних ресурсів та їх ритмічністю);

2) фактор – якісна ознака, а результат – кількісна (наприклад, взаємозв'язок між кваліфікацією робітників та їх виробітком);

3) фактор і результат – якісні ознаки (наприклад, взаємозв'язок між категорією працівників та їх освітній рівень, кваліфікацією робітників та продуктивністю їх праці);

4) фактор і результат – кількісні показники (наприклад, взаємозв'язок між виробітком робітників та рівнем середньої заробітної плати).

Ознаки, за якими здійснюється розподіл одиниць сукупності на групи, називаються *групувальними ознаками* або *основою групування*.

*Групувальною ознакою* називається ознака, за якою проводиться розбиття одиниць сукупності на окремі групи. В основу групування можуть бути покладені як кількісні, так і якісні ознаки. Усю сукупність ознак групувань можна поділити на дві групи: факторні й результативні.

*Факторними* називаються ознаки, під впливом яких змінюються інші ознаки, іменовані *результативними*. Взаємозв'язок виявляється в тому, що зі зміною значення факторної ознаки систематично змінюється *середнє* значення результативного.

Групування *за кількістю ознак* бувають:

- *прості* – групи утворені за однією ознакою;
- *складні* – поділ сукупності на групи проводиться за двома або більше ознаками, узятими у комбінації (приклад – групування родин України за місцем проживання і кількістю дітей). На практиці будують складні групування не більше, ніж за трьома ознаками.

### 2.3. Принципи вибору групувальної ознаки та утворення груп

Всі ознаки, за якими можна здійснити групування, класифікують таким чином:

1. За формою вираження:
  - *атрибутивні* – характеризують властивість, якість явища і не мають кількісного виразу (стать, професія, освіта);
  - *кількісні* (варіаційні) – набувають різного кількісного вираження у певних одиницях досліджуваної сукупності (кількість працівників, обсяг виробництва, вартість основних виробничих фондів);
  - *дискретні* – описуються цілими числами, без проміжних значень (розряд працівників);
  - *безперервні* – можуть набувати різного значення але в певних межах, тобто виражатися не лише цілими числами, а й дробовими (так, рівень оплати праці – в гривнях чи в гривнях з копійками);
2. За характером варіативності:
  - *альтернативні* – коли одним одиницям властиві певні якісні параметри, а іншим – ні (наприклад, випущена продукція може бути якісною чи неякісною);
  - *варіаційні* – мають багато кількісних значень (наприклад, вартість основних виробничих фондів, чисельність робітників тощо);
3. За роллю ознаки у взаємозв'язку досліджуваних явищ:
  - *факторні* – впливають на інші ознаки;
  - *результативні* – їх розмір і динаміка формуються під впливом інших ознак.

Після обробки кількісних значень, їх систематизації, отримують певний цифровий ряд, який називається статистичним рядом. Він має дві форми: ряд розподілу і динамічний ряд.

*Ряд розподілу* – це впорядкований за певною варіюючою ознакою розподіл сукупності, розташованої в певному порядку (зростання, спадання тощо). *За характером розподілу* варіаційні ряди можуть бути симетричні і асиметричні. Ряд розподілу, де частоти спочатку наростають, а потім спадають, називається *симетричним*. Ряд розподілу, в якому частоти розташовані несиметрично від середини, називається *асиметричним або скошеним*.

За виглядом групувальної ознаки *ряди розподілу* поділяються на:

- 1) *Атрибутивні* – групувальна ознака якісна. Ряд розподілу одиниць сукупності, в основу якого покладено якісні ознаки, називається

*атрибутивним*. Прикладом атрибутивного ряду розподілу може бути розподіл населення на міське і сільське (табл. 2.1).

Таблиця 2.1 – Приклад атрибутивного ряду розподілу

Ознака зросту	Низькі	Середні	Високі
Кількість студентів	10	56	15

2) *Варіаційні* – групувальна ознака кількісна (табл. 2.2, табл. 2.3, табл. 2.4). Ряд розподілу одиниць сукупності за ознакою, що має кількісне вираження, називається *варіаційним*. Варіаційний ряд розподілу має свої особливості. Він складається з двох елементів: варіант і частот.

Таблиця 2.2 – Дискретний варіаційний ряд

Курс	1	2	3	4	5
Кількість студентів	60	54	48	47	47

Таблиця 2.3 – Інтервальний ряд

Зріст, см	150–170	170–190	190–210	210–230
Кількість студентів	10	22	34	15

Таблиця 2.4 – Інтервальні ряди, зображені в дискретному вигляді

Зріст, см	160	180	200	220
Кількість студентів	10	22	34	15

*Варіантами* називають числові значення розмірів кількісної ознаки. Числа, які відповідають цим варіантам, називаються *частотами*. Частоти можуть виражатися як в абсолютних, так і у відносних одиницях (наприклад у відсотках). Відповідно до варіації ознаки, варіаційні ряди розподілу поділяються на:

а) *Дискретні* – значення ознаки змінюється дискретно (див. табл.2.2). В *дискретному ряді* розподілу кількісна ознака набуває тільки цілих значень.

б) *Інтервальні* – значення ознаки змінюється безперервно (табл. 2.3). Коли значення варіантів ряду виражено у вигляді інтервалу, такий ряд розподілу називається *інтервальним*. При необхідності інтервальні ряди зображують в дискретному вигляді, для цього переходять до середин інтервалів (табл. 2.4).

Накопичення часток у міру зростання (зменшення) ознаки інтер-

вального ряду називається кумулятивною частотою. Приклад формування кумулятивної частоти інтервального ряду наведено в табл. 2.5.

Таблиця 2.5 – Ранжований дискретний ряд розподілу

Групи сімей – варіанта	Кількість сімей – частота	(Накопичена) Кумулятивна частота
1	80	80
2	20	80+20=100
3	350	100+350=450
4	45	450+45=495
5	5	495+5=500

Залежно від характеру розподілу одиниць сукупності інтервали можуть бути рівними або нерівними, відкритими або закритими, обмеженими або необмеженими (приклад ряду розподілу з рівними відкритими інтервалами наведено в табл. 2.6).

Таблиця 2.6 – Ряд розподілу з рівними відкритими інтервалами

Зріст, см	менше 160	160–190	190–200	більше 200
Кількість студентів	10	22	34	15

Після визначення групувальної ознаки важливим кроком є розподіл одиниць сукупності на групи. Для цього визначають кількість утворених груп та розмір інтервалу.

Кількість груп здебільшого залежить від того, яку ознаку покладено в основу групування. Кількість груп тісно пов'язана з обсягом сукупності. Орієнтовно оптимальна кількість груп визначається за стандартними процедурами, зокрема за формулою, яка запропонована американським вченим Стерджессом

$$k = \sqrt{n} \quad (2.1)$$

де  $n$  – обсяг сукупності;  $k$  – кількість інтервалів.

Інтервали груп використовують лише в разі значної кількості дискретної ознаки, при безперервній зміні кількісної ознаки. Інтервали являють собою каркас групувань. На практиці вони утворюються за такими формальними *принципами*: рівності інтервалів, кратності інтервалів, рівності частот.



Залежно від ступеня варіативності групувальної ознаки, характеру розподілу статистичної сукупності застосовують рівні або нерівні інтервали. Коли варіація ознаки виявляється у порівняно вузьких межах і розподіл більш-менш рівномірний, використовують однакові інтервали.

Значення інтервалу в разі групування із застосуванням рівних інтервалів визначають за формулою

$$\text{—————}, \quad (2.2)$$

де  $h$  – розмір інтервалу;  $x_{\max}$ ,  $x_{\min}$  – максимальне і мінімальне значення ознаки;  $K$  – кількість груп.

Інтервали при групуванні можуть бути відкритими (з однією межею – верхньою або нижньою) і закритими (що мають верхню та нижню межі).

Рівні інтервали – інтервали з однаковою різницею між верхньою і нижньою границями кожного проміжку.

Нерівні інтервали – інтервали з різними різницями між верхньою і нижньою границями в різних проміжках.

Групування за однією ознакою називають *простим*, а за двома і більше ознаками – *комбінаційним*.

У комбінаційних групуваннях ознаки ієрархічно впорядковуються за змістом чи за вагомістю. Групи, утворені за першою ознакою, поділяються на підгрупи за другою, а ті можуть поділитися на підгрупи за третьою тощо. На кожному етапі використовується лише одна ознака, тобто відбувається послідовне описування груп.

Іноді доводиться перегруповувати дані, передусім щоб забезпечити порівнянність структур двох сукупностей за однією тією ж самою ознакою. Результат перегруповання називають *вторинним групуванням*. Перегрупування виконують або об'єднанням, або розбиттям інтервалів первинного групування.

Розрізняють два *способи вторинного групування*:

1) *просте укрупнення інтервалів* (наприклад, при порівнянні галузей за розміром основних фондів, оскільки групи підприємств, утворені в кожній з галузей, переважно не збігаються);

2) *перегрупування за часткою окремих груп у загальному їх підсумку* (наприклад, при зіставленні підприємств за рівнем продуктивності праці, оскільки групи мають різні інтервали рівня продуктивності).

## 2.4. Табличне відображення статистичних даних

Результати статистичного зведення подаються у формі статистичних таблиць, макети яких розроблюються одночасно з програмою обробки даних.

Статистичні таблиці – спосіб раціонального, наочного, систематизованого викладу та аналізу цифрових характеристик суспільних явищ і процесів. Статистичні таблиці є засобом наочного виразу результатів дослідження. Значення таблиць визначається тим, що вони дозволяють ізольовані статистичні дані розглядати спільно, достатньо повно і точно охоплюючи складну природу явищ. Статистичні таблиці дають змогу найбільш стисло, компактно, без будь-яких зайвих пояснень викласти зведену обробку статистичних матеріалів. За рахунок того, що об'єкти та їх показники розташовують за певною системою, яка дає змогу внести їх назви у вигляді заголовків, досягають переваги наочності.

*Статистичною таблицею* називається таблиця, яка містить зведену числову характеристику досліджуваної сукупності за однією або декількома істотними ознаками, взаємопов'язаними логікою економічного процесу. Якщо із статистичної таблиці вилучити всі слова і цифри, то вийде графлена сітка. Вертикальні стовпці її називаються графами, а горизонтальні – рядками. Якщо записати заголовки граф і рядків, то вийде макет таблиці.

*Макет статистичної таблиці* – це комбінація горизонтальних та вертикальних граф, на перетині яких утворюються клітинки. Ліві бокові клітинки призначені для словесних заголовків – переліку складових сукупності та системи показників, решта – для числових даних. Основний зміст таблиці розкриває її назва. Складену, але не заповнену цифрами таблицю називають *макетом таблиці*.

За логічним змістом статистична таблиця розглядається як «статистичне речення». *Статистичним підметом* називають те, про що йдеться в таблиці (тобто об'єкт дослідження: перелік елементів сукупності, їх групи, окремі територіальні одиниці або часові інтервали). *Статистичний присудок* – це числові підсумки, які характеризують статистичний підмет (тобто система показників).

На практиці складають робочі (допоміжні) та заключні таблиці. У статистичній таблиці виділяють підмет та присудок.

*Підмет таблиці* – це перелік одиниць сукупності або групи, тобто об'єкт вивчення. *Підмет* – це статистична сукупність, ті об'єкти чи

частини їх, які характеризуються показниками. Підмет розташовується зліва у вигляді назв рядків.

*Присудок таблиці* – це цифрові дані, що характеризують підмет. Присудок – це показники, що характеризують статистичну сукупність. Присудок формує заголовки граф і складає їх зміст. Статистичний присудок, об'єктивно перебуваючи в діалектичному взаємозв'язку із статистичним підметом, має бути поданий таким чином, аби за допомогою системи його показників можна було б одержати повну характеристику відокремлених груп, охарактеризувати їх істотні особливості. Показники в присудку розташовують у логічній послідовності. Присудок таблиці має бути не набором випадкових цифр, а певною системою думок, висловлених мовою цифр.

Приклад формування статистичної таблиці наведено в табл. 2.7.

Використовують різні таблиці, які залежно від побудови статистичного підмета поділяються на:

1) *Прості* – в них підмет не поділяється на окремі групи, а становить перелік будь-яких об'єктів чи територіальних одиниць.

Таблиця 2.7 – Складові статистичної таблиці

Підмет	Присудок		Заголовки граф				
А			1	2	3	4	5
Перелік (групи) одиниць сукупності							

Прості таблиці бувають:

- перелікові (підмет – перелік одиниць, складових об'єкта вивчення);
- територіальні (дається перелік територій, областей, міст та ін.);
- хронологічні (у підметі приводяться періоди часу або дати).

2) *Групові* – коли підмет поділяється на окремі групи за однією з істотних ознак. З таких таблиць отримують більше інформативного матеріалу порівняно з текстом для аналізу досліджуваних явищ. У підметі дається не перелік одиниць сукупності, а їх групи.

3) *Комбінаційні* – статистичний підмет поділяється за двома або більше ознаками, взятими у комбінації. Їх пізнавальна сторона полягає в тому, що з'являється можливість простежити вплив на ознаки присудка не одного, а двох і більше факторів, тобто ознак, які лягли в основу комбінованого групування. Кожна з груп, на які розбивається підмет, у свою чергу розбивається на підгрупи.

Таблиці за характером присудка поділяються на два авиди:

1) *проста розробка* – використовується лише 1-2 окремо взяті ознаки.

2) *складна розробка* – використовується комбінація ознак.

Практикою вироблено певні *правила* складання і оформлення таблиць:

1. Таблиця повинна бути невеликою за розміром, простою для читання та аналізу, не слід перевантажувати її зайвими подробицями.

2. Кожна таблиця повинна мати докладну назву. Загальний заголовок таблиці повинен коротко характеризувати її зміст, а бокові заголовки – зміст підмета. Заголовки мають бути чіткими, лаконічними, змістовними, без скорочень, з яких можна буде зрозуміти таке:

а) яке коло питань висвітлює або ілюструє таблиця;

б) які географічні межі статистичної сукупності наведено в таблиці;

в) період часу, за який наведено дані;

г) одиниці вимірювання, якщо вони однакові для всіх табличних клітинок.

3. Рядки та графи доцільно нумерувати. Це полегшує користування таблицею, показує спосіб розрахунку цифр в графах, дає можливість краще орієнтуватись. Перші графи, що містять підмет, позначаються великими буквами алфавіту; графи, що містять присудок, нумеруються арабськими цифрами. Заголовки рядків підмета і присудка повинні бути сформульовані коротко, точно і ясно. Всі слова в заголовках записуються, якщо можливо, повністю. Заголовки граф слід сформулювати так, щоб був зрозумілим економічний зміст даної величини і порядок її розрахунку.

4. Значення показників, якщо можливо, слід округлювати у межах рядка чи графи з однаковим ступенем точності (до 0,1; до 0,01 і таке інше).

5. Ознаки, які наводяться в підметі і присудку, повинні бути розташовані в логічному порядку з урахуванням необхідності розглядати їх спільно. Звичайний принцип розміщення – від часткового до загального, тобто спочатку показують доданки, а в кінці підбивають підсумки (якщо це необхідно). Коли приводяться не всі доданки, а лише найбільш важливі з них, застосовується протилежний принцип – спочатку показують загальні підсумки, а потім виділяють найбільш важливі частини («Зокрема», «З них»). Слід розрізняти «РАЗОМ» і «ВСЬОГО»: «РАЗОМ» є підсумком для певної частини сукупності, а «ВСЬОГО» –

підсумком для сукупності в цілому.

6. Таблиця може супроводжуватись примітками, в яких вказуються джерела наведених даних, детальніше розкривається зміст показників, даються інші пояснення, а також формули для розрахунків у випадку, якщо таблиця містить дані, отримані в результаті обчислень.

7. Якщо в таблиці наводяться відсотки зростання, то у багатьох випадках доцільно відсотки від 300 і більш замінювати відношеннями в разях. Наприклад, писати не «1000 %», а «в 10 разів».

8. При оформленні таблиць зазвичай застосовуються такі умовні позначення:

- знак «тире» (–) – коли явище відсутнє;
- «х» – якщо явище не має осмисленого змісту;
- трикрапка (...) – коли відсутні відомості про розмір явища (або робиться запис «Немає відомостей» або «н/д»). Якщо відомості є, але числове значення менше від прийнятої в таблиці точності, воно виражається дробовим числом (0,0).

Приклад сформованої комбінаційної таблиці наведено в табл. 2.8.

Таблиця 2.8 – Групування банків України за величиною статутного фонду на 1 березня 2013 року, млн грн

Розмір статутного капіталу	Прибутковість активів	Кількість банків	Питома вага в загальній кількості банків, %
5,0–7,5	1,5–2,0	2	11
	2,0–2,5	1	23
	2,5–3,0	5	11
<b>Разом</b> (підсумок частини сукупності)		<b>8</b>	<b>45</b>
7,5–10,0	1,5–2,0	7	10
	2,0–2,5	2	25
	2,5–3,0	3	20
<b>Разом</b> (підсумок частини сукупності)		<b>12</b>	<b>55</b>
<b>ВСЬОГО</b> (в цілому за сукупністю)		<b>20</b>	<b>100</b>

Таблиці-матриці. *Таблиця спряженості* – це таблиця, яка містить зведену числову характеристику досліджуваної сукупності за двома та більше атрибутивними ознаками, або комбінації атрибутивних і кількісних ознак. Застосовується при вивченні соціальних явищ. Таблиці

спряженості бувають великих розмірів. Найбільш простий вид матриці (у цьому випадку, таблиці спряженості) – 2x2. Прикладом таблиці спряженості є табл. 2.9.

Таблиця 2.9 – Приклад таблиці спряженості

	$B_1$	$B_2$	Всього
$A_1$	$A_{11}$	$A_{12}$	$A_{10}$
$A_2$	$A_{21}$	$A_{22}$	$A_{20}$
Всього	$A_{01}$	$A_{02}$	$A_{00}$

Внутрішнє цифрове наповнення таблиці:  $A$  – це частоти, що володіють одночасно  $i$ -м значенням однієї ознаки і  $j$ -м значенням іншої. Розрізняють два види матриць:

- прямокутна (розмірність  $m \times n$ );
- квадратна (розмірність  $n \times n$ ). Якщо число рядків строго дорівнює числу стовпців ( $m=n$ ), то матриця називається квадратною порядку  $n$ .

Матриці і аналіз явищ і процесів на їх основі становлять базу матричного моделювання і дозволяють досліджувати взаємозв'язки між економічними об'єктами.

### 2.5. Графічне відображення статистичних даних

Статистичні графіки – спосіб раціонального, наочного, систематизованого викладу та аналізу цифрових характеристик суспільних явищ і процесів. Статистичні графіки є засобом наочного вираження результатів дослідження. Статистичні графіки дають змогу найбільш стисло, компактно, без будь-яких зайвих пояснень, викласти зведену обробку статистичних матеріалів. За рахунок того, що об'єкти та їх показники розташовують за певною системою, яка дає змогу здійснити їх аналіз, досягають переваги наочності.

*Графіками* в статистиці називають умовні зображення числових величин і їх співвідношень у вигляді різних геометричних образів – крапок, ліній, плоских фігур і таке інше. Кожний графік складається із графічного образу, допоміжних елементів та експлікації.

1) *Графічний образ* – це сукупність крапок, ліній і фігур, за допомогою яких зображуються статистичні дані. Ці знаки утворюють властиву язику тканину графіка, його основу, мову графіки.

2) *Допоміжними елементами* графіка є:

- поле графіка – простір, у якому розміщуються геометричні

знаки. Поле графіка характеризується його форматом, тобто розміром і пропорціями (співвідношенням сторін);

– просторові орієнтири, що визначають розташування геометричних знаків у полі графіка. Просторові орієнтири задаються системою координатних сіток або контурних ліній, які ділять це поле на частини. У більшості випадків у статистичних графіках застосовується система прямокутних (декартових) координат, але нерідко зустрічаються і кругові графіки, побудовані за принципом полярних координат.

3) *Експлікація графіка* складається з пояснення:

– предмета, зображеного на графіку (його назви);  
– сутнісного значення кожного знака, застосованого на даному графіку. Без експлікації графік не можна прочитати і зрозуміти. Назва графіка повинна коротко і точно розкривати його складові та сутність. Пояснювальні тексти можуть розташовуватися в межах графічного образу або поруч із ним (ярлики), а також можуть виноситись за його межі (ключ).

*Система координат* необхідна для розміщення геометричних знаків на полі графіка. Найпоширенішою системою координат при побудові статистичних графіків є система прямокутних координат. При цьому найкраще співвідношення масштабу по осях абсцис і ординат, яке дорівнює «1,62:1» і називається «Золотим перетином».

*Масштабні орієнтири* визначаються масштабом і масштабною шкалою. Масштаб – спосіб переведення числової величини в графічну. Масштабна шкала – лінія, окремі крапки якої можуть бути прочитані як певні числа. Шкали бувають рівномірними і нерівномірними. Масштаб рівномірної шкали – це довжина відрізка, прийнятого за одиницю виміру.

Статистичні графіки можна класифікувати за різними ознаками:

*За складом* можна виділити графіки порівняння, графіки різних відносних величин (структури, динаміки тощо) графіки варіаційних рядів, графіки розміщення по території, графіки взаємозалежних показників.

*За характером графічного образу* розрізняють графіки крапкові, лінійні, об'ємні і поверхневі (стовпчикові, смугові, квадратні, кругові, секторні, фігурні).

Класифікацію статистичних графіків за формою графічного образу та способом і завданням побудови наведено на рис. 2.1 та рис. 2.2 відповідно.

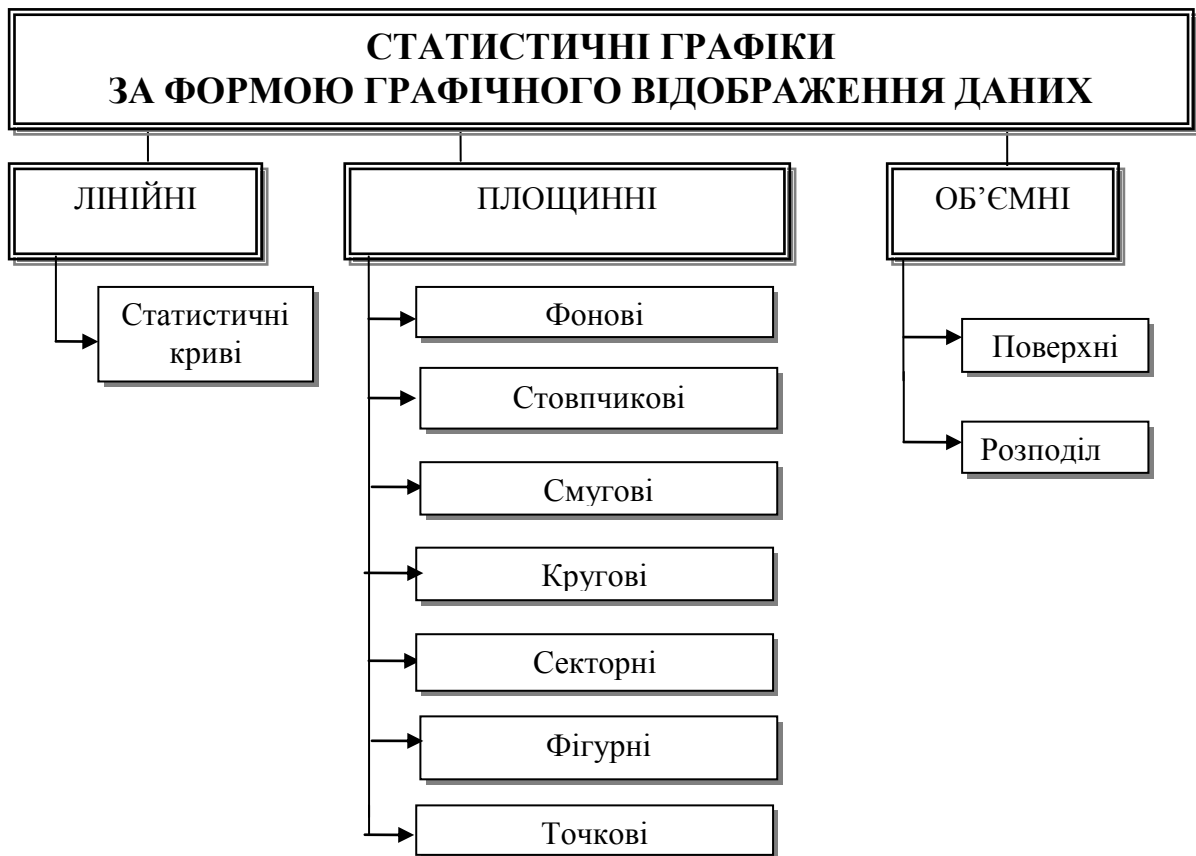


Рисунок 2.1 – Класифікація статистичних графіків за формою графічного образу

Найбільш простими і в той же час достатньо широко розповсюдженими в економіко-статистичному аналізі є *діаграми*, а саме:

– *Кільцева діаграма* – різновид секторної діаграми; використовується для наочного зображення структури соціально – економічних явищ за двома і більше ознаками статистичних даних. Можливо одночасно відображати структуру декількох ознак.

– *Радіальні діаграми* – будуються в полярних координатах для наочного зображення циклічних соціально-економічних процесів. Використовуються для вивчення і аналізу сезонних коливань і бувають двох видів: *замкнуті і спіральні*.

– *Шарові діаграми (з областями)* – для графічного зображення зміни у часі явища, яке досліджується. Використовуються для одночасного зображення процесу змін соціально – економічного явища у часі як у цілому, так і окремих його складових частин, тобто зміни його структури.

– *Знаки Варзара* – застосовуються, коли потрібно наочно відобразити сукупність основних ознак досліджуваних явищ і взаємозв'язок між



ними. Використовуються для одночасного зображення трьох величин, пов'язаних між собою таким чином, що одна величина є добутком двох інших. Знак Варзара являє собою прямокутник, у якого один співмножник прийнятий за основу, інший – за висоту, а вся площа дорівнює добутку. Обидва показники відкладаються на шкалах (кожний на своїй), третій (результат) зображується у вигляді прямокутника в полі графіка.

– *Фігурні діаграми* – спосіб відображення статистичних даних у вигляді фігур, символів, рисунків певного масштабу; доцільні при застосуванні різних розмірів одного символу пропорційно величині показника та різної кількості однакових за розміром символів.

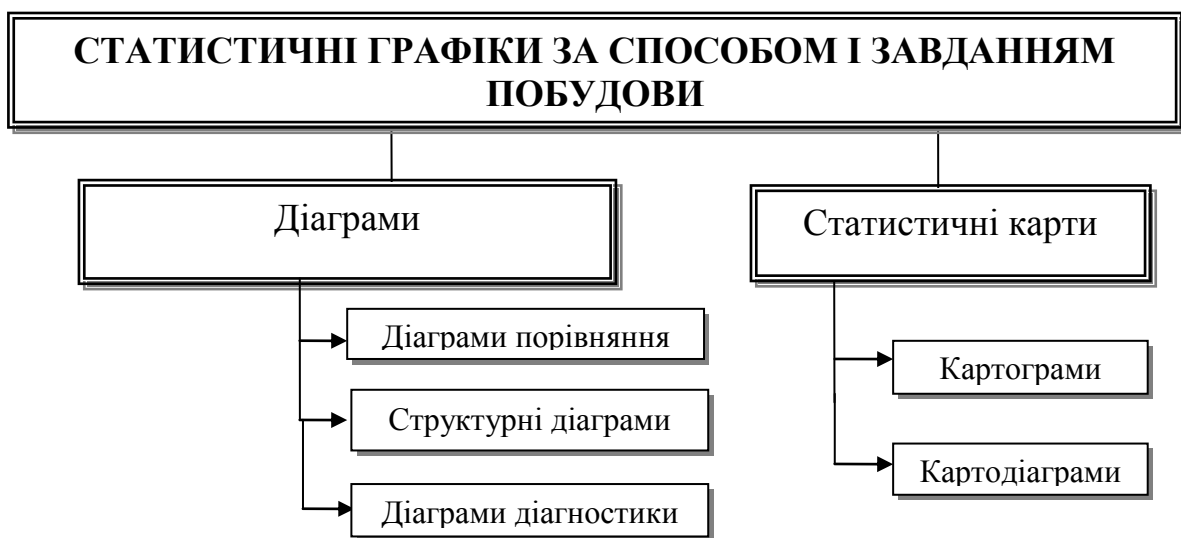


Рисунок 2.2 – Класифікація статистичних графіків за способом і завданням побудови

Для оцінки географічного розміщення явищ, порівняльного аналізу за територіями застосовуються статистичні карти, які містять у собі картограми й картодіаграми.

– *Картограма* – статистична карта, на якій розподіл досліджуваної ознаки по території відображено умовно, залежно від величини статистичної ознаки. Картограми бувають двох видів: фонові; точкові.

– *Картодіаграма* – графічне зображення закономірностей просторового розміщення статистичних показників за допомогою діаграмних фігур.

Іноді для цілей порівняльного аналізу по регіонах і країнах використовують квадратні, кругові, фігурні діаграми. Приклади застосування діаграм:

1) *Лінійні діаграми* – найчастіше застосовуються для характеристи-

ки динаміки, тобто для оцінки зміни явища в часі (приклад застосування лінійної діаграми наведено на рис. 2.3).

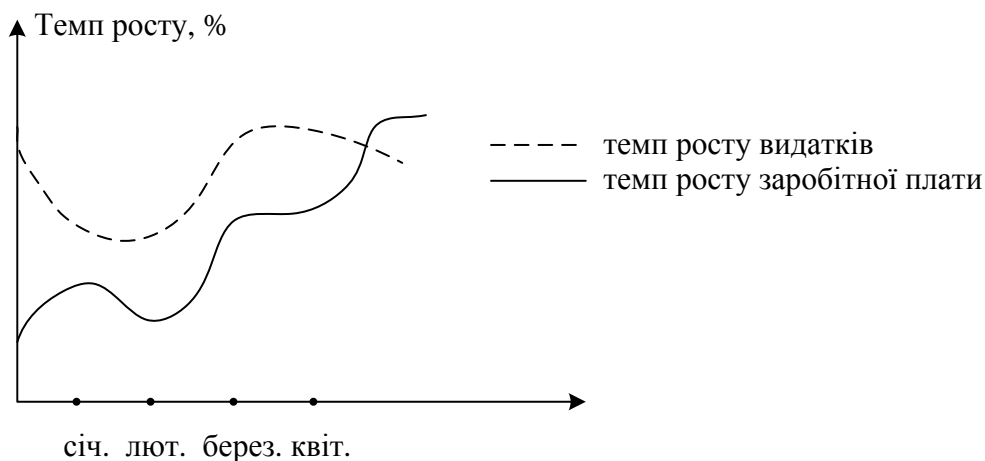


Рисунок 2.3 – Динаміка співвідношення темпів зростання витратів і заробітної плати

2) *Стовпчикові діаграми* – використовуються для порівняльної оцінки розміру явищ (приклад застосування стовпчикової діаграми наведено на рис. 2.4).

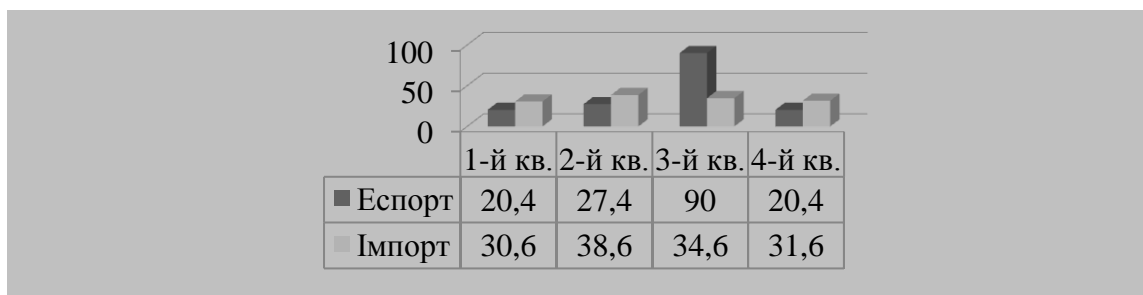


Рисунок 2.4 – Динаміка експорту-імпорту в Україні в 2013 р. (по кварталах)

3) *Секторні діаграми* – використовуються для характеристики структури соціально-економічних явищ (приклад застосування секторної діаграми наведено на рис. 2.5). Коло ділиться на сектори пропорційно питомій вазі частин у цілому. Розмір кожного сектора визначається за величиною кута з урахуванням того, що 1 % відповідає 3,6 градусам.

4) *Смугові діаграми* – зручні для застосування, коли окремі об'єкти порівняння характеризуються протилежними за знаком показниками (приклад застосування смугової діаграми наведено на рис. 2.6).

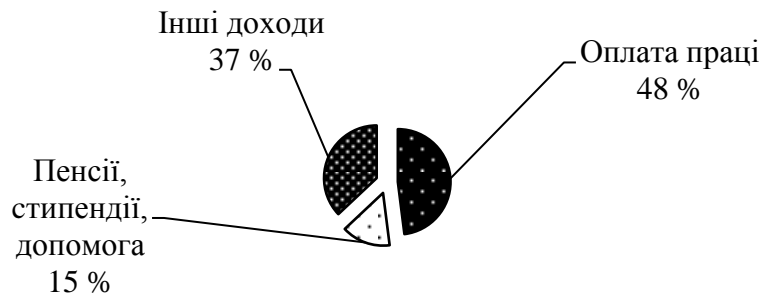


Рисунок 2.5 – Структура грошових доходів населення України

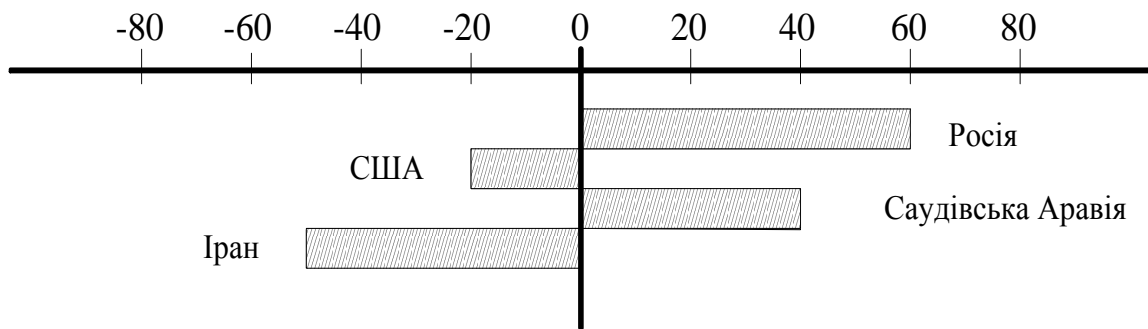


Рисунок 2.6 – Видобуток нафти в 2013 році у порівнянні з 2000 роком, %

📖 Література до теми 2: [4; 5; 7–9; 15–21; 23; 25–29].

### **Контрольні запитання до теми 2**

1. Що називається статистичним зведенням?
2. Що називається статистичним угрупованням?
3. Які існують види статистичних угруповань?
4. Що називається статистичною таблицею?
5. Класифікація таблиць.
6. Що є рядом розподілу?
7. Групувальна ознака, її види.
8. За допомогою якої формули визначається число груп і як вона називається?
9. Як визначити розмір інтервалу при групуванні з рівними інтервалами?
10. Які види групувань за кількістю ознак групування Ви знаєте?

11. Як називається групування за однією ознакою?
12. Що таке структурне групування?
13. Що таке типологічне групування?
14. Що таке аналітичне групування?
15. Що таке комбінаційне групування?
16. Що таке багатовимірне групування?
17. Дайте визначення поняття статистичної таблиці.
18. Що ви розумієте під присудком статистичної таблиці?
19. Що ви розумієте під підметом статистичної таблиці?
20. Які види статистичних графіків Ви знаєте?

## Тема 3. СТАТИСТИЧНІ УЗАГАЛЬНЮЮЧІ ПОКАЗНИКИ

- 3.1. Види та функції статистичних показників.
- 3.2. Абсолютні та відносні статистичні показники.
- 3.3. Середні величини.
- 3.4. Основні правила і особливості застосування середніх для розрахунків економічних показників

### ***3.1. Види та функції статистичних показників***

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

*Статистичний показник* – це узагальнююча характеристика суспільних явищ і процесів, в якій поєднується їх кількісна та якісна ознаки. Показник є мірою відображення певної властивості соціально-економічного явища чи процесу, яка має кількісну і якісну визначеність.

*Якісний зміст* показника визначається суттю явища і відображається у його назві. *Кількісна сторона* подається числом та його вимірником. Єдність якісного і кількісного відображення властивостей показника забезпечує його модель, тобто правило побудови показника. Модель показника розкриває його структуру. У моделі обґрунтовуються одиниці вимірювання, технологія збирання даних та обчислювальні операції. Модель показника має надзвичайно важливе значення для забезпечення *вірогідності статистичної інформації*.

*Адекватність* – здатність показника відобразити ту властивість явища, яка досліджується. *Адекватність* або *відповідність* показника полягає у його спроможності відобразити ті властивості явищ, які передбачені програмою статистичного дослідження. Для однієї категорії можна побудувати декілька показників. Наприклад, рентабельність виробництва можна оцінити на основі валового або чистого прибутку відносно рівня капіталу чи активів.

*Точність та повнота* обліку залежить від організації збору та обробки даних. *Точність* оцінки залежить від структури показника, організації статистичного спостереження та обробки отриманих даних. Таким чином, щоб показник відповідав своєму призначенню і виконував покладені на нього функції, потрібно логічно та статистично обґрунтувати модель формування і складу показника.

Статистичні показники виконують такі *функції*: пізнавальну; управлінську; контрольну; стимулюючу.

Показники є базою для аналізу та прогнозування соціально-економічного розвитку країни, її окремих галузей і регіонів, стану та розвитку досліджуваних явищ, напряму та інтенсивності процесів, які відбуваються у суспільстві. Вивчаючи явища кількісно, економіст аналізує їх, пізнає якісно, проникає в сутність. У цьому виявляється *пізнавальна функція* показників. *Управлінська функція* статистичних показників полягає в тому, що вони є важливим елементом процесу управління на всіх його рівнях. З розвитком ринкових відносин підвищується значення *контрольної та стимулюючої функції* показників.

Існує така *класифікація показників*:

1. За суттю досліджуваних явищ:
  - *об'ємні показники* – характеризують розміри явищ, процесів (наприклад, обсяг виробництва, реалізації);
  - *якісні показники* – характеризують кількісні співвідношення, характерні властивості досліджуваних явищ (наприклад, продуктивність праці);
2. За ступенем агрегування явищ:
  - *індивідуальні показники* – відображають розміри ознаки окремих одиниць сукупності;
  - *загальні показники* – виражають розміри ознаки окремих груп показників або всієї сукупності;
3. За часовою ознакою:
  - *інтервальні показники* – виражають розміри кількісної ознаки за певні періоди часу (обсяг капітальних вкладень за тиждень, декаду, місяць, квартал, рік);
  - *моментні показники* – виражають розміри кількісної ознаки на певний момент. Моментні показники дають кількісну характеристику явищ на певний момент часу (поголів'я великої рогатої худоби на початок кожного року, перепис багаторічних насаджень на початок року, чисельність працівників на перше число кожного місяця);
4. За способами обчислення:
  - *первинні* – визначають статистичне спостереження і подаються у формі абсолютних величин. Первинні показники отримують в результаті зведення матеріалів статистичного спостереження у формі абсолютних величин (кількість заводів галузі, вартість їх основних виробничих фондів і випущеної продукції);

– *похідні* – обчислюються на базі первинних, мають форму середніх або відносних величин. На основі первинних даних обчислюють вторинні (середню вартість основних виробничих фондів і випуск продукції в середньому на один завод, а при визначенні інтенсивності використання основних виробничих фондів – похідні показники другого порядку – фондовіддачу, фондоємність).

Статистичні показники пов'язані між собою, тому їх розглядають в певному взаємозв'язку. Для характеристики діяльності підприємства необхідно розглядати декілька показників, які перебувають в певному взаємозв'язку і утворюють систему статистичних показників.

*Система показників* – це єдиний комплекс характеристик складного предмета пізнання. Системі показників властиві такі риси:

- *всебічність* кількісного відображення явищ;
- *організаційний взаємозв'язок* окремих показників.

Систему показників визначають як ієрархічну структуру, на верхньому рівні якої знаходиться узагальнюючий (інтегральний) показник, а на нижньому – часткові показники, які об'єднуються у блоки.

### ***3.2. Абсолютні та відносні статистичні показники***

*Абсолютні величини* – це показники, які виражають розміри суспільно-економічних явищ і процесів в конкретних умовах місця і часу. Визначають їх методами статистичного спостереження і зведення вихідної інформації. Абсолютні величини широко використовуються для аналізу та прогнозування.

Абсолютні величини є джерелом формування статистичної інформації, за їх допомогою оцінюють усі сторони суспільного життя. У статистиці розрізняють такі *типи абсолютних величин*:

1) *за способом вираження розмірів досліджуваних явищ*:

- *індивідуальні* – характеризують кількісні ознаки окремих одиниць сукупності (наприклад, рівень виробітку окремого робітника за конкретний період);
- *підсумкові* – характеризують розмір ознаки сукупності, одержаної від додавання значень ознак окремих одиниць сукупності.

2) *залежно від різних причин і завдань аналізу*:

- *натуральні* – відповідають природним або споживчим властивостям предмета і виражаються у фізичних одиницях ваги, довжини тощо;
- *умовно-натуральні* – у разі виробництва однорідної, але неодна-

кової продукції. Суть цього полягає в тому, що один з продуктів беруть за одиницю, решту порівнюють з ним на основі обчислених коефіцієнтів;

– *вартісні* – дають змогу узагальнити та зіставити різноманітні явища; їх застосовують при виробництві різнойменної продукції для визначення загального обсягу продукції. Облік продукції в такий спосіб застосовують для обчислення результатів виробництва та для вимірювання вартості продукції. Абсолютні величини, як узагальнюючі показники, характеризують сукупність за її чисельністю (число працівників, кількість магазинів, музеїв, підприємств) і обсягом (випуск продукції, фонд заробітної плати, обсяг товарообороту і таке інше).

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

*Індивідуальні* – виражають розміри кількісних ознак окремих одиниць досліджуваної сукупності.

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

Абсолютні статистичні величини виражають розміри явищ в таких одиницях виміру як: вага, об'єм, площа, довжина, вартість та ін. Абсолютні статистичні величини завжди іменовані, вони мають певну розмірність, певні одиниці виміру. Одиниці виміру абсолютних величин – важливий елемент статистичного дослідження, вони можуть бути простими, складними та умовними.

У статистиці використовується велика кількість різноманітних одиниць виміру, які можна об'єднати в три групи: *натуральні, вартісні і трудові*.

*Натуральними* називаються одиниці виміру, які виражають розміри конкретних явищ у фізичних вимірниках (тоннах, кілограмах, метрах, гектарах, літрах, кубометрах і ін.). Натуральні одиниці виміру можуть бути *простими, складними і умовно-натуральними*.

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

У ряді випадків статистика використовує *умовно-натуральні оди-*



*ниці виміру*. Такі одиниці використовуються для зведення декількох різновидів однакової споживної вартості. Одну з них беруть за еталон, а всі інші перераховують за допомогою спеціальних перевідних коефіцієнтів в одиниці виміру взятого еталона. Перерахунок в умовно-натуральні одиниці здійснюють за формулою

$$(3.1)$$

де  $Y$  – кількість умовно-натуральних одиниць;  $e$  – кількість еталонних одиниць;  $n$  – кількість одиниць сукупності, які відрізняються від еталонних;  $k$  – коефіцієнт перерахунку нееталонних одиниць сукупності в еталонні.

Наприклад, завод виготовив 200 чотиривісних вагонів та 400 двовісних. Загальну кількість вагонів потрібно перерахувати у двовісні, тобто  $200 \cdot 2 + 400 = 800$ , оскільки один чотиривісний вагон дорівнює за своєю місткістю двом двовісним.

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

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

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

Соціально-економічний аналіз потребує різного роду порівняння. Статистичні показники порівнюються у часі (за єдиним об'єктом), у просторі (між об'єктами) та співвідношенні різних ознак того ж самого об'єкта. Абсолютна величина, взята сама по собі, не завжди дає правильну оцінку явища.

В багатьох випадках тільки в порівнянні з іншою абсолютною величиною дана величина виявляє свою істинну значущість. Результатом порівняння є *відносна статистична величина*, яка характеризує міру кількісного співвідношення різнойменних чи однойменних показників.

*Відносні статистичні величини* – це показники, які виражають кількісні співвідношення між явищами суспільно-економічного життя. Їх визначають як частку від ділення двох абсолютних величин.

При обчисленні відносних величин слід мати на увазі, що чисельник – це показник, який вивчається. Його називають *звітною величиною*. Величину, з якою зіставляють інші величини (знаменник), називають *основою*, або *базою порівняння*, *базисною величиною*.

Відносні величини мають велике значення під час аналізу соціально-економічних явищ, оскільки абсолютні величини не завжди дають змогу правильно оцінити явище з огляду на його динаміку, склад, інтенсивність поширення тощо. Залежно від бази порівняння відносні величини можуть виражатись у формі:

- а) коефіцієнтів – якщо базу порівняння беруть за одиницю;
- б) відсотків (%) – якщо базу порівняння беруть за 100 %;
- в) проміле (‰) – якщо за базу порівняння взято 1000 ‰;
- г) продецеміле (‰<sub>00</sub>) – якщо база порівняння становить 10 000 ‰<sub>00</sub>;
- д) просантиміле (‰<sub>000</sub>) – якщо базу порівняння взято за 100 000 ‰<sub>000</sub>.

Порівняння статистичних величин використовують для вирішення таких *аналітичних задач*: оцінювання інтенсивності явищ; характеристика структури сукупності; оцінювання виконання плану; порівняння характеристик об'єктів; оцінювання дотримання норм і стандартів; оцінювання динаміки тощо. У статистиці введено такі правила позначення показників для розрахунку відносних величин:

- $Y_0$  – фактичний рівень показника в базисному періоді;
- $Y_1$  – фактичний рівень показника у звітному періоді;
- $Y_{пл}$  – плановий рівень показника у звітному періоді;
- $Y_{ф.дог}$  – фактичний рівень договірних зобов'язань;
- $Y_{пл.дог}$  – плановий рівень договірних зобов'язань.

Відповідно до виду аналітичних завдань виділяють наступні *види відносних величин*:

1) *Відносна величина виконання договірних зобов'язань* – показник, який визначають відношенням фактичного обсягу договірних зобов'язань до обсягу зобов'язань, передбачених договором. Даний показник характеризує ступінь виконання підприємством своїх договірних зобов'язань:

$$\text{—} \quad (3.2)$$

2) *Відносна величина динаміки* – показує відношення досягнутого

рівня розвитку явища до рівня, який існував до того, або відносно еталона, норми, стандарту. Характеризує розвиток явища в часі і просторі. Відносні величини динаміки ще мають назву *темпи зростання*. Вони характеризують напрям та швидкість зміни явищ у часі, темпи їх розвитку. Визначаються як відношення показників за звітний період та минулий період:

$$\text{—} \quad (3.3)$$

3) *Відносна величина планового завдання* – характеризує відношення величини показника, встановленого на плановий період, до величини показника, досягнутого у плановому періоді, або до норми, стандарту, еталона:

$$\text{—————} \quad (3.4)$$

4) *Відносна величина виконання плану* – характеризує виконання плану за певний період:

$$\text{—} \quad (3.5)$$

*Приклад:* Обсяг товарообігу в першому півріччі становить 100 тис. грн, а в другому – 80 тис. грн. Визначити темп динаміки товарообігу, величину виконання плану і виконання плану, якщо планований обсяг товарообігу на друге півріччя склав 90 млн грн.

Розрахунок:

$$Y_1 / Y_0 = 80 / 100 = 0,8 \text{ або } 80 \% \text{ (динаміка товарообігу);}$$

$$Y_{\text{пл}} / Y_0 = 90 / 100 = 0,9 \text{ або } 90 \% \text{ (планового завдання);}$$

$$Y_1 / Y_{\text{пл}} = 80 / 90 = 0,89 \text{ або } 89 \% \text{ (виконання плану).}$$

*Висновок:* Обсяг товарообігу в другому півріччі скоротився на 20 % у порівнянні з першим. Незважаючи на те що планований обсяг товарообігу в другому півріччі повинен був скласти 90 % від рівня першого півріччя, план другого півріччя був виконаний на 89 %.

