

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Є. О. Семенов, В. Ф. Райко, О. І. Ільїнська

ЧИСЛОВІ МЕТОДИ АНАЛІЗУ З ОХОРОНИ ПРАЦІ

Текст лекцій

**для студентів освітньо-кваліфікаційного рівня бакалавр
спеціальності 263 «Цивільна безпека»**

Затверджено
редакційно-видавничою радою
університету,
протокол № 1 від 19.02.2020 р.

Харків
НТУ «ХП»
2020

УДК 331.45:519.22(075)

С30

Рецензенти: *Михайлова Є.О.* – канд. техн. наук, доцент кафедри природоохоронних технологій, екології та безпеки життєдіяльності, ХНЕУ ім. С. Кузнеця.

Семенов Є. О.

С30 Текст лекцій із дисципліни «Числові методи аналізу з охорони праці» для студентів освітньо-кваліфікаційного рівня бакалавр спеціальності 263 «Цивільна безпека» / Є.О. Семенов, В.Ф. Райко, О. І. Ільїнська – Харків: НТУ «ХП», 2020. – 84 с.

Розглянуто основні базові питання за програмою дисципліни «Числові методи аналізу з охорони праці»: первинна обробка та представлення її результатів; абсолютні, відносні та середні величини; показники варіації; вибіркове спостереження; статистичні методи вимірювання взаємозв'язків; кореляційний аналіз; статистичне вивчення динаміки.

Призначено для студентів ВНЗ спеціальностей з охорони праці, інженерно-технічних працівників.

Табл.20. Рис.2. Бібліогр.13 назв.

УДК 331.45:519.22(075)

© Семенов Є. О., Райко В. Ф., Ільїнська О. І., 2020

© НТУ «ХП», 2020

Вступ

Для виробничої діяльності підприємств характерні промислові (виробничі, професійні) ризики, пов'язані з низькою кваліфікацією працівників, незадовільним станом організації охорони праці у виробничих підрозділах підприємства, несправністю інструменту і устаткування (верстатів, приладів, складних технічних комплексів тощо), а також ризики пов'язані з порушенням правил і норм о охорони праці при експлуатації виробничих будівель і споруд. Отже вивчення, оцінка і зменшення ризиків мають велике значення у господарській діяльності кожного підприємства, організації і установи.

Передумовою для створення ефективної системи охорони здоров'я та виробничої безпеки є оцінка ризиків. У процесі оцінки ризиків на підприємстві повинен розроблятися комплекс заходів, що проводяться з метою оцінки стану умов праці на робочих місцях, виявлення шкідливих і небезпечних виробничих факторів, а також приведення умов праці у відповідність до діючого законодавства.

Числові методи аналізу з охорони праці – дисципліна, що вивчає різні методи аналізу виробничого травматизму. Використання числових методів оптимізує розробку дієвих профілактичних заходів з врахуванням конкретних виробничих умов суб'єкта господарювання.

Метою викладання дисципліни є отримання знань з використання числових методів для аналізу інформації, узагальнення та розробки профілактичних заходів з охорони праці.

Об'єктами вивчення курсу є:

- методи аналізу виробничого травматизму;
- методи математичної статистики;

– використання числових методів для розробки профілактичних заходів.

Предметом вивчення курсу є елементи засади апарату числових методів, закони, що діють у сфері масових випадкових подій та явищ, методи систематизації, опрацювання й аналізу масових числових даних.

Задачами курсу “Числові методи аналізу з охорони праці” є вивчення студентами основ:

- числових методів аналізу виробничого травматизму;
- групових та топографічних методів аналізу травматизму;
- комплексних методів математичної статистики;
- методів фізичного та математичного моделювання;
- використання числових методів для розробки профілактичних заходів.

Курс складається з 90 годин, з яких – лекцій 16 годин, практичних занять –16 годин, самостійна робота студентів – 48 годин. Вивчення курсу закінчується складанням іспиту.

ЛЕКЦІЯ 1. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

1.1 Сутність статистичного спостереження.

1.2 План статистичного спостереження. Основні визначення.

1.3 Форми, види і способи статистичного спостереження.

1.1 Сутність статистичного спостереження

Статистичне спостереження – це спланована, науково організована реєстрація масових даних про соціальні, економічні та інші явища і процеси.

Статистичні дані – це масові системні кількісні характеристики соціально-економічних явищ та процесів.

Статистичне спостереження може бути первинним або вторинним.

Первинне спостереження – це реєстрація даних, що надходять безпосередньо від об'єкта спостереження.

Вторинне спостереження – це збирання раніше зареєстрованих та оброблених даних.

Статистична інформація повинна відповідати певним вимогам:

- вірогідність (відповідність реальному стану);
- повнота за обсягом та змістом;
- своєчасність;
- порівнянність за часом або у просторі;
- доступність.

Джерелами статистичної інформації є:

- звітність (вторинне спостереження);
- спеціально організоване статистичне спостереження – реєстрація даних, що надходять безпосередньо від об'єкта (первинне спостереження);
- реєстри – первинне постійне спостереження.

Звітність – це форма організації статистичного спостереження, при якій інформація регулярно поступає в статистичні органи від підприємств, установ та організацій у вигляді обов'язкових звітів.

Звітність має певні характерні ознаки:

- обов'язковість надання звітів за раніше обумовленими програмою та строком, встановленим до цього;
- обґрунтованість інформації документами первинного обліку і бухгалтерського обліку;
- юридична сила звітів і відповідальність.