Між відносними величинами *планового завдання, виконання плану і динаміки* існує співзалежність

$$\text{;} \quad (3.6)$$

$$\frac{f_i}{f} \quad (3.7)$$

5) *Відносна величина структури* – показник, що характеризує склад досліджуваної сукупності. Показник є відношенням абсолютної величини кожного складового елемента до абсолютної величини всієї сукупності, тобто відношення частини до цілого. За його допомогою оцінюють структурні зрушення, тобто зміни у складі сукупності за певний період часу:

$$\frac{f_i}{f} \quad (3.8)$$

де  $f_i$  – частота появи ознаки у сукупності.

6) *Відносна величина координації* – це співвідношення окремих частин певної сукупності. Вона вказує на те, в скільки разів порівнювана частина сукупності є більшою або меншою від тієї частини, яку взято за базу порівняння. Відносні величини координації характеризують співвідношення між складовими частинами цілого:

$$\frac{f_i}{f_j} \quad (3.9)$$

*Приклад 1.* Чисельність міського населення дорівнює 34,8 млн осіб, а сільського – 16,5 млн осіб. База порівняння – чисельність сільського населення. Відносна величина координації визначається як  $34,8 / 16,5 = 2,1$ . Це означає, що чисельність міського населення більш ніж у два рази перевищує чисельність сільського населення.

*Приклад 2.* Загальний обсяг товарообігу становить 500 млн грн. Для промислових товарів він становить 300 млн. грн, а по продовольчих – 200 млн грн. Визначити структуру товарообігу і величину координації промислових і продовольчих товарів.

Відносна величина структури визначається як відношення частини до цілого:

$$\frac{f_i}{f} \quad \text{тобто} \quad \frac{300}{500} = 0,6 \quad \text{і} \quad \frac{200}{500} = 0,4$$

Відносна величина координації – визначається як відношення однієї частини сукупності до іншої:

тобто

*Висновок:* Частка товарообігу для промислових товарів становить 60 % від загального обсягу товарообігу, а для продовольчих – 40 %. Співвідношення двох товарів групи виглядає в такий спосіб: реалізація промислових товарів в 1,5 рази перевищує реалізацію продовольчих товарів.

7) *Відносна величина порівняння (наочності)* – показує співвідношення однойменних величин, які стосуються різних об'єктів, різних територій, але за той же самий період. Наприклад, зіставляють чисельність населення, розміри територій, посівних площ, обсяг промислової продукції між окремими країнами, областями, районами.

*Приклад 1.* Населення м. Києва 3,5 млн. осіб, населення м. Москви – 10,5 млн. осіб. Порівняти чисельність населення міст у відносному вираженні.

Розв'язання:

$$10,5 / 3,5 = 3$$

Отже населення м. Москви в 3 рази більше від населення м. Києва.

*Приклад 2.* Існують наступні дані щодо кількості студентів в групах: ЕК51=25 осіб; ЕК52=25 осіб; ЕК53=28 осіб. Проаналізувати існуючі співвідношення наповненості груп.

Розв'язання:

тобто

*Висновок.* Кількість студентів у групі ЕК51 становить 100 % групи ЕК52. Кількість студентів у групах ЕК51 і ЕК52 відповідно становить 89 % студентів групи ЕК53.

8) *Відносна величина порівняння зі стандартом* – співвідношення фактичних значень показника з певним еталоном – стандартом, нормативом, оптимальним рівнем. Для показників, що не мають визначеного еталона, базою порівняння може бути максимальне чи мінімальне значення або розрахована середня за сукупністю в цілому.

9) *Відносна величина інтенсивності розвитку* – показники, які характеризують ступінь поширення, розвиток явища в певному середовищі. За їх допомогою вимірюють інтенсивність його поширення, насиченості певного середовища даним явищем. Відносна величина інтенсивності розвитку характеризує співвідношення двох різнойменних, але пов'язаних між собою величин. Відносні величини інтенсивності розвитку вказують на те, скільки одиниць однієї сукупності припадає на одиницю іншої сукупності. Наприклад, забезпеченість лікарями на 1000 осіб, забезпечення підприємствами громадського харчування на 10000 осіб, захворюваність чи злочинність на 100 осіб та інше.

*Приклад.* У місті А народилось 100 тис. дітей. Чисельність населення 5 млн осіб. Визначити коефіцієнт народжуваності.

Розв'язання:

---

---

*Висновок.* Таким чином, у місті А на кожну тисячу осіб доводиться 20 новонароджених дітей.

Ефективність використання статистичних показників значною мірою залежить від дотримання цілої низки вимог і, насамперед, від урахування специфіки та умов розвитку суспільно-економічних явищ та процесів, а також комплексного застосування абсолютних та відносних величин в економіко-статистичних дослідженнях.

### **3.3. Середні величини**

Вивчаючи суспільні явища з метою виявлення характерних, зако-

номірних рис у конкретних умовах місця та часу, статистика використовує середні величини. Статистика вивчає сукупності за варіаційними ознаками, зміна яких виявляється в зміні кількісних значень окремих одиниць сукупностей. Для цього розраховують середні величини, які потім порівнюють за різними базами порівняння.

*Середня величина* – це узагальнююча кількісна характеристика досліджуваної ознаки в досліджуваній сукупності, у якій знаходить вираження дійсність загальних умов і закономірностей розвитку досліджуваного явища.

Масові явища і процеси формуються під впливом двох груп причин:

1. Загальних для всіх одиниць сукупності, що визначають стан масового процесу. Вони формують типовий рівень для одиниць даної якісно однорідної сукупності і пов'язані з сутністю досліджуваного явища.

2. Індивідуальні, що формують специфічні особливості окремих одиниць масової сукупності, а отже, їх відхилення від типового рівня. Ці причини не пов'язані з природою досліджуваного явища, їх називають випадковими причинами.

При обчисленні середньої величини за масою одиниць вплив випадкових причин взаємопогашається, і середня, абстрагуючись від індивідуальних особливостей окремих одиниць сукупності, виражає загальні властивості, властиві всім одиницям.

*Принципова сутність* статистичного пізнання полягає в погашенні випадкового, викликаного дією індивідуальних причин, і у виявленні закономірностей, обумовлених загальними причинами.

*Важливість методу середніх величин* пояснюється можливістю переходу від одиничного до загального, від випадкового до закономірного і має широке застосування в статистичному дослідженні.

Середня величина завжди *іменована*, має ту ж розмірність (одиницю виміру), що і ознака в окремих одиниць сукупності. *Основною умовою* наукового використання середньої величини є якісна однорідність сукупності, за якою обчислена середня.

*Приклад неоднорідної сукупності.* Чи можна назвати величину 500 \$ середньою зарплатою трьох осіб з індивідуальними заробітками 1200 \$, 200 \$ і 100 \$. Зрозуміло, що за рівнем своєї зарплати ці люди належать до різних категорій працівників, і некоректним буде використання даної величини для характеристики середньої зарплати обстежених осіб. Тобто така сукупність є неоднорідною, і наведене середнє значення не

відображає реальний рівень.

Використання середніх має ґрунтуватися на позиціях діалектичного розуміння категорій загального та індивідуального, масового та одиничного. У кожному випадку слід дотримуватись *вимог*, що передбачають:

1) *Визначення середньої на підставі масових даних*. Індивідуальні значення досліджуваної ознаки окремих одиниць сукупності можуть бути різними. Для того щоб отримати науково обґрунтовану типову величину, обчислювати середню треба за даними, до яких залучається якнайбільше одиниць сукупності. Ця вимога пов'язує середні величини із законом великих чисел;

2) *Якісну однорідність, одноманітність сукупності, для якої визначають середню*. Це означає, що не можна застосовувати середні до таких сукупностей, окремі частини яких підлягають різним законам розвитку відносно осереднюваної ознаки.

При обчисленні середніх у соціально-економічних явищах необхідно визначати *логічну формулу середньої*. Чисельником логічної формули є обсяг значень ознаки, що варіює, а знаменником – обсяг сукупності.

Середня розрахована за сукупністю в цілому, називається загальною середньою, а середня, обчислена для кожної групи, груповою середньою. Загальна середня відображає загальні риси досліджуваного явища, а групова середня дає характеристику розміру явища, що складається в конкретних умовах даної групи. Наприклад, народжуваність у середньому по Євросоюзу і по окремих країнах, що входять до його складу, дуже різняться.

Середні величини поділяються на два великі класи:

- степеневі середні;
- структурні середні.

До степеневих середніх відносяться такі найбільш відомі та часто застосовувані види, як середня геометрична, середня арифметична і середня квадратична. Степеневі середні застосовуються при розрахунку середнього значення статистичної сукупності за певний проміжок часу. Як структурні середні розглядаються мода і медіана.

Степеневі середні залежно від вихідних даних можуть бути *простими і зваженими*. *Прості* застосовуються у випадку, якщо окремі значення досліджуваної ознаки застосовуються лише один раз з одиниць досліджуваної сукупності. При розрахунку *зважених степеневих*



середніх у якості ваги виступає частота прояву тієї або іншої ознаки. Іноді на практиці зручно застосовувати частотність ознаки, тобто відношення частоти появи індивідуальної ознаки до суми частот.

Загальні формули розрахунків степенних середніх мають показник ступеня  $m$ . Залежно від того, якого значення він набуває, розрізняють наступні види степеневих середніх (табл. 3.1).

Таблиця 3.1 – Формули степеневих середніх

Вид середньої	Показник ступеня $m$	Проста	Зважена
Середня гармонійна	$m = -1$	$\frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$	$\frac{1}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}}$
Середня геометрична	$m \rightarrow 0$	$\sqrt[m]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$	$\sqrt[m]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n}}$
Середня арифметична	$m = 1$	$\frac{x_1 + x_2 + \dots + x_n}{n}$	$\frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n}$
Середня квадратична	$m = 2$	$\sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$	$\sqrt{\frac{f_1 x_1^2 + f_2 x_2^2 + \dots + f_n x_n^2}{f_1 + f_2 + \dots + f_n}}$
Середня кубічна	$m = 3$	$\sqrt[3]{\frac{x_1^3 + x_2^3 + \dots + x_n^3}{n}}$	$\sqrt[3]{\frac{f_1 x_1^3 + f_2 x_2^3 + \dots + f_n x_n^3}{f_1 + f_2 + \dots + f_n}}$

Якщо розрахувати всі види середніх для однакових вихідних даних, то значення їх виявляться неоднаковими. Це пояснюється дією *правила маторантності степеневих середніх*: зі збільшенням показника ступеня збільшується і відповідна середня величина:

$$x_1 < x_2 < x_3 < x_4 < x_5 < \dots \quad (3.10)$$

Найчастіше на практиці визначають середню арифметичну і середню гармонійну зважені. Більш докладно з застосуванням прикладів розглянемо найбільш використовувані види степеневих середніх величин.

### 3.3.1. Середня арифметична

Оскільки для більшості соціально-економічних явищ характерна адитивність обсягів (виробництво продукції, витрати на виробництво тощо), то найпоширенішою середньою є середня арифметична.

*Середня арифметична* – це таке значення ознаки, яке б мала кожна одиниця сукупності, якби загальний підсумок усіх значень ознаки був рівномірно розподілений між всіма одиницями сукупності. Для того щоб розрахувати середню арифметичну, потрібно скласти всі окремі варіанти (індивідуальні значення ознаки) і суму поділити на їх кількість. Середня арифметична буває двох видів: проста та зважена.

За первинними незгрупованими даними обчислюється *середня арифметична проста*

$$\bar{X} = \frac{\sum X_i}{n} \quad (3.11)$$

де  $X_i$  – варіанта (значення) осереднюваної ознаки;  $n$  – показник ступеня.

У великих за обсягом сукупностях окремі значення ознаки можуть повторюватись. У такому разі їх можна об'єднати в групи, а обсяг значень ознаки визначити як суму добутків варіант  $x_i$  на відповідні до них частоти  $f_i$ . Такий процес множення у статистиці називається *зважуванням*, а число елементів сукупності з однаковими варіантами – *вагами*. Значення ознаки осереднюється за формулою *середньої арифметичної зваженої*, яка розраховується для згрупованих даних.

*Зважена середня* розраховується за згрупованими даними й має загальний вигляд

$$\bar{X} = \frac{\sum X_i f_i}{\sum f_i} \quad (3.12)$$

де  $X_i$  – варіанта осереднюваної ознаки або серединне значення інтервалу, у якому вимірюється варіанта;  $m$  – показник ступеня середньої;  $f_i$  – частота, яка показує, скільки разів зустрічається  $i$ -те значення осереднюваної ознаки.

Середня арифметична для інтервального ряду носить умовний характер. Як варіанти використовують середні значення інтервалів. Використання інтервальної середньої застосовується при відсутності первинних даних.

Середня арифметична має певні *властивості*, які розкривають її сутність, а саме:

1) алгебраїчна сума відхилень окремих варіант ознаки від середньої дорівнює нулю:

$$\sum_{i=1}^m (x_i - \bar{x}) \cdot f_i = 0, \quad (3.13)$$

тобто в середній взаємно компенсуються додатні та від'ємні відхилення окремих варіант.

2) сума квадратів відхилень окремих варіант ознаки від середньої менша, ніж від будь-якої іншої величини:

$$\sum_{i=1}^m (x_i - \bar{x})^2 \cdot f_i \rightarrow \min. \quad (3.14)$$

3) якщо всі варіанти збільшити або зменшити на ту саму величину  $a$  або в  $a$  разів, то середня зміниться аналогічно.

4) значення середньої залежить не від абсолютних значень ваг, а від пропорцій між ними. При пропорційній зміні всіх ваг, середня не зміниться. Згідно з цією властивістю замість абсолютних ваг – частот  $f_i$  – можна використовувати відносні ваги у вигляді часток  $d_i$  (3.15) або відсотків  $100d_i$ ; таким чином маємо *середню арифметичну зважену спрощену* (3.16).

Розрахунок частки:

$$d_i = \frac{f_i}{\sum f_i}, \quad (3.15)$$

у частках одиниці:

$$\bar{X} = \sum_{i=1}^m x_i d_i,$$

або у відсотках:

$$\bar{X} = \frac{\sum_{i=1}^m x_i d_i}{100}. \quad (3.16)$$

*Математичні властивості середньої арифметичної:*

1) Алгебраїчна сума відхилень всіх варіант від середньої дорівнює 0

(див. формулу 3.13).

2) Якщо одну із варіант збільшити або зменшити на певну величину  $k$ , то і середня зміниться на таку ж величину  $k$

$$\frac{\sum (x \pm k)f}{\sum f} = \bar{x} \pm k. \quad (3.17)$$

3) Якщо кожному варіанту розділити чи помножити на довільне число  $k$ , то і середня збільшиться або зменшиться на те ж саме число  $k$

$$\frac{\sum (x \cdot k)f}{\sum f} = \bar{x} \cdot k. \quad (3.18)$$

4) Якщо частоти всіх варіант помножити чи поділити на довільне число  $k$ , то середня не зміниться

$$\frac{\sum x(f \cdot k)}{\sum (f \cdot k)} = \bar{x}. \quad (3.19)$$

5) Сума квадратів відхилень варіант від середньої менша за будь-яку іншу величину (див. формулу 3.14).

### 3.3.2. Середня хронологічна

Середня хронологічна розраховується як середня у хронологічному ряді, тобто моментному ряді. Моментні показники замінюються середніми як півсума значень на початок та кінець періоду. Якщо моментів більше ніж два, а інтервали між ними рівні, то в чисельнику до півсуми крайніх значень додають усі проміжні, а знаменником є число інтервалів, яке на одиницю менше від числа значень ознаки. Таку формулу називають середньою хронологічною:

$$\bar{X} = \frac{\frac{x_1 + x_n}{2} + x_2 + x_3 + \dots + x_{n-1}}{n-1}, \quad (3.20)$$

або

$$\bar{X} = \frac{\frac{1}{2}x_1 + x_2 + x_3 + \dots + \frac{1}{2}x_n}{n-1}. \quad (3.21)$$

### 3.3.3. Середня гармонійна

Статистичні середні завжди виражають якісні властивості суспільних явищ та процесів. Під час досліджень важливо правильно вибрати тип середньої відповідно до природи взаємозв'язків явищ та їх ознак. У статистичних дослідженнях використовують *середню гармонійну*, яка за своїми властивостями застосовується, коли загальний обсяг ознаки формується як сума обернених значень варіантів, наприклад, коли треба вагу (тобто добуток варіантів на частоту) поділити на варіанти, або помножити на обернене до них число. Середня гармонійна обернена до середньої арифметичної:

– середня гармонійна *проста*:

$$\bar{X} = \frac{n}{\sum_1^n \frac{1}{x}}; \quad (3.22)$$

– середня гармонійна *зважена*:

$$\bar{X} = \frac{\sum f_i}{\sum \frac{f_i}{x_i}}. \quad (3.23)$$

Таким чином, середня гармонійна у статистиці – це перетворена середня арифметична, яку застосовують у разі, коли чисельність сукупності невідома, а варіанти зважуються обсягом ознаки.

### 3.3.4. Середня геометрична

*Середня геометрична* застосовується у тих випадках, коли загальне значення визначається не як сума індивідуальних значень ознак, а як їх добуток, тобто, якщо визначальна властивість сукупності формується як добуток індивідуальних значень ознаки, то використовується *середня геометрична*

$$\bar{x} = \sqrt[n]{x_1 x_2 x_3 \dots x_n} = \sqrt[n]{\prod_1^n x_i}. \quad (3.24)$$

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

$$\bar{x} = \sqrt[n]{\prod_{i=1}^m x_i^{n_i}}, \quad (3.25)$$

яка використовується для розрахунків середніх коефіцієнтів зростання.

### 3.3.5. Середня квадратична

Використовується для визначення характеристики варіації ознак (наприклад, дисперсії) та для узагальнення ознак, виражених лінійними мірами площ (наприклад, при обчисленні середнього діаметра).

Середня квадратична буває двох видів:

- проста

$$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}}; \quad (3.26)$$

- зважена

$$\bar{x} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}}. \quad (3.27)$$

### 3.3.6. Структурні середні

Вказані вище середні є узагальнюючими характеристиками сукупностей за тією чи іншою варіаційною ознакою. Водночас структуру цих сукупностей характеризують особливими показниками, які називаються у статистиці *структурними середніми величинами*. Зокрема, це мода та медіана.

*Мода* – це величина ознаки, яка найчастіше трапляється в даній сукупності. У варіаційному ряді це варіанта, яка має найбільшу частоту.

Моду широко використовують у комерційній діяльності, в соціологічних дослідженнях, коли вивчають ринковий попит, реєструють ціну, встановлюють рейтинг популярності осіб чи товарів тощо.

В інтервальному варіаційному ряді для визначення *модальної величини* ознаки, що знаходиться в певному інтервалі, формула має такий вигляд

$$M_o = x_0 + h \frac{f_m - f_{m-1}}{\underbrace{f_m - f_{m-1}} + \underbrace{f_m - f_{m+1}}}, \quad (3.28)$$

де  $x_0$  – нижня межа модального інтервалу;  $h$  – ширина або шаг інтервалу;  $f_m$  – частота модального інтервалу;  $f_{m-1}$  – частота інтервалу, що передує модальному;  $f_{m+1}$  – частота наступного, що йде за модальним, інтервалу.

*Медіана* – це варіанта, що є серединою впорядкованого варіаційного ряду, тобто ділить його на дві рівні частини: одна частина має значення варіаційної ознаки менше, ніж середня, а друга – більше. Медіана вказує на значення варіаційної ознаки, якого досягла половина одиниць сукупності.

Для визначення медіанного значення ознаки спочатку визначають номер медіанної одиниці ряду за наступною формулою

$$N_{Me} = \frac{n+1}{2}, \quad (3.29)$$

де  $n$  – обсяг сукупності (кількість варіантів сукупності).

Після розрахунку номера медіани по графі накопичених (кумулятивних) частот методом порівняння знаходять значення частоти *ближче більше від розрахованого*. Варіанта, яка містить таку частоту, і буде медіаною даної сукупності. Можна також місце медіани у дискретному ряді визначати за сумою всіх частот, яку необхідно поділити на два і до отриманого результату додати 0,5. У тому разі, коли сума частот парна, місце медіани є дробовим числом, але оскільки дробових варіантів не буває, то медіана лежить у середині сусідніх варіантів.

Для обчислення медіани в інтервальному ряді визначають медіанний інтервал. Відповідає такому, кумулятивна частота якого дорівнює або перевищує половину суми частот. Кумулятивні частоти визначають за допомогою поступового підсумовування частот, розпочинаючи з інтервалу з найменшим значенням ознаки. *Медіану* розраховують за наступною формулою

$$M_e = x_0 + h \frac{\sum \frac{1}{2} f - \sum f_{m_e-1}}{f_{m_e}}, \quad (3.30)$$

де  $x_0$  – нижня межа медіанного інтервалу;  $h$  – ширина або шаг інтервалу;  $f_{me}$  – частота медіанного інтервалу;  $\sum f_{m_e-1}$  – сума кумулятивних частот інтервалу, який передує медіанному інтервалу;  $f$  – частоти ряду.

В аналізі закономірностей розподілу використовуються також інші порядкові структурні характеристики: квартилі, децилі, перцентилі.

Аналогічно зі знаходженням медіани у варіаційних рядах розподілу можна розрахувати значення ознаки в одиниць, які ділять ряд на чотири рівні частини, на 10 або на 100 частин. Ці величини називаються відповідно «квартилі», «децилі», «перцентилі».

Квартилі  $Q$  – це значення варіант, які ділять упорядкований ряд за обсягом на чотири рівних частини, децилі  $D$  – на десять рівних частин, перцентилі  $P$  – на сто рівних частин. Отже, в ряду розподілу визначаються 3 квартилі, 9 децилів, 99 перцентилів. Медіана є водночас другим квартилем, п'ятим децилем та п'ятидесятим перцентилем. Розрахунок квартилів, децилів та перцентилів ґрунтується на кумулятивних частотах (частках).

Квартилі являють собою значення ознаки, яка ділить ранжовану сукупність на 4 рівновеликі частини. Розрізняють квартиль нижній  $Q_1$ , що відокремлює  $j$  частину сукупності з найменшими значеннями ознаки, і квартиль верхній  $Q_3$ , що відсікає  $j$  частину з найбільшими значеннями ознаки. Це означає, що 25 % одиниць сукупності будуть менші за величиною  $Q_1$ ; 25 % одиниць будуть укладені між  $Q_1$  і  $Q_2$ ; 25 % – між  $Q_2$  і  $Q_3$ , а інші 25 % перевищують  $Q_3$ . Середнім квартилем  $Q_2$  є медіана.

Для розрахунків квартилей по інтервальному варіаційному ряду використовуються наступні формули

$$Q_1 = x_{Q_1} + h \frac{1/4 \sum f - \sum f_{Q_1-1}}{f_{Q_1}}; \quad (3.31)$$

$$Q_3 = x_{Q_3} + h \frac{3/4 \sum f - \sum f_{Q_3-1}}{f_{Q_3}}, \quad (3.32)$$

де  $x_{Q_1}$  – нижня границя інтервалу, який містить нижній квартиль (інтервал визначається за накопиченою частотою, перший перевищує 25 %);  $x_{Q_3}$  – нижня границя інтервалу, що містить верхній квартиль (інтервал визначається за накопиченою частотою, перший перевищує 75 %);  $h$  – величина інтервалу;  $\sum f_{Q_1-1}$  – накопичена частота інтервалу, що передує інтервалу, який містить нижній квартиль;  $\sum f_{Q_3-1}$  – накопичена частота інтервалу, що передує інтервалу, який містить верхній квартиль;  $f_{Q_1}$  – частота інтервалу, який містить нижній квартиль;



$f_{Q_3}$  – частота інтервалу, який містить верхній кuartиль.

Крім кuartилей, у варіаційних рядах розподілу можуть визначатися децилі – варіанти, що ділять ранжований варіаційний ряд на десять рівних частин. Перший дециль  $d_1$  ділить сукупність у співвідношенні від 1/10 до 9/10, другий дециль  $d_2$  – у співвідношенні від 2/10 до 8/10 і т. д.

Обчислюються вони за формулами:

$$d_1 = x_{d_1} + h \frac{1/10 \sum f - \sum f_{d_1-1}}{f_{d_1}}; \quad (3.33)$$

$$d_2 = x_{d_2} + h \frac{2/10 \sum f - \sum f_{d_2-1}}{f_{d_2}}. \quad (3.34)$$

Значення ознаки, що ділять ряд на сто частин, називаються перцентилями. Розрахунок перцентилей здійснюється аналогічно попереднім. Співвідношення медіани, кuartилей, децилей і перцентилей наведено на рис. 3.1.

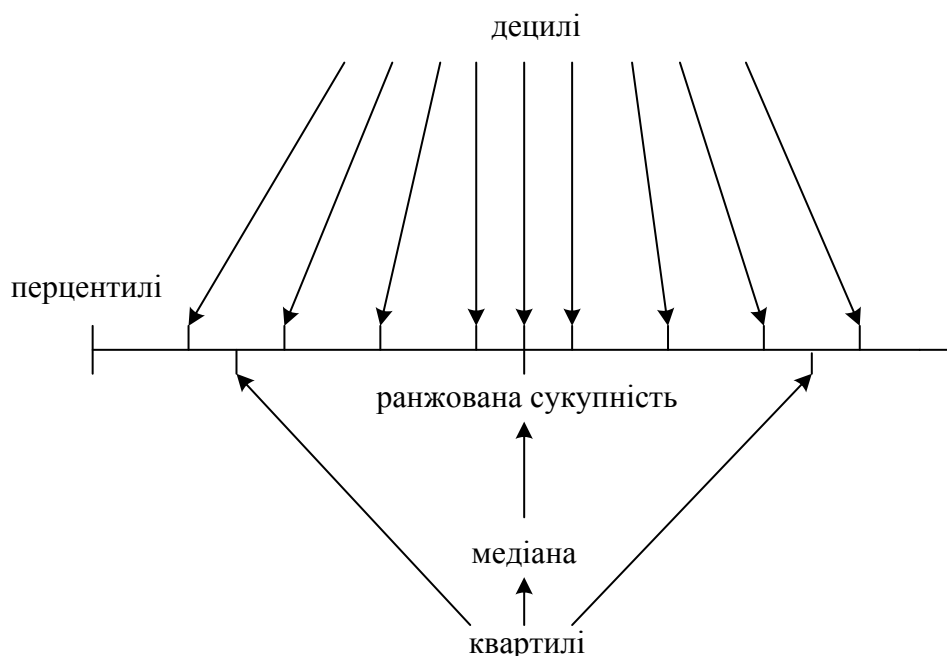


Рисунок 3.1 – Співвідношення медіани, кuartилей, децилей і перцентилей

### ***3.4. Основні правила і особливості застосування середніх для розрахунків економічних показників***

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

Головна умова наукового використання середньої полягає в тому, що середні характеристики повинні розраховуватись на основі масового узагальнення фактів, оскільки тільки тоді вони відображають сутність явища, на значення якого не впливають поодинокі фактори. Ця умова пов'язує статистичні середні із законом великих чисел.

Іншою важливою умовою застосування середніх у статистиці є якісна однорідність всіх одиниць сукупності. Вона полягає в тому, що не можна обчислювати середню з неоднорідної сукупності, окремі елементи якої підпорядковані різним законам розвитку стосовно осереднюваної ознаки.

Середня величина тільки тоді відобразить типовий розмір ознаки та її загальні риси, коли це загальне реально існуватиме, і всі його елементи будуть якісно однорідні і типові. Застосування методу середніх в статистиці тісно і нерозривно пов'язане з методом групувань. Загальні середні потрібно доповнювати груповими середніми в тих випадках, коли варіаційна ознака суттєво відрізняється в окремих групах, і в порівнюваних групах існує різне співвідношення груп. Особливого значення набуває доповнення загальної середньої груповими середніми при вивченні взаємозв'язку і взаємозалежності одних показників ознак від інших. При використанні середніх потрібно пам'ятати, що середні величини не можуть і не повинні підміняти індивідуальні показники, а повинні доповнюватись вивченням кращих і гірших одиниць сукупності.

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

Розглянемо на окремих прикладах застосування середніх залежно від існуючих вихідних даних.

Середня арифметична проста обчислюється, коли обсяг

осереднюваної ознаки, утворюється як сума значень з окремих одиниць.

Приклад логічної формули:

$$\text{Ср.}_{\%} = \frac{\text{затрати на випуск}}{\text{кількість виробленої продукції}}$$

$$\text{Ср.}_{\frac{3}{n}} = \frac{\text{фонд оплати праці}}{\text{чисельність робітників}}$$

Приклад. Є такі дані про виробіток робочими продукції за зміну:

№ робочого	1	2	3	4	5	6	7	8	9	10
Випущено виробів за зміну	16	17	18	17	16	17	18	20	21	18

У даному прикладі ознака, що варіює (змінюється), – випуск продукції за зміну. Чисельні значення ознаки (16, 17 і т. д.) називають варіантами. Визначимо середній виробіток продукції робітниками даної групи:

— \_\_\_\_\_ —

Проста середня арифметична застосовується у випадках, коли є окремі значення ознаки, тобто дані не згруповані. Якщо дані наведено у вигляді рядів розподілу або групувань, то середня обчислюється інакше.

Приклад. Є такі дані про заробітну плату робітників–відрядників:

Місячна заробітна плата (варіанта – $x$ ), грн	Число робітників $n$ (або $f$ )	$x \cdot n$
$x = 1100$	$n = 2$	2200
$x = 1300$	$n = 6$	7800
$x = 1600$	$n = 16$	25600
$x = 1900$	$n = 12$	22800
$x = 2200$	$n = 14$	30800
РАЗОМ	50	89200

Як видно з прикладу, середня залежить не тільки від значень ознаки, але і від їх частот, тобто від складу сукупності, від її структури.

За даними дискретного ряду розподілу видно, що ті самі значення ознаки (варіанти) повторюються декілька разів. Так, варіанта  $x=1100$

зустрічається в сукупності 2 рази, а варіанта  $x=1600$  – 16 разів і т. д. Число однакових значень ознаки в рядах розподілу називається частотою або вагою і позначається символом  $n$  (або  $f$ ).

Фонд заробітної плати для кожної групи робітників дорівнює добутку варіанти на частоту, а сума цих добутків дає загальний фонд заробітної плати всіх робітників. Обчислимо середню заробітну плату одного робітника  $\bar{x}$ :

$$\bar{x} = \frac{\sum x_n}{n}$$

Статистичний матеріал у результаті обробки може бути наведений не тільки у вигляді дискретних рядів розподілу, але й у вигляді інтервальних варіаційних рядів з закритими або відкритими інтервалами. Розглянемо розрахунки середньої арифметичної для таких рядів.

Змінимо в умові завдання склад робітників і обчислимо середню в змінній структурі.

*Приклад.* Обчислити середній виробіток продукції одним робітником за зміну. Є такі дані:

Групи робітників за кількістю виробленої продукції за зміну, шт.	Кількість робітників $n$	Середина інтервалу $x$	$x \cdot n$
3–5	10	4	40
5–7	30	6	180
7–9	40	8	320
9–11	15	10	150
11–13	5	12	60
РАЗОМ	100		750

У даному ряді варіанти осереднюваної ознаки (продукція за зміну) відображені не одним числом, а у вигляді інтервалу «від – до». Робітники першої групи виробляють продукцію від 3 до 5 шт. за зміну, робітники другої групи – від 5 до 7 шт. і т. д. Таким чином, кожна група ряду розподілу має нижнє і верхнє значення варіант, або закриті інтервали. Обчислення середньої за згрупованими даними здійснюється за формулою середньої арифметичної зваженої.

Щоб застосувати цю формулу, необхідно варіанти ознаки виразити одним числом (дискретним). За таке дискретне число береться середня арифметична проста, розрахована з верхнього і нижнього значення

інтервалу. Так, для першої групи дискретна величина становитиме:

$$X=(3 + 5) / 2 = 4.$$

Подальші розрахунки проводяться звичайним методом визначення середньої арифметичної зваженої

Отже, усі робітники виготовили 750 шт. виробів за зміну, а кожний у середньому виготовив 8 шт.

Перетворимо розглянутий ряд розподілу в ряд з відкритими інтервалами.

*Приклад.* Є наступні дані про виробництво продукції за зміну:

Групи робітників за кількістю виробленої продукції за зміну, шт.	Число робітників $n$	Середина інтервалу $x$	$x \cdot n$
до 5	10	4	40
5–7	30	6	180
7–9	40	8	320
9–11	15	10	150
понад 11	5	12	60
РАЗОМ	100		750

У таких рядах умовно величина інтервалу першої групи береться на рівні величині інтервалу наступної групи, а величина інтервалу останньої групи – на рівні величині інтервалу попередньої. Подальший розрахунок аналогічний викладеному вище.

У практиці економічної статистики іноді доводиться обчислювати середню за груповими середніми або за середніми окремими частинами сукупності (частками середніх). У таких випадках за варіанти  $x$  беруться групові або часткові середні, на основі яких обчислюється загальна середня, як середня арифметична зважена.

*Приклад.* Визначити середній відсоток виконання плану з випуску продукції для групи заводів на основі таких даних:

Номер заводу	Випуск продукції за планом, млн грн	Виконання плану, %
1	18	100
2	22	105
3	25	90
4	20	106
5	40	108
РАЗОМ	125	–

У цьому завданні варіанти (відсоток виконання плану) є не індивідуальними, а середніми по заводу. Вагами є випуск продукції за планом. При обчисленні середнього відсотка виконання плану слід використовувати формулу середньої арифметичної зваженої.

Проводячи розрахунок, варіанти  $x$  краще відобразити у вигляді коефіцієнта:

— \_\_\_\_\_

Таким чином, середній відсоток виконання плану по випуску продукції по групі заводів складає 102,4 %.

*Приклад 1.* Фонд заробітної плати на підприємстві склав 85 500 грн. Працівників 10 осіб. Визначити середню заробітну плату.

Для розрахунку використовується середня арифметична:

\_\_\_\_\_

*Приклад 2.* На підприємстві 12 працівників. У трьох стаж роботи становить 3 роки, у двох – 4 роки, у чотирьох – 5 років, у двох – 6 років, в одного – 7 років. Визначити середній стаж працівників.

У даному випадку для розрахунку використовується арифметична зважена середня:

\_\_\_\_\_

Поряд із середньою арифметичною, у статистиці застосовується середня гармонійна величина, *обернена* до середньої арифметичної. Як і середня арифметична, вона може бути простою і зваженою. Якщо в логічній формулі невідомий підсумковий знаменник, то розрахунки проводяться за формулою середньої гармонійної. Формула середньої гармонійної також використовується при розрахунках середньогрупових значень ознак.

*Приклад 3.* Бригада токарів була зайнята обточуванням однакових деталей протягом 8-годинного робочого дня. Перший токар витратив на виготовлення однієї деталі 12 хв., другий – 15 хв., третій – 11 хв., четвертий – 16 хв. і п'ятий – 14 хв. Визначити середній час, витрачений на виготовлення однієї деталі.

На перший погляд здається, що завдання легко вирішується за допомогою формули середньої арифметичної простої:

Отримана середня була б вірною, якби кожний робітник зробив тільки по одній деталі. Але протягом дня окремими робітниками була виготовлена різна кількість деталей. Для визначення кількості деталей, виготовлених кожним робітником, скористаємось наступним співвідношенням (логічною формулою):

Кількість деталей, виготовлених кожним робітником, визначається відношенням усього часу роботи до середнього часу, витраченого на одну деталь. Тоді середній час, необхідний для виготовлення однієї деталі, становитиме:

Це розв'язання можна подати інакше, використовуючи формулу середньої гармонійної:

Таким чином, для таких випадків використовується формула для розрахунку середньої гармонійної простої (формула 3.22).

*Приклад 4.* Витрати виробництва і собівартість одиниці продукції на трьох заводах характеризуються наступними даними:

Номер заводу	Витрати виробництва, тис. грн	Собівартість одиниці продукції, грн
1	200	20
2	460	23
3	110	22

Обчислимо середню собівартість виробу на трьох заводах. Головною умовою вибору форми середньої є вихідні дані. Логічна формула має такий вигляд:

тобто

Таким чином, використовується формула (3.23) для розрахунків середньої гармонійної зваженої.

*Приклад 5.* Є інформація про три банки. Визначити середній залишок внесків щодо трьох банків.

№ банку	Порівняльний залишок за одним внеском, тис. грн	Загальна сума залишків за внесками, тис. грн
I	1,67	1897,8
II	2,80	5040,0
III	3,25	6987,5

— це логічна формула (в даних в прямому вигляді відсутній знаменник), тобто використовується середня гармонійна:

$$\bar{X} = \frac{\sum f_i}{\sum \frac{f_i}{X_i}} = \frac{1897,8 + 5040,0 + 6987,5}{\frac{1897,8}{1,67} + \frac{5040,0}{2,8} + \frac{6987,5}{3,25}} = 2,74 \text{ тис.грн.}$$

Для моментних рядів використовується середня хронологічна.

*Приклад 6.* Визначити середньооблікову чисельність працівників лікарні, якщо відомі такі дані:

Дата	Чисельність, осіб
На 01.01.13	75,0
На 01.04.13	75,1
На 01.07.13	75,2
На 01.10.13	75,5
На 01.01.14	75,8

Використовується середня хронологічна (3.22)

$$\bar{X} = \frac{\frac{1}{2}x_1 + x_2 + x_3 + \dots + \frac{1}{2}x_n}{n-1} = \frac{\frac{75}{2} + 75,1 + 75,2 + 75,5 + \frac{75,8}{2}}{4} = 75,3 \text{ осіб.}$$



Середня геометрична застосовується у випадку, якщо необхідно визначити середній темп динаміки. Середня геометрична використовується також для визначення рівновіддаленої величини від максимального і мінімального значення ознаки.

Середня квадратична – застосовується при визначенні середньо-квадратичного відхилення в показниках варіації.

📖 Література до теми 3:[4; 5; 7; 8; 15–21; 23; 25–27].

### ***Контрольні запитання до теми 3***

1. Що називається статистичним показником?
2. Що таке абсолютні величини, які одиниці їх вимірювання?
3. Що називається відносним показником?
4. Форми подання відносних величин.
5. Види відносних величин.
6. Дайте визначення поняття середньої величини.
7. Форми обчислення середніх величин.
8. Види середніх величин.
9. Поняття моди, медіани.
10. У яких випадках зважені й не зважені середні рівні між собою?
11. У чому полягає правило маторантності середніх?
12. У яких випадках використовується середня гармонійна?
13. У яких випадках використовується середня арифметична зважена?
14. У яких випадках використовується середня хронологічна?
15. У яких випадках використовується середня квадратична?
16. У яких випадках використовується середня арифметична проста?
17. Чи зміниться середня величина, якщо усі ваги зменшити на деяку постійну величину?
18. Коли в розрахунках середньої величини використовуються формули середньої арифметичної, а коли – середньої гармонійної?
19. Наведіть формули та пояснить, коли використовується в розрахунках середня арифметична проста, а коли – середня зважена.
20. Чи зміниться середня арифметична величина, якщо:
  - а) всі варіанти зменшити на 4;
  - б) всі варіанти збільшити в 10 разів?

## Тема 4. РЯДИ РОЗПОДІЛУ. АНАЛІЗ ВАРІАЦІЙ ТА ФОРМИ РОЗПОДІЛУ. АНАЛІЗ КОНЦЕНТРАЦІЇ, ДИФЕРЕНЦІАЦІЇ ТА ПОДІБНОСТІ РОЗПОДІЛІВ

- 4.1. Поняття варіації та її основні показники.
- 4.2. Поняття та види рядів розподілу.
- 4.3. Характеристики форми розподілу.
- 4.4. Види та взаємозв'язок дисперсій.
- 4.5. Різновиди та характеристики форм розподілів.
- 4.6. Оцінка і аналіз нерівномірності та подібності розподілів.

### 4.1. Поняття варіації та її основні показники

Середні величини, як узагальнюючі показники, характеризують сукупності за варіаційною ознакою, вказують на їх типовий рівень у розрахунку на одиницю однорідної сукупності. Проте середня величина не пояснює, як групуються навколо неї окремі значення; чи лежать вони поблизу, чи істотно відхиляються від середньої. Чим менше відхилення, тим однорідніша сукупність, а отже, тим більш надійні та типові характеристики центру розподілу. Вимірювання ступеня коливання ознаки, її варіації є невід'ємною складовою аналізу закономірностей розподілу.

Термін «варіація» в перекладі з латинської – зміна, коливність, різниця. У статистиці *варіація* – це кількісні зміни ознаки в межах однорідної сукупності, які зумовлені впливом різних факторів. Розрізняють варіації випадкові і систематичні. Аналіз систематичної варіації дає змогу оцінити залежність зміни ознаки від суттєвих для неї чинників.

У системі показників варіації найпростішим є *показник розмаху варіації*, який визначають як різницю між найбільшим та найменшим значеннями варіант

$$R = x_{\max} - x_{\min}. \quad (4.1)$$

Недоліком цього показника є те, що він фіксує лише крайні відхилення і зовсім не враховує відхилень решти варіант від їх середнього значення. Квартильний розмах варіації  $R_Q = Q_3 - Q_1$  тотожний 50 % сукупності.

Узагальнюючу характеристику може дати лише середня величина, зокрема середня відхилень варіант від їх середньої, яка називається

середнє лінійне відхилення. Формула середнього лінійного відхилення:

1) для незгрупованих даних

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n}; \quad (4.2)$$

2) для згрупованих даних

$$\bar{d} = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i}. \quad (4.3)$$

Середнє лінійне відхилення не завжди характеризує розсів варіант.

Ступінь варіації об'єктивніше характеризує показник середнього квадрата відхилень (дисперсія). Його розраховують як середню арифметичну з суми квадратів відхилень окремих варіант від їх середньої:

1) для незгрупованих даних

$$D_x = \sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}; \quad (4.4)$$

2) для згрупованих даних

$$D_x = \sigma_x^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}. \quad (4.5)$$

Корінь квадратний із середнього квадрата відхилень варіант від їх середньої (дисперсії) називають *середнім квадратичним відхиленням*:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}. \quad (4.6)$$

Середнє квадратичне відхилення є мірою надійності середньої. Чим менше дане відхилення, тим повніше середня арифметична відображує всю сукупність.

Усі розглянуті показники варіації – розмах варіації, середнє лінійне та квадратичне відхилення, дисперсія – є абсолютними показниками варіації. Для того щоб забезпечити порівняння варіаційних рядів, потрібно обчислити показники, які характеризують варіацію, виражену в стандартних величинах, наприклад у процентах.

При порівнянні варіації різних ознак або варіації однієї ознаки в різних сукупностях використовують відносні характеристики – *коефіцієнти варіації*, які визначають як відношення середнього відхилення до середнього значення:

1) лінійний коефіцієнт варіації

$$V_{\bar{d}} = \frac{\bar{d}}{x} \cdot 100 \% ; \quad (4.7)$$

2) квадратичний коефіцієнт варіації

$$V_{\sigma_x} = \frac{\sigma_x}{x} \cdot 100 \% ; \quad (4.8)$$

3) кватильний коефіцієнт варіації

$$V_{\rho} = \frac{R_{\rho}}{2Me} \cdot 100 \% . \quad (4.9)$$

Коефіцієнт варіації є певною мірою критерієм типовості середньої. Якщо коефіцієнт дуже великий, то це означає, що середня характеризує сукупність за ознакою, яка суттєво змінюється в окремих одиницях. Типовість такої середньої сумнівна, тобто невелика. Існує певне визначення типовості ряду розподілу за допомогою визначення рівня квадратичного коефіцієнта варіації – чим більший коефіцієнт варіації, тим відносно більший розкид і менша однорідність досліджуваних об'єктів.

Шкала однорідності ряду розподілу:

$V_{\rho} < 10 \%$  – значна однорідність;

$10 \% < V_{\rho} < 20 \%$  – середня однорідність;

$20 \% < V_{\rho} < 33 \%$  – незначна однорідність;

$V_{\rho} > 33 \%$  – неоднорідна інформація.

При цьому інформацію необхідно виключити або відкинути нетипові спостереження, які звичайно бувають у перших і останніх ранжованих рядах розподілу.

Інколи водночас із коефіцієнтом варіації як відносний показник обчислюють *коефіцієнт осциляції*

$$K_o = \frac{R}{x} \cdot 100 \% . \quad (4.10)$$

Цей коефіцієнт характеризує відносну коливність крайніх значень ознаки навколо середньої.

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

#### **4.2. *Поняття та види рядів розподілу***

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

*Ряд розподілу* складається з двох елементів:

- 1) *Варіанти* – значення груповальної ознаки  $x_i$ .
- 2) *Частоти* –  $f_i$ .

Побудова рядів розподілу базується на принципах статистичного групування. Ряди розподілу можна утворити або за атрибутивними, або за кількісними (варіаційними) ознаками. Варіаційні ряди, варіанти яких мають чисельний вираз, поділяються на дискретні та інтервальні. У першому випадку варіанти являють собою дискретні числа, а у другому – інтервали групування, які, у свою чергу, можуть бути закритими або відкритими, рівними і нерівними, а останні – зростаючими або спадаючими. При побудові атрибутивних рядів розподілу варіанти потрібно розташувати за логічною послідовністю.

Розрізняють ряди розподілу з абсолютними, відносними та накопиченими частотами. Накопичені частоти називають *кумулятивними*. Абсолютні накопичені частоти є абсолютними числами, а відносні – питомою вагою або часткою кожної групи. Частки прийнято позначати через  $W_i$  і розраховувати за формулою

$$\text{---} \quad (4.11)$$

Ряди розподілу з абсолютними частотами характеризують зміст сукупності, а з відносними – їх структуру. Ряди розподілу з кумулятивними частотами вказують на кількість або питому вагу одиниць із значенням ознаки, меншим від заданої. Кумулятивні частоти знаходять підсумуванням їх по групах.

*Щільність розподілу* – це кількість одиниць сукупності, що припадає на одиницю величини інтервалу групувальної ознаки. Розрізняють абсолютну  $f_d^a$  та відносну  $f_d^b$  щільності, які відповідають наступним формулам:

$$f_d^a = \frac{f}{h} \quad (4.12)$$

або

$$f_d^a = \frac{W_i}{h}, \quad (4.13)$$

де  $f$  – частота;  $W_i$  – частка (питома вага);  $h$  – розмір інтервалу з формули (2.2).

*Інтерполяція* в рядах розподілу визначає, скільки одиниць сукупності (або її частки) мають значення ознаки, менші від заданих. Для інтерполяції використовують як абсолютні, так і відносні накопичені (кумулятивні) частоти. Поглиблений аналіз закономірностей розподілу передбачає характеристику зазначених особливостей сукупності:

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

### **4.3. Характеристики форми розподілу**

Однорідність сукупності є передумовою використання статистичних методів (середніх величин, регресійного аналізу тощо). *Однорідними* є такі сукупності, елементи яких мають спільні властивості і належать до одного типу, класу.

За формою ряди розподілу поділяються на:

- одновершинні (одномодальні);
- двовершинні;
- багатoverшинні.

*Однорідність сукупності* – це характеристика, яка означає, що всі елементи сукупності мають спільні властивості (риси) і належать до одного типу або класу; є певна наявність у них загального в істотному,

головному. Показником однорідності сукупності є одновершинність розподілу (одномодальність). Багатовершинність свідчить про неоднорідний склад сукупності, про різнотипність окремих складових. У такому разі необхідно перегрупувати дані, виділити однорідні групи. Наявність двох і більше вершин свідчить про неоднорідність сукупності, про поєднання в ній груп з різними рівнями ознаки. *Критерієм однорідності* сукупності вважається квадратичний коефіцієнт варіації, який завдяки властивостям середнього квадратичного відхилення в симетричному розподілі повинен становити не більше ніж 33 %.

Розподіл якісно однорідних сукупностей переважно одновершинний. Серед одновершинних розподілів є симетричні та асиметричні, гостро- та плосковершинні.

Якщо варіанти рівновіддалені від центру значень ознаки, то такий варіаційний ряд є *симетричним*. Якщо вершина розподілу зміщена, тобто частоти по обидва боки від центра змінюються неоднаково, тоді варіаційний ряд є *асиметричним* або скошеним. Якщо вершина зміщена ліворуч, маємо *правосторонню* асиметрію і навпаки.

Центром тяжіння статистичної сукупності є типовий рівень ознаки, узагальнююча характеристика всього розмаїття її індивідуальних значень. Такою характеристикою є *середня величина*, яка може бути середньою арифметичною, модою, медіаною. Крім того, в аналізі розподілу використовують кватилі і децилі. У разі чіткої асиметрії ряду для вивчення економічних явищ середнє значення ознаки доповнюється модою чи медіаною.

В асиметричному розподілі між середньою арифметичною, медіаною та модою є певні розбіжності:

- при правосторонній асиметрії  $\bar{x} > Me > Mo$  ;
- при лівосторонній  $\bar{x} < Me < Mo$  .

Стандартизовані відхилення характеризують напрям і міру якісності розподілу. *Коефіцієнт асиметрії*  $A$  є відношенням різниці між середньою арифметичною і модою чи медіаною до середнього квадратичного відхилення

$$A = \frac{\bar{x} - Mo}{\sigma_x} \quad \text{або} \quad A = \frac{\bar{x} - Me}{\sigma_x}. \quad (4.14)$$

Коефіцієнт асиметрії коливається в межах від  $-3$  до  $+3$ . У симетричному розподілі  $A = 0$ , при правосторонній асиметрії  $A > 0$ , при

лівосторонній –  $A < 0$ .

Криві, зображені на рис. 4.1 (а, б, в), дозволяють ілюструвати симетрію і два найбільш поширених види асиметрії розподілу. При симетричному розподілі (а) середня арифметична, мода і медіана рівні між собою. Для асиметричних кривих (б, в) ці статистичні величини неоднакові. Причому середня арифметична і медіана зміщені від центра в бік довшої вітки кривої. Оскільки середня арифметична «чутлива» до «точного» положення більш віддалених від моди точок кривої, а медіана «нечутлива», то середня зрушена більше, ніж медіана. У цьому випадку медіана знаходиться між модою і середньою арифметичною. Як видно з рисунку, напрямок асиметрії геометрично встановлюється дуже просто. Кількісна форма ступеня асиметрії вимагає знаходження її алгебраїчної міри.

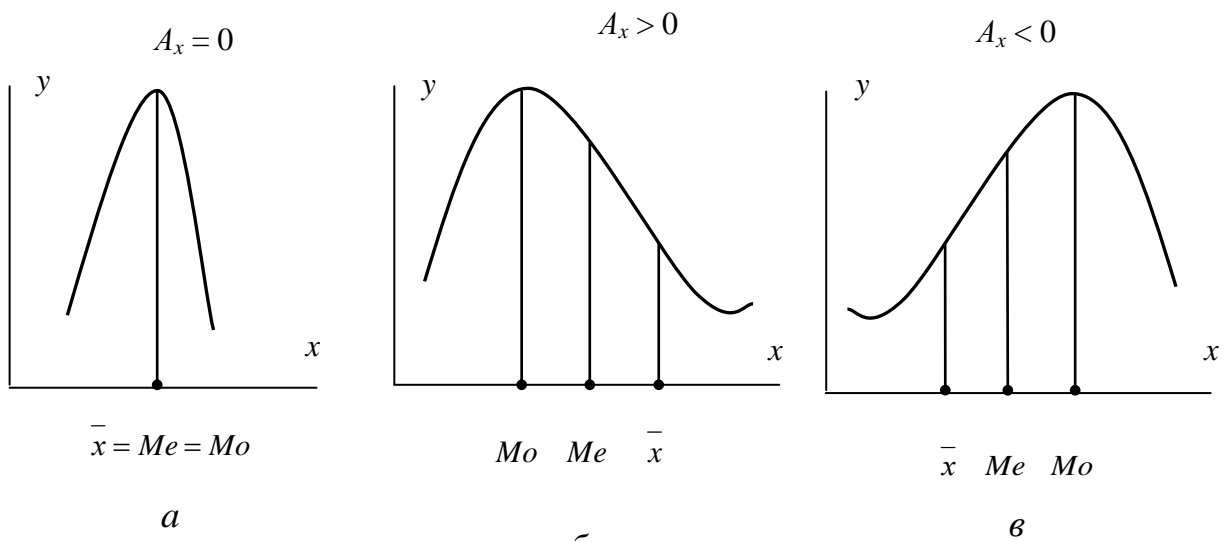


Рисунок 4.1 – Форми розподілу при різних значеннях коефіцієнта асиметрії:

а – симетричний розподіл; б та в – асиметричні криві

Крутість варіаційного ряду, тобто його високовершинність (гостровершинність) або низьковершинність (плосковершинність) називають *ексцесом*, тобто це ступінь розсередження елементів сукупності навколо центра розподілу. Позитивний ексцес відповідає гостровершинним рядам, а від’ємний – більш плосковершинним.

Для встановлення міри відхилення від нормального розподілу вивчають показник *ексцесу*. Він характеризує відхилення від нормального розподілу варіант з «виступанням» або «падінням» вершини кривої



розподілу. При «виступанні» вершини ексцес називають додатним, а при її «падінні» – від’ємним. Узагальнюючою характеристикою є *моменти*, які розраховуються як середня арифметична  $k$ -го ступеня відхилень варіантів  $x$  від деякої сталої  $A$

$$\mu_k = \frac{\sum_1^m (x_j - \bar{x})^k f_j}{\sum_1^m f_j} = \frac{\sum_1^m (x_j - \bar{x})^k}{n}. \quad (4.15)$$

Для порівняння ступеня асиметрії різних розподілів використовують стандартизований момент:  $A = \frac{\mu_3}{\sigma^3}$ . Вважають, якщо  $A < 0,25$  – асиметрія низька, якщо  $A$  не перевищує  $0,5$  – середня, при  $A > 0,5$  – висока.

Для кількісного виміру гостровершинності використовується центральний момент четвертого порядку. Відношення останнього до середньоквадратичного відхилення в четвертому степені називають коефіцієнтом гостровершинності  $E_x$  або ексцесом. Форми «вершинності» наведено на графіку (рис. 4.2).

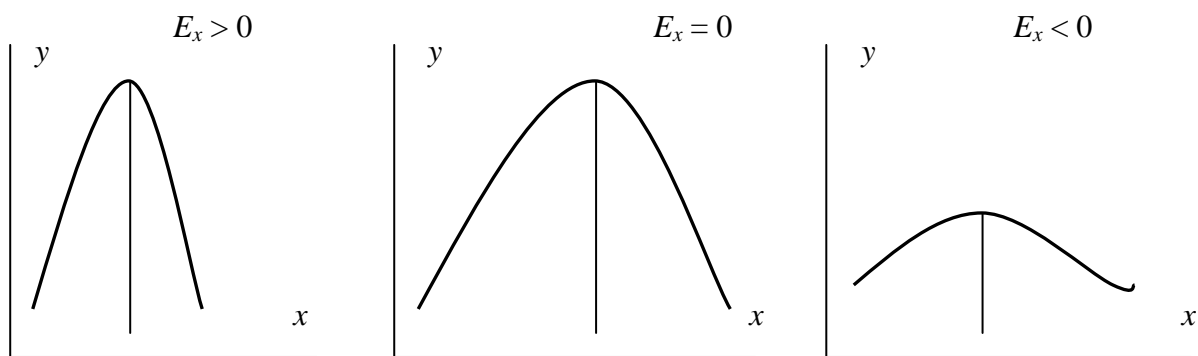


Рисунок 4.2 – Форми розподілу для різних значень ексцесу

Для вимірювання ексцесу використовують стандартизований момент четвертого порядку:  $E_x = \frac{\mu_4}{\sigma^4}$ . У симетричному розподілі  $E_x = 3$ , в разі гостровершинного –  $E_x > 3$ , для плосковершинного –  $E_x < 3$ .

Для визначення міри концентрації елементів сукупності обчислюють *коефіцієнт концентрації*



натуральних логарифмів ( $e \approx 2,1782$ ).

Функція  $F(X)$  табульована, її значення знаходять за спеціальною таблицею. Нормоване відхилення визначають за формулою

$$t = \frac{x - \bar{x}}{\sigma_x}, \quad (4.19)$$

де  $\sigma_x$  – середнє квадратичне відхилення.

Після обчислення теоретичних частот виникає потреба перевірити запропоновану гіпотезу на відповідність або невідповідність того чи іншого теоретичного закону розподілу, прийнятого за математичну модель для емпіричного розподілу. У статистиці використовується декілька показників, за допомогою яких можна оцінити, наскільки фактичний розподіл узгоджується з нормальним. Такі показники називаються *критеріями згоди*. Критерії згоди – це певна величина, що оцінює досліджуване явище з певною ймовірністю.

У статистиці застосовуються критерії згоди Пірсона, Колмогорова, Ястремського, Романовського, Фішера та інші. Одним із основних та найпоширеніших показників є критерій Пірсона та Колмогорова.

Англійський вчений К. Пірсон запропонував критерій, статистичну характеристику якого обчислюють за формулою

$$\chi^2 = \sum \frac{f - f'}{f'}, \quad (4.20)$$

де  $f, f'$  – відповідно фактичні та теоретичні частоти.

За спеціальними таблицями визначають ймовірність досліджуваного значення  $\chi^2$  залежно від кількості ступенів свободи

$$k = K - r, \quad (4.21)$$

де  $K$  – кількість груп;  $r$  – кількість обмежених зв'язків.

Якщо фактичне значення  $\chi^2$  менше від табличного, то це означає, що при прийнятому рівні значущості розбіги між фактичними та теоретичними частотами вважають випадковими, виходячи з чого приймають гіпотезу про тип закону розподілу.

*Приклад.* Розглянемо доведення гіпотези про правильність визна-

чення типу розподілу, за допомогою критерію Пірсона (табл. 4.1).

Таблиця 4.1 – Вихідні дані для визначення типу розподілу

Урожай- ність культури, ц/га	Кіль- кість госпо- дарств	$x$	$ x - \bar{x} $	$t = \frac{ x - \bar{x} }{\sigma}$	$f(t)$	$f^i$	$f - f^i$	$(f - f^i)^2$	$\frac{(f - f^i)^2}{f^i}$
40–42	4	41	5,72	2,22	0,0339	2	2	4	2,00
42–44	7	43	3,72	1,44	0,1415	11	-4	16	1,45
44–46	28	45	1,72	0,67	0,3187	25	3	9	0,36
46–48	35	47	0,28	0,11	0,3965	31	4	16	0,52
48–50	16	49	2,28	0,88	0,2709	21	-5	25	1,19
50–52	6	51	4,28	1,66	0,1006	8	-2	4	0,50
52–54	4	53	6,28	2,43	0,0208	2	2	4	2,00
Разом	100	–	–	–	–	100	–	–	8,02

Розрахунок:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{4672}{100} = 46,72 \text{ ц/га};$$

$$\sigma = \sqrt{6,6414} = 2,58 \text{ ц/га};$$

$$\sum f \frac{i}{\sigma} = 100 \frac{2}{2,58} = 77,5;$$

$$f^i = \sum f \frac{i}{\sigma} f = 77,5 \cdot 0,0339 = 2;$$

$$\chi_{\phi}^2 = \sum \frac{(f - f^i)^2}{f^i} = 8,02;$$

$$k = K - r = 7 - 3 = 4;$$

$$\chi_{табл}^2 = \Phi = 0,99 \Rightarrow 13,28.$$

*Висновок.* Оскільки фактичний критерій  $\chi_{\phi}^2$  набагато менший, ніж табличне значення  $\chi_{табл}^2$  ( $8,02 < 13,28$ ), то з імовірністю 0,99 можна вважати доведеним, що тип розподілу добрано правильно, тобто розподіл господарств за врожайністю культури нормальний.

*Критерій згоди Колмогорова*  $\lambda$  оцінює близькість фактичного та теоретичного розподілів за величиною  $D$ , тобто за максимальною різницею накопичених (кумулятивних) часток (частот) фактичного і теоретичного розподілів.

Критерій Колмогорова визначають за формулою

$$\lambda = D\sqrt{n}, \quad (4.22)$$

де  $D$  – абсолютна максимальна різниця кумулятивних часток  $D = \max|Sd - Sd'|$  або частот  $D = \max|Sf - Sf'|$  емпіричного та теоретичного розподілів;  $n$  – кількість спостережень (кількість одиниць сукупності).

Якщо розподіл задано в частотах, то

$$\lambda = \frac{D}{\sqrt{n}}. \quad (4.23)$$

*Приклад.* Розглянемо методику розрахунку даного показника за допомогою даних наведених в табл. 4.2.

Таблиця 4.2 – Вихідні дані для розрахунку  $\lambda$

Номер групи	Нагромаджена частота (кумулятивна)		Відхилення $ Sf - Sf' $
	емпіричні $Sf$	теоретичні $Sf'$	
1	4	2	2
2	11	13	2
3	39	38	1
4	74	69	5
5	90	90	0
6	96	98	2
7	100	100	0

Розрахунок:

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

*Висновок.* За спеціальною таблицею ймовірностей для критерію згоди Колмогорова визначаємо –розрахованому значенню 0,5 відповідає ймовірність 0,9639, тобто з ймовірністю 0,9639 можна стверджувати про нормальний розподіл господарств за врожайністю культури.

Отже, для перевірки гіпотези про відповідність чи невідповідність теоретичного закону розподілу емпіричному можна використовувати будь-який з наведених критеріїв, які забезпечують дослідження законів розподілу з різною точністю, надійністю і трудомісткістю.

*Характеристики центра розподілу.* Для характеристики середнього значення ознаки у варіаційному ряді використовуються середня арифметична, мода і медіана (відповідно формули 3.12, 3.28, 3.30).

#### 4.4. Види та взаємозв'язок дисперсій

Дисперсія посідає особливе місце у статистичному аналізі соціально-економічних явищ. Дисперсія є важливим елементом інших статистичних методів, зокрема дисперсійного аналізу.

Дисперсія, як було вказано вище (формула 4.5), – це середній квадрат відхилень індивідуальних значень ознаки від середньої.

Розглянемо структуровану сукупність, яка поділена на групи за факторною ознакою «х». Відповідна їй результативна ознака «у» також поділена на групи. У кожній групі розраховано групову середню арифметичну результативної ознаки  $\bar{y}_j$ ; частоту  $f_j$  групи «j». Загальна середня результативних ознак усієї сукупності  $\bar{y}$ .

У кожній групі обчислюється *внутрішньогрупова дисперсія*

$$\sigma_j^2 = \frac{\sum_1^{f_j} (y_j - \bar{y}_j)^2}{f_j}, \quad (4.24)$$

де  $\bar{y}_j$  – значення ознаки елементів «j» групи сукупності, їх кількість  $f_j$ .

Для всіх груп у цілому, тобто сукупності, обчислюється *середня з внутрішньогрупових дисперсій*

$$\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j}, \quad (4.25)$$

де  $m$  – кількість груп, на яку поділена сукупність.

Різними є групові середні  $\bar{y}_j$ . Міжгрупова дисперсія обчислюється за формулою:

$$\delta^2 = \frac{\sum_1^m (y_j - \bar{y})^2 \cdot f_j}{\sum_1^m f_j}, \quad (4.26)$$

$$\sum_1^m f_j = n, \quad (4.27)$$

де  $n$  – кількість елементів сукупності.

Загальна дисперсія сукупності за *правилом складання дисперсій*

$$\sigma^2 = \overline{\sigma_j^2} + \delta^2. \quad (4.28)$$

Отже, загальна дисперсія складається з двох частин. Перша характеризує внутрішньогрупову, друга – міжгрупову варіацію. Внутрішньогрупова дисперсія вимірює розшарування ознак у межах кожної з груп і пояснює дію різнобічних факторів у межах групи. Міжгрупова дисперсія вимірює розшарування ознак, яке викликано фактором, за яким сукупність поділено на окремі групи.

Взаємозв'язок дисперсій називається *правилом розкладання варіації*. Правило розкладання варіації є основою вимірювання щільності зв'язку.

Міжгрупова дисперсія – це результат впливу фактора, який покладено в основу групування, внутрішньогрупова – інших факторів, окрім групувального. Відношення міжгрупової дисперсії до загальної характеризує частку варіації результативної ознаки  $y$ , яка пов'язана з варіацією групувальної ознаки. Це відношення називають *кореляційним* і розраховують за наступною формулою

$$\eta^2 = \frac{\delta^2}{\sigma^2}. \quad (4.29)$$

Кореляційне відношення вимірює вплив факторної ознаки « $x$ », за якою поділена генеральна сукупність на групи, на результативну « $y$ », або щільність зв'язку між факторною та результативною ознаками.

Альтернативна ознака – це ознака, яка одній частині сукупності властива, а іншій – не властива. Хай альтернативна ознака має два значення: одиницю та нуль:

$x_1 = \langle 1 \rangle \rightarrow d_1$  – частка ознак, що мають значення «1»,

$x_2 = \langle 0 \rangle \rightarrow d_0$  – частка ознак, що мають значення «0».

Таким чином, дисперсія визначається за формулою

$$\sigma^2 = \frac{\sum_{i=1}^2 (x_i - \bar{x})^2 d_i}{d_1 + d_0}. \quad (4.30)$$

Звідси

$$\sigma^2 = \frac{d_1(-d_1) + (-d_1)d_0}{d_1 + d_0} = \frac{d_1 d_0^2 + d_0 d_1^2}{d_1 + d_0} = d_1 d_0. \quad (4.31)$$

Середнє арифметичне значення розраховують таким чином:

$$\bar{X} = \frac{x_1 d_1 + x_2 d_0}{d_0 + d_1}, \quad (4.32)$$

$$\bar{X} = \frac{1 \cdot d_1 + 0 \cdot d_0}{d_0 + d_1} = \frac{d_1 \cdot 1}{1} = d_1. \quad (4.33)$$

Максимальне значення альтернативної ознаки становить 0,25.

Ми розглянули визначення показників варіації ознаки для дискретних ранжированих рядів. Ці показники для інтервальних рядів визначаються аналогічно, але замість значення варіанти  $x_i$  використовується середнє значення кожного інтервалу  $x_i$ , яке відрізняється від середньої арифметичної. Це призводить до появи систематичної похибки при розрахунках дисперсії. В. Ф. Шеппард встановив, що похибки у розрахунках дисперсії, викликана застосуванням згрупованих даних, становить  $1/12$  квадрата величини інтервалу (тобто  $h^2/12$ ), як у напрямку заниження, так і у напрямку завищення величини дисперсії.

Виправлення (поправка) Шеппарда повинна застосовуватися, якщо розподіл близький до нормального, відноситься до ознаки з безперервним характером варіації, побудований на великій кількості вихідних даних ( ). Однак, виходячи з того, що в ряді випадків обидві похибки, діючи в протилежних напрямках, нейтралізуються і компенсують одна одну, можна іноді відмовитись від введення поправок.

Чим менше значення дисперсії і середнього квадратичного відхилення, тим однорідніша сукупність і тим більш типовою буде середня величина.

Дисперсія має ряд *властивостей*, які дозволяють спростити її обчислення, а саме:

- ✓ Дисперсія постійної величини дорівнює 0.
- ✓ У разі заміни частот частками дисперсія не зміниться.



✓ Якщо всі значення ознаки у сукупності зменшити чи збільшити на певну постійну величину дисперсія, не зміниться.

✓ Якщо всі значення ознаки змінити у  $K$  разів, то дисперсія зміниться в  $K^2$  разів.

✓ Сума квадратів відхилень ознаки від її середньої (дисперсія) завжди менша від суми квадратів відхилень ознаки від будь-якого іншого числа

$$\sum_{i=1}^n (x_i - \bar{x})^2 < \sum_{i=1}^n (x_i - a)^2, \quad \bar{x} \neq a. \quad (4.34)$$

Середній квадрат відхилень (дисперсія) може бути обчислений за формулами (4.4), (4.5) та за формулою наступного вигляду

$$D_x = \sigma_x^2 = \overline{x^2} - \bar{x}^2. \quad (4.35)$$

#### ***4.5. Різновиди та характеристики форм розподілів***

Для відображення особливостей форм розподілу застосовуються криві розподілу. Розрізняють емпіричні і теоретичні криві розподілу. Емпірична крива – це по суті полігон чи гістограма розподілу. Криві розподілу якісно однорідних сукупностей, як правило, одновершинні, серед яких розрізняють симетричні і асиметричні. Асиметрія може бути правосторонньою чи лівосторонньою. Крутизна, гостровершинність кривої розподілу називають ексцесом. Мова про гостровершинність емпіричної кривої може йти лише в тому випадку, коли вона порівнюється з теоретичною кривою нормального розподілу (крива Гаусса). Крива нормального розподілу є симетричною (тобто її ліве крило дорівнює правому). Вона характеризує закономірність розподілу в тих випадках, коли відхилення варіант  $(x_1, x_2, \dots, x_n)$  від середньої в одну сторону зустрічаються також часто, як і в іншу сторону.

У практиці соціально-економічних явищ нормальна крива в чистому вигляді зустрічається вкрай рідко. Частіше зустрічаються криві розподілу, вершина яких скошена вліво або вправо.

Розглянемо типи розподілу :

- 1) симетричний розподіл;
- 2) правостороння асиметрія;
- 2) лівостороння асиметрія.

Показником, що характеризує міру асиметрії, є коефіцієнт асиметрії. Для характеристики крутизни (ексцесу) емпіричної кривої використовується показник, що називається коефіцієнтом ексцесу. Ексцес кривої нормального розподілу завжди дорівнює 3 ( $E_x=3$ ). Якщо  $E_x > 3$ , то ексцес вищий нормального, а якщо  $E_x < 3$ , то він нижчий від нормального.

Дискретна випадкова величина (біноміальний розподіл) описується *схемою Бернуллі*: якщо випадкова подія  $A$  в  $n$  незалежних випадках зустрілася  $m$  разів, то  $p$  – ймовірність появи події  $A$  у кожному випадку. Формула Бернуллі дозволяє оцінити ймовірність того, що серед  $n$  взятих довільно елементів виявиться  $m$  сподіваних. Даний розподіл характеризується двома параметрами: середнім числом сподіваного результату (математичне очікування) і дисперсією частоти події  $A$  в  $n$  незалежних випадках

$$\text{-----} \tag{4.36}$$

і має вигляд

$$P_n(m) = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}, \tag{4.37}$$

$$m = 0, \dots, n.$$

Граничним випадком біноміального розподілу є формула Пуассона

$$P_n(m) = \frac{a^m}{m!} \exp\{-a\}, \tag{4.38}$$

$$a \approx n \cdot p$$

$$m = 0, 1, 2, \dots$$

Випадкова величина розподілена за законом Пуассона, якщо вона має рахункову множину можливих значень 0, 1, 2, з ймовірностями  $P_n(m)$ . Коли у схемі Бернуллі існує ймовірність появи події  $A$  (величина  $p$  дорівнює сотим чи тисячним часткам одиниці), тобто коли частина випадків дуже мала, то розподіл частот таких рідких подій у  $n$  випадках стає несиметричним і зазвичай описується формулою Пуассона. Розподіл характеризується одним параметром – середньою величиною, яка дорівнює  $a$ , середнє квадратичне відхилення в даному випадку також

дорівнює  $a$ . Для такого розподілу характерний високий рівень варіації. Зі зростанням значень  $a$  розподіл прямує до нормального закону. Розподіл Пуассона є моделлю, яку можна використовувати для опису випадкової кількості появи визначених подій у фіксованому проміжку часу. Безперервний розподіл – це рівномірний розподіл на відрізку  $[0;1]$

$$p(x) = \begin{cases} 1, & 0 \leq x \leq 1; \\ 0, & x \notin [0,1]. \end{cases} \quad (4.39)$$

Безперервний розподіл можна розповсюджувати на випадок відрізка  $[0;1]$ , тоді ймовірність має значення в будь-якій точці відрізка і дорівнює  $1/(b-a)$ . Математичне очікування розподілу дорівнює  $(b-a)/2$ , дисперсія дорівнює  $(b-a)^2/12$ . Безперервний експонентний (показовий) розподіл має вигляд

$$p(x) = \begin{cases} \lambda \cdot \exp(-\lambda x), & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (4.40)$$

де  $\lambda$  – параметр експонентного розподілу.

Математичне очікування дорівнює  $1/\lambda$ , а дисперсія –  $1/\lambda^2$ .

Розподіл Максвелла (безперервний розподіл) має вигляд

$$p(x) = \frac{2}{\sqrt{2\pi}} \cdot \frac{t^2}{a} \exp\left\{-\frac{t^2}{2}\right\}, \quad (4.41)$$

$$t = x/a$$

і описує асиметричні розподіли. У цій формулі параметр  $a$  дорівнює середньому арифметичному, помноженому на величину 0,6267. Характерною ознакою розподілу Максвелла є рівність середнього квадратичного відхилення величини  $0,674a$ . Крива розподілу за формулою нагадує нормальний розподіл, але починається від нуля, крутіше піднімається з малих значень випадкової величини і потім, досягнувши максимуму, більш полого спадає у бік великих значень. Такий розподіл виникає, наприклад, при побудові розподілу осіб і популяції за їх відстанями від оптимального фенотипу, що обернено пропорційні до їх фенотипової цінності.

Розподіл Шарльє (безперервний) має вигляд

$$p(x) = p(x) - Axp'(x) + Bxp''(x) + Cxp'''(x) + Dxp''''(x), \quad (4.42)$$

де  $p(x)$  – щільність нормального розподілу;  $p'(x)$  – похідна відповідного порядку щільності нормального розподілу  $p(x)$ ;  $Ax$  – асиметрія;  $Bx$  – ексцес.

Розподіл Шарльє описує асиметричний розподіл з вираженням ексцесом, що виникає при порушенні форми кривої, характерної для нормального розподілу. Така крива розподілу є асиметричною, її звоноподібна вершина стає пікоподібною, чи трапецієподібною. За допомогою розподілу такого виду «формується» порушення нормальної форми розподілу.

*Гамма-розподіл* (безперервний) має вигляд

$$p(x) = e^{-x/\beta} \exp\left\{-\frac{x}{\beta}\right\} / \beta^\alpha \Gamma(\alpha), x > 0, \quad (4.43)$$

де  $\Gamma(\alpha)$  – гамма-функція. Її визначення за Ейлером задається співвідношенням

$$\Gamma(\alpha) = \int_0^{\infty} \exp[-t] t^{\alpha-1} dt. \quad (4.44)$$

Основні властивості гамма-функції:  $\Gamma(1)=1$ ,  $\Gamma(x+1)=x \Gamma(x)$ .

Гамма-функція являє собою двопараметричний розподіл, де  $\alpha$  – параметр форми, а  $\beta$  – параметр масштабу. Математичне очікування дорівнює  $\alpha\beta$ , дисперсія задається співвідношенням  $\alpha\beta^2$ , мода дорівнює  $(\alpha-1)\beta$  при  $\alpha \geq 1$ . Гамма-функція є безперервним аналогом негативного біноміального розподілу.

При  $\alpha=1$  гамма-розподіл збігається з показовим; при  $\alpha=n$ ,  $\beta=1/n$  гамма-розподіл називається ерлангівським розподілом з параметрами  $(n, \mu)$  і описує розподіл тривалості інтервалу часу до появи  $n$  подій процесу Пуассона з параметром  $\mu$ .

#### **4.6. Оцінка і аналіз нерівномірності та подібності розподілів**

У загальному вигляді графічно всі види рядів розподілу можна відобразити у вигляді гістограми або полігону (де ось ординат  $OY$  – результативна ознака, ось абсцис  $OX$  – факторна ознака).

Графічне зображення рядів розподілу дає наочне уявлення про закономірності розподілу. Розглянемо більш докладно застосування графічного методу відображення рядів розподілу. Дискретний ряд відображається на графіку у вигляді ламаної лінії – *полігону розподілу* (*полігон частот*) (рис. 4.3).

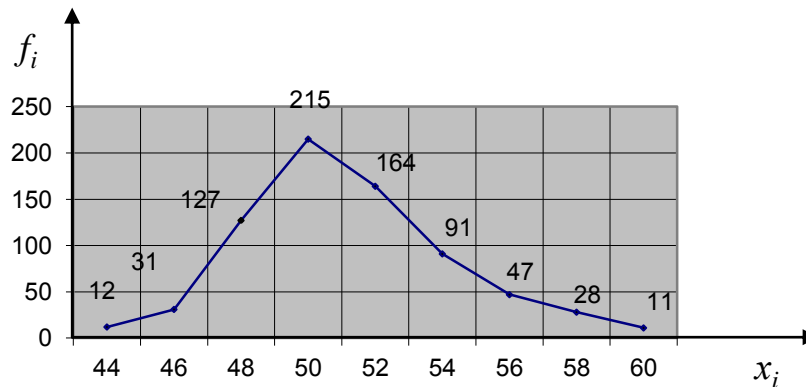


Рисунок 4.3 – Графічне відображення дискретних рядів розподілу

*Полігон* – графічне зображення варіаційного ряду в прямокутній системі координат, при якому величина ознаки відкладається на осі абсцис, а частоти або частки (щільність розподілу) – на осі ординат. Частіше полігон застосовується для зображення дискретного варіаційного ряду, однак може використовуватись і для інтервального ряду.

Інтервальні ряди відображаються у вигляді *гістограм розподілу* (тобто стовпчикових діаграм), при цьому основою кожного прямокутника служить величина відповідного інтервалу, а його висота – це частотна характеристика (рис. 4.4). Будь-яка гістограма може бути перетворена в полігон розподілу, для цього необхідно з'єднати між собою відрізками прямої середини її прямокутників.

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

Для характеристики рядів розподілу (дискретних і інтервальних) застосовують графіки накопичених частот або кумуляти. *Кумулятивна крива (кумулята)* – графічне зображення варіаційного ряду, побудована за накопиченими частотами або частками.

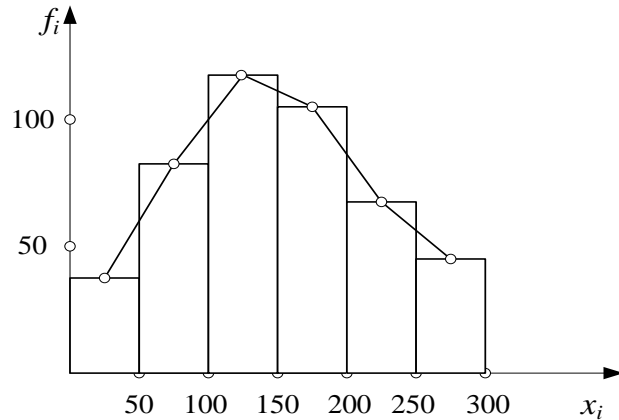


Рисунок 4.4 – Графічне відображення інтервальних рядів розподілу

*Накопичена (кумулятивна) частота* – це сума частот даного і всіх попередніх інтервалів. *Кумулята* дозволяє визначити, яка частина сукупності, що має значення ознаки, що вивчається, не перевищує заданої межі, а яка частина, навпаки, перевищує цю межу.

Приклад побудови кумуляти частот варіаційного ряду наведено на рис. 4.5 на основі даних табл. 4.3.

Таблиця 4.3 – Розподіл господарств за врожайністю зернових

Врожайність, га	Число господарств $f_i$	Накопичена частота $\sum f_i$
До 6	2	2
6–10	8	10 (2+8)
10–14	17	27 (10+17)
14–18	12	39 (12+27)
18–22	6	45 (6+39)
Понад 22	2	47 (25+2)
Разом	47	47

*Огіва* – графічне зображення варіаційного ряду, аналогічне кумуляті, але на ось ординат наносяться значення ознаки, а на ось абсцис – накопичені частоти (рис. 4.6.)

*Крива концентрації Лоренца* – це крива, що використовується як показник ступеня оцінки концентрації розподілу частот. Приклад графічного відображення кривої Лоренца наведено на рис. 4.7.

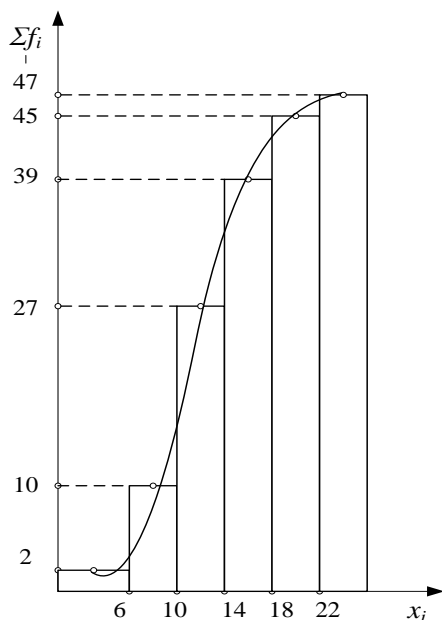


Рисунок 4.5 – Графічне відображення кумуляти

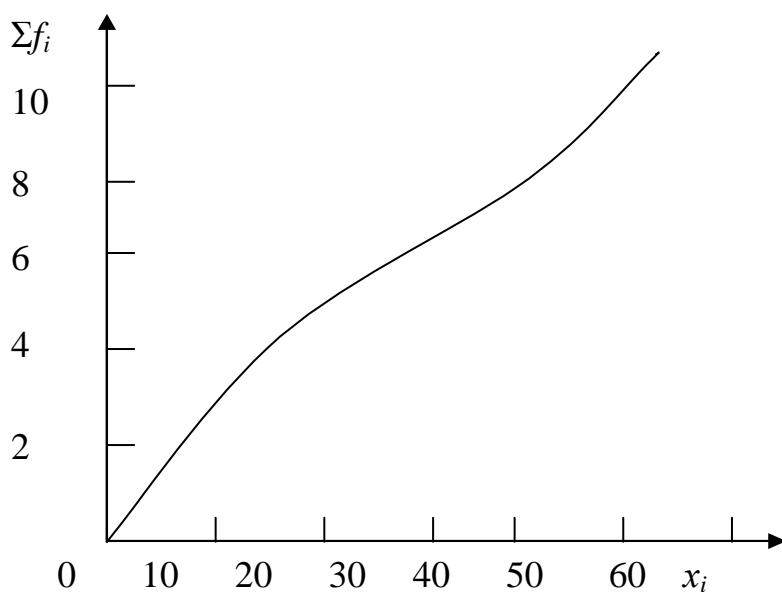


Рисунок 4.6 – Графічне зображення Огіви

До інших видів графічного зображення рядів розподілу належать: *показникова крива, крива Парето і антимода*. Прикладом *показникового розподілу* є термін служби предметів, які вибувають через аварійність, – довговічність посуду в підприємствах громадського харчування, тривалість телефонних викликів, час між простоями верстатів і т. д. *Крива Парето* показує розподіл доходів відповідно до їх величини. *Антимода* – абсциса нижньої точки, розміщеної в центральній частині

$U$ -подібного розподілу ознаки. По обидві сторони антимода частоти зростають. Приклад графічного відображення цих кривих наведено на рис. 4.8.

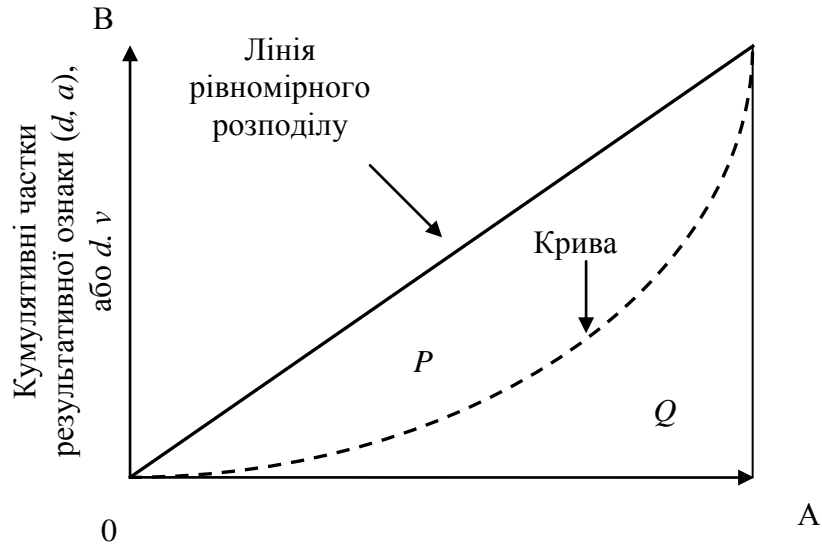


Рисунок 4.7 – Графічне відображення кривої концентрації Лоренца

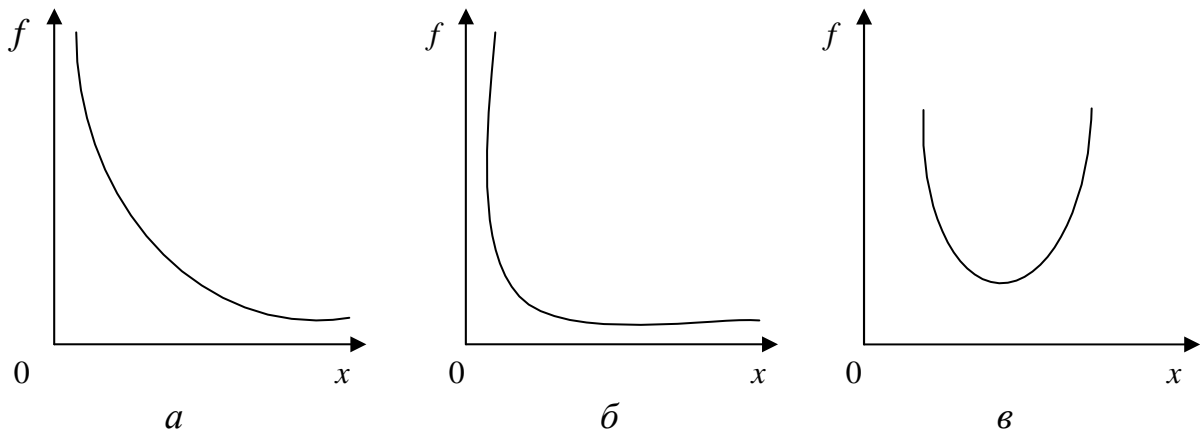


Рисунок 4.8 – Інші види графічного відображення рядів:  
*a* – показникова крива; *б* – крива Парето; *в* – антимода

Оцінку та аналіз нерівномірності рядів розподілу зручніше виконувати за допомогою графічного методу.

На першій стадії використовується діаграма «казусів» – на діаграмі наносять значення ознаки, що змінюється, у тій послідовності, у якій вони отримані. Тобто наносять первинні дані (рис. 4.9).



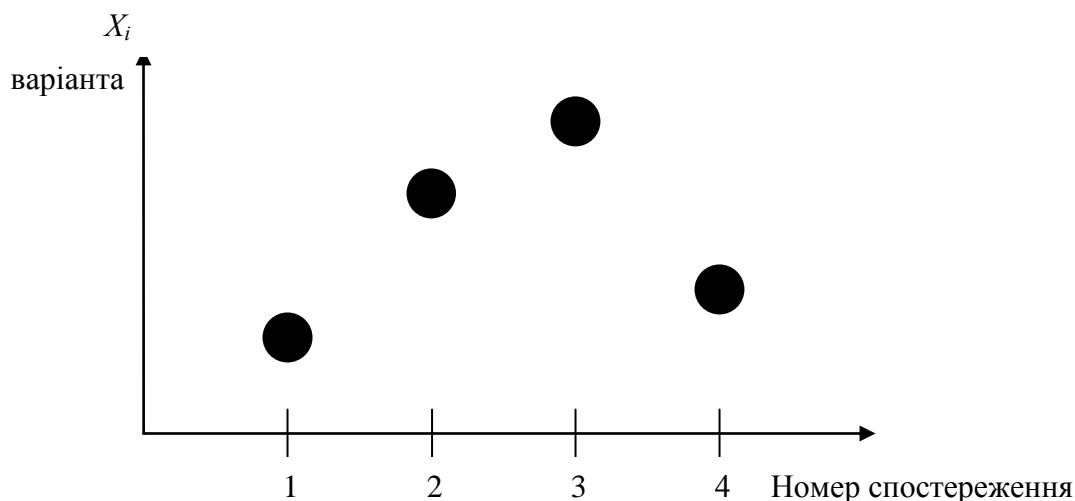


Рисунок 4.9 – Діаграма «казусів»

На другій стадії будується *ранжований ряд* – ряд, у якому окремі значення ознаки розташовані у порядку зростання, тобто за ранжиром.

Точки поєднуються відрізками і утворюють ламану лінію, яка називається *огівою*. За нахилом огіви можливо отримати уяву про ступінь різномірності сукупності (рис. 4.10).

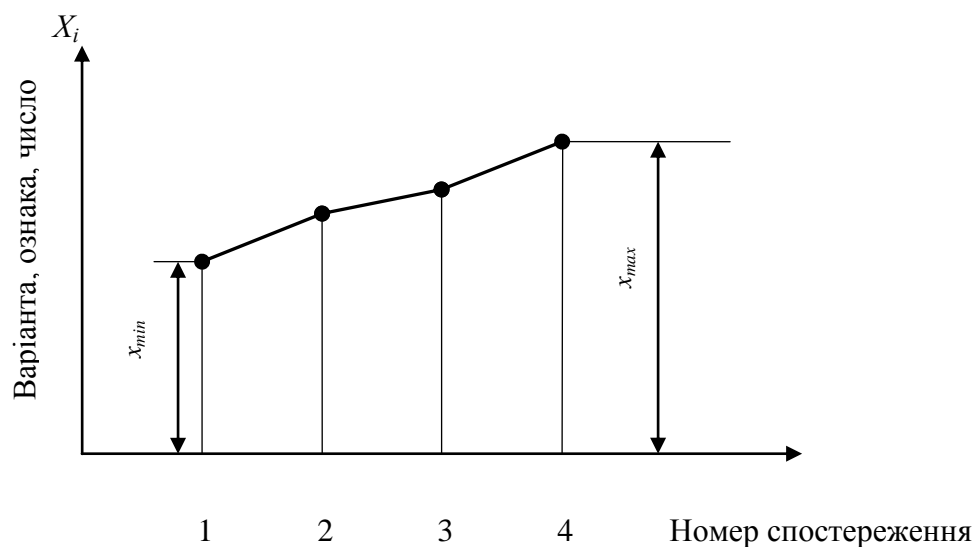


Рисунок 4.10 – Огіва

На третій стадії вивчення ряду розподілу будується *полігон* – графік, на якому ряд зображений у вигляді лінійної діаграми. На осі абсцис відкладають значення ознаки  $X_j$  у порядку зростання, а на осі ординат – частоту відповідної варіанти (ознаки)  $f_j$ . Точки поєднуються

відрізками та утворюють ламану лінію (рис. 4.11).

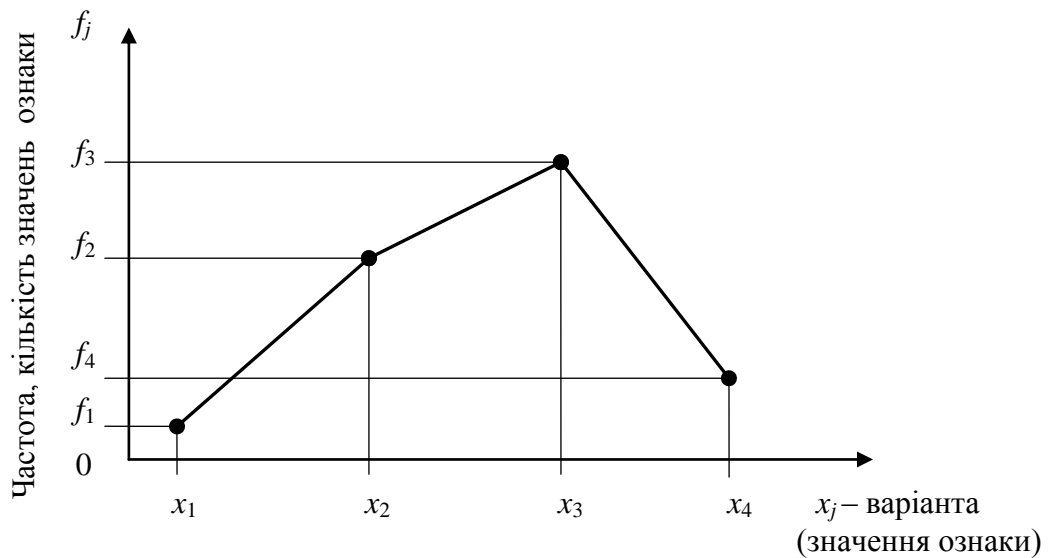


Рисунок 4.11 – Полігон

Для інтервальних рядів розподілу будується *гістограма* – графік, на якому ряд розподілу зображується у вигляді розташованих один біля одного стовпчиків. Висота стовпчика пропорційна частоті, тобто кількості ознак, що потрапили до інтервалу. Ширина стовпчика дорівнює ширині інтервалу розподілу (рис. 4.12).

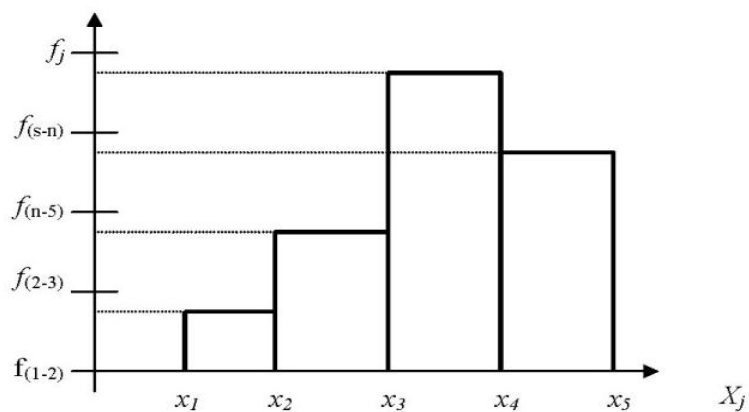


Рисунок 4.12 – Гістограма

Щоб уникнути впливу нерівномірного інтервалу розподілу на висоту стовпчиків, висоту стовпчика зображають пропорційно не частотам, а *густині розподілу*.

*Густина розподілу* – кількість випадків, що припадає на одиницю ширини інтервалу ознаки, яка змінюється. Частота не має розміру – вона

вимірюється у кількості випадків. На відміну від частоти, густина розподілу має розмір:

$$\text{Густина розподілу} = \text{кількість випадків} / \text{розмір інтервалу}. \quad (4.45)$$

На четвертій стадії вивчення ряду розподілу будується кумулята – графік, на якому на осі абсцис відкладають значення ознаки у порядку зростання, а на осі ординат – відповідний підсумок накопичених частот (часток) (рис. 4.13).

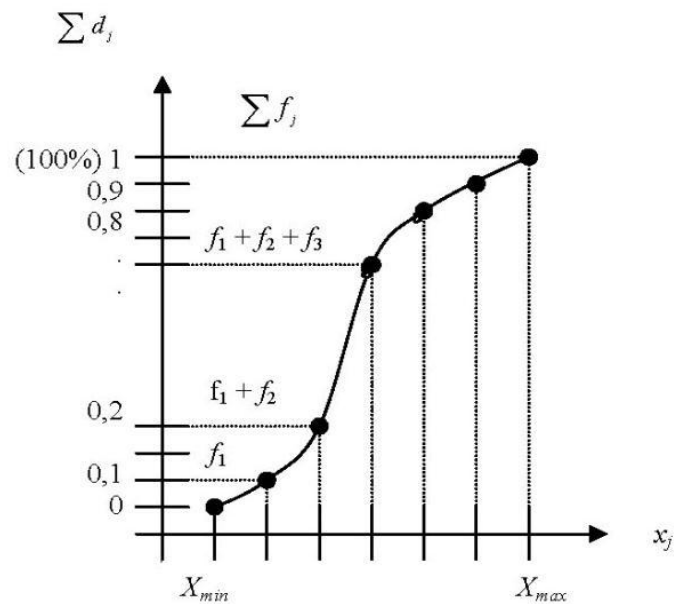


Рисунок 4.13 – Кумулята

Накопичені частоти підраховують шляхом послідовного складання. Ознаці, яка має максимальне значення, відповідає відносна частка 1 або 100 %, а підсумок частот дорівнює  $n$  (кількості одиниць усієї сукупності) (рис. 4.14).

На п'ятій стадії вивчення ряду розподілу будується інтегральна функція розподілу (рис. 4.15).

Графічне зображення закону розподілу Гаусса у диференціальній формі наведено на рис. 4.14, а графічне зображення закону розподілу Гаусса у інтегральній формі – на рис. 4.15.