Звіт надається за затвердженою формою, яка має наступні обов'язкові реквізити:

- форма та дата затвердження;
- адреса звіту;
- термін звітності;
- назва особи, яка звітує;
- назва і адреса вищої інстанції по відношенню до особи, що звітує;
- показники звіту;
- підписи відповідальних осіб.

1.2. План статистичного спостереження. Основні визначення

Статистичне спостереження здійснюється в 3 етапи:

1. Підготовка спостереження. Тут постають і вирішуються основні методологічні та організаційні питання: де, хто, коли проводить спостереження і що для цього необхідно. Та цьому етапі складається докладний план статистичного спостереження.

2. Реєстрація статистичних даних. На цьому етапі здійснюється безпосередній процес збирання даних, який потребує чіткої дії, координаності всіх виконавчих служб.

3. Формування бази даних. Цей етап передбачає контроль та нагромадження даних спостереження, а також їх збереження. На цьому етапі відпрацьовуються система оперативного доступу та пошуку даних.

Статистичне спостереження проводиться у відповідності з планом статистичного спостереження, що містить *програмно-методологічні та організаційні питання*.

Програмно-методологічні питання плану надають відповіді на наступні запитання:

- для чого проводиться спостереження (мета спостереження);
- що спостерігається (об'єкт спостереження);
- які складові частини об'єкта (одиниця сукупності);
- яке джерело інформації (одиниця спостереження);
- на які питання планується одержати відповіді (програма спостереження).

Організаційні питання спостереження надають відповіді на наступні запитання:

- хто проводить спостереження;
- де проводиться спостереження (місце спостереження);
- за допомогою чого проводиться спостереження (матеріально-технічне забезпечення обстеження);
- який спосіб забезпечення точності результатів спостереження (система контролю та пробні обстеження);
- коли проводиться спостереження (час, період обстеження).

Основні визначення. *Мета спостереження* – це одержання статистичних даних, які є підставою для розрахунку узагальненої характеристики стану та розвитку явища або процесу. Метою спостереження підприємств галузі є об'єктивна характеристика стану та прийняття рішення про подальший розвиток підприємств.

Об'єкт спостереження – сукупність явищ, що підлягають обстеженню. Щоб уникнути різного тлумачення результатів спостереження може застосовуватись набори кількісних та якісних обмежувальних ознак.

Час спостереження – *об'єктивний час* – час, до якого відносять дані спостереження.

Якщо об'єкт спостереження – це процес, то визначають інтервали часу, протягом якого накопичуються дані.

Якщо об'єкт спостереження – стан, то обирається *критичний момент* – *момент часу*, станом на який реєструються дані. Критичний момент є різновидом об'єктивного часу.

Суб'єктивний час – період спостереження – час, протягом якого здійснюється реєстрація статистичних даних. В окремих випадках суб'єктивний час може співпадати з об'єктивним часом.

Статистична сукупність – це множина елементів, яка складається головним чином під дією загальних причин та умов, які пов'язані загальними характерними рисами, загальними важливими ознаками, загальністю процесів розвитку.

Окремі елементи статистичної сукупності зветься *одиницями сукупності*.

Одиниця сукупності – первинний елемент об'єкта спостереження, що є носієм ознак, які підлягають реєстрації.

Ознака – це властивість, яка відбиває сутність, характер, особливість одиниці статистичної сукупності.

За формою виразу ознаки розрізняють: описові (атрибутивні) та кількісні, котрі у свою чергу бувають дискретні та неперервні.

Одиниця спостереження – це первинна одиниця статистичної сукупності, від якої одержують інформацію.

Програма спостереження – перелік запитань, на які слід одержати відповіді у результаті спостереження і які мусять бути узгоджені з метою спостереження. До програми спостереження входить *статистичний інструментарій* – це набір статистичних формулярів, інструкцій та пояснень щодо спостереження.

Статистичний формуляр – це обліковий документ, що містить характеристику об'єкта спостереження та статистичні дані про нього.

Він може мати форму статистичного звіту, переписного або опитувального листа, анкети, картки або простого бланку.

1.3. Форми, види і способи статистичного спостереження

Форми статистичного спостереження

З точки зору організації статистичного спостереження розрізняють три основні організаційні форми спостереження: звітність, спеціально організоване статистичне спостереження та реєстри.

Звітність – це форма спостереження, при якій кожний суб'єкт ді-

яльності регулярно подає дані в державні органи статистики та відомства у вигляді документів (звітів) спеціально затвердженої форми.

Розрізняють *типову і спеціалізовану* звітність.

Типова звітність має єдину форму і зміст для всіх підприємств і організацій незалежно від форми власності та відомчого підпорядкування.

Спеціалізована звітність властива тим підприємствам чи окремим виробництвам, що мають свої специфічні властивості.

Залежно від рівня затвердження та призначення звітність поділяється на *зовнішню та внутрішню*.

За періодичністю подання звітність буває *поточна* (тижнева, місячна, квартальна), яка охоплює показники поточної діяльності суб'єктів, і *річна*.

Звітність характеризується такими властивостями, як *обов'язковість, систематичність, вірогідність*.

Обов'язковість – подання звітів обов'язкове для усіх зареєстрованих суб'єктів діяльності з додержанням уніфікованої форми, затвердженого переліку показників, із зазначенням реквізитів підзвітного об'єкту.

Систематичність передбачає регулярне, своєчасне складання та подання звітності в затвержені терміни.