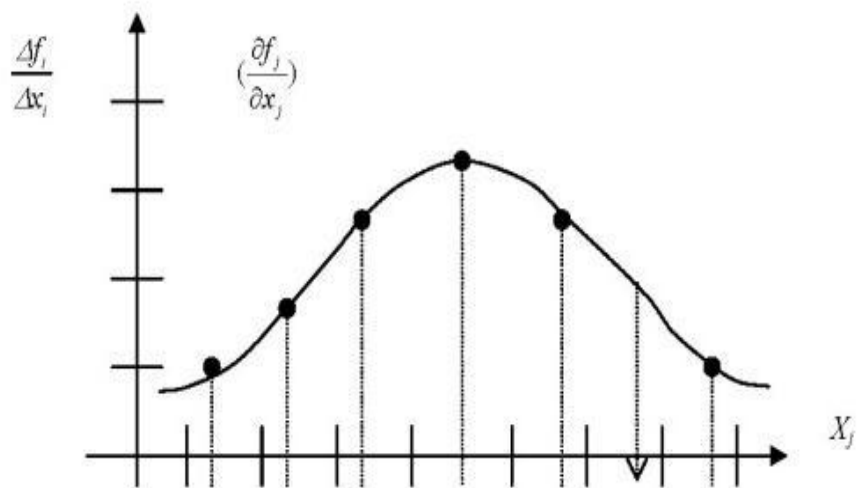


Рисунок 4.14 – Густина розподілу (щільність ймовірностей, диференціальна функція розподілу)

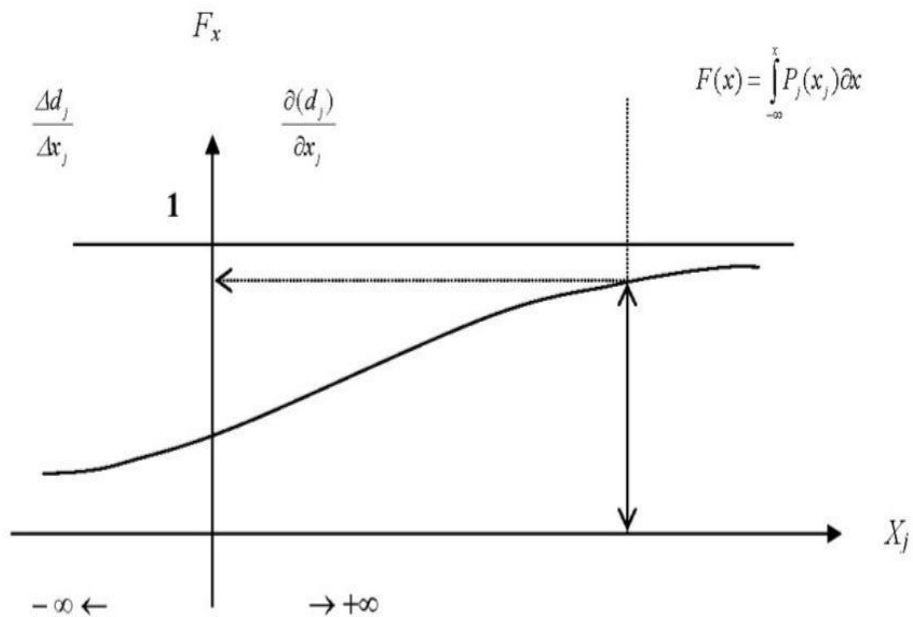


Рисунок 4.15 – Інтегральна функція розподілу (емпірична функція розподілу)

У теорії ймовірностей і статистиці геометричний розподіл визначається як будь-який з двох розподілів ймовірностей:

- дискретна випадкова величина  $X$  має геометричний розподіл з параметром  $p$ , якщо вона збігається з кількістю випробувань до першого успіху в нескінченній послідовності випробувань Бернуллі з імовірністю успіху в одному випробуванні

(4.46)

де  $k = 1, 2, 3, \dots$

• величина  $Y = X - 1$ , що дорівнює кількості неуспіхів до першого успіху.

(4.47)

де  $k = 0, 1, 2, 3, \dots$

Який з цих розподілів називати геометричним – питання згоди і зручності. Ці два різні геометричні розподіли не можна плутати один з одним. Очікувана величина геометричного розподілу випадкової величини  $X$  та її похибка розраховуються за формулами

— (4.48)

Так само очікувана величина геометричного розподілу випадкової величини  $Y$  та її похибка розраховуються за формулами

— (4.49)

*Оцінка параметра.* Для обох варіантів геометричного розподілу параметр  $p$  може оцінюватися через порівняння очікуваної величини. Це метод моментів, який в даному випадку проводить оцінки максимальної ймовірності.

Припустимо, для першого варіанта  $k_1, k_2, \dots, k_n$ , коли  $k_i \geq 1$  для  $i = 1, 2, \dots, n$ . Тоді параметр  $p$  може бути оцінений як

$$\hat{p} = \left(\frac{1}{n} \sum_{i=1}^n k_i\right)^{-1}; \tag{4.50}$$

$$p \sim \text{Beta}\left(\alpha + n, \beta + \sum_{i=1}^n (k_i - 1)\right); \tag{4.51}$$

$$\hat{p} = \left(1 + \frac{1}{n} \sum_{i=1}^n k_i\right)^{-1}; \quad (4.52)$$

$$p \sim \text{Beta}(\alpha + n, \beta + \sum_{i=1}^n k_i). \quad (4.53)$$

*Інші властивості.* Функція вірогідності  $X$  і  $Y$  відповідно

$$\text{—————} \quad (4.54)$$

$$\text{—————} \quad (4.55)$$

Подібно до неперервного аналога (показниковий розподіл), геометричний розподіл має властивість відсутності пам'яті. Це означає, що кількість попередніх «невдач» не впливає на кількість наступних «невдач». Таким чином, геометричний розподіл – це єдиний дискретний розподіл з такою властивістю.

Серед всіх дискретних ймовірних розподілів на  $\{1, 2, 3, \dots\}$  з даною очікуваною величиною  $\mu$  геометричний розподіл  $X$  з параметром  $p = 1/\mu$  є одним. Геометричний розподіл числа  $Y$  невдач перед першим успіхом є нескінченно діленим. Для будь-якого додатного цілого  $n$ , існують незалежні тотожно розподілені випадкові величини  $Y_1, \dots, Y_n$ , сума яких має такий самий розподіл, як і  $Y$ .

Як було зазначено вище, нормальний розподіл має *ексцес*  $E_x=3$ ; гостровершинний розподіл має *ексцес*  $E_x>3$ ; плосковершинний розподіл має *ексцес*  $E_x<3$ . Більш детально форма розподілу характеризується моментом розподілу (див. формулу 4.15).

Протилежною характеристикою однорідності сукупності є *нерівномірність розподілу* ознак між складовими сукупності, яка оцінюється *коефіцієнтом концентрації* (див. формулу 4.16).

*Коефіцієнт локалізації* характеризує для однієї сукупності співвідношення часток двох різних розподілів ознак однієї групи « $j$ »

$$L_j = \frac{D_j}{d_j} \cdot 100 \%. \quad (4.56)$$

Порівняння двох структур *різних сукупностей*, що поділені на однакову кількість груп «*m*», здійснюється за допомогою *коефіцієнта подібності (схожості)* структури двох сукупностей

$$P = 1 - \frac{1}{2} \sum_{j=1}^m |d_j - d_k|, \quad (4.57)$$

де  $d_j$  – частка ознак однієї сукупності у групі,  $d_k$  – частка ознак другої сукупності у такій самій групі. Якщо структури однакові, то тоді  $P = 1$ ; якщо протилежні, то  $P = 0$ .

Зміна окремих складових (часток) однієї сукупності свідчить про структурні зрушення, що відбуваються з сукупністю в часі.

*Інтенсивність структурних зрушень* оцінюється за допомогою середнього відхилення часток

$$l_d = \frac{\sum_{j=1}^m |d_{j1} - d_{j0}|}{m}, \quad (4.58)$$

або *середнього квадратичного відхилення часток*

$$d_d = \sqrt{\frac{\sum_{j=1}^m |d_{j1} - d_{j0}|^2}{m}}, \quad (4.59)$$

де  $m$  – число складових (груп) сукупності;  $d_{j1}$  – частки поточного періоду;  $d_{j0}$  – частки базисного періоду.

📖 Література до теми 4: [4–8; 15–21; 23; 25–28].

#### **Контрольні запитання до теми 4**

1. Що називають рядом розподілу?
2. Дайте визначення варіанти, частоти і частки.
3. Чим варіаційний ряд розподілу відрізняється від атрибутивного?
4. Що таке діаграма «казусів», огіва, полігон, гістограма?
5. Що таке кумулята?
6. Що таке густина розподілу, диференціальна крива розподілу,

інтегральна крива розподілу?

7. Що є варіацією ознаки та від чого залежить її розмах?
8. Що є середнім лінійним відхиленням?
9. Що є середнім квадратичним відхиленням?
10. Що є середнім квадратом відхилення (дисперсія)?
11. Наведіть коефіцієнти варіації.
12. Назвіть властивості дисперсії.
13. У чому полягає правило складання дисперсій?
14. Як розраховується і який зміст має внутрішньогрупова дисперсія?
15. Як розраховується і який зміст має середня з внутрішньогрупових дисперсій?
16. Як розраховується і який зміст має міжгрупова дисперсія?
17. Як розраховується загальна дисперсія?
18. Дайте визначення поняття медіани.
19. Чи зміниться дисперсія, якщо усі варіанти осереднюваної ознаки зменшити у 2 рази?
20. Назвіть елементи варіаційного ряду розподілу та охарактеризуйте кожний з них.



## Тема 5. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

5.1. Поняття і сутність вибіркового методу, причини і умови його застосування.

5.2. Способи відбору одиниць генеральної сукупності.

5.3. Визначення помилки вибіркової середньої.

5.4. Мала вибірка.

5.5. Способи розрахунків характеристик вибірки.

### ***5.1. Поняття і сутність вибіркового методу, причини і умови його застосування***

Статистичне дослідження може здійснюватися за даними несучільного спостереження, основна мета якого полягає у визначенні характеристик досліджуваної сукупності по її обстеженій частині. Одним з найбільш поширених у статистиці методів, який використовує несучільне спостереження, є *вибірковий метод*.

*Вибіркове спостереження* є найпоширенішим з усіх видів несучільного спостереження. Під час проведення вибіркового спостереження вивчаються не всі одиниці досліджуваного об'єкта, а лише окрема відібрана частина цих одиниць. Однак спостереження організоване таким чином, що ця частина відібраних одиниць відображає всю сукупність наче у зменшеному масштабі.

Загальну сукупність однорідних об'єктів (одиниць), з якої здійснюється відбір для дослідження у статистиці, називають *генеральною сукупністю*. Відібрана для обстеження частина генеральної сукупності має назву *вибірки* або *вбіркової сукупності*. На підставі результатів дослідження вибіркової сукупності можна зробити висновки про властивості генеральної сукупності, тобто теоретичні характеристики замінити статистичними.

До теоретичних (аналітичних) характеристик належать узагальнюючі показники генеральної сукупності – середній розмір ознаки  $\bar{X}$ , частка  $P$ , генеральна дисперсія  $\sigma^2$ , які називаються *параметрами*.

Статистичними характеристиками (оцінками) є характеристики вибіркової сукупності – середня вибіркова  $\bar{X}_v$ , вибіркова частка  $W$ , вибіркова дисперсія  $\sigma_v^2$ , які називаються *статистиками*.

Отже, завдання вибіркового спостереження – за результатами

обстеженої частини надати характеристику усїєї сукупності в цілому.

*Переваги вибіркового спостереження перед суцільним:*

- практичність (іноді немає можливості чи недоцільно здійснювати суцільне спостереження, наприклад, контролювати якість книжкової продукції – підручників, дитячої книги тощо);
- висока точність результатів за рахунок зниження помилок реєстрації (наприклад, соціологічні чи інші наукові дослідження, що вимагають професійних знань сутності явища);
- проведення статистичних досліджень, пов'язаних із знищенням зразків (перевірка якості харчової продукції, екологічного стану та інше);
- мінімальні затрати, швидкість дослідження, детальне вивчення кожної одиниці спостереження.

**Зверніть увагу!** Ці переваги вибіркового спостереження мають місце лише за умов дотримання наукових принципів його проведення, що дає можливість отримати репрезентативну вибірку, яка відтворює характеристики генеральної сукупності. Порушення цих принципів, коли спостереженню підлягають одиниці, відібрані на основі суб'єктивного погляду дослідника, призводять до того, що результати такого спостереження мають відношення не до усїєї генеральної сукупності, а тільки до тієї її частини, яка була піддана спостереженню.

Вибірка вважається *репрезентативною*, якщо кожний елемент генеральної сукупності має однакову ймовірність (рівну можливість) потрапити у вибірку і якщо відбір здійснюється строго випадковим чином за схемами, розробленими математичною статистикою, а також, коли обсяг вибіркової сукупності є достатнім. Вибіркове спостереження здійснюється у такій послідовності:

- обґрунтування мети, складання програми;
- визначення необхідної частки (5 %, 25 %, 50 % або ін.) одиниць для обстеження та способу їх відбору до вибіркової сукупності;
- формування вибірки;
- узагальнення даних спостереження і визначення їхніх вибіркових характеристик (оцінок відповідних характеристик генеральної сукупності);
- обчислення похибок вибірки, перевірка надійності результатів;
- поширення вибіркових даних на всю досліджувану сукупність (висновки про параметри генеральної сукупності).

Точність результатів вибіркового спостереження залежить від

способу відбору одиниць, варіації ознаки в сукупності і від кількості відібраних одиниць.

Під вибіркоvim розуміється метод статистичного дослідження, при якому узагальнюючі показники сукупності, яка підлягає вивченню, встановлюються по деякій її частині на основі положень випадкового відбору. При вибіркоvim методі обстеженню піддається порівняно невелика частина всієї сукупності, що вивчається (зазвичай до 5–10 %, рідше до 15–25 %). При цьому статистична сукупність, з якої проводиться відбір частини одиниць, називається *генеральною сукупністю*. Відібрана з генеральної сукупності деяка частина одиниць, що піддається обстеженню, називається *вибірковою сукупністю* або просто *вибіркою* (рис. 5.1).

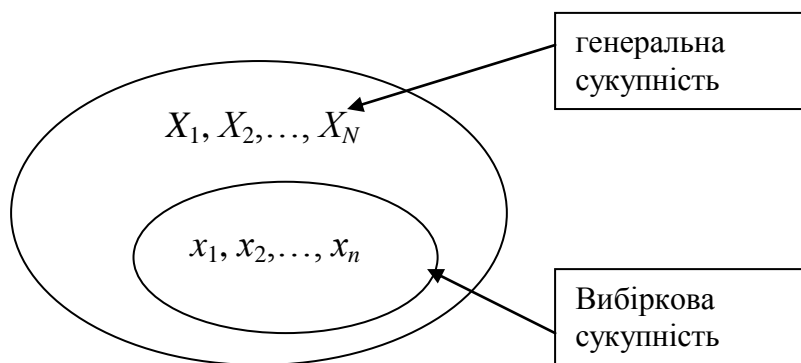


Рисунок 5.1 – Характеристики сукупності

Значення вибіркового методу полягає в тому, що при мінімальній чисельності обстежуваних одиниць здійснення дослідження відбувається в коротші терміни і з мінімальними витратами праці і коштів. Це підвищує оперативність отримання і достовірність статистичної інформації, зменшує помилки реєстрації.

У проведенні ряду досліджень вибіркоvim метод є єдино можливим, наприклад, при контролі якості продукції (товару), якщо перевірка супроводжується знищенням або розкладанням на складові частини обстежуваних зразків (визначення цукристості фруктів, клейковини печеного хліба, встановлення носкості взуття, міцності тканин на розрив і т. ін.).

Здійснення дослідження соціально-економічних явищ вибіркоvim методом складається з ряду послідовних етапів:

- 1) обґрунтування (відповідно до завдань дослідження) доцільності

застосування вибіркового методу;

2) складання програми проведення статистичного дослідження вибіркоvim методом;

3) вирішення організаційних питань збору і обробки початкової інформації;

4) встановлення частки вибірки, тобто частини тих одиниць генеральної сукупності, що підлягають обстеженню;

5) обґрунтування способів формування вибіркової сукупності;

6) здійснення відбору одиниць з генеральної сукупності для їх обстеження;

7) фіксація у відібраних одиницях (пробах) ознак, що вивчаються;

8) статистична обробка отриманої у вибірці інформації з визначенням узагальнювальних характеристик ознак, що вивчаються;

9) визначення кількісної оцінки помилки вибірки;

10) розповсюдження узагальнювальних вибіркових характеристик на генеральну сукупність.

У генеральній сукупності частка одиниць, що мають ознаку, яка підлягає вивченню, називається *генеральною часткою* (позначається як  $p$ ), а середня величина варіюючої ознаки, що вивчається, – *генеральною середньою* (позначається як  $\bar{x}$ ). У вибірковій сукупності частку ознаки, що вивчається, називають *вибірковою часткою*, або *частістю* (позначається  $\omega$ ), а середню величину у вибірці – *вибірковою середньою* (позначається як  $\bar{x}$ ).

*Приклад.* Під час контрольної перевірки якості хлібобулочних виробів здійснено 5 %-ве вибіркоче обстеження партії нарізних батонів з борошна вищого гатунку. При цьому з 100 відібраних у вибірку батонів 90 од. відповідали вимогам стандарту. Середня вага одного батона у вибірці складала 500,5 г при середньому квадратичному відхиленні  $p$ . На основі отриманих у вибірці даних потрібно встановити можливі значення частки стандартних виробів і середньої ваги одного виробу у всій партії.

Перш за все, встановлюються характеристики вибіркової сукупності. Вибіркова частка, або частість  $\omega$  визначається як відношення одиниць, що мають ознаку  $m$ , тобто ознаку, яка вивчається, до загальної чисельності одиниць вибіркової сукупності  $n$

$$\omega = \frac{m}{n}. \quad (5.1)$$

Оскільки з 100 виробів, що потрапили у вибірку  $n$ , 90 од. опинилися стандартними  $m$ , то показник частоти становить:  $\omega = 90 : 100 = 0,9$ .

Середню вагу виробу у виборці  $x = 500,5$  г визначено зважуванням, але отримані показники частоти (0,9) і середньої величини (500,5 г) характеризують частку стандартної продукції і середню вагу одного виробу лише у вибірці. Для визначення відповідних показників для всієї партії товару треба встановити можливі при цьому значення помилки вибірки.

Завдання вибіркового спостереження полягає в тому, щоб за вибірковими характеристиками оцінити відповідні генеральні характеристики. Система правил відбору одиниць і способів характеристики досліджуваної сукупності складає зміст вибіркового методу.

Розрізняють такі *способи формування* вибіркових сукупностей (рис. 5.2):

- випадковий (простий випадковий) відбір;
- механічний (систематичний);
- комбінований;
- серійний (гніздовий);
- типовий (районований);
- багатоступінчастий та багатофазний;

Усі види відбору (крім механічного) можуть бути проведені повторним чи неповторним методом, який визначає можливість для вже відібраної одиниці продовжувати участь в процедурі відбору. Механічний відбір завжди є неповторним.

*Повторним* називають відбір, при якому відібрана одиниця повертається назад у генеральну сукупність і може повторно брати участь у вибірці. При такому відборі шанси (ймовірність) кожної окремої одиниці потрапити у вибірку залишаються постійними.

*Безповторним* є відбір, у якому відібрана одиниця не повертається до генеральної сукупності і в подальшій вибірці участі не бере. В цьому випадку ймовірність одиниці потрапити до вибірки постійно змінюється – для одиниць, що залишилися, вона зростає.

При достатньо великому обсязі генеральної сукупності, оцінки її характеристик, отримані на основі неповторної вибірки, істотно не відрізняються від оцінок, отриманих на основі повторної вибірки.

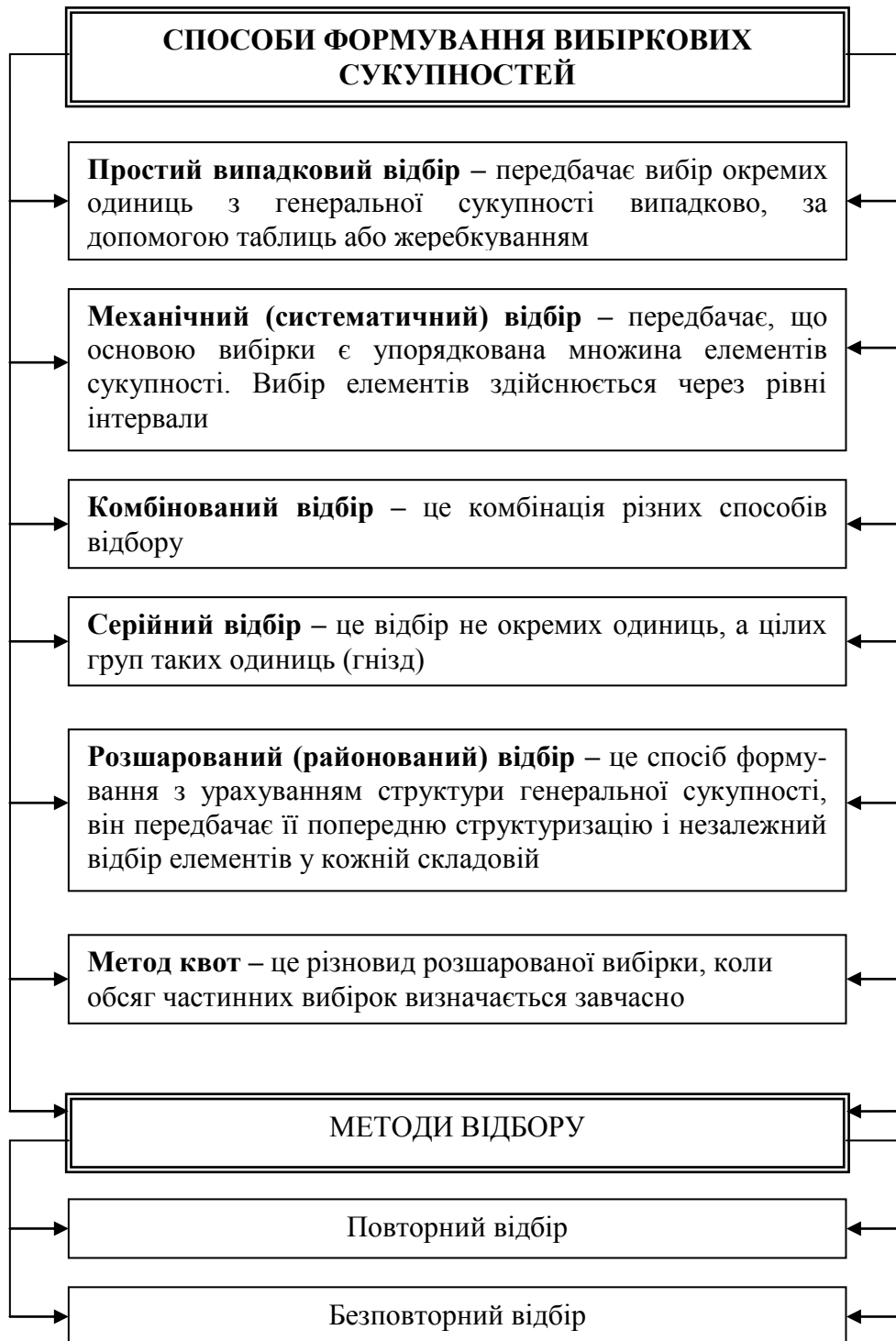


Рисунок 5.2 – Різновиди вибірок та методів

*Випадкова вибірка* – це класичний спосіб формування вибіркової сукупності, який передбачає попередню, досить складну підготовку до формування вибірки. Дає хороші результати за умов однорідності сукупності.

Простий випадковий відбір проводиться жеребкуванням або за допомогою таблиць випадкових чисел без будь-яких елементів системності (без попереднього розчленування генеральної сукупності на окремі групи). Для жеребкування на кожну одиницю генеральної сукупності необхідно заготовити відповідну фішку, шар або картку з інформацією щодо окремої одиниці сукупності. При використанні таблиць випадкових чисел усі елементи сукупності мають бути пронумеровані. Необхідно також встановити чіткі межі генеральної сукупності таким чином, щоб присутність чи відсутність у ній окремих одиниць не викликала сумніву. Так, наприклад, при обстеженні студентів необхідно вказати, чи будуть братися до уваги особи, що перебувають в академічній відпустці. У великих за обсягом сукупностях така робота здебільшого недоцільна, а часом і неможлива, тому, незважаючи на простоту власно випадкового відбору, на практиці застосовуються інші різновиди випадкових вибірок, які являють собою видозмінення простої випадкової вибірки.

Однією з найпоширеніших форм у практиці вибіркового спостереження є *механічний відбір*. Це послідовний вибір одиниць, який здійснюється за списками чи безпосередньо на місці дослідження через рівні проміжки в розташуванні цих одиниць у генеральній сукупності. Розташування одиниць сукупності буває впорядкованим (ранжованим) або неупорядкованим (випадковим) відносно досліджуваної ознаки. Наприклад, список співробітників підприємства в порядку зменшення заробітної плати або список студентів за алфавітом. Генеральна сукупність розбивається на рівні за чисельністю одиниць групи (інтервали) і з кожної групи для обстеження вибирається одна одиниця. При цьому задається початок, крок відліку і обсяг вибірки.

Початок відліку залежить від способу розташування одиниць в генеральній сукупності. У випадку їх неупорядкованого розташування він може бути встановлений довільно або жеребкуванням з сукупності одиниць першого інтервалу. При впорядкованому розташуванні за початок відліку беруть одиницю, яка розташована всередині першого інтервалу, щоб уникнути систематичної помилки вибірки.

Крок відліку визначається співвідношенням обсягів генеральної та вибіркової сукупностей  $h = N / n$ , наприклад, з генеральної сукупності обсягом 1000 од. Обстеженню підлягає 100 од. ( $1000/100=10$ ), це означає, що з кожних 10 од. буде обрано для обстеження 1 од. (кожну 10-ту), крок відліку дорівнює 10.

Механічний відбір дуже зручно застосовувати у випадках, коли не

можна заздалегідь скласти список одиниць генеральної сукупності. Наприклад, у маркетингових дослідженнях можна спостерігати кожного 10-го покупця, що увійшов до магазину; при дослідженні самостійної роботи студентів кожного 5-го відвідувача читального залу бібліотеки та комп'ютерного класу. В такому разі вибірка береться з сукупності, яка формується поступово чи з практично нескінченної сукупності.

У *комбінованих вибірках*, які є найбільш поширеними, поєднуються різні способи відбору. На практиці можна комбінувати типову і серійну вибірки, коли серії відбирають у встановленому порядку з декількох типових груп. Можлива також комбінація серійного і простого випадкового відборів, при якій окремі одиниці відбирають всередині серії у випадковому порядку. Комбінують також суцільне і вибіркове спостереження, де вибіркове уточнює, підтверджує результати суцільного чи є його пробним варіантом.

*Серійний відбір* зручно використовувати у тих випадках, коли одиниці сукупності об'єднані у невеликі групи чи серії, які розглядаються як одне ціле. Серії складаються з одиниць, які пов'язані або територіально (райони, селища), або організаційно (фірми, акціонерні товариства тощо). Як серії можуть розглядатися упаковки з визначеною кількістю готової продукції, партії товару, студентські групи, творчі колективи та інше.

Вибіркова сукупність серій формується за схемами механічної або простої випадкової неповторної вибірки. Якщо серія потрапила до вибірки, то обстежують усі без винятку елементи серії, тобто проводять суцільне їх обстеження, а результати розповсюджують на всю сукупність. Цей спосіб відбору застосовують при перевірці якості продукції того чи іншого підприємства, роботи відділу, при атестації ВНЗ (у відібраних групах студентів роблять суцільне вимірювання залишкових знань).

Застосовуючи серійну вибірку, слід звертати увагу на те, що випадкова похибка цієї вибірки виходить дещо більшою, ніж при інших способах відбору. Щоб цьому запобігти, необхідно збільшити обсяг вибірки.

*Типовий (розшарований, районований) відбір* орієнтований на забезпечення представництва у вибірці відповідних типових груп генеральної сукупності. Формування вибірки йде з урахуванням структури генеральної сукупності. На відміну від простого випадкового та механічного відборів, які проводяться в цілому за генеральною сукупністю,



типовий передбачає розділення (розшарування) її на якісно однорідні групи за якоюсь важливою типовою ознакою. Наприклад, при обстеженнях населення такими групами можуть бути райони, соціальні, вікові чи освітні групи. Потім з кожної групи здійснюють незалежний відбір елементів простим випадковим чи механічним способом пропорційно питомій вазі (обсягу) групи у загальній сукупності.

Перевагою типового відбору є отримання більш достовірних результатів порівняно з попередніми відборами. Оскільки до вибірки потрапляють представники усіх типових груп, ймовірність отримати більшу точність вибірових даних тут вища. Застосовувати цей метод відбору краще у разі неоднорідності одиниць досліджуваної сукупності.

*Багатоступінчастий відбір* характеризується тим, що на усіх етапах, крім останнього, проводиться лише відбір, а спостереження досліджуваних одиниць відбувається тільки на останньому етапі. Одиниці відбору на попередніх етапах – це зазвичай організаційні об'єднання одиниць спостереження, їх певні групи.

Отже, спочатку відбувається поступовий вибір з генеральної сукупності великих груп одиниць, потім груп, менших за обсягом і так доти, поки не оберуть ті одиниці, які підлягатимуть обстеженню. Етапів може бути два, три і більше. На різних етапах вибірки використовують різні одиниці і різні способи відбору. Наприклад, під час обстежень рівня знань студентів ВНЗ відбір населених пунктів, у яких є вищі навчальні заклади, можна здійснювати типовим способом, ВНЗ – теж типовим чи випадковим, курсів, груп – серійним чи механічним, студентів – механічним чи випадковим.

При *багатофазній вибірці* на усіх фазах зберігається однакова одиниця відбору, а кожна фаза відрізняється від попередньої обсягом вибірки та програмою обстеження. З генеральної сукупності формується первинна вибірка, а з первинної – підвибірка і т. д. На кожній наступній фазі обсяг підвибірки зменшується, а програма обстеження розширюється. Наприклад, під час дослідження ставлення студентів до організації навчального процесу в університеті на 1-й фазі досліджують 25 % від загальної чисельності студентів за короткою програмою (три запитання), на 2-й фазі – 15 % за ширшою програмою, яка містить ще два додаткових запитання, на 3-й фазі – 5 % за повною програмою з сьома запитань. Вибіркові оцінки кожної фази використовуються як додаткова інформація на наступних фазах, що підвищує точність результатів і є важливою властивістю багатофазної вибірки. Похибки багатоступінчастої та

багатофазної вибірок розраховують окремо для кожного етапу чи фази.

## **5.2. Способи відбору одиниць генеральної сукупності**

Залежно від того, скільки разів відібрані для обстеження одиниці беруть участь у відборі, розрізняють повторний та неповторний відбори. При повторному відборі обстежені одиниці «повертаються» у генеральну сукупність і знову беруть участь у відборі. При неповторному відборі одиниці, що потрапили у вибірку, більше не беруть участі у відборі. Таким чином, кожна одиниця може бути відбранаю лише один раз. При формуванні вибіркової сукупності використовують також наступні види відбору:

- індивідуальний, при якому у вибірку сукупність вибирають по одній одиниці з генеральної сукупності;
- груповий або серійний, при якому вибирається група (серія) одиниць;
- комбінований, тобто такий що поєднує перші два види відбору.

Розрізняють *чотири основних способи формування вибіркової сукупності*:

1) власне випадковий відбір (повторний чи неповторний), при якому вибірка сукупність формується виключно випадково (методом жеребкування, за таблицями чисел тощо);

2) механічний (систематичний) відбір, при якому у вибірку сукупність потрапляють одиниці з певними порядковими номерами. При цьому всі одиниці генеральної сукупності спочатку впорядковуються та їм присвоюються порядкові номери. Далі визначається пропорція відбору та крок. Наприклад, пропорція відбору  $1/20$ , отже, крок (різниця між порядковими номерами) становить 20. Далі з першої групи випадковим чином визначається перший порядковий номер, а наступні – шляхом додавання кроку відбору. Наприклад, з перших 20 одиниць обрано 7-му, тоді наступні одиниці – 27, 47, 67 і т. д. Цей спосіб відбору є неповторним.

3) типовий відбір, який передбачає, що генеральна сукупність поділяється на однорідні групи і з кожної групи випадковим або механічним способом формується вибірка сукупність. Якщо з кожної групи відбирається однаковий процент одиниць, то типовий відбір називається пропорційним, а якщо однакова кількість одиниць – непропорційним. Типовий відбір може бути повторним і неповторним.

4) серійний відбір, при якому у вибірку сукупність відбираються групи одиниць (серії), які надалі обстежуються суцільно.

У статистичній практиці застосовується відбір у часі, наприклад моментне спостереження, що передбачає реєстрацію ознак на певний момент часу, як правило, через рівні інтервали.

Уявлення про ймовірність подій лежать в основі одного з найвідоміших методів формування вибірки – *методу простого ймовірного відбору*. Ймовірність – це відношення кількості подій, на які сподіваються, до кількості можливих.

*Приклад.* Загальна чисельність робітників цеху становить 1000 осіб; з них 200 – висококваліфіковані, 300 – середньокваліфіковані, 500 – низькокваліфіковані. Для опитувань необхідно відібрати 100 осіб, які репрезентують думку загалу. Що у цьому разі треба робити? Уявимо 1000 кульок, із яких 200 – червоні, 300 – чорні, а 500 – білі. Для забезпечення рівноймовірності відбору кожної кульки їх треба, поклавши в урну, перемішати. Але при цьому можливі варіанти: а) кожну відібрану кульку відкладають, потім беруть іншу і т. д., тобто йдеться про ймовірнісно-безповторний відбір; б) вийняту із урни кульку фіксують і знову кладуть до урни для участі у повторному відборі. Це ймовірнісно-повторювальний вибір. Замість кульок можна використати картки, фішки, жетони тощо.

Серед різновидів простого ймовірного відбору достатньо поширеними є такі форми вибірки: *вибірка «першого зустрічного»*; *поштове опитування*; *метод «снігової кулі»*. Наприклад, методом опитувань першого зустрічного досить часто користуються журналісти теле- і радіомовлення. Нерідко вдаються до опитувальних листів у періодичних виданнях (поштове опитування). Але цей спосіб формування вибірки недоцільно вважати рівноймовірним. Інтерв'юер, наприклад, підсвідомо надає перевагу певному типу респондентів, а різні верстви населення – певним засобам інформації, тобто порушується принцип рівноймовірності подій. Зібрана у такий спосіб інформація, безумовно, відображає погляди та думки окремих верств населення, але вибірка за таких умов має характер стихійної.

Метод «снігової кулі» використовують тоді, коли потрібно, скажімо, опитати декілька сотень представників громадських організацій, а реальна можливість дає змогу провести інтерв'ю лише з десятима. Тоді пошук інших респондентів можна вести за допомогою цих десяти. Припустимо, що кожен з них погодився повідомити додатково двох

своїх знайомих, ті – ще по двоє і т. д. Таким чином, вже на п'ятому етапі вибіркова сукупність становитиме:  $10+20+40+80+160=310$ .

Перевага методу простого ймовірного відбору та його різновидів полягає ще й у тому, що він достатньо простий, тобто не потребує глибокого методологічного аналізу, а також доступний в реалізації. Для формування вибірки цим методом достатньо мати спрощену вихідну попередню інформацію про генеральну сукупність (перелік, список чи опис її основних елементів).

Однак недоліків у цього методу, мабуть, значно більше: успішне його використання можливе тільки для генеральних сукупностей не менше 1000 одиниць; виготовлення ж карток чи куль для масивів із 10, 100 і більше тисяч одиниць ускладнює процедуру і робить її нерентабельною. А зусилля на складання та доопрацювання переліків, списків на великі сукупності здебільшого не виправдовуються отриманням лише формальних ознак вибірки.

Наступним є *метод систематичного (механічного) відбору*. Сутність його полягає у визначенні так званого кроку вибірки. При цьому усі елементи генеральної сукупності зводяться до загального списку (переліку), з нього через рівні інтервали («кроки») відбирають групу респондентів. Крок вибірки розраховують за формулою

$$K = M / H, \quad (5.2)$$

де  $K$  – крок вибірки;  $M$  – генеральна сукупність;  $H$  – вибіркова сукупність.

Однією з головних переваг цього методу є достатньо проста техніка відбору, яка за невеликої вибірки дає змогу відобразити параметри великих і навіть дуже великих генеральних сукупностей. Серед недоліків – феномен зміщення ознак генеральної сукупності. Тоді інтервали відбору з невизначеною періодичністю можуть деформувати модель.

Три наступні вибірки – *гніздова, серійна та стратифікована* – пов'язані з поняттям районування генеральної сукупності. Районування *генеральної сукупності* – процес розподілу досліджуваного об'єкта на складові частини відповідно до мети та завдань дослідження. Якісні ознаки складових генеральної сукупності визначають особливості вибірки.

У гніздовій вибірці за одиницю відбору беруть групи чи колективи, далі проводять суцільне опитування в кожному з них.

Наприклад, досліджуючи колективи навчальних закладів, за одиницю беруть учнівські класи, студентські групи та ін. Зручність, доступність та простота відокремлення гнізд роблять цей метод раціональним та економним. Але ця доступність багато в чому підступна: формальний підхід до відокремлення обмеженої кількості одиниць (гнізд) може призвести до спотворення якісної картини досліджуваного об'єкта.

Метод серійної вибірки також належить до простих схем відбору, а покладені в його основу вимоги виділення замість гнізд статистичних серій роблять його гнучкішим. Оскільки під серією розуміють сукупність статистично відмінних одиниць, дослідник може будувати вибірку з урахуванням детальних ознак структурної організації об'єкта.

*Приклад.* Припустимо, що в досліджуваній генеральній сукупності 2000 осіб, із них 300 – висококваліфіковані робітники, 1000 – мають середні розряди, а 700 – низькокваліфіковані. Якщо до вибірки планується включити 200 осіб, то із кожної серії слід відібрати необхідну кількість згідно з формулою

$$P = C_1 \cdot H / M, \quad (5.3)$$

де  $P$  – кількість респондентів для відбору в кожній із серій;  $C_1$  – показник серії;  $M$  – генеральна сукупність;  $H$  – вибіркова сукупність.

Загальний показник вибіркової сукупності розраховують за формулою

$$P = P_1 + P_2 + P_3. \quad (5.4)$$

Для нашого прикладу маємо  $P = 30 + 100 + 70 = 200$  осіб.

При формуванні стратифікованої вибірки вихідною одиницею є однорідна частина генеральної сукупності, аналіз якої проводять за однією чи кількома ознаками. Оскільки коливання ознак всередині однорідної страти менші, то виділені одиниці будуть якісно однорідними, тобто репрезентативно спроможними. Страти можна виділяти за однією, двома чи більше ознаками, пов'язаними з історичними, природними, психічними, економічними та іншими аспектами діяльності досліджуваного об'єкта.

Процедура формування стратифікованої вибірки складається з таких етапів:

а) районування вибірки;

б) стратифікація вибірки.

Спочатку необхідно провести попередній розподіл генеральної сукупності на страти відповідно до визначених критеріїв, а потім вже з кожної страти сформувати вибірку, тобто здійснити так званий стратифікований відбір. Якщо обсяг вибірки із страти збігається з розміром страти генеральної сукупності, тоді йдеться про стратифіковану вибірку з пропорційним розташуванням. Якщо обсяг вибірки із страти пропорційний у страті середньоквадратичному відхиленню ознаки і обернено пропорційний витратам на формування вибірки, то це стратифікована вибірка з оптимальним розміщенням.

Вдалий розподіл сукупності на однорідні групи дає змогу відобразити специфіку змістово-сенсових параметрів досліджуваного об'єкта. Якщо страти не збігаються з фактичним розподілом досліджуваних ознак, то це призводить до зсунення вибірки.

*Метод квотної вибірки* є одним з найточніших і найпоширеніших у соціологічних дослідженнях. Застосовують його, як правило, коли до початку дослідження є відомими широкі дані про контрольні ознаки елементів генеральної сукупності. Провідні поняття, які використовуються при цьому, – «квота» та «параметр квоти». У перекладі з латини квота означає «частина», «частка», «пропорція» або «норма» і трактується як сукупність даних про конкретну контрольну ознаку. Параметр квоти – окремі показники контрольованої ознаки, які можуть бути подані у формі чисел, понять чи знаків. Задають квоти за залежними (взаємопов'язаними) та незалежними параметрами. Чисельність квотних ознак не повинна бути великою. На практиці вона не перевищує 2-3 показники.

До переваг квотної вибірки належать доступність та зручність її структури, незначні витрати на збирання інформації, а також короткі терміни її формування. До недоліків – неможливість точного виміру відхилень, зумовлених процесом цілеспрямованого відбору одиниць.

Формування вибірки методом багатоступеневого опрацювання даних генеральної сукупності відносять до складних форм роботи з досліджуваним об'єктом. Процес її формування охоплює декілька відносно самостійних етапів, на кожному з яких використовують той чи інший із раніше розглянутих методів. Так, при дослідженні складних об'єктів на першому етапі формування вибірки проводять відбір великих одиниць (міст, районів, підприємств чи установ) гніздовим методом, на другому етапі можливе квотування тих чи інших ознак, а на завер-

шальному – ймовірний відбір респондентів за заданими параметрами.

Розрізняють випадкові та комбіновані багатоступеневі вибірки. Якщо в багатоступеневих вибірках вихідні одиниці визначаються одним із методів ймовірного відбору, то їх відносять до розряду випадкових, а якщо іншими методами (крок, серія, квота) – до комбінованих. Ця ознака має суттєві відмінності, які так чи інакше впливають на зміст дій та відповідних операцій при формуванні вибірок.

### **5.3. Визначення помилки вибіркової середньої**

Між характеристиками вибіркової сукупності та аналітичними характеристиками (параметрами) генеральної сукупності, як правило, існують деякі розбіжності, які називаються *помилками*. Загальна величина можливої помилки вибіркової характеристики складається з помилок двоякого роду: помилки реєстрації і помилки репрезентативності.

*Помилки реєстрації* властиві будь-якому статистичному спостереженню взагалі. Порівняно з суцільними спостереженнями небезпека виникнення помилок реєстрації у вибіркових спостереженнях повинна бути меншою завдяки більш кваліфікованому і ретельному здійсненню вибірки.

*Помилки репрезентативності* виникають від того, що вибірка не точно відтворює склад генеральної сукупності, оскільки вона є випадковою величиною, тому і оцінки характеристик є випадковими величинами, які можуть не збігатися з відповідними параметрами генеральної сукупності. Через виникнення помилки репрезентативності поділяються на систематичні та випадкові.

*Систематичні (тенденційні)* помилки репрезентативності виникають при порушенні принципу випадковості відбору (упереджений вибір елементів, недосконалий вибір одиниці відбору тощо). Систематичні помилки мають односторонній, тенденційний напрям і призводять до зсунення (викривлення) результатів обстеження в той чи інший бік. Їх ще визначають як помилки зміщення (зсунення).

*Приклад.* Якщо для визначення середньої успішності потоку студентів взяти вибірку студентів, які сидять на останніх партах, то середня успішність вибірки буде меншою за середню успішність всього потоку. Якщо вибірку сформувати зі студентів, які сидять в перших рядах, то середня успішність у такій вибірці буде вищою від середньої успішності потоку. При такому нерівномірному відборі статистична оцінка успіш-

ності потоку буде заздалегідь занижена або завищена.

Систематичних помилок важливо уникати, оскільки незсуненість – одна з вимог до будь-якої вибіркової оцінки.

*Випадкові* помилки репрезентативності зумовлені тим, що розподіл відібраної сукупності одиниць недостатньо точно відтворює розподіл одиниць генеральної сукупності. На відміну від систематичних, випадкові помилки властиві вибіркового спостереженню та неминуче виникають і при додержанні принципу випадковості відбору, але вони не носять тенденційного характеру і не ведуть до зміщення оцінок. Випадкові помилки репрезентативності *залежать від* таких факторів:

- способу відбору;
- обсягу вибірки (обернено пропорційна залежність) – чим більше одиниць відібрано, тим менша помилка (але слід звернути увагу, що це стосується лише репрезентативної вибірки!); у випадку зсуненої вибірки навпаки, збільшення її обсягу збільшує ймовірність отримання помилкових висновків;
- ступеня варіації досліджуваної ознаки у генеральній сукупності (прямо пропорційна залежність) – чим більша варіація ознаки, тим більша у середньому і помилка вибірки.

Визначення можливої і фактично допущеної помилки вибірки (випадкової) має велике значення при застосуванні вибіркового методу. Величина помилки характеризує ступінь надійності результатів вибірки і є необхідною для оцінки параметрів генеральної сукупності.

Для однієї конкретної вибірки можна визначити:

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

*Помилка вибірки* – це об'єктивно виникаюча розбіжність між характеристиками вибірки і генеральної сукупності. Вона залежить від ряду чинників: ступеня варіації ознаки, що вивчається; чисельності вибірки; методу відбору одиниць у вибіркочуву сукупність; прийнятого рівня достовірності результату дослідження.

При випадковому повторному відборі *середня (стандартна) похибка вибірки*  $\mu$  є середнім квадратичним відхиленням вибіркових оцінок від значення параметра в генеральній сукупності і обчислюється за формулою



$$\mu = \sqrt{\frac{\sigma^2}{n}}, \quad (5.5)$$

де  $\mu$  – середня помилка вибіркової середньої;  $\sigma^2$  – генеральна дисперсія (на практиці замінюється вибірковою дисперсією – незсуненою оцінкою),  $n$  – обсяг вибіркової сукупності.

При безповторному відборі вона розраховується за формулою

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}, \quad (5.6)$$

де  $N$  – чисельність генеральної сукупності.

Отже, середня помилка вибірки для середньої величини збільшується із зростанням варіації генеральної сукупності і зменшується із збільшенням обсягу вибірки. Для вирішення практичних завдань обчислення тільки середньої помилки вибірки недостатньо, тому визначають граничний для певної ймовірності розмір (інтервал) вибіркової помилки.

*Визначення помилки вибіркової частки.* При повторному відборі середня помилка вибіркової частки розраховується за формулою

$$\mu = \sqrt{\frac{\omega(1-\omega)}{n}}, \quad (5.7)$$

де  $\omega = \frac{m}{n}$  – вибіркова частка одиниць, які мають ознаку, що вивчається;  $m$  – число одиниць, які мають ознаку, що вивчається;  $n$  – чисельність вибірки.

При безповторному способі відбору середня помилка вибіркової частки визначається за формулою

$$\mu = \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}. \quad (5.8)$$

*Гранична помилка вибірки  $\Delta$*  пов'язана з середньою помилкою вибірки  $\mu$  відношенням

$$\Delta = t \cdot \mu, \quad (5.9)$$

де  $t$  – нормоване відхилення – “коефіцієнт довіри”, залежний від вірогідності, з якою гарантується гранична помилка вибірки;  $\mu$  – середня помилка вибірки.

При цьому  $t$  як коефіцієнт кратності середньої помилки вибірки залежить від значення вірогідності  $P$ , з якою гарантується величина граничної помилки вибірки.

Межі довірчого інтервалу визначаються на основі точкової оцінки та *граничної помилки вибірки*. Гранична помилка дає можливість з’ясувати, у яких межах знаходиться величина генеральної середньої

$$X_B - \Delta \leq X_G \leq X_B + \Delta. \quad (5.10)$$

Значення ймовірності, з якою можна стверджувати, що помилка  $\Delta$  не перевищує величини середньої помилки вибірки  $\mu$  для різних значень  $t$ , обчислені та подані у спеціальних математичних таблицях. На практиці найчастіше використовують ймовірність 0,95 та 0,954 (відповідні їм величини  $t$  становлять 1,96 та 2,00). Відбір вважається задовільним, якщо гранична помилка репрезентативності потрапляє у межі  $\pm 2-5\%$ .

Наведені вище формули застосовують при простому випадковому та механічному відборі у разі повторної вибірки. Для інших видів відбору застосовують аналогічні формули з необхідними поправками на метод.