Вірогідність – дані, наведені у звітності, мають відповідати дійсності і виключати будь-які викривлення.

Спеціально організоване статистичне спостереження – охоплює ті сторони суспільного життя, які не відобразились у звітності. До них належать: переписи, одноразові обліки, спеціальні обстеження, опитування.

Переписи проводяться періодично або одноразово і дають повну характеристику масового явища станом на якусь дату або певний момент часу.

Обліки – суцільні спостереження масових даних, які ґрунтуються на даних огляду, опитування та документальних записів.

Спеціальні спостереження – переважно не суцільне обстеження масових явищ згідно з певною тематикою, що виходить з певною темати-

кою, що виходить за межі звітності.

Опитування – це, як правило, не суцільне спостереження з метою вивчення думок, мотивів, оцінок, що реєструються зі слів респондентів.

Статистичний реєстр – це список або перелік одиниць певного об'єкта спостереження зі значенням необхідних ознак, який складається та оновлюється під час постійного відстеження.

Види та способи статистичного спостереження

За часом реєстрації даних статистичне спостереження поділяється на поточне, періодичне та одноразове.

Поточне спостереження полягає у системній реєстрації фактів по мірі їх виникнення або збирання фактів щодо безперервного процесу.

Періодичне спостереження проводиться регулярно, через певні проміжки часу.

Одноразове спостереження проводять епізодично по мірі виникнення потреби в дослідженні явища чи процесу з метою вирішення певних завдань.

За способами одержання статистичних даних спостереження поділяється на безпосереднє, документальне і опитування.

Документальний облік ґрунтується на даних різноманітних документів первинного обліку.

Опитування респондентів – це таке спостереження, при якому відповіді на питання формуляра записують зі слів респондента. Воно буває експедиційне, самореєстраційне, кореспондентське і анкетне.

За повнотою охоплення одиниці сукупності спостереження поділяють на суцільне і несуцільне.

При *суцільному* спостереженні обстеженню і реєстрації підлягають усі без винятку елементи сукупності, а при *несуцільному* спостереженні обліку підлягають не всі елементи, а тільки їх частина.

Несуцільне спостереження поділяють на такі види: спостереження основного масиву, вибіркове, монографічне, анкетне, моніторинг.

Спостереження основного масиву – охоплює переважно частину елементів сукупності, обсяг значень істотної ознаки у яких визначає розмір явища, тобто при обстеженні відкидають певну кількість одиниць, які не можуть істотно вплинути на характеристику сукупності в цілому.

При вибірковому спостереженні також обстежують певну вибірково одібрану частину.

Монографічне спостереження передбачає детальне обстеження лише окремих типових елементів сукупності.

Моніторинг – це спеціально організоване систематичне спостереження за станом певного середовища. Прикладом моніторингу є моніторинг рівня радіаційного забруднення на територіях, що постраждали в наслідок аварії на ЧАЕС.

Контрольні запитання

1. Назвіть вимоги, що ставляться до статистичної інформації.
2. Вкажіть джерела статистичної інформації.
3. Назвіть етапи проведення будь-якого статистичного спостереження.
4. Вкажіть основні організаційні форми спостереження с точки зору організації статистичного спостереження.
5. На які види поділяється спеціально організоване статистичне спостереження?
6. Якими властивостями характеризується звітність?
7. Як поділяється за способами одержання статистичних даних спостереження?
8. Чим відрізняється несуцільне спостереження від суцільного?
9. На які види поділяється несуцільне спостереження?
10. Що таке моніторинг?

ЛЕКЦІЯ 2. ПЕРВИННА ОБРОБКА ТА ПРЕДСТАВЛЕННЯ ЇЇ РЕ- ЗУЛЬТАТІВ

2.1 Суть статистичного зведення.

2.2 Статистичні групування. Види групувань.

2.3 Ряди розподілу.

2.4 Статистичні таблиці.

2.5 Графічні способи зображення статистичних даних. Види графіків.

2.1. Суть статистичного зведення

Другою стадією статистичного дослідження є статистичне зведення і групування, оскільки після збору даних, ми повинні їх звести, згрупувати для обробки.

Зведення – це комплекс послідовних операцій по узагальненню конкретних поодиноких факторів, які утворюють сукупність, для виявлення типових рис і закономірностей, що належать досліджуваному явищу в цілому.

За зведеними даними розраховуються узагальнюючі показники, виконується порівняльний аналіз, а також аналіз причин групових відмінностей, вивчаються взаємозв'язки між ознаками.

Зведення може бути *простим* або *складним*.

Просте зведення – це простий підрахунок підсумків первинних статистичних даних.

Складне зведення передбачає групування, види груповальної ознаки, встановлення меж групування, підрахунок групових і узагальнюючих підсумків, а також викладення результатів зведення у вигляді таблиць чи графіків.

За організацією робіт і формою обробки даних розрізняють *централізоване* і *децентралізоване* зведення.

2.2 Статистичні групування. Види групувань

Одним основних і найбільш поширених методів обробки і аналізу первинної статистичної інформації групування.

Під групуванням в статистиці розуміють розподіл одиниць статистичної сукупності на групи, однорідні в якому-небудь суттєвому відношенні.

Воно є якраз той метод і та стадія, пропустивши яку ми не можемо застосовувати інші методи. Тому в статистиці групування використовується для вирішення різних завдань, таких як, наприклад:

- визначення і вивчення структури і структурних зрушень сукупності;
- виявлення соціально-економічних типів явищ і процесів;
- виявлення і характеризування зв'язків і залежностей між явищами та їх ознаками (таке дослідження має назву аналітичної функції групування).