Таким чином, гранична помилка вибірки при безповторному відборі визначається за наступними формулами:

- для визначення похибки вибіркової частки

$$\Delta_{\omega} = t \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}; \quad (5.11)$$

- для визначення середньої похибки вибірки

$$\Delta_x = t \sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)}. \quad (5.12)$$

Гранична помилка вибірки при повторному відборі визначається за формулами:

- для визначення похибки вибіркової частки

$$\Delta_\omega = t \sqrt{\frac{\omega(1-\omega)}{n}}; \quad (5.13)$$

- для визначення середньої похибки вибірки

$$\Delta_x = t \sqrt{\frac{\sigma_x^2}{n}}. \quad (5.14)$$

*Обсяг вибірки  $n$*  визначають до проведення вибіркового спостереження, застосовуючи алгебраїчне перетворення формули граничної похибки вибірки, звідки

$$n = \frac{t^2 \cdot \sigma^2}{\Delta^2}. \quad (5.15)$$

Отже, достатній обсяг вибірки залежить від ступеня однорідності генеральної сукупності  $\sigma^2$ , ймовірності, з якою гарантується результат  $t^2$  і необхідної точності вибіркової оцінки  $\Delta^2$ . На практиці ці значення встановлюють приблизно на основі попередніх або аналогічних досліджень.

Для поширення вибірових характеристик на генеральну сукупність необхідно спочатку оцінити результати спостереження з точки зору самої можливості поширення.

*Істотність* вибірових середніх, часток і дисперсій можна оцінити методом статистичної перевірки гіпотез, який дає можливість визначити ймовірність прийняття помилкового рішення на основі відповідних критеріїв (правил) узгодженості. Для цього необхідно:

- сформулювати нульову гіпотезу  $H_0$  – припущення про те, що вибіркова сукупність узгоджується з генеральною, і відповідну до неї альтернативну гіпотезу  $H_1$ ;

– обрати рівень істотності  $\alpha$  (ймовірність ризику відхилення нульової гіпотези, найчастіше береться таким, що дорівнює 0,05 або 0,01) та відповідної до нього ймовірності  $(1 - \alpha)$ ;

– визначити межі довірчих значень (при яких запропонована гіпотеза  $H_0$  приймається) за відповідними формулами критеріальних статистик та критичні межі (при яких нульова гіпотеза відхиляється);

– прийняти те чи інше рішення на основі порівняння фактичного та критичного (табличного) значень критерію. При нормальному розподілі ознаки у генеральній сукупності використовують таблиці Лапласа.

Отже, якщо вибіркові дані суперечать гіпотезі  $H_0$  (фактичне значення критерію перевищує критичне), вона відхиляється. Коли ці дані узгоджуються з гіпотезою  $H_0$ , вона приймається і результати вибіркового спостереження можна поширити на генеральну сукупність.

Існує декілька способів поширення вибіркових даних:

– спосіб прямого розрахунку – полягає у визначенні показників обсягу генеральної сукупності: середній розмір ознаки, обчислений в результаті вибіркового спостереження, треба помножити на чисельність одиниць генеральної сукупності;

– метод поправочних коефіцієнтів – застосовують тоді, коли вибірконе спостереження проводять з метою перевірки та уточнення результатів суцільного спостереження: дані суцільного та вибіркового спостережень зіставляють і обчислюють поправочні коефіцієнти.

**УВАГА!** Оприлюднюючи статистичні дані, які обчислені на основі вибірки, обов'язково слід пояснювати, що відповідні показники отримано розрахунковим способом на базі матеріалів вибіркового спостереження!

#### **5.4. Мала вибірка**

У практиці статистичного дослідження іноді доводиться зустрічатися з невеликими за обсягом, так званими малими вибірками. Під *малою вибіркою* розуміється несцільне статистичне обстеження, при якому вибірка сукупність утворюється з порівняно невеликої кількості одиниць генеральної сукупності. Обсяг малої вибірки зазвичай не перевищує 30 одиниць і може доходити до 4-5 одиниць. Хоча загальний принцип вибіркового обстеження (із збільшенням обсягу вибірки підвищується точність вибіркових даних) лишається в силі, іноді доводиться обмежуватися малою кількістю спостережень. Часто така

необхідність здійснення експерименту на основі малої вибірки в економічних дослідженнях виникає при вибірковій перевірці якості продукції, в науково-дослідній роботі та в інших випадках.

Розробка теорії малої вибірки була розпочата англійським статистиком В. С. Госсетом, який друкувався під псевдонімом Стьюдент. Ним було доведено, що і при малих вибірках характеристики вибіркової сукупності можна застосовувати для оцінки параметрів генеральної сукупності, але за певних умов. Такими умовами є нормальний розподіл ознаки у генеральній сукупності. Для оцінки результатів малої вибірки та визначення можливих границь її випадкової помилки користуються *критерієм Стьюдента*. Однак це не означає, що можна використовувати малу вибірку тоді, коли необхідна велика. В багатьох випадках розбіжності між знайденими межами можуть досягати значних розмірів, що навряд задовольнить дослідників. Тому малу вибірку слід застосовувати в статистичному дослідженні дуже обережно, з відповідним теоретичним та практичним обґрунтуванням. Необхідно також взяти до уваги і те, що точність результатів вибірки малого обсягу все ж нижча, ніж у великій вибірці.

Середня помилка малої вибірки  $\mu_{\text{м.в}}$  обчислюється за формулою

$$\mu_{\text{м.в}} \approx \sqrt{\frac{\sigma_{\text{м.в}}^2}{n}}, \quad (5.16)$$

де  $\sigma_{\text{м.в}}^2$  – дисперсія малої вибірки.

При визначенні дисперсії  $\sigma_{\text{м.в}}^2$  число ступенів свободи дорівнює  $n-1$

$$\sigma_{\text{м.в}}^2 = \frac{\sum (x_i - \tilde{x})^2}{n-1}. \quad (5.17)$$

Гранична помилка малої вибірки  $\Delta_{\text{м.в}}$  визначається за формулою

$$\Delta_{\text{м.в}} = t \cdot \mu_{\text{м.в}}. \quad (5.18)$$

При цьому значення коефіцієнта довіри  $t$  залежить не тільки від заданої довірчої вірогідності, але і від чисельності одиниць вибірки  $n$ .

Для окремих значень  $t$  і  $n$  довірча вірогідність малої вибірки визначається по спеціальних таблицях Стьюдента (частково наведено в табл. 5.1), в яких наведено дані розподілу стандартизованих відхилень

$$t = \frac{\tilde{x} - \bar{x}}{\sigma_{\text{м.в}}}. \quad (5.19)$$

Таблиця 5.1 – Таблиця Стьюдента

$n$	$t$				
	0,5	1,0	1,5	2,0	3,0
4	0,347	0,609	0,769	0,861	0,942
6	0,362	0,637	0,806	0,898	0,970
8	0,368	0,649	0,823	0,914	0,980
10	0,371	0,657	0,832	0,923	0,985
15	0,376	0,666	0,846	0,936	0,992
20	0,377	0,670	0,850	0,940	0,993

Оскільки при здійсненні малої вибірки довірча вірогідність практично набуває значення 0,95 або 0,99, то для визначення граничної помилки малої вибірки  $\Delta_{\text{м.в}}$  використовуються наступні дані розподілу Стьюдента (табл. 5.2)

Таблиця 5.2 – Дані розподілу Стьюдента

$n$	$S_t$	
	0,95	0,99
4	3,183	5,841
5	2,777	4,604
6	2,571	4,032
7	2,447	3,707
8	2,364	3,500
9	2,307	3,356
10	2,263	3,250
15	2,119	2,921
20	2,078	2,832

*Приклад.* Під час контрольної перевірки якості поставленої в торгівлю ковбаси отримані дані про вміст куховарської солі в пробах. За даними вибіркового обстеження потрібно встановити з вірогідністю 0,95 межу, в якій знаходиться середній відсоток вмісту куховарської солі в даній партії товару. Складаємо розрахункову таблицю (табл. 5.3) і за її підсумками визначаємо середню пробу малої вибірки:

Таблиця 5.3 – Вихідні дані для розрахунку прикладу

Проби	$x_i - \tilde{x}$	$(x_i - \tilde{x})^2$
4,3	0,2	0,04
4,2	0,1	0,01
3,8	0,3	0,09
4,3	0,2	0,04
3,7	-0,4	0,16
3,9	-0,2	0,04
4,5	0,4	0,16
4,4	0,3	0,09
4,0	-0,1	0,01
3,9	-0,2	0,04
$\sum$ 41,0	-	0,68

$$\tilde{x} = \frac{\sum x_i}{n} = \frac{41}{10} = 4,1.$$

Визначаємо дисперсію малої вибірки

$$\sigma_{\text{м.в.}}^2 = \frac{\sum (x_i - \tilde{x})^2}{n-1} = \frac{0,68}{10-1} = 0,075.$$

Визначаємо середню помилку малої вибірки

$$\mu_{\text{м.в.}} \approx \sqrt{\frac{\sigma_{\text{м.в.}}^2}{n}} = \sqrt{\frac{0,075}{10}} = \pm 0,087.$$

Виходячи з чисельності вибірки ( $n=10$ ) і заданої вірогідності  $S_i=0,95$ , встановлюємо за розподілом Стьюдента (див. табл. 5.2.) значення коефіцієнта довіри:  $t=2,263$ .

Гранична помилка малої вибірки становить:

$$\Delta_{\text{м.в.}} = t \cdot \mu_{\text{м.в.}} = 2,263 \cdot (\pm 0,087) \approx \pm 0,2.$$

Отже, з вірогідністю 0,95 можна стверджувати, що у всій партії ковбаси вміст куховарської солі знаходиться в таких межах:

$$\bar{x} = \tilde{x} \pm \Delta_{\text{м.в.}} = 4,1 \pm 0,2.$$

Тобто в межах від 3,9 до 4,3.

### 5.5. Способи розрахунків характеристик вибірки

При нормальному розподілі перевірка гіпотез здійснюється з використанням таблиць Лапласа. Якщо вибірка малого обсягу, то для перевірки критерію використовується  $t$ -розподіл Стьюдента. Порівнюючи

вибірки парами, цей критерій можна застосовувати для будь-якої кількості вибірок.

*Процедура перевірки гіпотез використовується* при порівнянні вибірових характеристик (середньої, частки, дисперсії) з відповідними нормативами, порівнянні характеристик двох вибірових сукупностей, оцінюванні істотності розбіжностей двох розподілів, у регресійно-кореляційному аналізі.

Способи розрахунків характеристик вибірки застосовується для дослідження деякої множини чи загальної сукупності предметів, елементів, подій або явищ з метою визначення їх кількісних характеристик, таких, як *частка* елементів, що мають певні властивості, *середня величина* деякого виміру, *частота* будь-якої події.

Елементи вибірового спостереження наведено в табл. 5.4.

Таблиця 5.4 – Елементи вибірового спостереження

Характеристики (параметри та оцінки)	Генеральна сукупність (досліджувана сукупність, з якої проводять добір у вибірку)	Вибіркова сукупність (яка безпосередньо обстежується)
Обсяг сукупності (чисельність одиниць)	$N$	$n$
Кількість одиниць, які мають певне значення досліджуваної ознаки (частота)	$M$	$m$
Частка одиниць, які мають певне значення досліджуваної ознаки, %	$P = \frac{M}{N} \cdot 100 \%$	$w = \frac{m}{n} \cdot 100 \%$
Середнє значення ознаки	$\tilde{x} = \frac{\sum x_i}{N}$	$\tilde{x} = \frac{\sum x_i}{n}$
Дисперсія (міра варіації) кількісної ознаки $\sigma_{\text{ген}}^2 \approx \sigma_{\text{вибірки}}^2$	$\sigma_{\tilde{x}}^2 = \frac{\sum (x_i - \tilde{x})^2}{N}$	$\sigma_{\tilde{x}}^2 = \frac{\sum (x_i - \tilde{x})^2}{n}$
Дисперсія частки альтернативної (якісної) ознаки, де $q$ та $(1 - w)$ є часткою одиниць, яким не властива певна ознака $\sigma_{\text{частки}}^2 \leq 0,25$	$\sigma_P^2 = P \cdot q$	$\sigma_w^2 = w \cdot (1 - w)$

*Похибки вибірки* – це розбіжності між характеристиками генеральної та вибірової сукупностей. Виникають при порушенні принципів формування вибірки (систематичні помилки) та внаслідок випадковості відбору одиниць (випадкові помилки), при якому вибірка не відтворює точно структуру та показники генеральної сукупності.

Завданням вибірового методу є обчислення величин *випадкових*



помилки: стандартної (середньої) помилки  $\mu$  та граничної помилки  $\Delta = t \cdot \mu$  відповідно до методу та схеми відбору;  $t$  – коефіцієнт довіри, що залежить від ймовірності, з якою гарантується розмір граничної помилки. Найчастіше використовують наступні рівні ймовірності (табл. 5.5).

Таблиця 5.5 – Рівні ймовірності

Рівень ймовірності $p$	0,683	0,954	0,997	0,999
Коефіцієнт довіри $t$	1	2	3	4

Розрахункові формули граничних помилок наведено в табл. 5.6.

Таблиця 5.6 – Розрахункові формули граничних помилок

Метод відбору	Схема відбору	При визначенні середньої	При визначенні частки
Випадковий (за жеребом) і механічний (відбір через рівні інтервали)	Повторна	$\Delta_x = t \cdot \sqrt{\frac{\sigma^2}{n}}$	$\Delta_w = t \cdot \sqrt{\frac{w(1-w)}{n}}$
	Безповторна	$\Delta_x = t \cdot \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta_w = t \cdot \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Типовий (враховує структуру генеральної сукупності)	Повторна Безповторна	Замість загальної дисперсії $\sigma^2$ у відповідних формулах використовують середню з групових дисперсій $\bar{\sigma}^2 = \sum \sigma_i^2 / n$	Замість загальної дисперсії $w(1-w)$ використовують середню з групових дисперсій $\overline{w(1-w)} = \sum w_i(1-w_i) / n$
Серійний (відбираються не окремі одиниці, а цілі групи одиниць – серії)	Безповторна	$\Delta_x = t \cdot \sqrt{\frac{\delta_x^2}{s}}$ , де $\delta_x^2 = \frac{\sum (x_i - \tilde{x})^2}{s}$ – міжсерійна (міжгрупова) дисперсія	$\Delta_w = t \cdot \sqrt{\frac{\delta_w^2}{s}}$ , де $\delta_x^2 = \sigma^2 - \bar{\sigma}^2$ – різниця між загальною дисперсією та середньою з внутрішньо серійних дисперсій часток
	Безповторна	$\Delta_x = t \cdot \sqrt{\frac{\sigma_x^2}{s} \left(1 - \frac{s}{S}\right)}$ , де $s$ – число відібраних у вибірку серій	$\Delta_w = t \cdot \sqrt{\frac{\sigma_w^2}{s} \left(1 - \frac{s}{S}\right)}$ , де $S$ – число серій у генеральній сукупності

Гранична похибка дає можливість встановити межі довірчого інтервалу, тобто ті межі, в яких знаходиться величина генеральної середньої та частки:

– для середньої

$$\Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x \quad \text{або} \quad \tilde{x} - \bar{x} = \pm \Delta_x; \quad (5.20)$$

– для частки

$$x - \Delta_w \leq \bar{x} \leq \tilde{x} + \Delta_w \quad \text{або} \quad \tilde{x} - \bar{x} = \pm \Delta_w \quad (5.21)$$

Чим менший довірчий інтервал, тим точніша вибіркова оцінка.

При організації вибірки важливе значення має визначення необхідної її чисельності  $n$ . Остання залежить від варіації одиниць обстежуваної сукупності. Чим більша коливність, тим більшою повинна бути чисельність вибірки. Зворотний зв'язок існує між чисельністю вибірки та її граничною помилкою. Прагнення отримати меншу помилку вимагає збільшення чисельності вибіркової сукупності.

Формули визначення достатнього обсягу вибірки наведено в табл. 5.7.

Таблиця 5.7 – Формули визначення достатнього обсягу вибірки

Метод відбору	Схема відбору	При визначенні середньої	При визначенні частки
Випадковий і механічний	Повторна	$n = \frac{t^2 \cdot \sigma_x^2}{\Delta_x^2}$	$n = \frac{t^2 w(1-w)}{\Delta_w^2}$
	Безповторна	$n = \frac{Nt^2 \cdot \sigma_x^2}{N\Delta_x^2 + t^2 \sigma_x^2}$	$n = \frac{Nt^2 \cdot w(1-w)}{N\Delta_w^2 + t^2 w(1-w)}$
Типовий	Повторна	$n = \frac{t^2 \cdot \overline{\sigma_x^2}}{\Delta_x^2}$	$n = \frac{t^2 w(1-w)}{\Delta_w^2}$
	Безповторна	$n = \frac{Nt^2 \cdot \overline{\sigma_x^2}}{N\Delta_x^2 + t^2 \overline{\sigma_x^2}}$	$n = \frac{Nt^2 \cdot \overline{w(1-w)}}{N\Delta_w^2 + t^2 \overline{w(1-w)}}$
Серійний	Повторна	$s = \frac{t^2 \cdot \delta_x^2}{\Delta_x^2}$	$s = \frac{t^2 w_s(1-w_s)}{\Delta_w^2}$
	Безповторна	$s = \frac{St^2 \cdot \delta_x^2}{S\Delta_x^2 + t^2 \delta_x^2}$	$s = \frac{St^2 \cdot w_s(1-w_s)}{S\Delta_w^2 + t^2 w_s(1-w_s)}$

При визначенні необхідного обсягу вибірки величина дисперсій

для середньої  $\sigma_x^2$  та частки  $\sigma_w^2$  є невідомою. Ці значення встановлюються приблизно на основі аналогічних або пробних досліджень.

Для частки можна взяти максимальне значення  $\sigma_w^2 = 0,25$  (якщо ознака альтернативна).

Наведені вище формули використовуються лише у великих вибірках (при  $n \geq 30$ ). Малі вибірки, чисельність одиниць яких не перевищує 20–30, не відображають досить точно показників генеральної сукупності і тому рідко використовуються для встановлення надійних меж цих показників.

У табл. 5.8 наведено формули розрахунків точкових і інтервальних помилок вибірки, оцінок генерального параметра.

Таблиця 5.8 – Розрахунок точкових і інтервальних помилок вибірки

Схема відбору	Оціночний параметр			
	середня характеристика		частка ознаки	
	Точкова оцінка	Інтервальна оцінка	Точкова оцінка	Інтервальна оцінка
Повторна	$\sqrt{\frac{\sigma^2}{n}}$	$T \sqrt{\frac{\sigma^2}{n}}$	$\sqrt{\frac{\omega(-\omega)}{n}}$	$t \cdot \sqrt{\frac{\omega(-\omega)}{n}}$
Безповторна	$\sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$t \cdot \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{\omega(-\omega)}{n} \left(1 - \frac{n}{N}\right)}$	$t \cdot \sqrt{\frac{\omega(-\omega)}{n} \left(1 - \frac{n}{N}\right)}$

Слід відзначити, що все викладене вище відносно статистичних оцінок ґрунтується на припущенні, що вибіркова сукупність, параметри якої використовуються при оцінці, була одержана з використанням методу (способу) відбору, який забезпечує одержання вірогідності вибірки. При цьому, обираючи довірчу ймовірність оцінки, слід керуватися тим принципом, що вибір її рівня не є математичним завданням, а визначається конкретно вирішуваною проблемою.

📖 Література до теми 5: [4; 5–10; 15–21; 23; 25–28].

### **Контрольні запитання до теми 5**

1. Яке спостереження називається вибірковим?
2. Поняття генеральної сукупності.
3. Поняття вибірки.
4. Які показники підлягають оцінці при вибіркового обстеженні?
5. Способи відбору одиниць генеральної сукупності у вибірку.

6. Методи відбору одиниць генеральної сукупності у вибірку.
7. Середня помилка вибіркової середньої.
8. Середня помилка вибіркової частки.
9. Гранична помилка вибірки.
10. Визначення необхідного обсягу вибірки.
11. Чим відрізняється індивідуальний відбір від групового?
12. Яке число не повинна перевищувати чисельність одиниць вибіркового спостереження при малій виборці?
13. У якому відборі елементи вибираються через рівні інтервали?
14. При якому відборі одиницю, яка потрапила у відбір, не повертають у сукупність?
15. Як визначається гранична помилка серійним методом безповторної вибірки при визначенні частки?
16. Як визначається гранична помилка типовим методом повторної вибірки при визначенні середньої?
17. Які ви знаєте схеми відбору одиниць вибірки?
18. Які ви знаєте методи відбору одиниць генеральної сукупності?
19. Елементи вибіркового спостереження.
20. Визначте поняття довірчого інтервалу.

## Тема 6. ІНДЕКСНИЙ МЕТОД У СТАТИСТИЧНИХ ДОСЛІДЖЕННЯХ

6.1. Сутність, функції та класифікація індексів у статистичному аналізі.

6.2. Загальні (складні, агрегатні) індекси.

6.3. Середні індекси. Індекси середніх величин.

6.4. Базисні і ланцюгові індекси.

6.5. Системи індексів. Аналіз факторів розвитку соціально-економічних явищ індексним методом.

### ***6.1. Сутність, функції та класифікація індексів у статистичному аналізі***

У статистиці індекси, разом з середніми величинами, є найбільш поширеними статистичними показниками. З їх допомогою характеризується розвиток національної економіки в цілому і її окремих галузей, аналізуються результати виробничо-господарської діяльності підприємств і організацій, досліджується роль окремих чинників у формуванні найважливіших економічних показників, виявляються резерви виробництва. Індекси використовуються також в зіставленнях міжнародних економічних показників, визначенні рівня життя, моніторингу ділової активності в економіці і ін.

*Індексний метод* – найпоширеніший метод аналізу соціально-економічних явищ. Існують індекси врожайності, заробітної плати і т. д. Проте, індексний метод має істотний недолік, він адекватно вимірює тільки функціональні причинно-наслідкові залежності, які в економіці не переважають. Побудова індексів вимагає глибоких знань у специфіці досліджуваного явища.

*Індекс* (у перекладі з латинського index – показник) – це відносна величина, яка показує, у скільки разів рівень досліджуваного явища в даних умовах відрізняється від рівнів того ж явища в інших умовах.

Відмінність умов може виявлятися в часі (тоді говорять про індекси динаміки), у просторі (територіальні індекси), як база порівняння якого-небудь умовного рівня, наприклад планового показника, рівня договірних зобов'язань і таке інше. Відповідно розраховують індекс виконання зобов'язань, або – якщо плановий рівень порівнюється з рівнем попереднього періоду – індекс планового завдання.

Величина, зміна якої вивчається в даному конкретному випадку за допомогою індексу, називається *індексованою величиною*.

*Кожна індексована величина має позначення:*

–  $q$  – кількість (об'єм) будь-якого продукту в натуральному виразі;

–  $p$  – ціна одиниці товару;

–  $z$  – собівартість одиниці продукції;

–  $t$  – витрати часу на виробництво одиниці продукції (трудомісткість);

–  $w$  – виробіток продукції у вартісному виразі на одного працівника або в одиницю часу;

–  $v$  – виробіток продукції в натуральному виразі одного працівника або в одиницю часу;

–  $T$  – загальні витрати часу або чисельність працівників;

–  $\Pi$  – посівна площа;

–  $V$  – врожайність окремих культур;

–  $pq$  – загальна вартість проведеної продукції даного вигляду або загальна вартість проданих товарів даного вигляду (товарообіг, виручка);

–  $zq$  – витрати на виробництво всієї продукції;

– ВП – валовий збір окремої культури.

Щоб розрізнити, до якого періоду відносяться індексовані величини, прийнято біля символу індексу знизу справа ставити підрядкові знаки: 1 – для порівнюваних (поточних, звітних) періодів і 0 – для періодів, з якими проводиться порівняння. Якщо зміна явищ вивчається за ряд періодів, то кожен з періодів позначається відповідно підрядковими знаками 0, 1, 2, 3 і так далі.

Класифікацію індексів наведено на рис. 6.1.

За ступенем охоплення елементів сукупності розрізняють індивідуальні і зведені (загальні, складні, агрегатні) індекси.

*Індивідуальними* називаються індекси, що характеризують зміну тільки одного елемента сукупності. Вони є відносними величинами динаміки, виконання зобов'язань, порівняння. Вибір бази порівняння визначається метою дослідження. Індивідуальні індекси позначаються буквою « $i$ » і підрядковим знаком індексованого показника. Розрахунок індивідуальних індексів простий, їх визначають розрахунком відношення двох індексованих величин.



Рисунок 6.1 – Класифікація індексів

*Приклад.* Індивідуальний індекс фізичного обсягу розраховується за формулою

— (6.1)

де  $q_0$  і  $q_1$  – кількість виробленої продукції в базисному і поточному (звітному) періодах відповідно.

Індивідуальний індекс ціни розраховується за формулою

— (6.2)

де  $p_0$  і  $p_1$  – ціни одиниці продукції в базисному і поточному (звітному) періодах відповідно.

Індивідуальний індекс продуктивності праці розраховується за формулою

— (6.3)

де  $w_0$  і  $w_1$  – виробіток одиниці продукції у вартісному виразі на одного працівника або в одиницю часу в базисному і поточному (звітному) періодах відповідно.

Індивідуальний індекс трудомісткості розраховується за формулою

— (6.4)

де  $t_0$  і  $t_1$  – витрати часу на виробництво одиниці продукції в базисному і поточному (звітному) періодах відповідно.

Якщо індекси визначаються за ряд послідовних проміжків часу, то вони називаються *ланцюговими* або *базисними*.

З аналітичної точки зору індивідуальні індекси аналогічні темпам зростання і характеризують зміни індексованої величини в поточному періоді в порівнянні з базисним, тобто в скільки разів вона зросла (зменшилась) або скільки відсотків складає її зростання (зниження). Значення індексів виражаються в коефіцієнтах або відсотках. Якщо із значення індексу, вираженого у відсотках, відняти 100 %, тобто ( $i, \% - 100$ ), то отримана різниця покаже, на скільки відсотків зросла (зменшилась) індексована величина.

Основна перевага індивідуальних індексів – простота, *недолік* –



обмежена сфера застосування – тільки для одного однорідного явища.

## **6.2. Загальні (складні, агрегатні) індекси**

У економічних розрахунках для вимірювання динаміки складного явища найчастіше використовуються *загальні* індекси. Методика розрахунку загальних індексів складніша, ніж індивідуальних, і різноманітна залежно від характеру індексованих показників, наявності початкових даних і цілей дослідження.

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 гармонійні. Агрегатні індекси можуть бути розраховані як індекси змінного складу і індекси постійного (фіксованого) складу.

Агрегатний індекс є основною формою індексу. Агрегатним він називається тому, що його чисельник і знаменник являють собою набір безпосередньо непорівнянних, що не піддаються підсумовуванню елементів – суму добутоків двох величин, одна з яких змінюється (індексується), а інша залишається незмінною в чисельнику і знаменнику (вага індексу). Вага індексу служить для порівняння індексованих величин. Індекс в агрегатній формі, завдяки використанню універсальної ваги, дозволяє підсумувати товари різних одиниць виміру, які безпосередньому підсумовуванню не піддаються.

Типовим агрегатним індексом кількісних показників є індекс фізичного обсягу. Складність при побудові цього індексу полягає в тому, що обсяги різних видів продукції і товарів в натуральному виразі неспівмірності і безпосередньо підсумовуватись не можуть. Причиною неспівмірності є неоднорідність – відмінність натуральної форми і властивостей (одиниць виміру – літрів, кілограмів, штук).

Єдність різних видів продукції або різних товарів полягає в тому, що вони є продуктами суспільної праці, мають певну вартість і її грошовий вимір – ціну  $p$ . Кожний продукт має також собівартість  $Z$  і трудомісткість  $t$ . Ці якісні показники і можуть бути використані як загальна міра – коефіцієнт порівняння різнорідних продуктів. Помножуючи обсяг продукції кожного виду  $q$  на відповідну ціну, собівартість, трудомісткість одиниці продукції, отримують порівнянні показники, які можна підсумовувати ( $pq, zq, tq$ ).

Коефіцієнти порівняння забезпечують кількісну порівнянність,

дозволяють враховувати «вагу» продукту в реальному економічному процесі. Тому їх показники (співмножники, пов'язані з індексованими величинами), прийнято називати *вагами індексів*, а множення на них – зважуванням. Помножуючи кількість виробленої продукції (проданих товарів) на ціни, (які, як правило, виступають як міра неоднорідності продукції), отримуємо вартісний («ціннісний») вираз продукції кожного виду, який допускає підсумовування. Вартістю продукції є вироблення кількості продукції в натуральному виразі  $q$  на одиницю продукції  $p$ .

Відношення вартості продукції поточного періоду  $\sum p_1 q_1$  до вартості продукції базисного періоду  $\sum p_0 q_0$  є агрегатним індексом вартості продукції і товарообігу

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}, \quad (6.5)$$

де  $p_0 q_0$  – товарообіг базисного періоду,  $p_1 q_1$  – товарообіг звітного періоду.

Цей індекс показує, в скільки разів зросла (зменшилась) вартість продукції (товарообігу) звітного періоду в порівнянні з базисним, або скільки відсотків складає зростання (зниження) вартості продукції. Якщо з індексу вартості продукції відняти 1, то різниця ( $I_{pq} - 1$ ) покаже, на скільки змінилась вартість продукції в звітному періоді в порівнянні з базисним.

За допомогою агрегатних індексів можна не тільки розрахувати відносну зміну явища, що вивчається, але і отримати абсолютний приріст результативного показника по чинниках. Наприклад, абсолютний приріст вартості продукції розраховується, як різниця між чисельником і знаменником формули (6.5)

$$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0, \quad (6.6)$$

і показує, на скільки грошових одиниць змінилась вартість продукції в поточному періоді в порівнянні з базисним.

Значення індексу вартості продукції (товарообігу) залежить від двох чинників: зміни кількості продукції (обсягів) і цін.

Для того, щоб індекс охарактеризував зміну тільки одного чинника, потрібно усунути в його формулі вплив іншого чинника, зафіксувавши його як в чисельнику, так і в знаменнику на рівні того ж

самого періоду. Так, якщо продукцію (товари) порівнюваних періодів оцінювати за тими самими, наприклад, базисними цінами  $p_0$ , то такий індекс відобразить зміну одного чинника – індексованого показника  $q$  і буде агрегатним індексом фізичного обсягу продукції

$$I_q = \frac{\sum p_1 q_1}{\sum p_1 q_0}, \quad (6.7)$$

де  $p_1 q_0$  – умовна вартість товарів, реалізованих у базисному періоді за звітними цінами.

При цьому, абсолютний приріст вартості продукції за рахунок зміни фізичного обсягу продукції розраховується аналогічно формулі (6.6) і має вираз

$$\Delta q = \sum p_1 q_1 - \sum p_1 q_0, \quad (6.8)$$

що показує, на скільки грошових одиниць змінилась вартість продукції в результаті зростання (зменшення) її обсягу. При побудові  $I_q$  як ваги беруться зіставні, незмінні, фіксовані ціни на рівні базисного періоду, що дозволяє усунути їх вплив на зміну обсягу. Використання незмінних цін залежно від об'єкта дослідження дає можливість вивчити динаміку випуску сукупності вироблених товарів на окремому підприємстві, в окремих галузях промисловості і промисловості в цілому. Зіставні ціни не повинні сильно відрізнятися від цін, що діють (поточних). Тому їх періодично переглядають, переходять до нових зіставних цін.

В умовах ринкових відносин особливе місце відводиться індексу цін. За допомогою індексу споживчих цін здійснюється оцінка динаміки цін на товари виробничого і невиробничого споживання, перерахунок найважливіших вартісних показників з фактичних цін в зіставні. Індекс споживчих цін є загальним вимірником інфляції, використовується при коректуванні законодавчо встановленого мінімального розміру оплати праці, встановленні ставок податків і таке інше. Розрахувавши індекс цін, можна підрахувати економічний ефект від зміни цін. Індексованою величиною в даному індексі є ціна товару, а так званий вплив кількості проданих товарів повинен бути усунений, тоді прийнята кількість товарів одного з періодів береться як вага.

Перша спроба усунути недоліки індивідуальних індексів для

визначення загального індексу цін була зроблена французьким вченим Дюто в 1752 р. Він запропонував зведений індекс і свій запис індексу суми цін товарів

$$I_p = \frac{\sum p_1}{\sum p_0}. \quad (6.9)$$

Недоліком цього зведеного індексу є те, що він не враховує різницю цін на неоднакові товари і структуру товарообігу.

Більш досконалим індексом був індекс Карлі (1766 р.)

$$I_p = \frac{\sum i_p}{n}. \quad (6.10)$$

Він не залежить від рівня цін на окремі товари, однак він також не враховує структуру товарообігу. Індекси Дюто і Карлі у теперішній економіці не застосовуються, однак вони слугують базою для створення двох сучасних галузей індексів:

- індексу в агрегатній формі (Дюто);
- середніх індексів (Карлі).

Термін «агрегат» був запозичений з техніки, він означає з'єднання різнорідних механізмів у єдину машину. Вперше індекс в агрегатній формі був побудований в 1871 р. професором Ласпейресом (індекс з базисними вагами)

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}, \quad I_q = \frac{\sum p_1 q_1}{\sum p_1 q_0}, \quad (6.11)$$

де  $p_1 q_0$  – умовна вартість товарів, реалізованих у базисному періоді за звітними цінами.

Індекс Ласпейреса є основним індексом цін в умовах ринкової економіки для розрахунків динаміки вартості споживчого кошика. Індекс Ласпейреса показує, в скільки разів змінилися ціни в звітному періоді в порівнянні з базисним, але щодо тієї продукції, яка була реалізована в базисному періоді, і яку економію (перевитрату) можна було б отримати від зміни цін, тобто умовну економію (перевитрату)

$$\Delta p = \sum p_1 q_0 - \sum p_0 q_0. \quad (6.12)$$

Наступним індексом ціни в агрегатній формі був агрегатний індекс цін із звітними вагами, запропонований в 1874 році німецьким економістом Пааше:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0}, \quad (6.13)$$

де  $p_0 q_1$  – умовна вартість товарів, реалізованих у звітному періоді за базисними цінами.

Індекс цін Пааше показує, в скільки разів зріс (зменшився) в середньому рівень цін на масу товарів, реалізовану в звітному періоді або скільки відсотків складає його зростання (зменшення) в звітному періоді в порівнянні з базисним.

При побудові індексу цін у сфері реалізації як ваги індексу зазвичай беруть кількість товарів, продану в поточному (звітному) періоді. Це пояснюється тим, що такий розрахунок  $I_p$  дозволяє визначити не тільки відносну зміну цін, але і абсолютну економію (+) або абсолютну перевитрату (–) грошових коштів покупців в результаті зміни цін на товари.

Проте вказаний вибір вагів при побудові  $I_p$  не можна вважати обов'язковим. Наприклад, під час економічної кризи через підвищення цін ряд продуктів випадає із споживання малозабезпечених верств населення, тобто зменшення може буди зумовлене зменшенням обсягу продаж. В цьому випадку індекс  $I_p$ , розрахований за Пааше, невірно відобразить зміну цін на ті продукти, які випали із споживання. Тому в подібних випадках правильніше відобразить реальну зміну цін індекс, побудований щодо продукції базисного періоду (індекс Ласпейреса). В умовах високої інфляції зважування по вагах звітного періоду (індекс Пааше) вимагає щоквартального перерахунку інформації для формування системи вагів, що пов'язане з великими витратами часу, матеріальних і трудових ресурсів, тому вітчизняна і зарубіжна статистика віддає перевагу моделі Ласпейреса. Для характеристики динаміки цін на споживчому рівні розраховується зведений індекс споживчих цін

$$\text{ІСЦ} = \frac{\text{Вартість ринкового кошика базисного періоду в поточному році}}{\text{Вартість ринкового кошика базисного періоду в базисному році}} \cdot 100 \%,$$

тобто

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}. \quad (6.14)$$

Аналогічно розраховуються і інші агрегатні індекси показників.

Собівартість продукції (товарів, робіт, послуг) – найважливіший показник ефективності діяльності підприємства; він є вартісною оцінкою використуваних у процесі виробництва продукції (робіт, послуг) природних ресурсів, сировини, матеріалів, палива, енергії, основних фондів, трудових ресурсів, а також інших витрат на її виробництво і реалізацію. Вочевидь, чим економніше витрачаються матеріали, енергія, тим менші інші види матеріальних витрат; чим правильніше організована праця і його оплата, тим менша собівартість продукції. Собівартість є частиною відпускної ціни продукції, а отже, частиною вартості продукції. Зниження собівартості без збитку для її якості або зниження її питомої ваги в повній вартості продукції – найважливіша умова забезпечення конкурентоспроможності товару на ринку, джерело отримання додаткового прибутку. Агрегатний індекс собівартості розраховується за формулою

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}, \quad (6.15)$$

де  $\sum z_1 q_1$  – витрати на виробництво і реалізацію продукції звітного періоду;  $\sum z_0 q_1$  – витрати на виробництво тієї ж продукції, якби собівартість одиниці продукції залишалася на рівні базисного періоду.

Економія (перевитрата) від зниження собівартості одиниці продукції розраховується за формулою

$$\Delta z q = \sum z_1 q_1 - \sum z_0 q_1. \quad (6.16)$$

*Продуктивність праці* – результативність конкретної живої праці, ефективність доцільної діяльності людей для створення продукту протягом певного проміжку часу; вимірюється кількістю споживчих вартостей, проведених в одиницю часу, або кількістю часу, витраченою на одиницю продукції. Для характеристики рівня продуктивності праці в

статистичній практиці використовуються два показники: виробіток (у натуральному і вартісному виразі) і трудомісткість. Виробіток розраховується за формулою

$$W = \frac{q}{T}, \quad (6.17)$$

де  $q$  – кількість виробленої продукції;  $T$  – витрати робочого часу на виробництво продукції (або чисельність працівників).

Трудомісткість  $t$  відображає витрати праці на виробництво одиниці продукції

$$t = \frac{T}{q}. \quad (6.18)$$

Трудомісткість є показником, зворотним до продуктивності праці. Зниження трудомісткості свідчить про підвищення продуктивності праці. Динаміка продуктивності праці вивчається в статистиці за допомогою індексів продуктивності праці. Агрегатний індекс продуктивності праці

$$I_w = \frac{\sum t_0 q_1}{\sum t_1 q_1}, \quad (6.19)$$

де  $\sum t_0 q_1$  – умовна величина, що характеризує витрати праці на продукцію звітного періоду при рівні продуктивності праці базисного періоду;  $\sum t_1 q_1$  – фактичні витрати праці на продукцію звітного періоду.

Особливість цього індексу полягає в тому, що  $t_0$  знаходиться в чисельнику, а  $t_1$  – в знаменнику. Це пояснюється тим, що індексуються витрати праці на одиницю продукції, тобто величини, зворотні продуктивності праці (індивідуальний індекс продуктивності праці

$$i_n = \frac{1}{t_1}; \quad \frac{1}{t_0} = \frac{t_0}{t_1}).$$

Основні формули для розрахунків індивідуальних та загальних зведених індексів наведено в табл. 6.1.

Індексний метод є методом факторного аналізу для визначення впливу окремих факторів на рівень результативного показника (табл. 6.2).

Таблиця 6.1 – Формули основних індексів

Назва індексу	Індивідуальні індекси	Загальні (складні, агрегатні) індекси	
		Пааше	Ласпейреса
Фізичного обсягу	$i_q = \frac{q_1}{q_0}$	$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0}$	$I_q = \frac{\sum p_1 q_1}{\sum p_1 q_0}$
Ціни	$i_p = \frac{p_1}{p_0}$	$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$	$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}$
Вартості	$i_{pq} = \frac{p_1 q_1}{p_0 q_0}$	$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$	
Трудомісткості	$i_t = \frac{t_1}{t_0}$	$I_t = \frac{\sum t_1 q_1}{\sum t_0 q_1}$	
Продуктивності	$i_w = \frac{t_0}{t_1}$	$I_w = \frac{\sum t_0 q_1}{\sum t_1 q_1}$	
Витрат праці	$i_{tq} = \frac{t_1 q_1}{t_0 q_0}$	$I_{tq} = \frac{\sum t_1 q_1}{\sum t_0 q_0}$	
Витрат на виробництво	$i_{zq} = \frac{z_1 q_1}{z_0 q_0}$	$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0}$	
Собівартості	$i_z = \frac{z_1}{z_0}$	$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}$	
Питомої собівартості	$i_c = \frac{c_1}{c_0}$	$I_c = \frac{\sum c_1 q_1}{\sum c_0 q_1}$	

Таблиця 6.2 – Визначення впливу факторів та взаємозв'язок між ними

Показники	За Пааше	За Ласпейресом
Абсолютна зміна вартості в результаті зміни цін продукції	$\Delta p = \sum p_1 q_0 - \sum p_0 q_0$	$\Delta p = \sum p_1 q_1 - \sum p_0 q_1$
Абсолютна зміна вартості в результаті зміни фізичного обсягу продукції	$\Delta q = \sum p_1 q_1 - \sum p_1 q_0$	$\Delta q = \sum p_0 q_1 - \sum p_0 q_0$
Загальна абсолютна зміна вартості продукції	$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0$	
Взаємозв'язок: – абсолютних показників – відносних показників	$\Delta pq = \Delta p + \Delta q$ $I_{pq} = I_p \cdot I_q$	



### 6.3. Середні індекси. Індекси середніх величин

Середні індекси – це комбінація індексу в агрегатній формі й індивідуальних індексів. Застосовуються в тому випадку, коли відсутні дані у звітному або базисному періодах. Якщо відсутні дані про кількість проданих товарів, але зареєстровані показники виторгу і індекси цін на окремі товари, то на базі індексу Пааше можна розрахувати середній гармонійний індекс цін.

Виводимо його через індекс Пааше  $I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ . Є дані про товарообіг звітного періоду й індивідуальні індекси цін. Тоді, враховуючи, що  $i_p = \frac{p_1}{p_0}$ , можна розрахувати, що  $p_0 = \frac{p_1}{i_p}$ , а  $p_0 q_1 = \frac{p_1 q_1}{i_p}$ . Тобто в результаті математичних перетворень отримаємо

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}$$

Якщо є дані про динаміку фізичного обсягу проданих товарів, то можна на базі даних обороту за минулий період розрахувати середній індекс фізичного обсягу:

$$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0}, \quad i_q = \frac{q_1}{q_0}, \quad q_1 = i_q q_0,$$

тобто

$$p_0 q_1 = i_q p_0 q_0,$$

звідси

$$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_1} = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0}.$$

На динаміку якісних показників, рівні яких виражені середніми величинами, впливає зміна структури явища, що вивчається. Під зміною структури розуміють зміну частки окремих одиниць сукупності, з яких формуються середні, в загальній їх чисельності. Структурні зрушення в народному господарстві – це важливі процеси вдосконалення виробництва і велике додаткове джерело розвитку продуктивних сил

суспільства.

Таким чином, завдання полягає у визначенні ступеня впливу двох чинників – змін значень осереднюваного показника і змін структури явища – на загальну динаміку середньої. Це завдання вирішується за допомогою індексного методу, тобто шляхом побудови *системи взаємопов'язаних індексів*, в яку включаються три індекси: змінного складу, постійного (фіксованого) складу і структурних зрушень.

Вивчення сумісної дії вищезгаданих чинників на загальну динаміку середнього рівня здійснюється в статистиці за допомогою індексу змінного складу. *Індексом змінного складу* є відношення двох зважених середніх з тими, що змінюються (змінними вагами, що показують зміну індексованої середньої величини). Для будь-яких якісних показників  $X$  індекс змінного складу можна записати у вигляді

$$I_x = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0}, \quad (6.20)$$

де  $x_1, x_0$  – рівні осереднюваного показника в звітному і базисному періодах відповідно;  $f_1, f_0$  – ваги (частоти) осереднюваного показника в звітному і базисному періодах відповідно.

Щоб елімінувати (усунути) вплив зміни структури сукупності на динаміку середньої величини, використовують відношення середніх зважених з тими самими вагами (як правило, на рівні звітного періоду). Індекс, що характеризує динаміку середньої величини при тій самій фіксованій структурі сукупності, носить назву *індексу постійного (фіксованого складу)* і має вигляд

$$I_x = \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_1}{\sum f_1}. \quad (6.21)$$

Після скорочення на  $\sum f_1$  формула набуває вигляду формули агрегатного індексу якісного показника

$$I_x = \frac{\sum x_1 f_1}{\sum x_0 f_1}. \quad (6.22)$$

Індекс постійного (фіксованого) складу показує, як у звітному періоді в порівнянні з базисним змінилося середнє значення показника щодо однорідної сукупності за рахунок зміни тільки самої індексованої величини, тобто коли вплив структурного чинника усунений.

Для вимірювання впливу тільки структурних змін на досліджуваній середній показник обчислюють *індекс структурних зрушень* як відношення середнього рівня індексованого показника базисного періоду, розрахованого на звітну структуру, до фактичної середньої цього показника в базисному періоді

$$I_{\text{стр. зр.}} = \frac{\sum x_0 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0}. \quad (6.23)$$

Як ваги (частоти) індексів середніх величин  $X$  разом з абсолютними показниками  $f$  можуть використовуватися і відносні показники (частки)  $d$ . В останньому випадку індекси для будь-яких якісних показників  $X$  можна виразити в загальному вигляді наступними формулами:

$$I_x = \frac{\sum x_1 d_1}{\sum x_0 d_0}; \quad I_x = \frac{\sum x_1 d_1}{\sum x_0 d_1}; \quad I_{\text{стр. зр.}} = \frac{\sum x_0 d_1}{\sum x_0 d_0}, \quad (6.24)$$

де  $d_1, d_0$  – частки одиниць з визначеним значення ознаки в загальній сукупності в звітному і базисному періодах відповідно.

#### **6.4. Базисні і ланцюгові індекси**

Часто в ході економічного аналізу зміну індексованих величин вивчають не за два, а за ряд послідовних періодів. Отже, виникає необхідність побудови індексів за ряд послідовних періодів, які утворюють індексні системи. Такі системи характеризують зміни в явищі, що вивчається, протягом досліджуваного періоду часу.

Залежно від бази порівняння індекси можуть бути як ланцюгові (коли здійснюється порівняння рівня поточного періоду з попереднім), так і базисні (коли здійснюється порівняння рівня поточного періоду з базисним).

Ряди індивідуальних індексів:

– базисні:

$$i_{p_{1/0}} = \frac{P_1}{P_0}; \quad i_{p_{2/0}} = \frac{P_2}{P_0}; \quad i_{p_{3/0}} = \frac{P_3}{P_0}; \quad (6.25)$$

– ланцюгові:

$$i_{p_{1/0}} = \frac{P_1}{P_0}; \quad i_{p_{2/1}} = \frac{P_2}{P_1}; \quad i_{p_{3/2}} = \frac{P_3}{P_2}. \quad (6.26)$$

Між ланцюговими і базисними індивідуальними індексами існує взаємозв'язок, що дозволяє переходити від одних індексів до інших; множення послідовних ланцюгових індивідуальних індексів дає базисний індекс останнього періоду

$$i_{p_{3/0}} = i_{p_{1/0}} \cdot i_{p_{2/1}} \cdot i_{p_{3/2}} = \frac{P_1}{P_0} \cdot \frac{P_2}{P_1} \cdot \frac{P_3}{P_2} = \frac{P_3}{P_0}. \quad (6.27)$$

Відношення базисного індексу звітного періоду до базисного індексу попереднього періоду дає ланцюговий індекс звітного періоду

$$i_{p_{3/2}} = i_{p_{3/0}} \div i_{p_{2/0}} = \frac{P_3}{P_0} \div \frac{P_2}{P_0} = \frac{P_3}{P_2}. \quad (6.28)$$

Це правило дозволяє застосовувати так званий ланцюговий метод, тобто знаходити невідомий ряд базисних індексів на основі відомих ланцюгових і навпаки.

*Ланцюговий метод агрегатних індексів.* Як відомо, в кожному окремому індексі ваги в чисельнику і знаменнику обов'язково фіксуються на тому ж самому рівні. Якщо ж будувється ряд індексів, то ваги в ньому можуть бути або постійними, або змінними.