Відповідно до цих трьох функцій розрізняють різні види групування: структурні, типологічні і аналітичні.

Групування, в результаті якого виділяють однорідні групи або типи явищ, як вираз конкретного суспільного процесу називаються *типологічними*.

Прикладом типологічного групування є розподіл житлового фонду Львівської області за формами власності, який наведено в табл. 2.1.

Таблиця 2.1 – Розподіл житлового фонду Львівської області за формами власності (у відсотках)

Показники	1990 рік	2000 рік
Весь житловий фонд, у тому числі:	100,0	100,0
– державний та колективний житловий фонд	45,6	22,6
– приватний житловий фонд	54,4	77,4

Структурними групуваннями називаються групування, які харак-

теризують розподіл одиниць однотипної сукупності за будь-якою ознакою.

Типологічні і структурні групування дуже близькі один до одного: типологічні групування виділяють самі типи, а структурні – вказують питому вагу окремих типів у загальній масі.

Прикладом структурного групування може бути розподіл спільних підприємств регіону за чисельністю працюючих, який наведено в табл. 2.2.

Таблиця 2.2 – Розподіл спільних підприємств регіону за чисельністю працюючих

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

Аналітичні групування – це групування, які визначають взаємозв'язок між різними ознаками одиниць статистичної сукупності. За допомогою такого групування можна виявити певні взаємозв'язки між факторними і результативними ознаками. Наприклад, залежність між рівнем кваліфікації працівника та його заробітною працею. Аналітичні групування є дуже складними і для того, щоб зрозуміти, як вони будуються, необхідно чітко виділити факторні і результативні ознаки в досліджуваному явищі. Можливі змішання цих типів групування.

Явища суспільного життя тісно пов'язані один з одним. Взаємопов'язані ознаки можна поділити на *факторні і результативні*.

Ознака, яка впливає на інші та зумовлює їх зміни називається *факторною*.

Ознака, яка залежить від інших ознак і змінюється під їх впливом називається *результативною*.

За кількістю ознак покладених в основу групування розрізняють – *прості і складні*.

Прості групування – це такі групування, які здійснені на підставі однієї ознаки.

Складне групування може бути комбінаційним, якщо в його основі скомбіновано більше ознак, або *багатомірним* – якщо проводиться за певною множиною одночасно.

Комбінаційні групування дають можливість комплексного характеризування досліджуваного явища чи процесу, виявити взаємозв'язки та залежність між ними.

За формою вираження групувальні ознаки можуть бути *атрибутивними* та *кількісними*.

Окремим видом атрибутивних групувань є групування *альтернативною ознакою*, коли є лише 2 варіанти атрибутивної ознаки, причому один виключає інший. Приклад: розподіл на студентів, що отримують стипендію, і ти що ні.

Кількісні ознаки поділяються на *дискретні та інтервальні*.

Наступним кроком після визначення групувальної ознаки є розподіл одиниць сукупності на групи. Тут виникає питання, щодо кількості груп і величини інтервалу.

При групуванні за *атрибутивними ознаками* кількість груп дорівнює кількості атрибутивних ознак.

При групуванні за *кількісною ознакою* постає питання, щодо кількості груп і інтервалів групування.

Якщо відома чисельність сукупності N , то число груп n можна визначити за формулою Стерджеса:

$$n = 1 + 3,322 \cdot \lg N$$

Наприклад, при 200 одиницях сукупності число груп дорівнює:

$$n = 1 + 3,322 \cdot \lg 200 = 8,644 \approx 9$$

За величиною розрізняють інтервали рівні і нерівні.

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.

Величина інтервалу при групуванні із застосуванням рівних інтервалів визначають за формулою:

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

де h – величина інтервалу; x_{\max} – максимальне значення; x_{\min} – мінімальне значення; n – кількість груп у сукупності.

Розрізняють інтервали *відкриті* і *закриті*.

Відкритий інтервал – інтервал з відсутньою однією із границь. Наприклад, більше 100, менше 1.

Закриті інтервали – інтервали, в яких присутні всі границі.

Особливим видом групування є класифікація.

Класифікацією називається систематизований розподіл явищ і процесів (об'єктів) на визначені групи, класи, розряди на підставі їх подібності і розбіжності.

Класифікації відрізняються від групувань тим, що групувальною основою класифікації є якісна ознака, вони більш стійкі, сталі і стандартні.

2.3 Ряди розподілу

Результати зведення і групування можна оформляти у вигляді рядів розподілу, статистичних таблиць та графіків.

Статистичний ряд розподілу – це ряд, який характеризує розподіл одиниць сукупності по групах за будь-якою ознакою, різновидності якої розташовані у певному порядку.

Він складається з двох елементів: *варіантів і частот*.

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 (напр. у відсотках).

Залежно від статистичної групувальної ознаки ряди розподілу поділяють на *атрибутивні і варіаційні*.

Ряд розподілу одиниць сукупності, в основу якого покладено якісні ознаки називається *атрибутивним*. Прикладом атрибутивного ряду

розподілу може бути розподіл населення на міське і сільське.

Ряд розподілу одиниць сукупності за ознакою, що має кількісне вираження, називається *варіаційним*.

Відповідно до варіації ознаки, варіаційні ряди розподілу можуть бути *дискретними і інтервальними*.