#### Базисні індекси

Індекси цін Ласпейреса:

$$I_{p_{1/0}} = \frac{\sum P_1 q_0}{\sum P_0 q_0}; \quad I_{p_{2/0}} = \frac{\sum P_2 q_0}{\sum P_0 q_0}; \dots, \quad I_{p_{n/0}} = \frac{\sum P_n q_0}{\sum P_0 q_0}. \quad (6.29)$$

Індекси цін Пааше:

$$I_{p_{1/0}} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad I_{p_{2/0}} = \frac{\sum p_2 q_2}{\sum p_0 q_2}; \dots, \quad I_{p_{n/0}} = \frac{\sum p_n q_n}{\sum p_0 q_n}. \quad (6.30)$$

Індекси фізичного обсягу:

$$I_{q_{1/0}} = \frac{\sum p_0 q_1}{\sum p_0 q_0}; \quad I_{q_{2/0}} = \frac{\sum p_0 q_2}{\sum p_0 q_0}; \dots, \quad I_{q_{n/0}} = \frac{\sum p_0 q_n}{\sum p_0 q_0}. \quad (6.31)$$

#### *Ланцюгові індекси*

Індекси цін Пааше:

$$I_{p_{1/0}} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad I_{p_{2/1}} = \frac{\sum p_2 q_2}{\sum p_1 q_2}; \dots, \quad I_{p_{n/n-1}} = \frac{\sum p_n q_n}{\sum p_{n-1} q_n}. \quad (6.32)$$

Індекси цін Ласпейреса:

$$I_{p_{1/0}} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \quad I_{p_{2/1}} = \frac{\sum p_2 q_0}{\sum p_1 q_0}; \dots, \quad I_{p_{n/n-1}} = \frac{\sum p_n q_0}{\sum p_{n-1} q_0}. \quad (6.33)$$

Індекси фізичного обсягу:

$$I_{q_{1/0}} = \frac{\sum p_0 q_1}{\sum p_0 q_0}; \quad I_{q_{2/1}} = \frac{\sum p_0 q_2}{\sum p_0 q_1}; \dots, \quad I_{q_{n/n-1}} = \frac{\sum p_0 q_n}{\sum p_0 q_{n-1}}. \quad (6.34)$$

Отже, в базисних агрегатних індексах всі звітні дані зіставляються тільки з базисними (закріпленими) даними, а в ланцюгових – з попередніми (в даному випадку – суміжними) даними. Період вагів у всіх індексах цін Пааше узятий поточний (індекси *із змінними вагами*), в індексах фізичного обсягу і в індексах цін Ласпейреса – закріплений (індекси *з постійними вагами*). Постійні ваги (незмінні при переході від одного індексу до іншого) дозволяють виключити вплив зміни структури на значення індексу. Ряд агрегатних індексів з постійними вагами мають перевагу – зберігається взаємозв'язок між ланцюговими і базисними індексами. Наприклад, у ряді агрегатних індексів фізичного обсягу

$$\frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_0 q_2}{\sum p_0 q_1} \cdot \frac{\sum p_0 q_3}{\sum p_0 q_2} = \frac{\sum p_0 q_3}{\sum p_0 q_0} \quad (6.35)$$

або у ряді агрегатних індексів Ласпейреса

$$\frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_2 q_0}{\sum p_1 q_0} \cdot \frac{\sum p_3 q_0}{\sum p_2 q_0} = \frac{\sum p_3 q_0}{\sum p_0 q_0} \quad (6.36)$$

Таким чином, використання постійних вагів протягом ряду років дозволяє переходити від ланцюгових загальних індексів до базисних і навпаки.

У рядах агрегатних індексів із змінними вагами (ряд цін Пааше) множення ланцюгових індексів не дає базисний. Для таких індексів перехід від ланцюгових індексів до базисних (і навпаки) неможливий.

### **6.5. Системи індексів. Аналіз факторів розвитку соціально-економічних явищ індексним методом**

Індекси взаємозалежні в системі, подібно тому, як взаємозалежні індексні економічні показники. Розглянемо взаємозалежність індексів на такому прикладі: товарообіг можна розглядати як суму добутоків цін на товари на кількість товару:

$$\sum \text{товарообіг} = \sum \text{ціна}_{\text{товару}} \cdot \text{кількість};$$

$$I_{\text{товарообіг}} = I_{\text{цін}} \cdot I_{\text{фізичногообсягу}};$$

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum q_1 p_1}{\sum q_0 p_1}.$$

Подібним чином пов'язані між собою індекс обсягу виробництва з індексом цін і фізичного обсягу, а індекс валового збору з індексом врожайності і посівних площ. Іншим різновидом систем індексів є взаємодія між індексами змінного складу, постійного складу і структурних зрушень.

*Приклад 1.* Розрахунок системи індексів середньої собівартості одиниці продукції. Дані для розрахунку наведено в табл. 6.3.

Таблиця 6.3 – Дані для розрахунку

Під-приємства	Собівартість одиниці продукції		Обсяг виробництва		Витрати на виробництво		Звітний обсяг щодо базисної собівартості	Структура виробництва		Витрати в базисному періоді з розрахунку 100 тис. штук виробів	
	$z_0$	$z_1$	$q_0$	$q_1$	$z_0q_0$	$z_1q_1$		$z_0q_1$	$dq_0$	$dq_1$	$z_0dq_0$
А	2,4	2,3	200	220	480	506	528	44,44	45,83	106,67	110
Б	2,8	2,6	100	100	280	260	280	22,22	20,83	62,22	58,33
В	2,6	2,4	150	160	390	376	416	33,33	33,33	86,67	86,67
Разом	-	-	450	480	1150	1142	1224	100,00	100,00	255,56	255,00

Індекс змінного складу характеризує зміну індексованих показників під дією двох факторів:

1) застосування якісного показника в окремих варіантах ряду (у цьому випадку зміна собівартості одиниці продукції на окремих підприємствах);

2) зміни, внаслідок структурних зрушень кількісних показників (структура виробництва, у цьому випадку  $dq_0 = \frac{q_0}{\sum q_0}$ ;  $dq_1 = \frac{q_1}{\sum q_1}$ ).

Вплив кожного із цих факторів відображає відповідний індекс, який виводиться з індексу змінного складу шляхом закріплення одного з факторів на постійному рівні.

Індекс змінного складу характеризується співвідношенням двох середніх величин індексованого показника у звітному і базисному періодах:

$$I_{\bar{z}} = \frac{\bar{z}_1}{\bar{z}_0} = \frac{\sum z_1 q_1}{\sum q_1} \cdot \frac{\sum z_0 q_0}{\sum q_0} = \frac{1142}{480} \cdot \frac{1150}{450} = 0,9297 \text{ або } 92,97 \%$$

Таким чином, середні витрати на виробництво одиниці продукції знизились на 7,03 %.

Індекс зміни собівартості показує середню зміну середньої собівартості одиниці продукції в результаті зміни собівартості виробництва на окремих підприємствах у звітному періоді в порівнянні з базисним:

$$\bar{I}_z = \frac{\sum z_1 q_1}{\sum q_1} \cdot \frac{\sum z_0 q_1}{\sum z_0 q_1} = \frac{\sum z_1 q_1}{\sum z_0 q_1} = \frac{1142}{1224} = 0,933 \text{ або } 93,3 \% \text{ (аналог індексу Пааше).}$$

У результаті зміни собівартості одиниці продукції обсяг виробни-

цтва зменшився на 6,7 %.

Індекс структурних зрушень характеризує вплив змін структури виробництва на динаміку середньої величини (собівартість одиниці продукції). Так само розраховується індекс змінного складу шляхом закріплення якісного показника (собівартість одиниці продукції) на базисному рівні:

$$I_{dq} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{\sum z_0 q_1}{\sum z_0 q_0} : \frac{\sum q_1}{\sum q_0} = \frac{1224}{1150} : \frac{480}{450} = 0,9961 \text{ або } 99,61 \%$$

У результаті структурних зрушень у виробництві середня собівартість одиниці продукції знизилась на 0,39 %.

Між індексами постійного, змінного складу і структурних зрушень існує взаємозв'язок:  $I_{\bar{z}} = \bar{I}_z \cdot I_{dq}$ , що дозволяє розрахувати один з індексів, якщо відомі два інших.

$$\text{Перевірка: } 0,9297 \approx 0,933 \cdot 0,9961.$$

Система індексів дозволяє вимірювати не тільки відносні, але й абсолютні показники індексованих похідних від нього:

$$\begin{aligned} \Delta p q &= \Delta p q^p + \Delta p q^q, \\ \sum p_1 q_1 - \sum p_0 q_0 &= \underbrace{\left( \sum p_1 q_0 - \sum p_0 q_0 \right)}_{\Delta p q^p} + \underbrace{\left( \sum p_1 q_1 - \sum p_1 q_0 \right)}_{\Delta p q^q} \end{aligned} \quad (6.37)$$

Це різниці між чисельниками і знаменниками відповідних комплексних індексів. Дана система індексів дозволяє визначити складові загальної абсолютної зміни витрат на виробництво у звітному періоді в порівнянні з базисним:

$$\Delta z q = \sum z_1 q_1 - \sum z_0 q_0 = 1142 - 1150 = -8 \text{ тис. грн};$$

$$\Delta z q^z = \sum z_1 q_1 - \sum z_0 q_1 = 1142 - 1224 = -82 \text{ тис. грн};$$

$$\Delta z q^{dq} = \frac{\sum z_1 q_1}{100} \left( \sum z_0 d q_1 - \sum z_0 d q_0 \right) = \frac{1142}{100} (55,00 - 255,56) = -6,5 \text{ тис. грн};$$

$$\Delta z q^q = \left( \frac{\sum q_1}{\sum q_0} - 1 \right) \sum z_0 q_0 = \left( \frac{480}{450} - 1 \right) 1150 = 80,5 \text{ тис. грн}.$$

$$\text{Перевірка: } \Delta z q = \Delta z q^z + \Delta z q^{dq} + \Delta z q^q = -82 - 6,5 + 80,5 = -8 \text{ тис. грн}.$$

Особливе місце в економічному аналізі явищ і процесів посідає факторний аналіз. Факторний аналіз з використанням індексного методу





Приклад 2. Розрахунок та економічна інтерпретація індексів. Вихідні дані наведено в табл. 6.4.

Таблиця 6.4 – Вихідні дані для розрахунку

Продукція	Витрати сировини на одиницю продукції (питомі витрати), кг		Ціна 1 кг сировини, грн		Кількість виробленої продукції, шт	
	I кв. ( $m_0$ )	IV кв. ( $m_1$ )	I кв. ( $p_0$ )	IV кв. ( $p_1$ )	I кв. ( $q_0$ )	IV кв. ( $q_1$ )
А	20	22	100	160	40	50
Б	17	13	220	270	100	140
В	40	45	350	360	200	170

Результативний показник – витрати на виробництво

(6.42)

Індекс витрат сировини на одиницю продукції (питомих витрат):

\_\_\_\_\_

Індекс цін на сировину:

\_\_\_\_\_

Індекс кількості (фізичного обсягу) продукції:

\_\_\_\_\_

Індекс витрат на виробництво:

\_\_\_\_\_

Отже, витрати на виробництво зросли на 5,1 %, в тому числі за рахунок збільшення витрат сировини на одиницю продукції – на 8,3 %, за рахунок зростання цін – на 5,9 %, а під впливом зменшення кількості виробленої продукції витрати зменшились на 8,3 %.

Визначимо абсолютний приріст витрат на виробництво:

Прирости за рахунок змін:


– витрат сировини на одиницю продукції:

– цін на сировину:

– кількості (фізичного обсягу) продукції:

При цьому

Отже, факторний індексний аналіз є ефективним методом дослідження мультиплікативних детермінованих зв'язків при кількості факторних ознак два і більше.

 Література до теми 6: [4; 5; 7; 8; 15–21; 23–28].

***Контрольні запитання до теми 6***

1. Що таке статистичний індекс?
2. Яким чином класифікують статистичні індекси?
3. Що характеризують індивідуальні індекси?

4. Що характеризують загальні індекси цін, фізичного обсягу і товарообігу, як вони взаємопов'язані?
5. Що характеризує індекс цін змінного складу?
6. Що характеризує індекс цін постійного складу?
7. Що характеризує індекс цін структурних зрушень?
8. Що характеризують територіальні індекси?
9. Які індекси бувають за формою побудови?
10. Які індекси бувають за видом ваги-порівнювача?
11. Коли в розрахунках використовують індивідуальні індекси, а коли – зважені?
12. У чому полягає сутність правила побудови агрегатних індексів?
13. Наведіть формулу розрахунку середньозваженого індексу ціни та поясніть, що він характеризує.
14. Побудуйте індекс середньої собівартості фіксованого складу та поясніть, що він характеризує.
15. Запишіть формулу індексу середнього рівня ціни змінного складу та поясніть, що він характеризує.
16. Запишіть формулу індексу середнього рівня продуктивності праці структурних зрушень та поясніть, що він характеризує.
17. Що визначають за допомогою індивідуальних індексів?
18. Що таке базисні і ланцюгові індекси?
19. У чому полягає сутність середніх індексів?
20. У чому полягає сутність взаємозв'язку між індексами?

## Тема 7. СТАТИСТИЧНЕ ВИВЧЕННЯ ДИНАМІКИ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ

- 7.1. Основні поняття і види рядів динаміки.
- 7.2. Аналітичні та середні показники ряду динаміки.
- 7.3. Методи обробки динамічних рядів.
- 7.4. Виявлення тенденцій розвитку соціально-економічних явищ в динаміці.
- 7.5. Дослідження сезонних коливань в рядах динаміки.
- 7.6. Особливості дослідження взаємозв'язків в рядах динаміки.

### **7.1. Основні поняття і види рядів динаміки**

Динаміка (грецькою мовою *dinamik* – «сила, розвиток») – це процес розвитку явища в часі і просторі. Для того щоб відобразити цей процес, будують ряди динаміки (інша назва – динамічні ряди).

*Динамічним рядом (рядом динаміки)* називають ряд статистичних показників, які розташовані в хронологічній послідовності і які характеризують зміну явища в часі. Динамічний ряд складається з двох елементів:

- 1) статистичного показника (інша назва – рівень ряду) – характеризує величину явища, його розмір і найчастіше позначається через *у*;
- 2) моменту часу, або ряду періодів – показника, який характеризує певний час, у якому відповідний статистичний показник дійсний.

У класифікаційному аспекті ряди динаміки розрізняють за такими ознаками:

- Залежно від показників, які утворюють дану сукупність: *абсолютні, відносні і середні.*
- Залежно від часу, який визначений в динамічних рядах: *інтервальні і моментні.*

*Моментні динамічні ряди* характеризують рівень явища станом на певний момент часу. Рівні моментних динамічних рядів не слід підсумовувати, тому що кожний наступний рівень умовно або фактично містить у собі попередній. Приклад: показники чисельності населення на початок року, величина запасу матеріалів на початок періоду і т. д. Важлива аналітична відмінність моментних рядів від інтервальних полягає в тому, що сума рівнів інтервального ряду дає цілком реальний показник – загальний випуск продукції за рік, загальні витрати робочого часу, загальний обсяг продажів акцій і т. ін., а сума рівнів моментного ряду,

хоча іноді і підлягає підрахунку, але реального економічного змісту не містить.

*Інтервальні динамічні ряди* відображають масштаби явища (товарообіг, витрати, доходи і т. ін.) за певні періоди часу (дні, тижні, декади, місяці, квартали і т. і.). Показники інтервального ряду можна підсумувати. Така операція називається *укрупненням тимчасових інтервалів*. Різновидом інтервальних рядів є *ряди динаміки з наростаючими підсумками*. Вони застосовуються для оцінки динаміки виконання запланованих показників і поточних, порівняння результатів діяльності різних господарських суб'єктів. Кожний рівень такого ряду – це сума значень аналізованого показника за всі попередні періоди його реєстрації.

– Залежно від відстані між рівнями ряду динаміки: *рівні і нерівні* (тобто з *рівними і нерівними інтервалами*).

– Залежно від кількості статистичних показників: *одновимірні (ізолювані)* і *багатовимірні (комплексні)*. *Ізолюваний ряд динаміки* – якщо ведеться аналіз у часі одного показника. *Комплексний ряд динаміки* – коли в хронологічній послідовності дається система показників, пов'язаних між собою єдністю процесу або явища.

– Залежно від відстані між датами або інтервалами часу: розрізняють *повні і неповні хронологічні ряди*. *Повні ряди динаміки* – коли дати реєстрації або закінчення періодів розташовані один за одним з рівними інтервалами. *Неповні* – коли принцип рівних інтервалів не дотримується.

– Залежно від бази порівняння: *базисні і ланцюгові*. Якщо база порівняння постійна, то характеристики динаміки називаються *базисними*. Якщо база порівняння змінюється, то характеристики динаміки називаються *ланцюговими*.

Статистичне дослідження рядів динаміки передбачає:

- вимірювання інтенсивності розвитку тимчасового ряду;
- визначення загальної тенденції змін явищ у часі;
- аналіз причинно-наслідкової залежності в рядах динаміки;
- дослідження періодичних (циклічних і сезонних) коливань;
- прогнозування розвитку динамічних рядів.

## **7.2. Аналітичні та середні показники ряду динаміки**

При вивченні явища в часі виникає проблема опису інтенсивності зміни і розрахунків середніх показників динаміки. Вирішується вона

шляхом побудови відповідних аналітичних показників. У випадку, коли порівняння проводиться з періодом (моментом) часу, початковим у ряді динаміки одержують *базисні показники*. Якщо ж порівняння проводиться з попереднім періодом або моментом часу, то говорять про *ланцюгові показники*. Для характеристики інтенсивності зміни в динаміці розраховують наступні аналітичні показники:

1. Абсолютний приріст – різниця між рівнями ряду. Сума ланцюгових абсолютних приростів дорівнює базисному абсолютному приросту за відповідний період часу.

2. Коефіцієнт росту – відносна величина динаміки. Він показує, у скільки разів поточний рівень більший або менший від рівня, з яким здійснюється порівняння. Добуток ланцюгових коефіцієнтів росту дорівнює базисному коефіцієнту росту за весь аналізований період.

3. Темп росту – відносна величина динаміки у відсотковому вираженні. Він показує, в скільки разів поточний рівень більший або менший від порівнюваного. Різниця між темпом росту і коефіцієнтом в тому, що коефіцієнт виражається лише в частках, а темп росту – у відсотках. Добуток ланцюгових темпів росту дорівнює базисному темпу росту за весь аналізований період;

4. Темп приросту – показує, на скільки відсотків збільшився або зменшився поточний рівень у порівнянні із рівнем, прийнятим за базу порівняння.

5. Абсолютне значення одного відсотка приросту – якщо рівні ряду динаміки послідовно зростають у часі, то важливе значення має не тільки відсоток зміни показників, але і абсолютне значення одного відсотка приросту. Абсолютне значення одного відсотка дорівнює відношенню абсолютного приросту до попереднього (базисного) періоду до 100 %. Цей показник найчастіше розраховується для ланцюгового ряду.

6. Темп нарощування – якщо економіка постійно зростає, то для порівняльної оцінки інтенсивності росту застосовується такий показник, як темп нарощування, коли абсолютні ланцюгові прирости порівнюються з базисними рівнями.

7. Пункти росту – використовуються в тому випадку, коли проводиться порівняння достатньо віддалених у часі показників. Пункт росту (або пунктпроцент) – це різниця базових темпів росту (або приросту) в процентах або коефіцієнтах двох суміжних періодів.

Існують наступні позначення рівнів ряду:

$y_i$  – поточний рівень ряду;

$y_{i-1}$  – попередній рівень ряду;

$y_0$  – початковий рівень, або рівень, який узятий за базу порівняння.

Основні формули для розрахунку аналітичних показників динаміки наведено в табл. 7.1

Таблиця 7.1 – Аналітичні показники динамічних рядів

Показник	Базисний	Ланцюговий	Примітки
Абсолютний приріст	$\Delta_{\text{баз}} = y_i - y_0$	$\Delta_{\text{ланц}} = y_i - y_{i-1}$	$\sum \Delta_{\text{ланц}} = \Delta_{\text{баз}}$
Коефіцієнт росту	$K_{\text{р}}^{\text{баз}} = \frac{y_i}{y_0}$	$K_{\text{р}}^{\text{ланц}} = \frac{y_i}{y_{i-1}}$	$K_{\text{р}}^{\text{ланц} 1} \cdot \dots \cdot K_{\text{р}}^{\text{ланц} n} = K_{\text{р}}^{\text{баз}}$
Темп росту, %	$T_{\text{р}}^{\text{баз}} = \frac{y_i}{y_0} \cdot 100 \%$ $T_{\text{р}}^{\text{баз}} = K_{\text{р}}^{\text{баз}} \cdot 100 \%$	$T_{\text{р}}^{\text{ланц}} = \frac{y_i}{y_{i-1}} \cdot 100 \%$ $T_{\text{р}}^{\text{ланц}} = K_{\text{р}}^{\text{ланц}} \cdot 100 \%$	Добуток ланцюгових темпів росту дорівнює базисному темпу росту за весь аналізований період: $T_{\text{р}}^{\text{ланц} 1} \cdot \dots \cdot T_{\text{р}}^{\text{ланц} n} = T_{\text{р}}^{\text{баз}}$
Коефіцієнт приросту	$K_{\text{пр}}^{\text{баз}} = K_{\text{р}}^{\text{баз}} - 1$ $K_{\text{пр}}^{\text{баз}} = \frac{y_i - y_0}{y_0}$ $K_{\text{пр}}^{\text{баз}} = \frac{\Delta_{\text{баз}}}{y_0}$	$K_{\text{пр}}^{\text{ланц}} = K_{\text{р}}^{\text{ланц}} - 1$ $K_{\text{пр}}^{\text{ланц}} = \frac{y_i - y_{i-1}}{y_{i-1}}$ $K_{\text{пр}}^{\text{ланц}} = \frac{\Delta_{\text{ланц}}}{y_{i-1}}$	
Темп приросту, %	$T_{\text{пр}}^{\text{баз}} = K_{\text{пр}}^{\text{баз}} \cdot 100 \%$ $T_{\text{пр}}^{\text{баз}} = \frac{\Delta_{\text{баз}}}{y_0} \cdot 100 \%$ $T_{\text{пр}}^{\text{баз}} = \frac{y_i - y_0}{y_0} \cdot 100 \%$ $T_{\text{пр}}^{\text{баз}} = T_{\text{р}}^{\text{баз}} - 100 \%$	$T_{\text{пр}}^{\text{ланц}} = K_{\text{пр}}^{\text{ланц}} \cdot 100 \%$ $T_{\text{пр}}^{\text{ланц}} = \frac{\Delta_{\text{ланц}}}{y_{i-1}} \cdot 100 \%$ $T_{\text{пр}}^{\text{ланц}} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100 \%$ $T_{\text{пр}}^{\text{ланц}} = T_{\text{р}}^{\text{ланц}} - 100 \%$	
Абсолютне значення одного відсотка приросту	$A_{\text{пр}}^{\text{баз}} = \frac{y_0}{100}$	$A_{\text{пр}}^{\text{ланц}} = \frac{y_{i-1}}{100}$	$A_{\text{пр}}^{\text{ланц}} = \frac{\Delta_{\text{ланц}}}{T_{\text{ланц}}}$ $A_{\text{пр}}^{\text{баз}} = \frac{\Delta_{\text{баз}}}{T_{\text{баз}}}$
Темп нарощування	$T_{\text{н}} = \frac{\Delta_{\text{ланц}}}{y_0} \cdot 100 \% = \frac{y_i - y_{i-1}}{y_0} \cdot 100 \%$		
Пункти росту	$\Pi_{\text{р}} = T_{\text{баз} i} - T_{\text{баз} i-1}$		

Приклад розрахунку аналітичних показників динаміки наведено в табл. 7.2



Таблиця 7.2 – Аналітичні показники ряду динаміки

Роки	Всього побудовано ЖБК, млн м <sup>2</sup>	Абсолютний приріст Δу, млн м <sup>2</sup>		Коефіцієнти або темпи зростання		Темпи приросту (відсотки)		Абсолютне значення одного відсотка приросту, тис. м <sup>2</sup>	Пункти росту, пункто-проценти
		Порівняно з 2009 р.	Порівняно з попереднім роком	Порівняно з 2009 р.	Порівняно з попереднім роком	Порівняно з 2009 р.	Порівняно з попереднім роком		
2009	2,9	–	–	–	–	–	–	–	–
2010	2,4	–0,5	–0,5	0,8276	0,8276	–17,24 %	–17,24 %	290	–17,24
2011	2,1	–0,8	–0,3	0,7241	0,8750	–27,59 %	–12,5 %	240	–10,35
2012	1,9	–1	–0,2	0,6552	0,9048	–34,48 %	–9,52 %	210	–6,89
2013	1,8	–1,1	–0,1	0,6207	0,9474	–38,93 %	–5,26 %	190	–3,45

Для дослідження інтенсивності явищ ряду динаміки використовується цілий ряд середніх показників. Система середніх показників динаміки містить:

- середній рівень ряду;
- середній абсолютний приріст (середня швидкість зростання) – розраховується як середня арифметична з показників швидкості зростання за певний період або за окремі проміжки часу ;
- середній темп росту;
- середній темп приросту.

*Середній рівень ряду* – це показник, що узагальнює підсумки розвитку явища за одиничний інтервал або момент із наявної тимчасової послідовності. Розрахунки середнього рівня ряду динаміки визначаються видом цього ряду і величиною інтервалу, відповідного до кожного рівня. Алгоритм розрахунку середнього рівня ряду залежить від того, який це ряд (інтервальний чи моментний), а також які інтервали він містить (рівні чи нерівні).

Середній рівень *динамічного інтервального ряду з рівними періодами часу* розраховується за формулою

$$\bar{y} = \frac{\sum y_i}{n}, \quad (7.1)$$

де  $y_i$  – рівні ряду;  $n$  – число рівнів.

Якщо рівні *інтервального ряду динаміки змінюються нерівномірно*, то для розрахунків середніх хронологічних доцільно використовувати формулу

$$\bar{y} = \frac{\sum y_t t_i}{\sum t_i}, \quad (7.2)$$

де  $Y_t$  – рівень ряду динаміки в конкретний момент часу;  $t_i$  – тривалість періоду часу в плинні якого даний рівень не змінюється.

Середня хронологічна моментного ряду з рівновіддаленими моментами визначається за формулою

$$\bar{Y} = \frac{\frac{1}{2} Y_1 + Y_2 + Y_3 + \dots + Y_{n-1} + \frac{1}{2} Y_n}{n-1}. \quad (7.3)$$

Для динамічних моментних рядів з *нерівновіддаленими моментами* розрахунок здійснюють у два етапи:

1) визначають середню для кожного проміжку часу як напівсуму двох сусідніх рівнів ряду

$$\bar{Y}_1 = \frac{Y_0 + Y_1}{2}, \quad \bar{Y}_2 = \frac{Y_1 + Y_2}{2}, \quad \dots, \quad \bar{Y}_n = \frac{Y_{n-1} + Y_n}{2}, \quad (7.4)$$

2) визначають загальний середній рівень ряду, тобто розраховують середню з отриманих на першому етапі результату за формулою середньої арифметичної зваженої

$$\bar{Y} = \frac{\sum_1^n \bar{Y}_i t_i}{\sum_1^n t_i} \quad (7.5)$$

де  $t_i$  – тривалість  $i$ -го проміжку часу.

Також для моментних рядів з нерівними інтервалами використовують середню хронологічну зважену

$$\bar{Y}_{\text{хр.зв.}} = \frac{(Y_1 + Y_2)t_1 + (Y_2 + Y_3)t_2 + \dots + (Y_{n-1} + Y_n)t_{n-1}}{2 \sum t_{n-1}}. \quad (7.6)$$

Але частіше в таких випадках використовують середню арифметичну зважену (7.5).

Вибір формули для розрахунку середнього рівня ряду визначається характером вихідних даних, при цьому чисельник повинен мати реальний зміст.

Основні формули для розрахунку середніх показників динаміки наведено в табл. 7.3.

Таблиця 7.3 – Середні показники динамічних рядів

Середній показник	Базисний	Ланцюговий
Середній абсолютний приріст	$\bar{\Delta}_{\text{баз}} = \frac{\Delta_{\text{баз}}}{n-1} = \frac{y_n - y_1}{n-1}$ $n$ – кількість періодів	$\bar{\Delta}_{\text{ланц}} = \frac{\sum \Delta_{\text{ланц}}}{n}$ $n$ – кількість ланцюгових приростів за період
Середній коефіцієнт росту	$\overline{K_p^{\text{баз}}} = \sqrt[n]{K_p^{\text{баз}}}$	$K_p^{\text{ланц}} = \sqrt[n]{\prod K_p^{\text{ланц}}}$
Середній темп зростання, %	$\overline{T_{\text{зр}}^{\text{баз}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$ $n$ – кількість періодів або $\overline{T_{\text{зр}}^{\text{баз}}} = \overline{K_p^{\text{баз}}} \cdot 100 \%$	$\overline{T_{\text{зр}}^{\text{ланц}}} = \sqrt[n]{T_1 \cdot T_2 \cdot T_3 \cdot T_4 \cdot \dots \cdot T_n}$ $n$ – кількість ланцюгових темпів зростання або $\overline{T_{\text{зр}}^{\text{ланц}}} = \overline{K_p^{\text{ланц}}} \cdot 100 \%$
Середній коефіцієнт приросту	$\overline{K_{\text{пр}}^{\text{баз}}} = \overline{K_p^{\text{баз}}} - 1$	$\overline{K_{\text{пр}}^{\text{ланц}}} = \overline{K_p^{\text{ланц}}} - 1$
Середній темп приросту, %	$\overline{T_{\text{пр}}^{\text{баз}}} = \overline{T_{\text{зр}}^{\text{баз}}} - 100 \%$	$\overline{T_{\text{пр}}^{\text{ланц}}} = \overline{T_{\text{зр}}^{\text{ланц}}} - 100 \%$

Середній абсолютний приріст, а також середні темпи росту застосовуються в статистичному прогнозуванні явищ зі стабільною динамікою розвитку.

### 7.3. Методи обробки динамічних рядів

У процесі зміни явищ у часі поряд з кількісними змінами відбуваються процеси, що змінюють якісний зміст об'єкта дослідження. Основними причинами якісних змін є:

- 1) інфляція, коливання курсу валют;
- 2) зміна державних і адміністративних границь;
- 3) перехід на інші методології розрахунків порівнюваних показників;
- 4) використання інших одиниць виміру;
- 5) зміна критичного моменту або періоду реєстрації;
- 6) зміна переліку об'єктів, що входять до складу сукупності;
- 7) зміна споживчої вартості одиниць сукупності.

Безпосереднє порівняння рівнів динамічних рядів, не зведених до порівнянного виду, дає помилкові результати і призводять до непра-

вильних управлінських рішень. Існують різні способи зведення до порівнянності даних. Вплив інфляції і курсів валют усувають шляхом розподілу фактичних даних на відповідний базисний індекс (відносний показник динаміки) інфляції або курсів валют. Таким чином, ряд динаміки перераховується в порівнянні на (базисні) ціни і курси валют.

Рівні рядів динаміки в різних одиницях виміру перераховують в порівнянні одиниці. Найбільшу складність становить зведення до порівнянного виду показників, розрахованих за різними методиками. Складність полягає не тільки в додатковій трудомісткості перерахування рівнів минулих періодів за новою методикою, але і у відсутності для цього необхідної інформації.

При зміні адміністративно-територіальних границь і внаслідок інших причин, що відображаються у складі порівнюваних сукупностей, використовують метод змикання динамічних рядів, коли в період зміни наводяться одночасно два показники: у старих границях і в нових, і розраховується коефіцієнт співвідношення між ними, який застосовується потім для перерахування показників у старих границях до нових.

*Метод зімкнення рядів* – це об'єднання двох і більше рядів, що характеризують зміну того самого явища; він використовується тоді, коли показники динамічних рядів не можуть бути порівнянні. Змикання рядів проводять наступним чином: розраховують відношення останнього показника першого ряду до першого показника другого ряду і визначають коефіцієнт (для цього необхідно, щоб останній показник першого ряду і перший показник другого ряду мали однакові рівні, наприклад один період часу). Потім на цей коефіцієнт помножують всі рівні другого ряду або ділять всі рівні першого ряду (у міжнародній статистичній практиці прийнято визначати двома горизонтальними або вертикальними рисками показники року, на базі якого були зроблені ці розрахунки);

*Приклад.* Є дані про чисельність населення міста А, до якого в 2012 році було приєднано пригороди. Отримати новий ряд динаміки наявного населення міста після приєднання пригородів. Розрахувати показник зростання рівня населення (табл. 7.4).

Розрахунок коефіцієнта зростання:

$$K = \frac{121470}{103520} = 1,1734 \text{ або } 117,34 \% ;$$

$$95400 \cdot 1,1734 = 111942;$$

$$97888 \cdot 1,1734 = 114861.$$

Таблиця 7.4 – Формування нового ряду динаміки

	2008	2009	2010	2011	2012	2013
Без пригородів	95400	97888	103520	-	-	-
Після приєднання пригородів	-	-	121470	130456	132370	134500
Порівнянна чисельність	111942	114861	121470	130456	132370	134500

У випадку зміни споживчих властивостей об'єкта дослідження проводиться перерахування рівнів динамічного ряду в умовно-натуральні показники. Якщо склад сукупностей змінюється в результаті ціле-спрямованої діяльності для досягнення більш високих показників, ряди динаміки можуть не перераховуватись.

#### ***7.4. Виявлення тенденцій розвитку соціально-економічних явищ в динаміці***

Будь-який ряд динаміки теоретично може бути поданий як ряд, що містить такі складові:

- 1) тренд – основна тенденція розвитку динамічного ряду (до збільшення або зниження його рівнів);
- 2) циклічні (періодичні) коливання, у тому числі сезонні;
- 3) випадкові коливання.

Одним із завдань, які виникають при аналізі рядів динаміки, є встановлення закономірності зміни рівнів досліджуваного показника в часі. У деяких випадках ця закономірність, загальна тенденція розвитку об'єкта цілком ясно відображається рівнями динамічного ряду. В одних рядах динаміки рівням ряду властива тенденція до зниження, що не порушується протягом усього розглянутого періоду, в інших – систематичне збільшення рівнів ряду. Однак дуже часто доводиться досліджувати такі ряди динаміки, коли з їх рівнями відбуваються різноманітні зміни (то зростають, то зменшуються) і можна говорити лише про загальну тенденцію розвитку явища, або про тенденцію до росту, або до зниження. У цих випадках для визначення основної тенденції розвитку явища, досить стійкої протягом даного періоду, використовують особливі прийоми обробки рядів динаміки.

Рівні ряду динаміки формуються під сукупним впливом безлічі

довгостроково і короткочасно діючих факторів, у тому числі різних випадкових обставин. Виявлення основної закономірності зміни рівнів ряду забезпечує можливість їх кількісного вираження, деякою мірою вільного від випадкових впливів.

Виявлення основної тенденції розвитку (тренда) називається в статистиці також вирівнюванням тимчасового ряду, а методи виявлення основної тенденції – методами вирівнювання. Вирівнювання дозволяє характеризувати особливості змін в часі даного динамічного ряду в найбільш загальному вигляді як функції часу, припускаючи, що через час можна виразити вплив усіх основних факторів.

Таким чином, одним з головних завдань статистичного дослідження динаміки є визначення загальної тенденції розвитку динамічного ряду в часі або тренда.

*Тренд* (фактор часу) розглядається як сукупний результат дії безлічі різних причин, які умовно поєднуються в одну причину. Вважається, що лінія тренда може бути опуклою, ввігнутою або прямою. Але вона не повинна мати хвилеподібну форму, яку прийнято вважати результатом циклічних змін соціальних і економічних показників. Крім того, тренд не повинен змінювати напрямок протягом приблизно 10 років.

*Тенденція (або тренд)* – це основний напрямок розвитку явища, яке підлягає дослідженню (приклад графічного відображення тренда наведено на рис. 7.1).

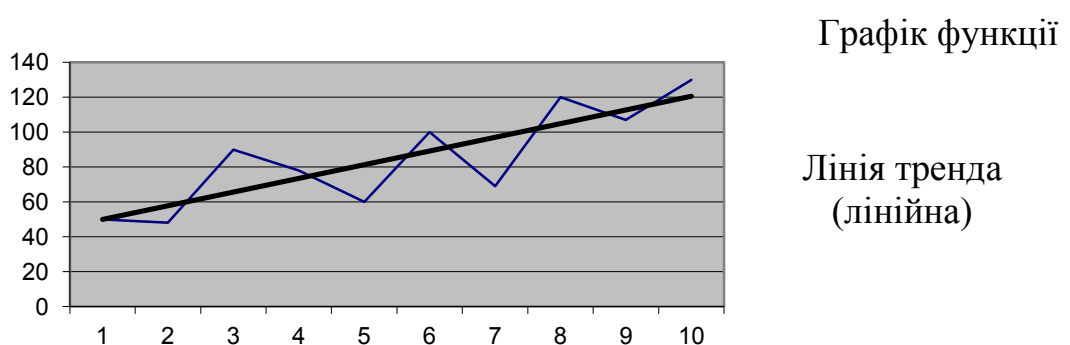


Рисунок 7.1 – Приклад графічного відображення тренда

Вивчення тренда включає два основні етапи:

- 1) перевірка ряду динаміки на наявність тренда;
- 2) вирівнювання тимчасового ряду і здійснення безпосереднього виділення тренда з екстраполяцією отриманих результатів.

Перевірка на наявність тренда в ряді динаміки може бути здійснена за декількома критеріями, включаючи:

– *Метод середніх*. Досліджуваний ряд динаміки розбивається на декілька інтервалів (звичайно на два), для кожного з яких визначається середня величина ( $\bar{y}_1, \bar{y}_2$ ). Висувається гіпотеза про істотну відмінність середніх. Якщо ця гіпотеза ухвалюється, то визнається наявність тренда.

– *Фазочастотний критерій знаків першої різниці (критерій Валліса і Мура)*. Сутність його полягає у такому: наявність тренда в динамічному ряді підтверджується, якщо цей ряд не містить, або містить у прийнятній кількості фази – зміни знака різниці першого порядку (абсолютного ланцюгового приросту).

– *Критерій Кокса і Стюарта*. Увесь аналізований ряд динаміки розбивають на три рівні за кількістю рівнів групи (якщо кількість рівнів ряду динаміки не ділиться на три, то недостатні рівні потрібно додати) і порівнюють між собою рівні першої і останньої груп.

– *Метод серій*. При цьому способі кожний конкретний рівень тимчасового ряду вважається таким, що належить до одного з двох типів: «А» або «В». Наприклад, якщо рівень ряду менший від медіанного значення, то вважається, що він має тип «А», а якщо ні, – то тип «В». Таким чином послідовність рівнів тимчасового ряду виступає як певна послідовність типів.

Розглянемо застосування методу серій на прикладі 1 (табл. 7.5).

Таблиця 7.5 – Число зареєстрованих шлюбів в Україні на 1000 жителів

Рік	Число зареєстрованих шлюбів, %	Рік	Число зареєстрованих шлюбів, %
1999	11,2	2006	9,6
2000	10,9	2007	9,7
2001	10,7	2008	9,8
2002	10,6	2009	9,9
2003	10,6	2010	9,5
2004	10,4	2011	9,4
2005	10,4	2012	9,1

Так, часовий ряд рівнів шлюбності має після впорядкування за зростанням на 7-му місці значення 10,4 і на 8-му місці – значення 9,6. Звідси медіана ряду становить  $(9,6 + 10,4) / 2 = 20 / 2 = 10,0$ .

Ряд типів виглядає так: «BBBBBBVAAAAAAAAA»

У послідовності типів, що утворилась, визначається кількість серій. *Серією* називається будь-яка послідовність елементів однакового типу, що граничить із елементами іншого типу. У даному прикладі кількість серій  $R$  дорівнює 2.

Розглянемо застосування методу серій на прикладі 2 (табл. 7.6).

Даний ряд має таку послідовність типів: «AABVVAVAVVVV»

У даному прикладі кількість серій становить  $R = 6$ .

Таблиця 7.6 – Обсяг продажів акцій на 15 найбільших біржах США за 2013 рік (млрд дол.)

МІСЯЦЬ	Рівень показника ( $B^t$ )	Послідовність типів
Січень	212,78	A
Лютий	122,08	A
Березень	709,98	B
Квітень	1602,61	B
Травень	651,83	B
Червень	220,80	A
Липень	327,68	B(274,24 = Me)
Серпень	217,12	A
Вересень	418,31	B
Жовтень	521,18	B
Листопад	396,2	B
Грудень	508,34	B

Якщо в динамічному ряді загальна тенденція до росту або зниження відсутня, то кількість серій є випадковою величиною, розподіленою приблизно за нормальним законом. Отже, якщо закономірності в змінах рівнів немає, то випадкова величина  $R$  буде знаходитись в довірчому інтервалі

$$\text{— — — —} \quad (7.7)$$

Параметр  $t$  призначається відповідно до прийнятого рівня довірчої інформації  $P$ .

Наприклад:

P	0,683	0,950	0,954	0,990	0,997
T	1	1,960	2	2,576	3
Середнє число серій					



$$\frac{1}{n} \sum_{i=1}^n (R_i - \bar{R})^2 \quad (7.8)$$

Середнє квадратичне відхилення числа серій

$$\frac{1}{n} \sum_{i=1}^n R_i \quad (7.9)$$

де  $n$  – число серій.

Вираз для довірчого інтервалу набуває вигляду

$$\bar{R} \pm t_{\alpha/2} \sqrt{\frac{1}{n} \sum_{i=1}^n (R_i - \bar{R})^2} \quad (7.10)$$

Отримані границі довірчого інтервалу округлюють до цілих чисел, зменшуючи нижню границю і збільшуючи верхню. У нашому випадку (для  $P = 0,954$ ) маємо такі результати:

для рівнів шлюбності (приклад 1):  $-3 \leq R \leq 12$ ;

для ряду обсягу продажів акцій (приклад 2):  $-3 \leq R \leq 10$ .

Як видно з прикладів 1 та 2, для ряду динаміки рівня шлюбності показник числа серій  $R = 2$  виходить за межі можливої випадкової поведінки, а отже, у зміні рівнів ряду є загальна закономірність, тенденція. Навпаки, для ряду обсягів продажу акцій число серій  $R = 6$  (з  $P = 0,954$ ) цілком укладається в межі випадкової поведінки і гіпотеза про наявність загальної закономірності зниження або зростання обсягів продажів у часі не може бути прийнята (з ймовірністю помилки 0,046).

Для безпосереднього виділення тренда існує декілька способів, вибір яких визначається метою дослідження і специфікою досліджуваного явища. Це такі способи:

- укрупнення інтервалу;
- ковзної середньої;
- аналітичного вирівнювання.

Сутність кожного зі способів полягає у згладжуванні випадкових одноразових коливань для виявлення загальної тенденції розвитку.

Розглянемо кожний спосіб окремо:

1) *Метод укрупнення інтервалів* – це підсумовування рівнів ряду за більш короткі проміжки часу з метою заміни їх більшими. Ряд динаміки розділяють на деяку достатньо велику кількість рівних інтервалів. Якщо середні рівні за інтервалами не дозволяють побачити тенденцію розвитку явища, то переходять до розрахунків рівнів за більші проміжки часу, збільшуючи довжину кожного інтервалу (одночасно зменшуючи

кількість інтервалів).

Наприклад, ряд, що містить дані про місячний випуск продукції, може бути перетворений у ряд квартальних даних. Знову утворений ряд може містити або абсолютні величини за укрупнені за тривалістю проміжки часу (ці величини одержують шляхом простого підсумовування рівнів первісного ряду абсолютних величин), або середні величини. При підсумовуванні рівнів або при виведенні середніх за укрупненими інтервалами, відхилення в рівнях, обумовлені випадковими причинами, взаємопогашаються, згладжуються і більш чітко виявляється дія основних факторів зміни рівнів (загальна тенденція).

Принцип цього прийому полягає в тому, що дані динамічного ряду об'єднують в групи за періодами, і для них розраховують середній показник на період 3, 5, 10 і більше років.

Розглянемо приклад застосування методу укрупнення інтервалів. На рис.7.2 наведено схему реалізації методу укрупнення інтервалів. Отже, згідно рис. 7.2, отримано дві точки для побудови лінії тренда і можна зробити висновок про загальну тенденцію ряду до зростання.

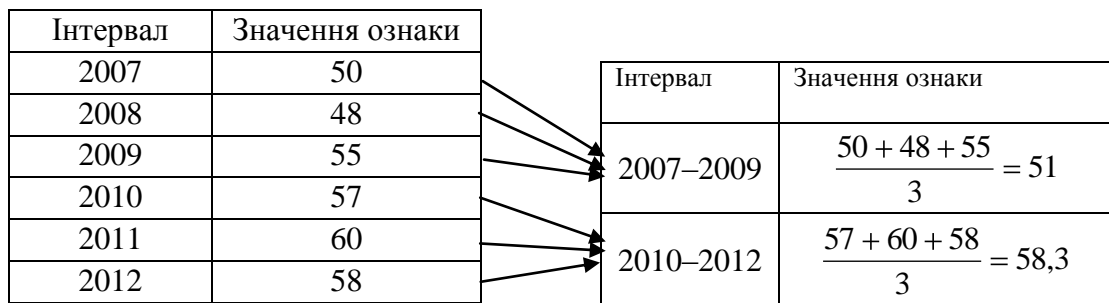


Рисунок 7.2 – Схема реалізації методу укрупнення інтервалів

2) *Спосіб ковзної середньої* – передбачає послідовне усереднення деякого постійного числа рівнів динамічного ряду за формулою простої середньої арифметичної. Число членів ковзної середньої звичайно прямо пропорційне чисельності і інтенсивності коливань рівнів динамічного ряду. У цьому методі вихідні рівні ряду замінюються середніми величинами, що їх одержують із даного рівня і декількох рівнів, що симетрично його оточують. Ціле число рівнів, за якими розраховується середнє значення, називають інтервалом згладжування. Для визначення ковзної середньої формуються укрупнені інтервали, які складаються з однакової кількості рівнів. Кожний наступний інтервал одержується, поступово зрушуючись від початкового рівня динамічного ряду на один рівень.

Тоді перший інтервал буде включати рівні  $Y_1, Y_2, \dots, Y_m$ ; другий – рівні  $Y_2, Y_3, \dots, Y_{m+1}$  і т. д. У такий спосіб інтервал згладжування нібито сковзає по динамічному ряду з інтервалом, що дорівнює 1. За сформованими укрупненими інтервалами визначають суму значень рівнів, на основі яких розраховуються ковзні середні. Отримана середня міститься посередині укрупненого інтервалу. Тому при згладжуванні ковзної середньої технічно зручніше укрупнений інтервал формувати з непарного числа рівнів ряду.

Для визначення ковзної середньої формують укрупнені інтервали, які складаються з однакового числа рівнів. Але за допомогою послідовних зсувів на одну дату (місяць, квартал, рік) абсолютні дані замінюють сумами (середніми) за визначені періоди (тобто 3, 5, 10 років).

Чим триваліший інтервал згладжування, тем сильніше осереднення, а виявлена тенденція розвитку виходить більш плавною. Найчастіше інтервал згладжування може складатися з трьох, п'яти або семи рівнів. Визначення основної тенденції розвитку методом ковзної середньої є лише емпіричним прийомом методу укрупнення інтервалів.

Приклад схеми реалізації методу ковзної середньої наведено на рис. 7.3.

Інтервал	Значення ознаки		Значення ковзної середньої ознаки
2007	50	→	49
2008	48		51,5
2009	55		56
2010	57		58,5
2011	60		59
2012	58		

Рисунок 7.3 – Схема реалізації методу ковзної середньої

Укрупнення інтервалу і метод ковзної середньої можуть розглядатися як важливий допоміжний засіб, який полегшує застосування інших методів, зокрема, більш строгих методів виявлення тенденції. Для того щоб подати кількісну модель, що виражає загальну тенденцію зміни рівнів динамічного ряду в часі, використовують аналітичне вирівнювання.

3) Аналітичне вирівнювання – це вибір рівняння прямої або кривої лінії, яка адекватно виражає загальну тенденцію розвитку динамічного ряду, і розрахунки параметрів цього рівняння здійснюються найчастіше

методом найменших квадратів. При виборі рівняння функції керуються специфікою досліджуваного явища та рядом формальних ознак. У цьому випадку фактичні рівні замінюються рівнями, обчисленими на основі певної кривої. Передбачається, що вона відображає загальну тенденцію змін в часі досліджуваного показника. Відхилення конкретних рівнів ряду від рівнів, відповідних до загальної тенденції, пояснюють дією факторів, що виявляються випадково або циклічно. У результаті приходять до трендової моделі

$$(7.11)$$

де  $f(t)$  – рівень, обумовлений тенденцією розвитку;  $E_t$  – випадкове і циклічне відхилення від тенденції.