В *дискретному ряді* розподілу кількісна ознака приймає тільки цілі значення. Приклад дискретного ряду показано в табл. 2.3.

Таблиця 2.3 – Приклад дискретного ряду розподілу

Групи сімей (варіанта)	Кількість сімей (частота)	Акумулятивна частка
1	80	80
2	20	100
3	350	450
4	45	495
5	5	500

Накопичення часток по мірі зростання (спадання) ознаки називається *аккумулятивна частка*.

Коли значення варіант ряду виражено у вигляді інтервалу, такий ряд розподілу називається *інтервальним*. Приклад інтервального ряду показано в табл. 2.4.

Таблиця 2.4 – Приклад інтервального ряду розподілу

Заробітна плата, грн.	Частка працюючих підприємства, %
5000 – 6000	15
6001 – 7000	20
7001 – 8000	26
8001 – 9000	23
9001 – 10000	9
10001 – 11000	7
Разом	100

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*.

Графічно ряди розподілу зображаються у вигляді гістограми або полігону (де ось ОУ – результативна ознака, ось ОХ – факторна ознака). Графічне зображення ряду розподілу показано на рис. 2.1.

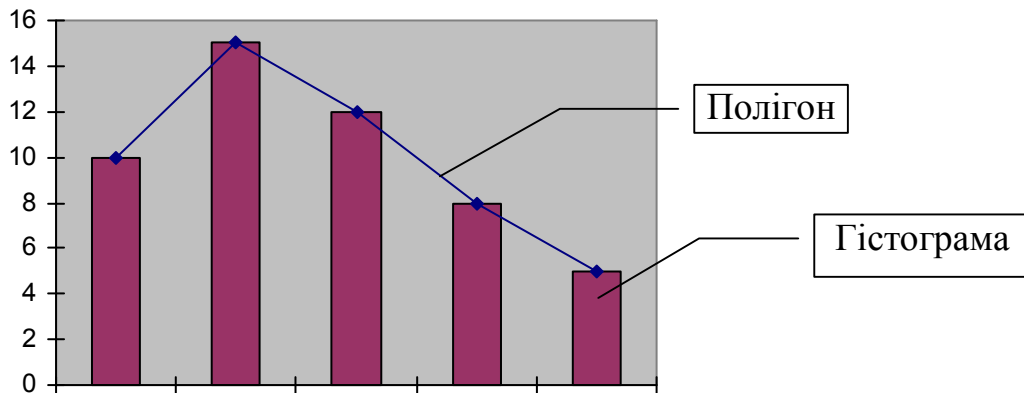


Рисунок 2.1 – Графічний ряд розподілу

2.4 Статистичні таблиці

Найчастіше всі зведення і групування оформлюються у вигляді статистичних таблиць.

Статистична таблиця – це форма найбільш раціонального, наукового і систематизованого викладу числових результатів зведення і обробки статистичних матеріалів. Статистична таблиця за своїм логічним змістом розглядається як «статистичне речення». «*Підмет*» – об'єкт дослідження, «*присудок*» – система показників, що характеризують об'єкт дослідження.

Підметом статистичної таблиці називається статистична сукупність або її частина, яка характеризується числовими показниками.

Присудком називається та частина, що вміщає показники, які характеризують досліджувану сукупність та її частини (тобто підмет).

Залежно від структури об'єкта дослідження (підмета) статистичні таблиці поділяються на: прості, групові і комбінаційні.

Простими називаються такі таблиці, в підметі яких міститься пе-

релік об'єктів, адміністративних і територіальних одиниць або ряд періодів, дат, охарактеризованих числовими показниками. Прості таблиці є найбільш поширеними.

Групові таблиці – це таблиці, підмет яких містить одиниці досліджуваного об'єкту, згрупованих за певною суттєвою ознакою.

Комбінаційні таблиці – це таблиці, в яких підмет побудований за двома і більше ознаками. Приклад комбінаційної таблиці наведено в табл. 2.5

Таблиця 2.5 – Групування банків України за величиною статутного фонду на 1.03.1999 р.

Розмір статутного фонду	Прибутковість активів	Кількість	Питома вага
5,0 – 7,5	1,5 – 2,0	2	11
	2,0 – 2,5	1	23
	2,5 – 3,0	5	11
Разом	–	8	45
7,5 – 10,0	1,5 – 2,0	7	10
	2,0 – 2,5	2	25
	2,5 – 3,0	3	20
Разом	–	12	55
В цілому по сукупності	–	20	100

За метою дослідження та призначення статистичні таблиці поділяють на: описово-інформаційні, аналітичні, типологічні, спеціального призначення.

Правила складання статистичних таблиць:

- 1) таблиця повинна бути компактною і мати тільки ті вихідні дані, які безпосередньо відображають досліджуване явище;
- 2) заголовок таблиці, назви граф і строчок повинні бути зрозумілими, чіткими, лаконічними і закінченими;
- 3) в графах допускаються скорочення тільки при необхідності;
- 4) таблиця повинна бути замкнута і мати підсумкову строку, яка може знаходитись на початку таблиці;

5) показники, що характеризують один одного, повинні міститися поруч;

б) графі нумерують арабськими цифрами, підмет – латинськими літерами;

7) якщо явище повністю відсутнє, то в клітинці, де має бути його кількісне значення ставиться тире. Якщо дослідник не може знайти відомості про певне явище, то в клітинку ставиться три крапки, або "н.в." – немає відомостей. Якщо дана клітинка не заповнюється, то в неї ставиться хрестик чи зірочка.

2.5 Графічні способи зображення статистичних даних. Види графіків

За загальним призначенням графіки поділяють на: аналітичні, ілюстративні та інформаційні.

За функціонально-цільовим призначенням поділяються: графіки групувань і рядів розподілу, ряди динаміки, графіки взаємозв'язку і порівняння.

За формою графічного сектору: лінійні, стовпчикові, стрічкові, секторні, кільцеві, радіальні та інші.

За видом поля розрізняють: діаграми, статистичні карти або картограми.

Важливим засобом наочного відображення та аналізу суспільних явищ є картограма та картодіаграма.

Картограма – це статистична карта, на якій розподіл досліджуваною ознаки по території відображено умовно (штрихуванням, фарбуванням) залежно від величини статистичної ознаки. Бувають фонові та точкові.

Картодіаграма – це графічне зображення закономірностей просторового розміщення статистичних показників.

Контрольні запитання

1. Назвіть в чому полягає суть статистичного зведення.
2. Вкажіть які існують види зведення.
3. Що розуміють під поняттям групування?
4. Вкажіть які існують види групувань.
5. Зазначте чим факторна ознака відрізняється від результативної?
6. Як знайти кількість груп при групуванні за атрибутивними ознаками?
7. Як знайти кількість груп при групуванні за кількісною ознакою?
8. Назвіть з яких елементів складається ряд розподілу.
9. Назвіть з яких елементів складається статистична таблиця.
10. Вкажіть як поділяються графіки залежно від форми графічного сектору.

ЛЕКЦІЯ 3. АБСОЛЮТНІ, ВІДНОСНІ ТА СЕРЕДНІ ВЕЛИЧИНИ

3.1 Абсолютні та відносні величини. Основні поняття і категорії.

3.2 Середні величини: поняття, види та способи їх обрахування.

3.1 Абсолютні та відносні величини. Основні поняття і категорії

Абсолютні величини (або абсолютні статистичні величини) виражають обсяги, розміри та рівні процесів і явищ.

Вони поділяються на *індивідуальні* та *сумарні*.

Індивідуальні виражають розміри кількісних ознак окремих одиниць сукупності, а *сумарні* характеризують величину тієї чи іншої ознаки усіх одиниць сукупності або окремих її груп, і отримують в результаті підсумування індивідуальних значень.

Абсолютні величини можуть бути виражені в натуральному вигляді (тони, метри, кілометри), в умовно-натуральному (в перерахунку на якусь умовну одиницю: умовне паливо, тощо), трудові (людино-години, людино-дні), комплексні (тонно-кілометри) і вартісний (в грошових одиницях).

Абсолютні статистичні показники можуть бути *моментні* та *інтервальні*. Моментні показники – неадитивні, а інтервальні – адитивні. Поняття адитивність означає – підсумування.

Відносні величини – це статистичні показники, які виражають кількісне співвідношення між явищами суспільного життя. Відносні величини є узагальнюючим показник, який дає міру співвідношення двох порівнювальних абсолютних величин, одна з яких береться з базового рівняння (називається базовою величиною), а ту, яку порівнюють з базовою – порівнювальна.

Якщо абсолютна величина показує, на скільки певне явище більше, то відносна величина показує в скільки разів це явище більше.

Відносна величина може бути виражена коефіцієнтом, або може бути виражена в процентах.

Приклади:

Коефіцієнт: $k = a_1 / a_2$ (в скільки разів явище a_1 більше/менше

явища a_2)

Темп зростання: $T_{зр.} = k \cdot 100 \%$ (у відсотках)

Темп приросту: $T_{пр.} = T_{зр.} - 1$ (або 100 %)

До відносних величин відносяться: *відносна величина планового завдання, відносна величина динаміки, відносна величина структури, відносна величина координації, інтенсивності розвитку і порівняння.*

Відносна величина планового завдання характеризує відношення величини показника, встановленої на плановий період, до величини показника, досягнутої до планового періоду або до якоїсь норми, стандарту, еталону:

$$y_{пл.з.} = \frac{y_{пл.}}{y_{0(норма)}}. \quad (3.1)$$

Відносна величина виконання плану – це така відносна величина, яка характеризує виконання плану за певний період:

$$y_{в.пл.} = \frac{y_1}{y_{пл.}}. \quad (3.2)$$

Відносна величина динаміки – це відносна величина, яка показує відношення досягнутого рівня розвитку явища до рівня, який існував до того, або відносно еталона, норми, стандарту. Характеризує розвиток явища в часі і просторі:

$$y_{\delta} = \frac{y_1}{y_0}. \quad (3.3)$$

Між відносними величинами планового завдання, виконання плану і динаміки існує спів залежність:

$$\begin{aligned} y_{пл.з.} \cdot y_{в.пл.} &= y_{\delta} \\ \frac{y_{пл.}}{y_0} \cdot \frac{y_1}{y_{пл.}} &= \frac{y_1}{y_0}. \end{aligned} \quad (3.4)$$

Відносними величинами структури називають такі величини, які характеризують відношення частки до цілого:

$$y_{стр.} = \frac{f_i}{\sum f_i} \quad (3.5)$$

Відносні величини координації характеризують співвідношення між складовими частинами цілого:

$$y_{\text{коор.}} = \frac{f_i}{f_j}; \quad (j \neq i) \quad (3.6)$$

Відносна величина порівняння – показує співвідношення односторонніх величин, що стосується різних об'єктів, різних територій, але за той же самий період.