Метою аналітичного вирівнювання динамічного ряду є визначення аналітичної або графічної залежності  $f(t)$ . На практиці за наявним рядом задають вид і знаходять параметри функції  $f(t)$ , а потім аналізують поведінку відхилень від тенденції. Функцію  $f(t)$  вибирають таким чином, щоб вона давала змістовне пояснення досліджуваного процесу.

Найчастіше при аналітичному вирівнюванні використовуються такі залежності:

А) лінійна –  $f(t) = a_0 + a_1t$  – використовується в тих випадках, коли у вихідному часовому ряді спостерігаються більш-менш постійні абсолютні ланцюгові прирости, які не виявляють тенденцію ні до збільшення, ні до зниження. Використовується для апроксимації даних методом найменших квадратів відповідно до рівняння.

Б) параболічна –  $f(t) = a_0 + a_1t + a_2t^2$  – використовується, якщо абсолютні ланцюгові прирости виявляють деяку тенденцію розвитку, але абсолютні ланцюгові прирости абсолютних ланцюгових приростів (різниці другого порядку) ніякої тенденції до розвитку не виявляють.

В) поліноміальна –  $f(t) = b + c_1x + c_2x^2 + c_3x^3 + \dots + c_6x^6$ , де  $b$  і  $c_1..c_6$  – константи.

Г) логарифмічна –  $y = c \ln x + b$ , де  $c$  і  $b$  – константи,  $\ln$  – функція натурального логарифма.

Д) експоненціальна залежність –  $y = ce^{bx}$ , де  $c$  і  $b$  – константи,  $e$  – основа натурального логарифма – застосовується, якщо у вихідному тимчасовому ряді спостерігається, або більш-менш постійне відносне зростання (стабільність ланцюгових темпів росту, темпів приросту, коефіцієнтів росту), або, за відсутності такої сталості, – стабільність у

зміні показників відносного росту.

Е) степенева –  $y = cx^b$ , де  $c$  і  $b$  – константи.

Оцінка параметрів  $a_0, a_1, a_2, \dots$  здійснюється такими методами:

- вибраних точок;
- найменших відстаней;
- найменших квадратів (МНК).

У більшості розрахунків використовується метод найменших квадратів, який забезпечує найменшу суму квадратів відхилень фактичних рівнів від вирівняних

$$\min \sum (y_t - f(t))^2. \quad (7.12)$$

Наприклад, якщо для розвитку явища характерні достатньо стабільні абсолютні ланцюгові прирости (тобто  $\Delta_{\text{ланц}} \approx \text{const}$ ), то вибирається рівняння лінійного тренду  $y_t = a_0 + a_1 t$ .

Якщо абсолютні ланцюгові прирости з часом поступово скорочуються, то для характеристики тренда застосовується напівлогарифмічна крива  $y_t = a_0 + a_1 \lg t$ .

Якщо явище розвивається з достатньо стабільними ланцюговими темпами росту, то для характеристики тренда застосовується показова функція  $y_t = a_0 \cdot a_1^t$ .

Якщо ланцюгові темпи приросту приблизно постійні ( $\Delta t \approx \text{const}$ ), то використовується парабола другого порядку  $y_t = a_0 + a_1 t + a_2 t^2$ .

З безлічі різноманітних функцій тренда з формально математичної точки зору найкращою вважається та, яка найменш вилучена від емпіричних рівнів ряду:  $\sum (y_t - y_i)^2 \rightarrow \min$ .

*Статистичні методи прогнозування.* Результати аналізу динамічних рядів використовуються для прогнозування шляхом екстраполяції, тобто знаходження рівнянь за границями динамічного ряду.

Існує короткострокове, середньострокове і довгострокове прогнозування. Поняття терміну прогнозування пов'язане зі специфікою досліджуваного явища. Для прогнозування валютних курсів довгостроковим є прогноз у межах 1 року, у той час як розвиток економіки здійснюється в довгостроковому плані на 5 і більш років. Короткостроковим є прогноз до 1 року, середньостроковим – до 3 років.

Залежно від строків прогнозування і особливостей розвитку явища

в прогностичний період використовують різні підходи. Якщо для явища (ряду динаміки) були характерні достатньо стабільні ланцюгові прирости (абсолютні), то прогнозування здійснюється за формулою

$$\bar{y}_{n+I} = y_n + I \cdot \bar{\Delta}, \quad (7.13)$$

де  $y_n$  – кінцевий рівень динамічного ряду;  $I$  – строк прогнозування;  $\bar{\Delta}$  – середньорічний абсолютний приріст.

Якщо для явища були характерні достатньо стабільні ланцюгові темпи росту, то прогнозування здійснюється за формулою

$$\bar{y}_{n+I} = y_n \cdot \bar{r}^I, \quad (7.14)$$

де  $\bar{r}$  – середній темп росту.

Найбільш точним і складним є прогнозування з використанням різних рівнянь трендів.

### **7.5. Дослідження сезонних коливань в рядах динаміки**

Змінюючись під спільним впливом систематичних і випадкових факторів, рівень ряду динаміки зазнає також впливів причин, обумовлених періодичністю коливань. Таким чином, при аналізі коливності динамічних рядів поряд з виділенням випадкових коливань виникає і проблема вивчення періодичних сезонних коливань.

У широкому розумінні до сезонних відносяться всі явища, які виявляють у своєму розвитку чітко виражену закономірність внутрішніх змін, тобто більш-менш повторювані рік у рік коливання рівнів. Виявлення сезонності завдають великої шкоди економіці, що пов'язано з нерівномірним завантаженням транспорту, використанням устаткування, робочої сили і т. ін.

*Сезонні коливання* – це стійкі внутрішньорічні коливання, які повторюються. Вони обумовлені природно-кліматичними і іншими факторами, які визначають нерівномірність виробництва і споживання в часі.

Найважливішими завданнями, які вирішуються в ході дослідження сезонності, є такі:

1. Визначення наявності сезонності, чисельний вираження прояву сезонних коливань і виявлення їх чинності і характеру в різних фазах річного циклу.

2. Характеристика факторів, що викликають сезонні коливання.

3. Оцінка наслідків, до яких приводить наявність сезонних коливань.

4. Математичне моделювання сезонності.

Для вимірювання сезонних коливань статистикою запропоновано різні методи. Найбільш прості і найчастіше вживані з них – це:

А) Метод абсолютних різниць – загальна середня  $\bar{y}$  визначається відношенням суми рівнів на їхню кількість. Потім визначається абсолютне відхилення рівнів ряду від загальної середньої

$$\bar{y} = \frac{\sum y_t}{n} \quad (7.15)$$

Б) Метод відносних різниць – метод відносних різниць є розвитком методу абсолютних різниць. Для знаходження відносних різниць абсолютні відхилення ділять на загальну середню і виражають у відсотках

$$\frac{y_t - \bar{y}}{\bar{y}} \quad (7.16)$$

При вивченні динаміки явищ виділяють звичайно чотири групи причин, що обумовлюють розмір і характер зміни рівнів ряду динаміки.

$$y = y_\epsilon + y_s + y_c + y_t, \quad (7.17)$$

де  $y_\epsilon$  – випадковий компонент;  $y_s$  – сезонний компонент;  $y_c$  – циклічна складова;  $y_t$  – тренд.

Логіка статистичного дослідження динамічного ряду полягає в послідовному визначенні окремих складових частин  $((y - y_t) - \text{адитивна модель})$ . Однак на практиці частіше застосовується виключення факторів не методом різниць, а методом співвідношень  $(y / y_t)$ .

Це дозволяє при послідовному проведенні аналізу виражати отримані на кожному етапі результати в порівнянному масштабі. Тобто ми заміняємо адитивну модель на мультиплікативну.

Якщо трендова складова визначається за однією з розглянутих

функцій, то циклічна складова розраховується звичайно за синусо-косинусоїдальною функцією (гармонікою Фур'є):  $y_c = a_0 + a_1 \cos kt + a_2 \sin kt$ , причому величина  $k$  – це ціле число, яке встановлюється прямо пропорційно інтенсивності циклічних коливань. Після визначення циклічної складової, розрахунок якої в умовах ринкової економіки, що розвивається, має важливе значення, визначається сезонний компонент.

Перші два способи припускають визначення різниць фактичних рівнів і рівнів, знайдених при виявленні основної тенденції розвитку. Застосовуючи метод абсолютних різниць, оперують безпосередньо розмірами цих різниць, а при використанні методу відносних різниць визначають відношення абсолютних розмірів зазначених різниць до вирівняного рівня. При виявленні основної тенденції використовують або метод ковзної середньої, або аналітичне вирівнювання. У деяких випадках у стаціонарних рядах можна використовувати різницю фактичних рівнів і середньомісячного рівня за рік.

Замість відносних різниць щомісяця можуть бути обчислені індекси сезонності.

В) Побудова індексів сезонності. *Індекс сезонності* показує, у скільки разів фактичний рівень ряду в момент або інтервал часу  $t$  більший від середнього рівня або рівня, що обчислюється за рівнянням тенденції  $f(t)$ . При аналізі сезонності рівні динамічного ряду показують розвиток явища по місяцях (кварталах) одного або декількох років. Для кожного місяця (кварталу) одержують узагальнений індекс сезонності як середню арифметичну з однойменних індексів кожного року.

Рівень сезонності оцінюється за допомогою :

- 1) індексів сезонності ;
- 2) гармонійного аналізу.

*Індекси сезонності* – це відносні величини координації, коли за базу порівняння прийнятий або середній рівень ряду, або рівень тенденції.

Якщо тренда немає або він незначний, то для кожного місяця (кварталу)

$$I_{t, \text{сез}} = \frac{y_t}{\bar{y}} \cdot 100 \%, \quad (7.18)$$

де  $y_t$  – рівень показника  $t$  за місяць (квартал);  $\bar{y}$  – загальний середній рівень показника  $t$ .



Для забезпечення стабільності показників можна обрати більший проміжок часу. У цьому випадку

$$I_{t, \text{ сез}} = \frac{\bar{y}_t}{y}, \text{ або } I_{t, \text{ сез}} = \frac{\sum i_{t, \text{ сез}}}{T}, \quad (7.19)$$

де  $\bar{y}_t$  – середній рівень показника по однойменних місяцях за ряд років;  $T$  – кількість років.

За наявності тренда індекс сезонності визначається на основі методів, які виключають вплив тенденції. Порядок розрахунків такий:

1) для кожного рівня визначають вирівняні значення за трендом  $f(t)$ ;

2) розраховують відношення  $i_t = \frac{y_t}{f(t)}$  ;

3) за необхідності визначають середнє з цих відношень для однойменних місяців (кварталів)

$$I_{t, \text{ сез}} = \frac{i_t^1 + i_t^2 + \dots + i_t^T}{T}. \quad (7.20)$$

Іншим методом вивчення рівня сезонності є гармонійний аналіз. Його виконують, подаючи часовий ряд як сукупність гармонійних коливальних процесів. Для кожної точки цього ряду формується вираз, записаний у вигляді формули

$$\text{—} \quad \text{—} \quad (7.21)$$

де  $t = 1, 2, 3, \dots, T$ ;  $y_t$  – фактичний рівень ряду в момент (інтервал) часу  $t$ ;  $f(t)$  – вирівняний рівень ряду в той же момент (інтервал)  $t$ ;  $a, b$  – параметри коливального процесу (гармоніки) з номером  $n$ , у сукупності з урахуванням розмаху (амплітуди) відхилення від загальної тенденції й зрушення коливальних відносно початкової точки.

Загальна кількість коливальних процесів, які можна виділити з ряду, що складається з  $T$  рівнів, дорівнює  $T/2$ . Зазвичай обмежуються меншою кількістю найбільш важливих гармонік.

Визначення сезонних коливальних дозволяє здійснити раціональне

внутрішньорічне і внутрішньомісячне планування, уникнути непотрібних втрат і використати всі наявні можливості. У більшості випадків статистичне дослідження рядів динаміки за короткі проміжки часу зводяться до вивчення сезонних коливань. Індикатором сезонних коливань є індекс сезонності, який визначається за формулою

$$I_{s_i} = \frac{y_i}{\bar{y}_i} \cdot 100 \%, \quad (7.22)$$

де  $y_i$  та  $\bar{y}_i$  – відповідно фактичне і вирівняне значення рівня динамічного ряду в  $i$ -й момент часу або в  $i$ -й періоді часу.

Залежно від способу вирівнювання вихідних даних розрізняють методи розрахунків індексу сезонності за простою середньою, ковзною середньою і аналітичного вирівнювання.

Розглянемо на прикладі використання цих методів застосування.

*Приклад 1.* Розрахунки індексів сезонності товарообігу за методом простої середньої. Вихідні дані для аналізу сезонності товарообігу наведено в табл. 7.7.

Таблиця 7.7 – Вихідні дані для аналізу сезонності товарообігу

Квартали	Товарообіг по роках, тис. грн			Середньоквартальні рівні товарообігу, тис. грн	Індекси сезонності, %
	2010	2011	2012		
1	11561	11919	12446	11975	102,9
2	8786	8832	9484	9034	77,6
3	10764	11323	11712	11266	96,8
4	13993	14176	14624	14264	122,6
Разом	45104	46250	48266	11635	-

Визначимо середньоквартальний рівень за 1-й квартал:

$$\bar{y}_1 = \frac{y_1^{10} + y_1^{11} + y_1^{12}}{3} = \frac{11561 + 11919 + 12446}{3} = 11975 \text{ тис. грн};$$

$$\bar{y}_2 = 9034 \text{ тис. грн};$$

$$\bar{y}_3 = 11266 \text{ тис. грн};$$

$$\bar{y}_4 = 14264 \text{ тис. грн}.$$

Середньоквартальний рівень за всі роки становить

$$\bar{y} = \frac{\sum y_i}{n} = \frac{45104 + 46250 + 48266}{12} = 11635 \text{ тис. грн}.$$

Індекси сезонності:

$$y_{s_1} = \frac{\bar{y}_1}{\bar{y}} \cdot 100 \% = \frac{11975}{11635} \cdot 100 \% = 102,9 \%;$$

$$y_{s_2} = 77,6 \%;$$

$$y_{s_3} = 96,8 \%;$$

$$y_{s_4} = 122,6 \%.$$

Індекси сезонності показують, що в 1 кварталі товарів продається приблизно на 2,9 % більше від середньоквартального рівня. У другому – на 22,4 % менше. У третьому – на 3,2 менше, а в четвертому – на 22,6 % більше від середньоквартального рівня. Отримані показники доцільно використовувати для внутрішньоквартального планування річного товарообігу.

Метод розрахунків індексів сезонності за простою середньою у розрахунку достатньо точний у випадках, коли аналізовані явища не мають стійкої інтенсивної тенденції росту або зменшення в часі. В інших випадках застосовують розрахунки індексу сезонності за ковзною середньою або за допомогою аналітичного вирівнювання.

*Приклад 2.* Розрахунки індексу сезонності за методом ковзної середньої (чотиричленної). На основі даних табл. 7.7 складемо табл. 7.8.

Таблиця 7.8 – Аналіз сезонності товарообігу методом ковзної середньої

Роки	Квартали	Товарообіг, тис. грн	Ковзна чотиричленна сума, тис. грн	Ковзна чотири- членна середня, тис. грн	Індекси сезонності, %
2010	1	11561	45104	11321	95,1
	2	8786			
	3	10764	45462		
	4	13993	45580	11371	123,1
2011	1	11919	46067	11447	<b>104,1</b>
	2	8832	46250	11540	76,5
	3	11323	46777	11628	97,4
	4	14176	47429	11776	120,4
2012	1	12446	47818	11906	<b>104,5</b>
	2	9484	48266	12011	79,0
	3	11712			
	4	14624			

Визначимо ковзні чотиричленні суми і середні, та індекси сезонності для кварталів 2010 року:

$$11561 + 8786 + 10764 + 13993 = 45104 \text{ тис. грн;}$$

$$8786 + 10764 + 13993 + 11919 = 45462 \text{ тис. грн;}$$

$$\left\langle \begin{matrix} 45104 \\ 45462 \end{matrix} \right\rangle 8 = 11321 \text{ тис. грн;}$$

$$\left\langle \begin{matrix} 45462 \\ 45580 \end{matrix} \right\rangle 8 = 11371 \text{ тис. грн;}$$

$$\frac{10764}{11321} \cdot 100 = 95,1 \% ; \quad \frac{13993}{11371} \cdot 100 = 123,1 \%$$

Далі визначимо індекс сезонності для кожного кварталу. Отримані індекси сезонності для кожного року і кварталу використовуються для розрахунків середніх індексів для кожного кварталу за методом простої середньої:

$$\bar{y}_{s_1} = \frac{y_{s_1}^{2010} + y_{s_1}^{2011} + y_{s_1}^{2012}}{n} = \frac{104,1 + 104,5}{2} = 104,3 \%$$

$$\bar{y}_{s_2} = \frac{76,5 + 79}{2} = 77,8 \%$$

$$\bar{y}_{s_3} = \frac{95,1 + 97,4}{2} = 96,3 \%$$

$$\bar{y}_{s_4} = \frac{123,1 + 120,4}{2} = 121,6 \%$$

*Приклад 3.* Визначення індексу сезонності методом аналітичного вирівнювання. За тенденцію розвитку товарообігу вибираємо лінійний тренд вигляду  $y_t = a_0 + a_1 t$ . Для розрахунків параметрів тренда використовуємо систему рівнянь

$$\begin{cases} \sum y = na_0 + a_1 \sum t; \\ \sum yt = a_0 \sum t + a_1 \sum t^2. \end{cases} \quad (7.23)$$

Оскільки, показник часу  $t$  являє собою ряд чисел, кожне з яких на 1 більше від попереднього, то систему рівнянь можна буде спростити штучно, підібравши ряд  $t$  таким чином, щоб сума  $t$  дорівнювала 0 ( $\sum t = 0$ ). У цьому випадку отримуємо наступні системи рівнянь:

$$\begin{cases} \sum y = na_0; \\ \sum yt = a_1 \sum t^2; \end{cases} \quad \begin{cases} a_0 = \frac{\sum y}{n}; \\ a_1 = \frac{\sum yt}{\sum t^2}. \end{cases} \quad (7.24)$$

Таблиця 7.9 – Аналіз сезонності товарообігу методом аналітичного вирівнювання

Роки	Квартали	Товарообіг, тис. грн	Умовні номери кварталів $t$	$y \cdot t$	$y_t$	Індекси сезон- ності $I$ , %
2010	1	11561	-11	-127171	10624	108,8
	2	8786	-9	-79074	10807	81,3
	3	10764	-7	-75348	10991	97,9
	4	13993	-5	-69965	11175	125,2
2011	1	11919	-3	-35757	11359	104,9
	2	8832	-1	-8832	11543	76,5
	3	11323	1	11323	11727	96,6
	4	14176	3	42528	11911	119,0
2012	1	12446	5	62230	12095	102,9
	2	9484	7	66388	12279	77,2
	3	11712	9	105408	12463	94,0
	4	14624	11	160864	12646	115,6

Для нашого прикладу (див. табл. 7.9):

$$a_0 = \frac{\sum y}{n} = \frac{139620}{12} = 11635;$$

$$\sum t^2 = \frac{n(n^2 - 1)}{3} = \frac{12(12^2 - 1)}{3} = 572;$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{52594}{572} = 91,95;$$

$$y_t = 11635 + 91,95 \cdot t.$$

Підставляючи в рівняння умовні значення  $t$ , одержимо теоретичні значення рівнів ряду динаміки  $y_t$ .

Далі за простою середньою розраховуємо середні індекси сезонності:

$$\bar{y}_{s_1} = \frac{y_{s_1}^{2010} + y_{s_1}^{2011} + y_{s_1}^{2012}}{n} = \frac{108,8 + 104,9 + 102,9}{3} = 105,5 \%;$$

$$\bar{y}_{s_2} = \frac{81,3 + 76,5 + 77,2}{3} = 78,3 \%;$$

$$\bar{y}_{s_3} = \frac{97,9 + 96,6 + 94,0}{3} = 96,2 \%;$$

$$\bar{y}_{s_4} = \frac{125,2 + 119,0 + 115,6}{3} = 119,9 \%.$$

Отримані індекси сезонності можна відобразити на графіку (див. рис. 7.4) у вигляді сезонної хвилі.

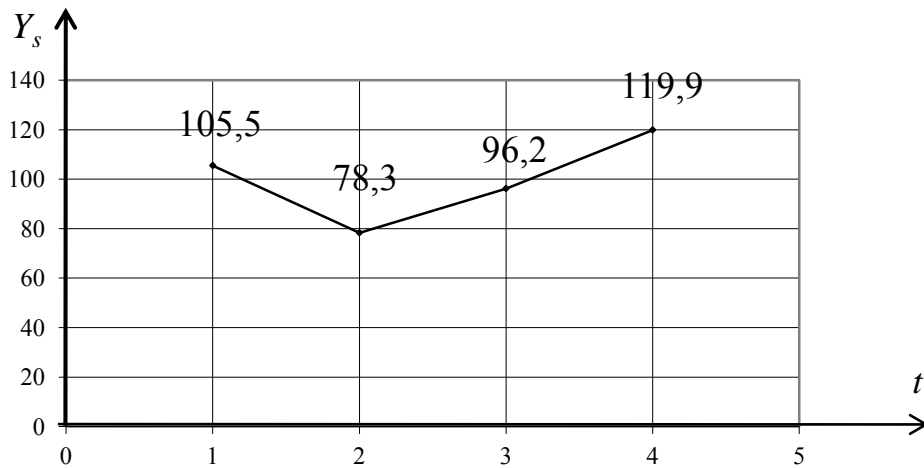


Рисунок 7.4 – Сезонна хвиля

### 7.6. Особливості дослідження взаємозв'язків в рядах динаміки

При аналізі рядів динаміки виникає необхідність дослідження взаємозв'язків між ознаками. Іноді досліджувати взаємозв'язки можна тільки в рядах динаміки. Це, в першу чергу, стосується багатofакторного кореляційного аналізу, коли число одиниць сукупності повинне не менш ніж у вісім разів перевищувати число факторів, включених у регресійну модель. Тому застосовується метод «заводо-років», коли аналізу підлягають динамічні ряди. Однак безпосереднє визначення тісноти зв'язку при цьому методі можливе тільки за відсутності автокореляції, тобто залежності наступних рівнів ряду від попередніх. Внаслідок автокореляції наявність синхронних коливань (тенденцій) розвитку рівнів двох показників може бути сприйнята як наявність зв'язку між ними.

Тому дослідження рядів динаміки завжди починається з визначення коефіцієнта автокореляції:

$$\frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t+1} - \bar{Y})}{\sqrt{\sum_{t=1}^{n-1} (Y_t - \bar{Y})^2 \sum_{t=1}^{n-1} (Y_{t+1} - \bar{Y})^2}}$$

(7.25)

Розраховані коефіцієнти автокореляції оцінюються на ймовірнісну надійність за допомогою критерію  $t$ -Стьюдента. Якщо фактична величина критерію  $t$  більша від табличної, то автокореляція має місце й розрахунки показників тісноти зв'язків можна здійснити одним зі спеціальних способів:

- корелювання відхилень від трендів;
- корелювання абсолютних різниць.

Коефіцієнт кореляції відхилень від трендів розраховується за формулою

$$\text{-----}, \tag{7.26}$$

де  $x_t, y_t$  – відповідно теоретичні значення рівнів факторної й результативної ознак, розраховані за допомогою рівнянь лінійних трендів виду

$$\tag{7.27}$$

де  $x_t, y_t$  – відповідно фактичні значення рівнів факторної й результативної ознак.

Для корелювання абсолютних різниць ланцюгові абсолютні прирости за факторними й результативними ознаками розраховується за формулами:

$$\tag{7.28}$$

а коефіцієнт кореляції – за формулою

$$\text{-----} \tag{7.29}$$

Деякі соціально-економічні явища або фактори впливають один на одного не відразу, а з деякою затримкою в часі, з тимчасовим лагом (запізнюванням). Наприклад, інвестиції в проект дають ефект після закінчення строку їх освоєння. Для визначення тісноти зв'язку подібних

явищ динамічні ряди факторної і результативної ознак зрушуються один до одного на величину динамічного лага.

Під взаємозалежними рядами динаміки розуміють такі, у яких рівні одного ряду якою мірою визначають рівні іншого. Наприклад, ряд, який відображає внесення добрив на 1 га, пов'язаний з динамічним рядом врожайності; ряд рівнів середнього виробітку пов'язаний з рядом динаміки середньої зарплати; ряд середньорічного поголів'я молочної худоби визначає річні рівні надоїв молока і т. д.

У найпростіших випадках для характеристики взаємозв'язку двох або більш рядів їх зводять до загальної основи, для чого беруть за базисні рівні за той самий період і обчислюють коефіцієнти випередження за темпами росту або приросту.

Коефіцієнти випередження за темпами росту – це відношення темпів росту (ланцюгових або базисних) одного ряду до відповідних за часом темпів росту (також ланцюгових або базисних) іншого ряду. Аналогічно розраховують і коефіцієнти випередження за темпами приросту.

Аналіз взаємозалежних рядів становить найбільшу складність при вивченні тимчасових послідовностей. Нерідко збіг загальних тенденцій розвитку буває викликаний не взаємним зв'язком, а іншими факторами, що не враховуються. Тому в рядах, що зіставляються, попередньо слід позбутися впливу існуючих у них тенденцій, а після цього провести аналіз взаємозв'язку за відхиленнями від тренда. Дослідження включає перевірку рядів динаміки на автокореляцію й установлення взаємозв'язків між ознаками.

Під *автокореляцією* розуміють залежність наступних рівнів ряду від попередніх. Перевірка на наявність автокореляції здійснюється за критерієм Дарбіна–Уотсона

$$\frac{E_t}{E_t}, \quad (7.30)$$

де  $E_t$  – відхилення фактичного рівня ряду в точці  $t$  від теоретично вирівняного значення.

Якщо  $K=0$ , то є повна позитивна автокореляція; при  $K=2$  автокореляція відсутня.

Перш ніж оцінювати взаємозв'язок, необхідно виключити автокореляцію. Найбільш простим методом виключення автокореляції є *метод кореляції перших різниць*: від вихідних рядів динаміки  $X$  и  $Y$  переходять



до нових, побудованих на основі перших різниць:

(7.31)

(7.32)

за  $\Delta x$  і  $\Delta y$  визначають напрямок і чинність зв'язку в регресії

(7.33)

При вивченні кореляційних зв'язків в багатовимірних рядах динаміки спостерігається певна залежність рівнів даного періоду від попереднього, внаслідок чого виникають певні методологічні особливості. В таких динамічних рядах фактором зміни рівнів виступає, крім інших, також час. Вплив даного рівня динамічного ряду на зміну наступного з плином часу і приводить до автокореляції.

У практиці статистичного аналізу рядів динаміки застосовують різні способи усунення автокореляції, такі, як спосіб різницевих перетворень (при лінійному тренді), спосіб відхилень тенденції (при нелінійній залежності), або введення змінної величини  $t$  в рівняння регресії  $y_t = f(x_1, x_2, x_3, \dots, t)$ , де вона відіграє роль чинника часу.

При застосуванні методу регресії для дослідження динамічних рядів виникає особливість, яка полягає в тому, що в рівнях динамічних рядів присутня авторегресія, яка виявляється таким самим чином, як і автокореляція. Авторегресія виражає залежність величини рівня динамічного ряду від попередніх значень рівня в певні моменти часу.

📖 Література до теми 7: [1–8; 15–21; 23; 25–28].

### ***Контрольні запитання до теми 7***

1. Що характеризує ряд динаміки?
2. Що є рівнем ряду динаміки?
3. Дайте визначення поняття інтервального ряду динаміки. Наведіть приклад.
4. Дайте визначення поняття моментного ряду динаміки. Наведіть приклад.
5. Які є види рівнів ряду динаміки? Дайте їм характеристику.
6. Назвіть типи показників ряду динаміки. Від чого вони залежать?

7. Які є види рядів динаміки залежно від способу вираження?
8. До якого виду слід відносити динамічний ряд, якщо його рівні виражають стан явища на відповідний момент часу?
9. Які є види рядів динаміки залежно від інтервалу між рівнями?
10. До якого виду слід відносити динамічний ряд, якщо його рівні виражають стан явища за певний термін часу?
11. Які є види рядів динаміки залежно від наявності основної тенденції процесу, який вивчається?
12. Що показує ланцюговий темп росту?
13. Чим відрізняється абсолютний приріст від темпу приросту?
14. Що показують середній абсолютний приріст, темп росту і темп приросту?
15. Який показник ряду динаміки слід визначити, щоб розрахувати, на скільки одиниць змінився рівень показника за період?
16. Який показник ряду динаміки слід визначити, щоб відповісти на запитання: у скільки разів у середньому щорічно збільшувався рівень показника за період?
17. За якою формулою треба визначати середній рівень інтервального ряду динаміки?
18. За яких умов середній рівень моментного ряду динаміки розраховується як середня хронологічна?
19. За яких умов середній рівень моментного ряду динаміки розраховується як середня модифікована, тобто середня арифметична зважена з усередненим рівнем?
20. Визначте аналітичні формули дослідження моментного ряду динаміки.

## **Тема 8. СТАТИСТИЧНІ МЕТОДИ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКІВ**

- 8.1. Зв'язки суспільних явищ і завдання їх статистичного вивчення.
- 8.2. Загальні методи вивчення зв'язків.
- 8.3. Кореляційний і регресійний методи аналізу зв'язку.
- 8.4. Нелінійні залежності.
- 8.5. Багатофакторний кореляційно-регресійний аналіз.
- 8.6. Непараметричні показники тісноти зв'язку.

### ***8.1. Зв'язки суспільних явищ і завдання їх статистичного вивчення***

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

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

Найчастіше в економічних дослідженнях зустрічаються стохастичні залежності, які відрізняються приблизністю, невизначеністю. Вони виявляються тільки в середніх рівнях при наявності значної кількості об'єктів (спостережень). Тобто кожній величині факторного показника (аргументу) може відповідати декілька значень результативного показника (функції).

Приклад: збільшення фондоозброєності дає різний приріст продуктивності праці на різних підприємствах навіть при дуже схожих умовах. Це пояснюється тим, що всі фактори, від яких залежить продуктивність праці, діють у комплексі, взаємозалежно. Залежно від того, наскільки оптимально взаємодіють різні фактори, буде неоднаковим ступінь впливу кожного з них на величину результативного показника.

Взаємозв'язок між досліджуваними факторами і результативним показником може бути виявлений, якщо для дослідження використовується велика кількість спостережень або об'єктів. Тоді відповідно до закону великих чисел вплив інших факторів на результативний показник згладжується. Це дає можливість встановити співвідношення між досліджуваними явищами.

У такий спосіб *кореляційний (ймовірнісний) зв'язок* – це неповна, ймовірнісна залежність між показниками, яка виявляється тільки в масі спостережень. Вирізняють парну і множинну кореляції.

*Парна кореляція* – це зв'язок між двома показниками, один з яких є факторним, а інший – результативним.

*Множинна кореляція* виникає від взаємодії декількох факторів на результативний показник.

Дослідження кореляційних співвідношень мають величезне значення в аналізі господарської діяльності: точніше обґрунтовуються плани і управлінські рішення, більш об'єктивно оцінюються підсумки діяльності підприємств і більш повно визначаються внутрішньогосподарські резерви.

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

*За характером залежності явищ* розрізняють функціональні і кореляційні зв'язки.

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

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

*За напрямом* розрізняють прямі і обернені зв'язки. *Прямий зв'язок* – це зв'язок, коли із зростанням факторної ознаки результативна також зростає. *При оберненому зв'язку* із збільшенням факторної ознаки результативна зменшується або, навпаки, із зменшенням факторної ознаки результативна зростає.

*За формою* зв'язок поділяється на прямолінійний і криволінійний. При *прямолинійній* кореляційній залежності рівним змінам середніх значень факторної ознаки відповідають приблизно рівні зміни середніх значень результативної ознаки. При *криволінійній* кореляційній залежності рівним змінам середніх значень факторної ознаки відповідають нерівні зміни середніх значень результативної ознаки.

Статистичне вивчення взаємозв'язків допомагає вирішенню на-

ступних завдань:

- а) визначення форми зв'язку;
- б) вимірювання тісноти (сила) зв'язку;
- в) виявлення впливу окремих чинників на результативну ознаку.

## 8.2. Загальні методи вивчення зв'язків

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

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

Однак ці методи дозволяють виявити тільки загальний характер і напрямок зв'язку. Основне завдання факторного аналізу – визначити ступінь впливу певного фактора на результативний показник.

Одним з поширених методів статистичного вивчення зв'язків суспільних явищ є балансовий метод, як спосіб аналізу зв'язків і пропорцій в економіці. *Статистичний баланс* являє собою систему показників, яка складається із двох сум абсолютних величин, пов'язаних між собою знаком рівності

$$a + б = в + г. \quad (8.1)$$

Цю балансову ув'язку можна відобразити через балансове рівняння

$$\text{залишок на початок періоду} + \text{надходження} = \text{видатки} + \text{залишок на кінець}. \quad (8.2)$$

Наведене балансове рівняння характеризує єдиний процес руху матеріальних ресурсів і показує взаємозв'язок і пропорції окремих елементів цього процесу.

*Метод порівняння паралельних рядів* полягає в тому, що отримані в результаті групування і статистичної обробки матеріали статистичного спостереження ранжуються паралельними рядами за факторною озна-

кою. Паралельно записуються значення результативної ознаки. Це дає можливість, порівнюючи їх, простежити співвідношення, виявити існування зв'язку і з'ясувати його напрям.

Кореляційний аналіз дозволяє кількісно виразити взаємозв'язок між показниками. *Необхідні умови застосування кореляційного аналізу:*

1) Наявність достатньої великої кількості спостережень про величину досліджуваних факторних і результативних показників (у динаміці або за поточний період за сукупністю однорідних об'єктів).

2) Досліджувані фактори повинні мати кількісний вимір і відображення в тих або інших джерелах інформації.

Застосування кореляційного аналізу дозволяє розв'язати наступні завдання:

1) Визначення зміни результативного показника під впливом одного або декількох факторів (в абсолютному вимірі), тобто, визначення на скільки одиниць змінюється величина результативного показника при зміні факторного на одиницю.

2) Встановлення відносного ступеня залежності результативного показника від кожного фактора.

*Кореляція* (від латинської *correlatio* – співвідношення) – це статистична залежність між випадковими величинами, що носить ймовірнісний характер.

Кореляційні зв'язки можна вивчати на якісному рівні з діаграм розсіяння емпіричних значень змінних  $X$  і  $Y$  (рис. 8.1) і відповідним чином їх інтерпретувати. Так, наприклад, якщо підвищення рівня однієї змінної супроводжується підвищенням рівня іншої, то йдеться про *позитивну* кореляцію або *прямий зв'язок* (рис. 8.1, *а, б*). Якщо ж зростання однієї змінної супроводжується зниженням значень іншої, то це *негативна* кореляція або *зворотний зв'язок* (рис. 8.1, *г, і*). Нульовою називається кореляція за відсутності зв'язку змінних (рис. 8.1, *в*). Проте нульова загальна кореляція може свідчити лише про відсутність *лінійної* залежності, а не взагалі про відсутність будь якого *статистичного* зв'язку.

В економічних дослідженнях здебільшого спостерігаються *нелінійні* зв'язки (рис. 8.1, *д*). Наприклад, зростання мотивації спочатку підвищує ефективність діяльності, а потім настає зниження продуктивності (ефект «перемотивації» – закон Йеркса–Додсона). Кількісна міра кореляційного зв'язку оцінюється за значеннями коефіцієнтів кореляції у межах від  $-1$  до  $+1$ . Від'ємні значення коефіцієнтів вказують на зворотний зв'язок, додатні – на прямий. Нульове значення може свідчи-

ти про відсутність зв'язку. Інтенсивність зв'язку (слабкий зв'язок – помірний – суттєвий – сильний) оцінюється за абсолютним значенням коефіцієнтів кореляції.

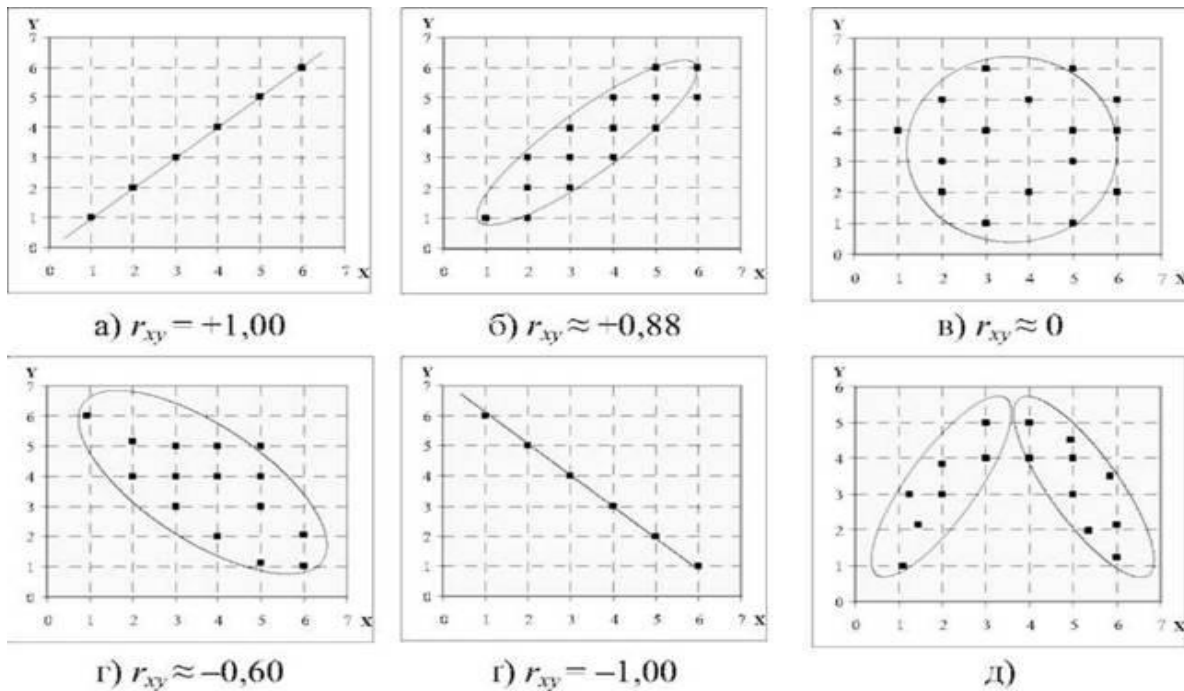


Рисунок 8.1 – Діаграми розсіяння емпіричних значень змінних X і Y: а – строга позитивна кореляція; б – сильна позитивна кореляція; в – нульова кореляція; г – помірна негативна кореляція; д – строга негативна кореляція; е – нелінійна кореляція

Найпростішим показником визначення рівня зв'язку між показниками є коефіцієнт Фехнера ( $K_f$ ), який розраховується за формулою

$$K_f = \frac{\Sigma Z - \Sigma H}{\Sigma Z + \Sigma H} \quad (8.3)$$

де  $\Sigma Z$ ,  $\Sigma H$  – число збігів та незбігів знаків відхилень значень фактичної і результативної ознак від своїх середніх, тобто  $\bar{x}$  та  $\bar{y}$ . При цьому фіксуються збіги та незбіги знаків у відхиленнях від середньої у різних пар значень ознак.

Якщо виконується нерівність  $\bar{x} > \bar{y}$  або  $\bar{x} < \bar{y}$ , то значенню присвоюється знак «+», в протилежному випадку – знак «-». У випадку, коли за обома показниками знаки однакові, має місце їх збіг, а коли вони різні – незбіг. Коефіцієнт Фехнера знаходиться в межах від -1 до +1.

Якщо  $\rho = 0$ , то зв'язок між показниками слабкий, а якщо  $\rho = \pm 1$  – зв'язок тісний. Цей коефіцієнт має додатне значення за наявності прямого зв'язку, а від'ємне – при оберненому.

Більш досконалим показником вважається коефіцієнт кореляції рангів Спірмена  $\rho$ , який визначається таким чином

$$\rho = 1 - \frac{\sigma \sum d^2}{n(n^2 - 1)}, \quad (8.4)$$

де  $d = r_x - r_y$  – різниця рангів факторного та результативного показників.

При цьому під рангом розуміють порядковий номер значення показника у порядку зростання або зменшення. Коефіцієнт кореляції рангів також змінюється від  $-1$  до  $+1$ . При  $\rho > 0$  зв'язок між показниками прямий, а при  $\rho < 0$  – обернений. Якщо  $\rho$  наближається до  $1$ , то між показниками існує тісний (сильний) зв'язок, а якщо  $\rho$  наближається до  $0$ , то вважається, що взаємозв'язок практично відсутній. Таким чином, наведені коефіцієнти дають можливість не тільки оцінити тісноту взаємозв'язку між факторною та результативною ознаками, але й визначити його напрямок (прямий чи обернений).

Англійський статистик Кенделл для визначення тісноти зв'язку між корельованими ознаками запропонував свою формулу коефіцієнта кореляції рангів – коефіцієнт кореляції Кендалла (*tau-b* Кендалла). У цьому методі одна змінна подається у вигляді монотонної послідовності в порядку зростання величин; іншій змінній привласнюються відповідні рангові місця. Кількість інверсій (порушень монотонності в порівнянні з першим рядом) використовується у формулі для кореляційних коефіцієнтів. Застосування коефіцієнта кореляції Кендалла є кращим, якщо у вихідних даних зустрічаються неузгодження.

Коефіцієнт кореляції Кендалла обчислюється за формулою

$$\tau_b = \frac{Q - P}{n(n-1)} \quad (8.5)$$

де  $n$  – кількість спостережень;  $Q$  – число узгоджених пар  $X_j Y_j$  і  $X_i Y_i$  для всіх комбінацій  $i$  і  $j$ .

Пари називаються неузгодженими, якщо для них виконується така умова:  $\text{sign}(X_j - Y_j)\text{sign}(X_i - Y_i) = -1$ , де  $\text{sign}$  означає «знак». Ця функція набуває значення  $+1$  для позитивного числа, та  $-1$  – для негативного.



Іншими словами, наведена умова означає, що збільшення  $X$  приводить до зменшення  $Y$ , і навпаки. Для перевірки значимості коефіцієнта кореляції Кендалла використовують спеціальні таблиці. Таким чином, коефіцієнт кореляції рангів Кенделла оцінює зв'язок між даними ознаками більш обережно, ніж коефіцієнт Спірмена.

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

Зв'язок між двома показниками можна наочно уявити, якщо побудувати графік (див рис. 8.1). Відобразивши на графіку точки, які відповідають значенням « $x$ » і « $y$ », отримаємо кореляційне поле, де за характером розміщення точок можна судити про напрямок і силу зв'язку. Якщо точки розташовані хаотично по всьому полю, це говорить про відсутність залежності між двома ознаками; якщо вони сконцентровані навколо осі, яка йде від нижнього лівого кута до верхнього правого, то це пряма залежність між досліджуваними ознаками; якщо точки будуть сконцентровані навколо осі, яка пролягає від верхнього лівого кута до нижнього правого, то маємо обернену залежність. При графічному способі залежність визначають за допомогою лінійки (протягнутої нитки) так, щоб по обох сторонах проведеної лінії була розташована приблизно однакова кількість точок. Якщо характер розташування точок вказує на наявність криволінійної залежності, то використовується лекало, або шляхом функціональних перетворень змінних нелінійна залежність зводиться до лінійного вигляду.

Числове значення коефіцієнта залежності функції — визначається за формулою

$$\text{—} \quad (8.6)$$

де  $y$  і  $x$  — похідні числові значення за графіком (рис. 8.2).

Відшукавши як мінімум дві точки, по них проводять пряму. Якщо лінійну пряму розташувати так, щоб точки приблизно однаково розташувались по обидві сторони від неї, то параметри лінійного рівняння можна визначити з рівняння

$$\frac{y - y_1}{y_2 - y_3} = \frac{x - x_1}{x_2 - x_3}, \quad (8.7)$$

де  $y_1x_1, y_2x_2$  – значення координат двох точок, які найкращим чином «вписуються» у пряму.

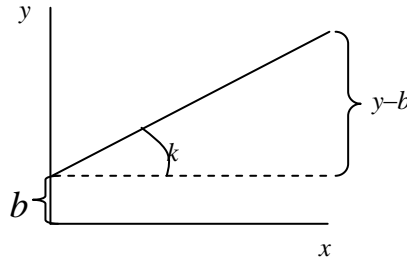


Рисунок 8.2 – Графік взаємозалежностей

Сутність *середньоарифметичного способу* полягає в знаходженні середньоарифметичного значення від всіх  $n$  значень  $Y$  і  $X$ , а також додаткових середніх для точок, чисельні значення яких менше  $x_{\text{ср.}}^{\text{B}}$  і  $y_{\text{ср.}}^{\text{B}}$  і більше  $x_{\text{ср.}}^{\text{H}}$  і  $y_{\text{ср.}}^{\text{H}}$ . У цьому випадку сума відхилень відстаней точок по обидві сторони від теоретичної лінії, визначених на осі ординат, дорівнює нулю. Кутовий коефіцієнт теоретичної лінії визначають за формулою

$$\text{—} \quad (8.8)$$

де  $y_{\text{ср.}}^{\text{B}}, x_{\text{ср.}}^{\text{B}}$  – верхня середня (координати точок);  $y_{\text{ср.}}^{\text{H}}, x_{\text{ср.}}^{\text{H}}$  – нижня середня.

Значення вільного члена  $b$  знаходять за графіком.

Основними недоліками графічного методу визначення математичної залежності слід вважати суб'єктивність «якнайкращого розташування» лінійної прямої. Якщо на тривалість виконання роботи впливають одночасно декілька чинників, то графічний метод можна застосовувати тоді, коли є можливість заздалегідь визначити залежність окремо за кожним чинником за умови, що значення решти чинників залишаються постійними.

*Метод статистичних групувань*, як спосіб виявлення кореляційної залежності, відноситься до числа найважливіших методів дослідження взаємозв'язків. Для виявлення залежності між ознаками за допомогою

цього методу матеріал статистичного спостереження групується за факторною ознакою, і для кожної групи вираховуються середні значення як факторної так і результативної ознаки. Порівнюючи зміни середніх значень результативної ознаки в міру зміни середніх значень факторної ознаки, виявляють характер зв'язку між ними.

Статистичні групування, проведені з метою виявлення і аналізу взаємозв'язків між ознаками, називаються аналітичними. Групування дозволяє також виявити одночасний вплив декількох чинників на результативну ознаку. Для цього здійснюють комбіновані групування, дані яких наводять у вигляді комбінованих таблиць.

Аналітичні групування характеризують лише загальні риси зв'язку, його тенденцію, але не дають кількісної оцінки його сили. На основі аналітичних групувань це завдання розв'язується за допомогою розрахунку *емпіричного кореляційного відношення*. Для кількісної оцінки зв'язку між явищами на базі матеріалів аналітичного групування вираховують коефіцієнт детермінації і емпіричне кореляційне відношення.

*Коефіцієнт детермінації* показує ступінь варіації ознаки під впливом чинника, покладеного в основу групування. Він визначається як відношення міжгрупової дисперсії до загальної. Коефіцієнт детермінації  $R^2$  показує частку зміни (варіації) результативної ознаки під дією факторної ознаки і розраховується за формулою

$$\frac{\text{Міжгрупові дисперсії}}{\text{Загальна дисперсія}} \quad (8.9)$$

Його значення може варіювати від 0 до 1. Чим ближче він до 1, тим адекватніше встановлена тенденція, і відповідно тіснішим є зв'язок вибраного тренда та динамічного ряду. Виходячи з величини коефіцієнта детермінації, у статистичній практиці прийнято застосовувати таку градацію відповідності тренда динамічному ряду: 0 – відсутність зв'язку; до 0,3 – слабка; від 0,3 до 0,6 – середня; від 0,7 до 0,9 – висока; від 0,9 до 1 – вибраний тренд повністю відповідає динамічному ряду. Потрібно також врахувати, що в Excel значення  $R^2$  не є повністю коректним, але для цілей визначення адекватності вибраної функції цим можна знехтувати.

Критерієм суттєвості і сили зв'язку між факторною і результа-

тивною ознаками виступає *емпіричне кореляційне відношення*.