Наприклад, населення м. Києва 3,5 млн. чол., населення Харкова – 1,75 млн. чол. Отже населення Києва в 2 рази більше за населення Харкова.

3.2 Середні величини: поняття, види та способи їх обрахування

Середня величина – це узагальнюючі показник, які характеризують рівень варіюючої ознаки в якісно однорідній сукупності.

Сукупність, яку ми збираємося характеризувати середньою величиною повинна бути:

- 1) якісно однорідною, однотипною;
- 2) складатися з багатьох одиниць.

Середні величини можуть бути *абсолютними* або *відносними* залежно від вихідної бази.

Середні можуть бути *прості* і *зважені*.

За допомогою середніх величин вирішуються наступні питання статистичного дослідження:

- характеристика досягнутого розвитку явища або процесу;
- порівняння показників, обчислених по різних сукупностях;
- характеристика розвитку (варіації) явища у часі і просторі;
- вивчення взаємозв'язків між показниками;
- інші інженерні завдання.

Види середніх величин та способи їх обрахування

Середні поділяються на два великі класи: степеневі (сюди належать середня арифметична, середня гармонійна, середня геометрична, середня квадратична тощо) і структурні.

Ступеневі середні. Одним з найбільш поширених видів середньої є середня арифметична. Вона застосовується у тих випадках, коли обсяг

варіюючої ознаки для всієї сукупності формується як сума значень ознаки у окремих одиниць сукупності, що вивчається.

Середня арифметична буває двох видів: проста і зважена.

Середньоарифметична проста:

$$\bar{x}_{ap.} = \frac{\sum x}{n}, \quad (3.7)$$

де n – кількість одиниць сукупності, x – варіююча ознака.

Вона застосовується в тому випадку, коли у нас варіююча арифметична ознака має різні значення, і є незгруповані дані.

Якщо ж ми маємо згруповані дані, або варіююча ознака зустрічається декілька раз, то застосовується *середня арифметична зважена*.

$$\bar{x}_{ap.} = \frac{\sum x \cdot f}{\sum f}, \quad (3.8)$$

де x – варіююча ознака, f – абсолютна кількість повторення варіюючої ознаки.

Приклад використання середньої арифметичної для розрахунку недискретного ряду. Вихідні дані для розрахунку наведено у вигляді таблиці 3.1

Таблиця 3.1 – Приклад використання середньої арифметичної

Групування робітників за розміром зарплати	Кількість робітників	Фонд заробітної плати
До 100	80	7200
100 – 120	250	27500
120 – 140	320	41600
140 – 160	230	34500
Понад 160	120	20400
Разом	1000	131200

Необхідно знайти середню заробітну плату робітників.

Перш за все ми повинні закрити верхні і нижні границі. Оскільки величина інтервалу в подальших групах дорівнює 20 од., перший інтервал записуємо "80 – 100", останній – "160-180". Потім знайдемо середню інтервалу, результати оформимо у вигляді таблиці 3.2.

Таблиця 3.2 – Результати розрахунків

Групування робітників за розміром зарплати (x)	Кількість робітників (f)	Середини інтервалу	Фонд заробітної плати
До 100	80	90	7200
100 – 120	250	110	27500
120 – 140	320	130	41600
140 – 160	230	150	34500
Понад 160	120	170	20400
Разом	1000		131200

Тоді середня арифметична зважена:

$$\bar{x}_{ар.зв.} = \frac{\sum xf}{\sum f} = \frac{90 \cdot 80 + 110 \cdot 250 + 130 \cdot 320 + 150 \cdot 230 + 170 \cdot 120}{80 + 250 + 320 + 230 + 120} = \frac{131200}{1000} = 131,2$$

Властивості середньої арифметичної

Середня арифметична має ряд математичних властивостей, важливих для спрощеного її обчислення. Найбільш важливі з них такі:

1. Добуток середньої на суму частот завжди дорівнює сумі добутку варіантів на частоти. Тобто:

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

2. Якщо від кожного значення варіанти відняти будь-яке довільне число A , то одержана середня арифметична зменшиться на те ж число.

Тобто:

$$x_a = \frac{\sum (x - A)f}{\sum f} = x - A.$$

3. Якщо до кожного значення варіанти додати будь-яке число A , то середня арифметична збільшиться на це ж число. Тобто:

$$x_a = \frac{\sum (x + A)f}{\sum f} = x + A.$$

4. Якщо кожне значення варіанти помножити на довільне число A ,

то середня арифметична збільшиться у стільки ж раз. Тобто:

$$x_A = \frac{\sum (x \cdot A)f}{\sum f} = x \cdot A.$$

5. Якщо кожне значення варіанти поділити на довільне число A , то середня арифметична зменшиться у стільки ж раз. Тобто:

$$x_a = \frac{\sum \frac{x}{A} f}{\sum f} = \frac{x}{A}.$$

6. Якщо всі частоти поділити або помножити на будь-яке число, то середня арифметична від цього не зміниться.

$$\frac{\sum x \frac{f}{A}}{\sum \frac{f}{A}} = \bar{x}.$$

7. Сума відхилень варіант від середньої величини завжди дорівнює нулю:

$$\sum (x - \bar{x})f = 0.$$

8. Сума квадратів відхилень варіант від середньої буде менше за будь-яку іншу величину:

$$\sum (x - \bar{x})^2 \Rightarrow \min .$$

Середня гармонічна – це обернена величина до середньої арифметичної, обчислена з обернених величин осереднюваних варіюючих ознак. За своїми властивостями середня гармонічна застосовується тоді, коли загальний обсяг ознаки формується як сума зворотних значень варіантів:

$$x_{\text{гарм}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}}. \quad (3.9)$$

Приклад. Використання середньої гармонічної. Автомобіль проїхав певну відстань (візьмемо її за 1) зі швидкістю 40 км/год. Назад він повертався зі швидкістю 60 км/год. Яка ж його середня швидкість?