Тіснота зв'язку між змінними характеризується ступенем відхилення (розсіяння) досліджуваних точок біля теоретичної лінії регресії. Чим ближче окремі спостереження розташовані до теоретичної лінії регресії, тим більше повна залежність  $y$  по  $x$ .

Кутовий коефіцієнт лінійного кореляційного зв'язку між  $y$  і  $x$ , який показує, на скільки одиниць в середньому зміниться функція, якщо аргумент збільшується (зменшується) на одиницю свого вимірювання, не може служити показником тісноти зв'язку між змінними. У цьому випадку його чисельне значення залежить від прийнятих одиниць вимірювання змінних.

Для оцінки тісноти зв'язку між змінними використовується емпіричне кореляційне відношення  $\eta_{y/x}^2$ , яке є часткою дисперсії (коливаності) функції  $y$  за рахунок впливу даного аргументу  $x$ . У даному випадку загальна (повна) дисперсія розкладається на дві частини: дисперсію усередині кожного інтервалу зміни функції  $\sigma_{y/x}^2$ , яка не залежить від впливу  $X$ , і дисперсію середніх значень функції  $\bar{y}$ , яка викликана впливом аргументу, тобто

$$\sigma_y^2 = \sigma_{y/x}^2 + \sigma_{\bar{y}}^2 \quad (8.10)$$

Звідси отримуємо формулу для оцінки тісноти зв'язку між змінними:

для незгрупованих даних:

$$\eta_{y/x}^2 = \frac{\sigma_{\bar{y}}^2}{\sigma_y^2} = \frac{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{\sigma_y^2}, \quad (8.11)$$

для згрупованих даних:

$$\eta_{y/x}^2 = \frac{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 m_i}{\sigma_y^2}, \quad (8.12)$$

де  $\bar{y}_i$  – розрахункове значення функції;  $\bar{y}$  – середнє значення функції за вибіркою;  $n$  – обсяг вибірки;  $m_i$  – число спостережень  $y$  в кожному інтервалі зміни.

Кореляційне відношення не залежить від одиниць вимірювання змінних, що вивчаються. Воно показує, яку частину загальної дисперсії  $\sigma_y^2$  можна віднести за рахунок зміни аргументу на одну  $\sigma_x^2$ .

При цьому характеристика  $\eta_{y/x}^2$  тим точніше визначає частку впливу  $x$  на загальну дисперсію  $y$ , чим менше варіюється залишкова дисперсія  $\sigma_{y/x}^2$  при кожному  $x$ . Якщо  $\eta_{y/x}=1$ , то має місце функціональна залежність  $y$  від  $x$ . Якщо  $\eta_{y/x}=0$ , то  $y$  кореляційно не залежить від  $x$ .

У разі лінійної залежності змінних дисперсію середніх значень функції можна записати у вигляді

$$\sigma_y^2 = \frac{\sum_{i=1}^k (y_i - \bar{y})^2 m_i}{n} = \frac{\sum_{i=1}^k (x_i + b - k\bar{x} - b)^2 m_i}{n} = \frac{k^2 \sum_{i=1}^k (x_i - \bar{x})^2 m_i}{\sum_{i=1}^k m_i}. \quad (8.13)$$

Спростивши це відношення, маємо

$$\delta_y^2 = K^2 \sigma_x^2. \quad (8.14)$$

Тоді показник тісноти зв'язку для випадку лінійної залежності буде мати вигляд

$$\eta_{y/n}^2 = \frac{\sigma_y^2}{\sigma_y^2} = \frac{k^2 \sigma_x^2}{\sigma_y^2}, \quad (8.15)$$

де

$$\eta_{y/x} = \frac{k\sigma_x}{\sigma_y}.$$

Отримане відношення служить для визначення вимірника тісноти зв'язку між змінними в разі їх лінійної залежності і має назву коефіцієнта кореляції  $r_{y/x}$ . Якщо замість  $k$  підставити формулу для його обчислень  $k = \frac{\bar{x}y - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}$ , то матимемо

$$k = \frac{(\bar{x}y - \bar{x} \cdot \bar{y})\sigma_x}{\sigma_x^2 \sigma_y} = \frac{\bar{x}y - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}. \quad (8.16)$$

Коефіцієнт кореляції показує, на яку частину середнього квадра-

тичного відхилення або  $\sigma_y$  змінюється функція  $y$ , якщо аргумент  $x$  збільшується (зменшується) на своє середньоквадратичне відхилення  $\sigma_x$ . Знак коефіцієнта кореляції збігається зі знаком коефіцієнта регресії, а його чисельне значення коливається в межах

$$-1 \leq r_{y/x} \leq 1. \quad (8.17)$$

Суттєвість коефіцієнта кореляції при заданому рівні значущості  $\alpha = 0,05$  або 5 % перевіряється за умовою

$$\frac{|r_{y/x}| \cdot \sqrt{n}}{1 - r_{xy}^2} \geq t_\alpha \geq 1,96, \quad (8.18)$$

де  $t_\alpha$  – значення, що визначається за умовою  $2\phi(t)$ .

*Приклад.* Визначити гарантійну вірогідність істотності коефіцієнта кореляції  $r_{y/x} = -0,745$  при  $n = 52$  спостережень.

Розв'язання:

$$t_2 = \frac{|r_{y/x}| \sqrt{n}}{1 - r_{y/x}^2} = \frac{|-0,745| \sqrt{52}}{1 - (-0,745)^2} = 12,07.$$

При  $\alpha = 0,05$  табличне значення  $t = 1,96$  значно менше від розрахованого. При  $t = 12,07$  довірчий інтервал в генеральній сукупності містить коефіцієнт кореляції з довірчою вірогідністю

$$P = 2\phi(12,07) = 100 \%$$

Лінійність (нелінійність) зв'язку між змінними перевіряється таким чином:

- шляхом порівняння абсолютних значень  $|r_{y/x}| = \eta_{y/x}$ ;
- статистично, з використанням довірчого інтервалу. При цьому кореляційне емпіричне відношення  $\eta_{y/x}$  повинне покриватися довірчим інтервалом для  $(r_{y/x})$

$$\left| r_{yx} \right| - \frac{1 - r_{y/x}^2}{\sqrt{n}} t_\alpha \leq \eta_{y/x} \leq \left| r_{y/x} \right| + \frac{1 - r_{yx}^2}{\sqrt{n}} t_\alpha \quad (8.19)$$

з довірчою вірогідністю  $2\phi(t_\alpha) = 1 - \alpha$ .

Отже, якщо нерівність задовольняється при  $\alpha = 0,05$ , то приймаємо гіпотезу лінійності, а якщо ж нерівність не задовольняється, то прийма-

ємо гіпотезу нелінійної кореляційної залежності між  $y$  і  $x$ .

Завершальним дослідженням парного рівняння є перевірка відповідності виведеного рівняння до реального процесу, який досліджується. Якщо отримано декілька рівнянь регресії, то кращим слід вважати те з них, яке ближче до закономірності по суті досліджуваного процесу. Слід оцінити ступінь близькості результатів розрахунків по кожному з отриманих рівнянь до звітних даних. Цей ступінь оцінюється за залишковою теоретичною дисперсією  $\tilde{\sigma}_{y/x}^2$ , яка характеризує розкид досліджених точок кореляційного поля навколо теоретичної лінії регресії під впливом чинників, що не враховані в отриманому рівнянні.

Залишкова теоретична дисперсія визначається з рівняння

$$\tilde{\sigma}_{y/x}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{h_i} (y_{ij} - \tilde{y}_i)^2}{n - p}, \quad (8.20)$$

де  $n$  – обсяг вибірки;  $p$  – число параметрів рівняння;  $k$  – число інтервалів;  $h$  – число спостережень в  $i$ -му інтервалі;  $y_{ij}$  – значення досліджених даних функціональної ознаки;  $\tilde{y}_i$  – розрахункове значення функціональної ознаки, обчислене за рівнянням регресії.

Якщо складено декілька рівнянь регресії і для кожного обчислена залишкова дисперсія, то з декількох рівнянь регресії, рівноцінних по суті, перевагу слід віддати тому, в якого залишкова дисперсія менша.

Для управління регресії з одним аргументом  $x$  залишкову дисперсію обчислюють за формулою

$$\tilde{\sigma}_{y/x}^2 = \sigma_y^2 (1 - \eta_{y/x}^2). \quad (8.21)$$

Одним з ознак узгодження початкових даних з отриманим рівнянням служить нормальність розподілу відхилень досліджених даних від розрахункових  $(y_{ij} - \tilde{y}_{xi})$  зі середнім значенням, що дорівнює нулю, і середнім квадратичним відхиленням, що дорівнює  $\sigma_{y/x}$ . Чим ближче розподіл цих відхилень до вказаного нормального закону, тим краще узгоджуються вихідні дані з виведеним рівнянням регресії. Обчисливши всі відхилення  $(y_{ij} - \tilde{y}_{xi})$ , перевіряють наявність нормального закону їх розподілу візуально і за критерієм згоди.

Таким чином, дослідження рівняння в тій або іншій формі полягає у визначенні його параметрів (із розв'язання системи нормальних рівнянь), підстановці в отримане рівняння всіх вихідних значень аргументу, розрахунку відповідних значень  $\tilde{y}_{ij}$ , порівнянні розрахункових значень  $\tilde{y}_i$  із вихідними значеннями, обчисленні суми квадратів відхилень розрахункових значень від вихідних і визначення залишкової дисперсії  $\tilde{\sigma}_{y/x}^2$ . Найкраща форма визначається за найменшою залишковою дисперсією. За такою схемою визначаються форми зв'язку  $y$  за всіма аргументами.

Існує також критерій адекватності, запропонований Фішером і заснований на порівнянні залишкової теоретичної дисперсії  $\tilde{\sigma}_{y/x}^2$  і загальної дисперсії  $\sigma_y^2$ . Розглядається відношення  $\frac{\sigma_{y/x}^2}{\sigma_y^2} = T_{\text{расч}}$  і порівнюється з табличним (для  $\phi(t)$  Фішера знайдено розподіл і складено спеціальну таблицю) при заданому рівні значущості і різних ступенях свободи.

Загальна дисперсія  $\sigma_y^2$  досліджених даних від їх середнього значення встановлюється з урахуванням числа ступенів свободи  $f = n - k$

$$\sigma_{y/x}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y})^2 m_i}{n - k}, \quad (8.22)$$

де  $k$  – число інтервалів у вибіркових даних.

Залишкова теоретична дисперсія  $\tilde{\sigma}_{y/x}^2$  встановлюється, як різниця розрахункових  $\tilde{y}_i$  і середніх інтервальних значень  $\bar{y}_i$  з урахуванням числа ступенів свободи  $d_1 = K - P$  і  $d_2 = n - K$ , де  $P$  – число параметрів рівняння. Якщо  $\bar{f}_{\text{рас}} \leq \bar{f}_{\text{табл}}$ , то при заданому рівні значущості складене рівняння регресії затверджується. Вірогідність помилки тим менша, чим більший рівень значущості  $\alpha$  %. У разі, коли чисельник  $\tilde{\sigma}_{y/x}^2$  менший від знаменника  $\sigma_y^2$ , міняємо їх місцями разом з відповідними ступенями свободи  $d_1 = K - P$  і  $d_2 = n - K$ .

### 8.3. Кореляційний і регресійний методи аналізу зв'язку

Основне завдання кореляційного і регресійного аналізу статистичних даних є виявлення залежності між досліджуваними ознаками у ви-



гляді певної математичної формули і встановлення за допомогою коефіцієнта кореляції порівняльної ознаки тісноти взаємозв'язку. Кореляційний і регресійний методи аналізу розв'язують два основних завдання:

1) визначають з допомогою рівняння регресії аналітичну форму зв'язку між варіацією ознак « $x$ » і « $y$ »;

2) встановлюють міру тісноти зв'язку між ознаками.

У практиці економіко-статистичних досліджень часто доводиться мати справу з прямолінійною формою зв'язку, яка виражається за допомогою рівняння регресії. Рівняння регресії характеризує зміну середнього рівняння результативної ознаки у залежності від зміни факторної ознаки  $x$ .

*Спосіб парної кореляції для вивчення стохастичних залежностей* є основним завданням кореляційного аналізу для визначення впливу факторів на величину результативного показника. Для вирішення цього завдання підбирається відповідний тип математичного рівняння, яке найкраще відображає характер досліджуваного зв'язку. Від правильного вибору рівняння регресії залежить хід розв'язку завдання і адекватність результатів розрахунків. *Обґрунтування рівняння зв'язку* робиться за допомогою зіставлення певних рядів, угруповання даних і лінійних графіків. Розміщення точок на графіку показує, яка залежність утворювалась між досліджуваними показниками: прямолінійна або криволінійна. Найбільш просте рівняння – це пряма

$$Y_x = a + bx, \quad (8.23)$$

де  $x$  – факторний показник;  $Y_x$  – результативний показник;  $a$  і  $b$  – параметри рівняння регресії, які потрібно відшукати.

У цьому випадку при зміні факторного показника спостерігається рівномірне зростання або зменшення результативного показника.

Значення коефіцієнтів  $a$  і  $b$  знаходять із системи рівнянь, отриманих способом найменших квадратів. Система рівнянь має вигляд

$$(8.24)$$

де  $n$  – кількість спостережень. Значення  $\sum x$ ,  $\sum y$ ,  $\sum xy$ ,  $\sum x^2$  розраховуються на основі фактичних вихідних даних.

Коефіцієнт  $a$  – const, величина результативного показника, яка не

пов'язана зі зміною даного фактора. Параметр  $b$  показує середню зміну результативного показника зі зростанням або зниженням величини фактора на одиницю виміру.

Підставивши в рівняння регресії відповідно значення  $X$  можна визначити вирівняні (теоретичні) значення результативного показника  $Y$ . Порівняння фактичного рівня результативного показника з розрахунковим дозволяє оцінити результат роботи деяких підприємств.

Для визначення тісноти зв'язку між факторними і результативними показниками використовується коефіцієнт кореляції. У випадку прямолінійної форми зв'язку між досліджуваними показниками коефіцієнт кореляції розраховується за наступною формулою

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} = \frac{\overline{xy} - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}. \quad (8.25)$$

Коефіцієнт кореляції може набувати значення від 0 до 1. Чим ближче його величина до 1, тим більш тісний зв'язок між досліджуваними явищами і навпаки. Якщо коефіцієнт кореляції возвести у квадрат, одержимо коефіцієнт детермінації  $d$ .

Приклад:  $d = 0,73$ , результативний показник на 73 % залежить від досліджуваного фактора, а на частку інших факторів доводиться 17 % приросту результативного показника.

Для визначення тісноти зв'язку при криволінійній залежності використовується нелінійний коефіцієнт кореляції, а кореляційне відношення

$$\eta = \sqrt{\frac{\delta_y^2 - \delta_{y_x}^2}{\delta_y^2}}, \quad (8.26)$$

де

$$\delta_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}; \quad \delta_{y_x}^2 = \frac{\sum (y_i - y_{x_i})^2}{n}.$$

Показник  $\eta$  є універсальним. Його можна використовувати при будь-якій формі залежності. Однак для визначення його величини спочатку необхідно розв'язати рівняння регресії і розрахувати вирівняні

значення результативного показника  $Y_x$ .

У висновку необхідно відзначити, що ця методика може бути використана для дослідження співвідношень між різними економічними показниками, що дозволяє глибше вивчати досліджувані явища, визначати місце і роль кожного показника в зміні рівня досліджуваного показника.

#### **8.4. Нелінійні залежності**

У практиці економічного аналізу найбільш часто використовують наступні нелінійні функції залежності: гіперболічну, параболічну другого порядку, напівлогарифмічну та деякі інші.

Аналогічно рівнянням прямолінійного зв'язку розв'язується рівняння зв'язку при *криволінійній залежності* між досліджуваними явищами.

У випадку, якщо при збільшенні одного показника значення іншого зростають до певного рівня, а потім починають знижуватися (приклад – залежність продуктивності праці робітників від віку), то для запису такої залежності найкраще підходить парабола другого порядку

$$Y_x = a + bx + cx^2. \quad (8.27)$$

Відповідно до вимог методу найменших квадратів для визначення параметрів  $a, b$  і  $c$  необхідно розв'язати наступну систему рівнянь:

$$(8.28)$$

Параметри  $a, b, c$  знаходять способом визначників або способом виключення.

*Спосіб визначників.* Знаходимо загальний визначник  $\Delta$  і часткові визначники  $\Delta a, \Delta b$  і  $\Delta c$ . Звідси

$$a = \frac{\Delta a}{\Delta}; \quad b = \frac{\Delta b}{\Delta}; \quad c = \frac{\Delta c}{\Delta}. \quad (8.29)$$

Параметри даного рівняння економічного змісту не мають. Якщо



багатьох змінних шляхом логарифмування або заміни змінних можна звести до лінійного вигляду.

Більшість математичних функцій, що використовуються для опису техніко-економічних показників, шляхом функціональних перетворень у по  $x$  (роздільно або одночасно) можуть бути зведені до лінійного вигляду. При цьому метод перетворень залежить від форми зв'язку. Гіпербола вигляду  $\bar{y} = \frac{k}{x} + b$  перетвориться в лінійну шляхом заміни (рис. 8.3).

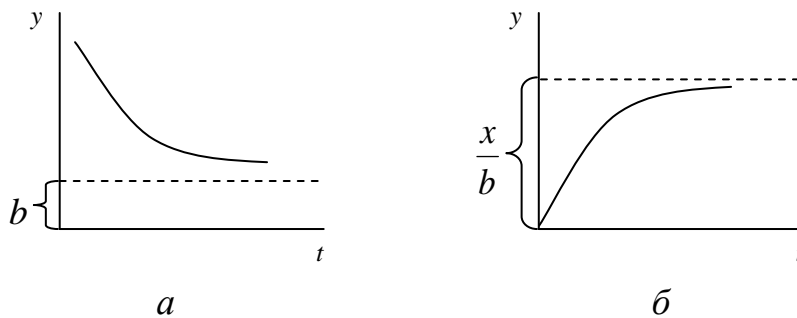


Рисунок 8.3 – Перетворення гіперболи в лінійну функцію

Статична функція вигляду  $\bar{y} = bk^x$  перетвориться в лінійну шляхом логарифмування. У результаті маємо  $\lg \bar{y} = \lg b + k \lg x$  (рис. 8.4).

Введемо таке позначення:  $\bar{y}' = \lg \bar{y}$ ;  $b' = \lg b$ ;  $x' = \lg x$ . У результаті маємо  $\bar{y}' = kx' + b'$ .

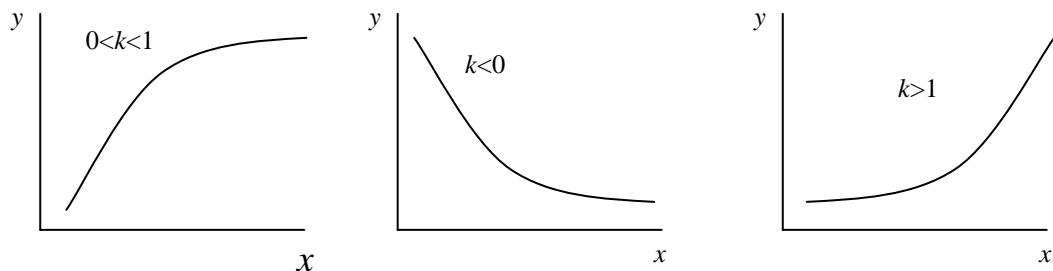


Рисунок 8.4 – Показова статична функція у вигляді лінійної залежно від значення  $k$

Показова функція вигляду  $y = be^{kx}$  перетвориться в лінійну логарифмуванням  $\lg \bar{y} = \lg b + kx \lg e$ . Введемо позначення:  $\bar{y}' = \lg \bar{y}$ ,  $b' = \lg b$ ,  $k' = k \lg e$ , при цьому  $\lg e = 0,4343$ . У результаті маємо  $y_1 x + b_1$  (рис. 8.5).

Теоретична лінія регресії може бути подана у вигляді плавної

кривої, яка кількісно виражає зв'язок між середніми інтервальними значеннями  $\bar{y}$  і відповідними значеннями  $X$  (аргументами). Процес знаходження невідомих параметрів теоретичної залежності є однією з важливих проблем теорії кореляції і регресії.

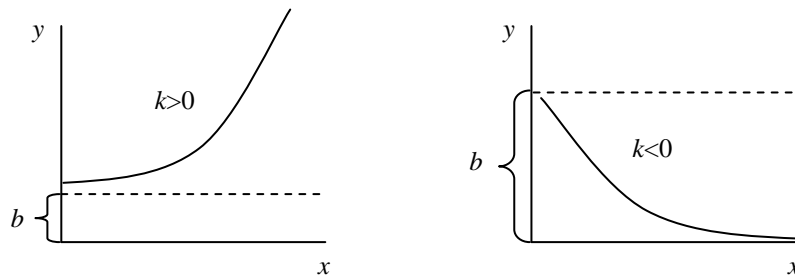


Рисунок 8.5 – Перетворення показової функції в лінійну функцію

*Етапи багатофакторного кореляційного аналізу:*

1. Визначення факторів, які впливають на досліджуваний показник і відбір з них найбільш істотних для кореляційного аналізу.
2. Збір і оцінка вихідної інформації (вірогідність, точність, однорідність і відповідність до закону нормального розподілу).
3. Вивчення характеру і моделювання зв'язку між факторами і результативним показником, тобто обґрунтування математичного рівняння, яке найбільш точно виражає сутність досліджуваної залежності.
4. Розрахунки основних показників зв'язку кореляційного аналізу.
5. Статистична оцінка результатів кореляційного аналізу і практичне їх застосування.

*Правила відбору факторів для багатофакторного кореляційного аналізу:*

1. Враховуються причинно-наслідкові зв'язки між показниками. Аналіз факторів, які перебувають тільки в математичних співвідношеннях з результативним показником, не має практичного змісту.
2. При створенні багатофакторної кореляційної моделі необхідно відбирати самі значущі фактори, які мають вирішальний вплив на результативний показник. Фактори, які мають критерій надійності за Стьюдентом менший від табличного, не рекомендується брати до уваги.
3. Усі фактори повинні бути кількісно вимірні, тобто мати одиницю виміру і інформація про них повинна знаходитись на обліку в звітності.
4. У кореляційну модель лінійного типу не рекомендується вклю-

чати фактори, зв'язок яких з результативним показником, має криволінійний характер.

5. Не рекомендується включати в кореляційну модель взаємозалежні фактори. Якщо парний коефіцієнт кореляції між двома факторами більше 0,85, то за правилами кореляційного аналізу один з них необхідно виключити, інакше це призведе до викривлення результатів аналізу.

6. Не можна включати в кореляційну модель фактори, зв'язок яких з результативним показником носить функціональний характер.

7. Чим більша сукупність об'єктів досліджується, тим точніші результати.

Однією з головних умов кореляційного аналізу є однорідність досліджуваної інформації щодо її розподілу близько середнього рівня. Якщо в сукупності є групи об'єктів, які значно відрізняються від середнього рівня, то це говорить про неоднорідність вихідної інформації. Критерієм однорідності служить середньоквадратичне відхилення і коефіцієнт варіації, які розраховуються за кожним факторним і результативним показниками.

Середньоквадратичне відхилення  $\delta = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$  – абсолютне відношення індивідуальних значень від арифметичної середньої зваженої  $\bar{x}$ .

Коефіцієнт варіації  $V_x = \frac{\delta}{\bar{x}}$  показує відносне відхилення окремих значень від арифметичної середньої зваженої  $\bar{x}$ .

Чим більший коефіцієнт варіації, тим відносно більший розкид і менша однорідність досліджуваних об'єктів.

Шкала мінливості варіаційного ряду:

$V_x < 10\%$  – незначна мінливість;

$10\% < V_x < 20\%$  – середня мінливість;

$20\% < V_x < 33\%$  – значна мінливість;

$V_x > 33\%$  – неоднорідна інформація.

При цьому інформацію необхідно виключити або відкинути нетипові спостереження, які звичайно бувають у перших і останніх ранжованих рядах вибірки. На підставі найвищого показника варіації можна визначити необхідний обсяг вибірки

$$\text{---} \tag{8.31}$$

де  $n$  – обсяг вибірки;  $V_x$  – варіація, %;  $t$  – показник надійності, який вибирається в таблиці відповідно до рівня ймовірності;  $p$ ,  $m$  – показник точності розрахунків (для економічних розрахунків допускається помилка 5–8 %).

Вихідна інформація повинна відповідати закону нормального розподілу. Згідно з цим законом основна маса досліджуваної інформації за кожним показником повинна бути згрупована близько до її середнього значення, а об'єкти з дуже маленькими значеннями або набагато більшими повинні зустрічатися, якомога рідше. Графік нормального розподілу має наступний вигляд (рис. 8.6).

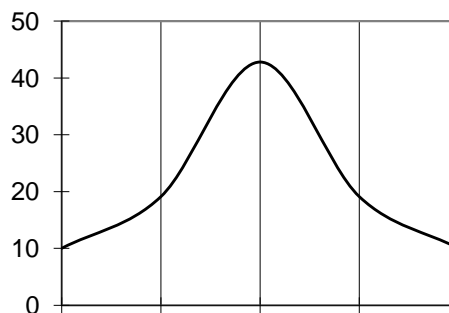


Рисунок 8.6 – Графік нормального розподілу

Показник асиметрії  $A$  і його помилка  $m_a$  розраховуються таким чином:

$$A = \frac{\sum (x - \bar{x})^3}{n\sigma^3}; \quad m_a = \sqrt{\frac{6}{n}}. \quad (8.32)$$

Показник ексцесу  $E_x$  і його помилка  $m_e$  розраховуються за такими формулами:

$$\frac{\sum (x - \bar{x})^4}{n\sigma^4} \quad \text{або} \quad \frac{\sum (x - \bar{x})^4}{n\sigma^4} \quad (8.33)$$

У симетричному розподілі  $A = 0$ ;  $A < 0$  – переважають дані з більшими значеннями, а з меншими зустрічаються значно рідше;  $A > 0$  – найчастіше зустрічаються дані з невеликими значеннями.

У нормальному розподілі  $E_x = 0$ ;  $E_x > 0$  – дані густо згруповані



близько до середньої, утворюючи гостровершинність;  $E_x < 0$  – крива розподілу буде плосковершинною.

Однак якщо  $k$  і  $g$  менше 3, то асиметрія і ексцес не мають істотного значення і досліджувана інформація підкоряється закону нормального розподілу.

Якщо зв'язок усіх факторних показників з результативним носить прямолінійний характер, то досліджують лінійну функцію

$$Y_x = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (8.34)$$

Якщо цей зв'язок носить криволінійний характер, то використовують статичну функцію

$$Y_x = b_0 \cdot b_1x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_n^{b_n}, \quad (8.35)$$

або логарифмічну

$$\lg Y_x = b_0 + b_1 \lg x_1 + b_2 \lg x_2 + \dots + b_n \lg x_n \quad (8.36)$$

У наведених моделях коефіцієнти мають певну інтерпретацію. У лінійній – коефіцієнти  $b_i$  показують, на скільки одиниць змінюється результативний показник зі зміною факторного на 1 в абсолютному вираженні у статичних і логарифмічних процесах.

У випадках, коли важко обґрунтувати форму залежності, розв'язок задачі можна провести за різними моделями і зрівняти результати. Адекватність різних моделей до фактичних залежностей перевіряється за критерієм Фішера – показником середньої помилки апроксимації і величини множинного коефіцієнта детермінації.

Критерій Фішера ( $F$  – відношення) має вигляд

$$\frac{\dots}{\dots} \quad (8.37)$$

де

$$\frac{\dots}{\dots} \quad (8.38)$$

де  $Y_{x_i}$  – індивідуальні значення результативного показника, розраховані

за рівнянням;  $\bar{Y}_x$  – середнє значення результативного показника, розраховане за рівнянням;  $Y_i$  – фактичні індивідуальні значення результативного показника;  $m$  – кількість параметрів у рівнянні зв'язку, з урахуванням вільного члена рівняння;  $n$  – кількість спостережень (обсяг вибірки).

Фактична величина  $F$ -відношення зіставляється з табличною при певному рівні ймовірності і кількості ступенів свободи  $\frac{m-1}{n-m}$ . Значення

повинне бути менше або дорівнювати , тобто

Для статистичної оцінки *точності* рівняння зв'язку використовується також середня помилка апроксимації

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \left( \frac{Y_{x_i} - \bar{Y}}{Y_i} \right)^2. \quad (8.39)$$

Чим менше теоретична лінія регресії (розрахована за рівнянням) відхиляється від фактичної, тем менше  $\bar{\varepsilon}$ . В економічних розрахунках допускається помилка 5–8 %.

Вплив кожного фактора на приріст результативного показника розраховується в такий спосіб

$$\Delta Y_{x_i} = b_i \cdot \Delta x_i. \quad (8.40)$$

Результати багатофакторного регресійного аналізу можуть бути також використані для планування і прогнозування рівня результативного показника. З цією метою в отримане рівняння зв'язку підставляють плановий (або прогнозований) рівень факторних показників.

При вивченні тісноти зв'язку треба мати на увазі, що величина коефіцієнтів кореляції залежить від обсягу. Відомо, що зі зменшенням кількості спостережень надійність коефіцієнтів кореляції знижується, і навпаки, при збільшенні спостережень надійність коефіцієнтів кореляції зростає.

Значимість коефіцієнтів кореляції перевіряється за критерієм Стьюдента

$$t = \frac{r}{\delta_r}, \quad (8.41)$$

де  $\delta_r$  – середньоквадратична помилка коефіцієнта кореляції, яка визначається за формулою

$$\delta_r = \frac{1-r^2}{\sqrt{n-1}}. \quad (8.42)$$

Якщо розрахункове значення  $t$  вище від табличного, то можна зробити висновок про те, що величина коефіцієнта кореляції є значимою. Табличні значення  $t$  знаходять по таблиці значень критерію Стюдента (приклад табл. 5.1). При цьому враховується кількість ступенів свободи  $V = n - 1$  і рівень довірчої ймовірності (в економічних розрахунках звичайно 0,05 або 0,01).

Наступний етап кореляційного аналізу – розрахунки рівняння зв'язку (регресії). Розв'язок проводиться кроковим способом. Спочатку до розрахунку приймається один фактор, який виявляє найбільш значимий вплив на результативний показник, потім другий, третій і т. д. На кожному етапі розраховується рівняння регресії множинного коефіцієнта кореляції і детермінації,  $F$  – відношення, стандартна помилка і інші. Величина їх на кожному етапі порівнюється з попередньою. Чим *вища* величина коефіцієнта множинної кореляції, детермінації і критерію Фішера і чим менша величина стандартної помилки, тим точніше рівняння зв'язку описує залежність, яка склалась між досліджуваними показниками. Якщо додавання наступних факторів не поліпшує оціночний показник зв'язку, то треба їх відкинути, тобто зупинитися на тому рівнянні, де ці показники найбільш оптимальні.

Коефіцієнти регресії в рівнянні зв'язку мають різні одиниці виміру, що робить їх непорівнянними, якщо виникає питання про порівняльну чинність впливу факторів на результативний показник. Щоб звести їх до порівнянного вигляду, усі змінні рівняння регресії збільшують на частку середньоквадратичного відхилення, тобто розраховують стандартизовані коефіцієнти регресії. Їх називають  $\beta$ -коефіцієнтами; вони пов'язані з коефіцієнтами регресії наступним співвідношенням

$$\beta_i = b_i \frac{\delta_{x_i}}{\delta_y}, \quad (8.43)$$

де  $\beta$ -коефіцієнти показують, що якщо величина фактора збільшиться на одне середньоквадратичне відхилення, то відповідна залежна змінна

збільшиться або зменшиться на частку свого середньоквадратичного відхилення. Зіставлення розрахованих  $\beta$ -коефіцієнтів дозволяє зробити висновок про порівняльний ступінь впливу кожного фактора на величину результативного показника. За аналогією можна зіставити і коефіцієнти еластичності, які розраховуються за формулою

$$\frac{\Delta Y}{Y} = \beta \cdot \frac{\Delta X}{X} \quad (8.44)$$

Коефіцієнт еластичності показує, на скільки відсотків у середньому змінюється результативний показник зі зміною факторного на 1 %.

Значення багатофакторного кореляційного аналізу дозволяє:

- 1) вивчити закономірності змін результативного показника залежно від поведінки різних факторів;
- 2) визначити вплив різних факторів на величину результативного показника;
- 3) встановити, які з факторів є основними, а які – другорядними;
- 4) у такий спосіб досягти більш об'єктивної оцінки діяльності підприємства, більш точно і повно визначити внутрішньогосподарські резерви і планові показники.

### **8.6. Непараметричні показники тісноти зв'язку**

Поряд з вивченням кореляційної залежності між кількісними показниками статистика встановлює також зв'язки і між якісними ознаками. При вивченні залежності між якісними ознаками встановлюють наявність зв'язку і вимірюють його тісноту.

Для вимірювання тісноти зв'язку між двома ознаками, які мають альтернативний вираз, застосовують коефіцієнт асоціації, запропонований статистиком Юлом. З метою розрахунку коефіцієнта асоціації використовують таблицю, яка складається з чотирьох комірок, позначених латинськими літерами  $a, b, c, d$ . Кожна з комірок відповідає відомій альтернативі тієї чи іншої ознаки.

Для аналізу взаємозв'язку між атрибутивними ознаками будуються спеціальні таблиці, що мають назву таблиць співзалежності. У тому випадку, коли утворюються по дві групи за факторною та результативною ознаками, або коли вони є альтернативними, для оцінки тісноти зв'язку визначають коефіцієнти асоціації та контингенції за такими формулами:

– коефіцієнт асоціації Юла

$$\frac{K_k}{a}, \quad (8.45)$$

де  $a$  – сильний зв'язок;  $K_k$  – слабкий зв'язок;  
– коефіцієнт контингенції Юла

$$\frac{K_k}{a}. \quad (8.46)$$

Зв'язок вважається підтвердженим, якщо  $a \geq 0,5$  або  $K_k \geq 0,5$ .

Коефіцієнт контингенції Юла завжди менший від коефіцієнта асоціації Юла. Для розрахунку названих коефіцієнтів використовують так звані тетрагоричні таблиці, які показують розподіл одиниць за факторною та результативною ознаками.

Для визначення тісноти зв'язку двох якісних ознак, кожна з яких складається тільки із двох груп, використовують коефіцієнти асоціації і контингенції. Для їх розрахунку будують чотириклітинну таблицю кореляції, яка виражає зв'язок між двома явищами, кожне з яких в свою чергу, повинне бути альтернативним, тобто складати з двох якісно відмінних один від одного значень ознаки (наприклад, хороший, поганий).

У тому випадку, коли утворюється більше, ніж дві групи по  $x$  та  $y$ , тісноту зв'язку оцінюють з допомогою коефіцієнтів співзалежності. Якщо кожна із якісних ознак складається більше ніж із двох груп, то для визначення тісноти зв'язку використовують коефіцієнт взаємної спряженості Пірсона. Цей коефіцієнт розраховують за формулою

$$\frac{q}{n} \quad (8.47)$$

де  $q$  – показник взаємної спряженості, визначається як сума відношень квадратів частот кожної групи до добутку кумулятивної частоти відповідної групи і ознаки. Для отримання значення  $q^2$  необхідно з отриманого значення вирахувати 1.

Коефіцієнт взаємної спряженості Чупрова визначають за формулою

$$\frac{q^2}{n}. \quad (8.48)$$

де  $K_1, K_2$  – число груп за кожною із ознак;

Коефіцієнт співзалежності Чупрова (при  $m_1 = m_2$ ) – за формулою

$$\frac{\sum_{i=1}^k \sum_{j=1}^k \frac{K_{ij}^2}{K_i K_j}}{K} \quad (8.49)$$

Коефіцієнт співзалежності Крамера (при  $m_1 \neq m_2$ ) розраховують за виразом

$$\frac{\sqrt{\chi^2 / m_{\min}}}{\sqrt{K - 1}} \quad (8.50)$$

де  $m_{\min}$  – менше з двох значень.

Значення  $\chi^2$  визначають за формулою

$$\chi^2 = \frac{K \sum_{i=1}^k \sum_{j=1}^k \frac{K_{ij}^2}{K_i K_j} - \sum_{i=1}^k K_i \sum_{j=1}^k \frac{K_{ij}^2}{K_j}}{m_{\min}} \quad (8.52)$$

де

$$K = \sum_{i=1}^k \sum_{j=1}^k \frac{K_{ij}^2}{K_i K_j} \quad (8.53)$$

Непараметричні методи вимірювання зв'язку застосовуються для перевірки умов використання методу найменших квадратів, незалежності розподілу ознак, однорідності вибірок, наявності тренда в рядах динаміки.

📖 Література до теми 8: [4; 5; 7; 9; 12–21; 23; 25–28].

### **Контрольні запитання до теми 8**

1. Які ознаки називаються факторними, а які – результативними?
2. Яка різниця між функціональним і кореляційним зв'язком?
3. Наведіть поняття прямих і обернених зв'язків.
4. Визначте поняття прямолінійних і криволінійних залежностей.
5. Визначте поняття статистичного балансу.
6. Суть методу порівняння паралельних рядів.
7. Оцінка зв'язку за коефіцієнтом Фехнера.

8. Оцінка зв'язку за допомогою коефіцієнтів кореляції рангів.
9. Сутність графічного методу виявлення кореляційної залежності.
10. Сутність та поняття методу статистичних групувань.
11. Що показують емпіричне кореляційне відношення і коефіцієнт детермінації?
12. Сутність кореляційно-регресійного аналізу.
13. В яких випадках використовують рівняння прямої?
14. Що показує коефіцієнт регресії?
15. Що показує коефіцієнт еластичності?
16. Поняття про нелінійні залежності.
17. Коли застосовують рівняння гіперболи?
18. Що таке багатofакторний кореляційно-регресійний аналіз?
19. Коли використовують непараметричні показники тісноти зв'язку?
20. Як визначається коефіцієнт асоціації, запропонований Юлом?

## СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Алексенко Л. М. Економічний словник / Л. М. Алексенко, В. М. Олексієнко, А. І. Юркевич. – К. : Вид. будинок «Максимум»; Тернопіль : Економічна думка, 2000. – 288 с.
2. Бек В. Л. Теорія статистики : навч. посіб. / В. Л. Бек. – К. : ТОВ «Центр учбової літератури», 2002. – 288 с.
3. Вашків П. Г. Теорія статистики : навч. посіб. / П. Г. Вашків, П. І. Пастер, В. П. Сторожук, Є. І. Ткач. – К. : Либідь, 2001. – 320 с.
4. Гетало А. В., Борух В. О. Економічна статистика : навч. посіб. / А. В. Гетало, В. О. Борух. – К. : ТОВ «УВПУ «Екс Об» , 2002. – 148 с.
5. Гончарук А. Г. Основи статистики : навч. посіб. / А. Г. Гончарук. – К. : ТОВ «Центр учбової літератури», 2004. – 288 с.
6. Горковий В. К. Статистика : підруч. / В. К. Горковий. – К. : Вища школа, 1995. – 415 с.
7. Громько Г. Л. Общая теория статистики / Г. Л. Громько. – М. : ИНФРА-М, 2000. – 139 с.
8. Елисеєва И. И., Юзбашев М. М. Общая теория статистики : учебн. / И. И. Елисеєва, М. М. Юзбашев ; под ред. И. И. Елисеєвой. – М. : Финансы и статистика, 2001. – 368 с.
9. Ефимова М. Р. Общая теория статистики : учебн. / М. Р. Ефимова. – М. : ИНФРА-М., 2000. – 416 с.
10. Ефимова М. Р. Практикум по общей теории статистики / М. Р. Ефимова, О. Н. Гапченко, Е. В. Петрова – М. : Финансы и статистика, 2000. – 280 с.
11. Єрина А. М. Статистика : навч.-метод. посібн. для самоств. вивч. дисц. / А. М. Єрина, Р. М. Моторін, А. В. Головач та ін. – К. : КНЕУ, 2002. – 448 с.
12. Єрина А. М. Економічна статистика : практикум / А. М. Єрина, О. К. Мазуренко, З. О. Кальян – К. : ТОВ «УВПУ «Екс Об» , 2002. – 232 с.
13. Закон України «Про державну статистику» // Відомості Верховної Ради України (ВВР), 1992, N 43, ст. 608.
14. Лугінін О. Є. Статистика національної та міжнародної економіки : навч. посіб. / О. Є. Лугінін, С. В. Фомішин. – Л. : Новий Світ-2000, 2011. – 471 с.
15. Лугінін О. Є. Статистика : підручн. / О. Є. Лугінін,



- С. В. Білоусова. – К. : Центр навчальної літератури, 2005. – 580 с.
16. Мармоза А. Т. Теорія статистики : навч. посіб. / А. Т. Мармоза. – К. : Ельга, Ніка-Центр, 2003. – 392 с.
17. Матковський С. О. Статистика : навч. посіб. / С. О. Матковський, Л. І. Гальків, О. С. Гринькевич, О. І. Сорочак. – Л. : Новий Світ-2000, 2011. – 429 с.
18. Овчарук Р. Ю. Теорія статистики : навч. посіб. / Р. Ю. Овчарук. – К. : Вікар, 2003. – 204 с.
19. Парфенцева Н. Міжнародні статистичні класифікації в Україні. Впровадження і використання / Н. Парфенцева. – К. : Основи, 2000. – 351 с.
20. Про затвердження Стратегії розвитку державної статистики на період до 2017 року : Розпорядження Кабінету Міністрів України від 20 березня 2013 р. № 145-р.
21. Про заходи щодо розвитку державної статистики : Указ Президента України від 22 листопада 1997 р. № 1299/97 // Статистика України. – 1998. – №1.
22. Сторожук В. П. Курс лекцій з дисципліни «Статистика». Частина 1. Теорія статистики / В. П. Сторожук, О. В. Кустовська, Є. І. Ткач, І. М. Шост та ін. ; за ред. Є. І. Ткача. – Тернопіль : Економічна думка, 2006 . – 224 с.
23. Теория статистики : учебн.; под ред. проф. Г. Л. Громько. – М. : ИНФРА-М, 2000. – 414 с.
24. Теория статистики : учебн.; под ред. проф. Р. А. Шмойловой. – М. : Финансы и статистика, 1998. – 576 с.
25. Уманець Т. В. Загальна теорія статистики : навч. посіб. / Т. В. Уманець – К. : Знання, 2006. – 239 с.
26. Уманець Т. В. Статистика : навч. посіб. / Т. В. Уманець, Ю. Б. Пігарев – К. : Вікар, 2003. – 623 с.
27. Харламов А. И. Общая теория статистики : учебн. / А. И. Харламов, О. Э. Башина, В. Т. Бабурин и др. ; под ред. А. А. Спирина, О. Э. Башиной. – М. : Финансы и статистика, 2000. – 440 с.
28. Щурик М. В. Статистика : навч. посіб. / М. В. Щурик. – Л. : Магнолія, 2006, 2013. – 545 с.

## ЗМІСТ

ПЕРЕДМОВА .....	3
Тема 1. ПРЕДМЕТ, МЕТОД І ОСНОВНІ ЗАВДАННЯ СТАТИСТИКИ. БАЗОВІ ПОНЯТТЯ, ТЕРМІНИ, ФОРМИ, СПОСОБИ І ВИДИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ .....	6
1.1. Статистика як суспільна наука .....	6
1.2. Предмет, метод і основні завдання статистики .....	9
1.3. Поняття, категорії та показники статистики .....	12
1.4. Базові поняття, терміни, форми, способи і види статистичного спостереження .....	14
Контрольні запитання до теми 1 .....	23
Тема 2. СТАТИСТИЧНЕ ЗВЕДЕННЯ, ГРУПУВАННЯ. СТАТИСТИЧНІ ТАБЛИЦІ І ГРАФІКИ.....	24
2.1. Сутність, організація і техніка статистичного зведення .....	24
2.2. Основні види та завдання статистичних групувань .....	26
2.3. Принципи вибору груповальної ознаки та утворення груп .....	30
2.4. Табличне відображення статистичних даних.....	34
2.5. Графічне відображення статистичних даних .....	38
Контрольні запитання до теми 2.....	43
Тема 3. СТАТИСТИЧНІ УЗАГАЛЬНЮЮЧІ ПОКАЗНИКИ .....	45
3.1. Види та функції статистичних показників .....	45
3.2. Абсолютні та відносні статистичні показники .....	47
3.3. Середні величини .....	54
3.4. Основні правила і особливості застосування середніх для розрахунків економічних показників .....	66
Контрольні запитання до теми 3.....	73
Тема 4. РЯДИ РОЗПОДІЛУ. АНАЛІЗ ВАРІАЦІЙ ТА ФОРМИ РОЗПОДІЛУ. АНАЛІЗ КОНЦЕНТРАЦІЇ, ДИФЕРЕНЦІАЦІЇ ТА ПОДІБНОСТІ РОЗПОДІЛІВ .....	74
4.1. Поняття варіації та її основні показники .....	74
4.2. Поняття та види рядів розподілу .....	77
4.3. Характеристики форми розподілу .....	78
4.4. Види та взаємозв'язок дисперсій.....	86
4.5. Різновиди та характеристики форм розподілів .....	89
4.6. Оцінка і аналіз нерівномірності та подібності розподілів .....	92
Контрольні запитання до теми 4.....	103
Тема 5. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ .....	105
5.1. Поняття і сутність вибіркового методу, причини і умови його	

застосування .....	105
5.2. Способи відбору одиниць генеральної сукупності .....	114
5.3. Визначення помилки вибіркової середньої .....	119
5.4. Мала вибірка .....	124
5.5. Способи розрахунків характеристик вибірки .....	127
Контрольні запитання до теми 5 .....	131
Тема 6. ІНДЕКСНИЙ МЕТОД У СТАТИСТИЧНИХ ДОСЛІДЖЕННЯХ .....	133
6.1. Сутність, функції та класифікація індексів у статистичному аналізі .....	133
6.2. Загальні (складні, агрегатні) індекси .....	137
6.3. Середні індекси. Індекси середніх величин .....	145
6.4. Базисні і ланцюгові індекси .....	147
6.5. Системи індексів. Аналіз факторів розвитку соціально-економічних явищ індексним методом .....	150
Контрольні запитання до теми 6 .....	155
Тема 7. СТАТИСТИЧНЕ ВИВЧЕННЯ ДИНАМІКИ .....	157
СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ .....	157
7.1. Основні поняття і види рядів динаміки .....	157
7.2. Аналітичні та середні показники ряду динаміки .....	158
7.3. Методи обробки динамічних рядів .....	163
7.4. Виявлення тенденцій розвитку соціально-економічних явищ в динаміці .....	165
7.5. Дослідження сезонних коливань в рядах динаміки .....	174
7.6. Особливості дослідження взаємозв'язків в рядах динаміки .....	182
Контрольні запитання до теми 7 .....	185
Тема 8. СТАТИСТИЧНІ МЕТОДИ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКІВ .....	187
8.1. Зв'язки суспільних явищ і завдання їх статистичного вивчення .....	187
8.2. Загальні методи вивчення зв'язків .....	189
8.3. Кореляційний і регресійний методи аналізу зв'язку .....	200
8.4. Нелінійні залежності .....	203
8.5. Багатофакторний кореляційно-регресійний аналіз .....	204
8.6. Непараметричні показники тісноти зв'язку .....	212
Контрольні запитання до теми 8 .....	214
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ .....	216

Навчальне видання

СТРИГУЛЬ Лариса Станіславівна  
АЛЕКСАНДРОВА Вікторія Олександрівна

**Статистика**  
Текст лекцій  
для студентів спеціальності 6.030509 «Облік і аудит»  
всіх форм навчання

Відповідальний за випуск  
Роботу рекомендував до видання

О. В. Манойленко  
П. Г. Перерва

Редактор Л. Л. Яковлєва  
Комп'ютерна верстка В. В. Мартинова

План 2013 р., поз. 110

Підписано до друку \_\_\_\_\_.\_\_\_\_\_.15. Формат 60 x 84 <sup>1/16</sup>.  
Папір офсетний. Гарнітура Таймс. Друк – офсетний.  
Ум. друк. арк. 8,5. Наклад 300 прим.  
Зам. № \_\_\_\_\_ Ціна договірна.

Надруковано у друкарні ФЛП Томенко Ю.І.  
м. Харків, вул. Руднева, 4  
тел. (057) 757-93-82