Для розрахунку використаємо середню гармонічну просту:

$$\bar{x}_{\text{гарм.}} = \frac{n}{\sum \frac{1}{x}} = \frac{1+1}{\frac{1}{40} + \frac{1}{60}} = 48 \text{ (км/год)}.$$

Середня гармонійна зважена застосовується тоді, коли показник, що виступає статистичною вагою “ f ”, відсутній і його слід додатково визначити на основі відомих варіант “ x ” і добутку варіант на частоту “ $x \cdot f$ ”, розраховується за формулою:

$$\bar{x}_{\text{гарм.}} = \frac{\sum \frac{z}{x}}{\sum \frac{z}{x}}; \quad z = x \cdot f. \quad (3.10)$$

Середня геометрична проста – це корінь з добутку величин динаміки.

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

$$\bar{x}_{\text{геом.}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\frac{x_n}{x_1}}. \quad (3.11)$$

Приклад. Нехай є два числа 4 та 16. Потрібно знайти число яке буде в стільки разів більше 4, і у стільки ж разів менше 16.

$$\frac{A}{B} = \frac{B}{C} \Rightarrow B^2 = A \cdot C \Rightarrow B = \sqrt{A \cdot C} = \sqrt{4 \cdot 16} = 8,$$

$$\bar{x}_{\text{геом.}} = \sqrt[3]{4 \cdot 8 \cdot 16} = \sqrt[3]{512} = 8.$$

Середня квадратична проста розраховується за формулою:

$$\bar{x}_{\text{кв.}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}, \quad (3.12)$$

Умовою використання є наявність незгрупованих даних. Середня квадратична характеризує також варіацію. Тому на практиці вона використовується, коли варіанти ряду подані у вигляді відхилень фактичних їх значень від середньої арифметичної чи іншої величини, яка виступає у ролі норми чи нормативу, стандарту.

Наприклад. Потрібно визначити середню сторону п'яти квадратів за умови, що їх площі становили: 25, 36, 64, 100 і 400 м². Тоді маємо середню сторону квадрату, м:

$$\bar{x}_{\text{кв.}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} = \sqrt{\frac{(25 + 36 + 64 + 100 + 400)}{5}} = \sqrt{\frac{625}{5}} = 11,18 .$$

Середня квадратична зважена розраховується за формулою:

$$\bar{x}_{\text{кв.}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 \cdot f_i}{\sum_{i=1}^n f_i}} . \quad (3.13)$$

Умовою використання середньої квадратичної зваженої є розбиття даних на групи.

Середні структурні

Середні арифметична і гармонічна є узагальнюючими характеристиками сукупностей за тією чи іншою варіручою ознакою. В той же час для характеристики структури цих сукупностей застосовуються особливі показники – моду і медіану.

До середніх структурних відносяться дві величини, які називаються "мода" і "медіана".

Мода (модальна величина) ряду – це значення варіанти, що найчастіше зустрічається в даній сукупності (ряді розподілу). У варіаційному ряді це буде варіанта, що має найбільшу частоту.

У дискретному ряді моду легко відшукати візуально, бо це варіанта, якій відповідає найбільша частота.

Іноколи зустрічаються ряди розподілу, в яких не одна, а дві варіанти однаково модальні, тобто мають найбільші частоти. Це значить, що є дві моди, і розподіл тут бімодальний. Бімодальні розподіли вказують на якісну неоднорідність сукупності за досліджуваною ознакою.

В інтервальному ряді легко відшукується лише модальний інтервал, а сама мода визначається за формулою:

$$M_o = x_0 + i \cdot \frac{(f_2 - f_1)}{(f_2 - f_1) + (f_2 - f_3)} , \quad (3.14)$$

де x_0 – це нижня межа модального інтервалу; i – величина інтервалу; f_2 – частота модального інтервалу; f_1 – частота перед модальним інтервалом (того, що передуює модальному); f_3 – частота поза модальним інтервалом (того, що йде після модального інтервалу).

Розрахуймо моду відповідно до табл. 3.2.

$$M_o = 120 + 20 \cdot \frac{(320 - 250)}{(320 - 250) + (320 - 230)} = 120 + 8,8 = 128,8$$

Медіаною називається така величина, що займає серединне положення у варіаційному ряду, в якому варіанти розташовані в зростаючому або спадаючому порядку.

Медіана – це значення ознаки в сукупності, що ділить ранжируваний ряд навпіл: половина варіант має значення, менші за медіани, а половина – значення, більші за медіани.

У *дискретному ряді* медіаною буде значення ознаки, для якого кумулятивна частота S_i дорівнює або перевищує половину обсягу сукупності $\sum_{i=1}^m f_i$ або кумулятивна частість $S_d \geq 0,5$.

В *інтервальному ряду* таким способом визначається медіанний інтервал. Конкретне значення медіани обчислюється за формулою:

$$M_e = x_0 + h_e \cdot \frac{0,5 \sum_{j=1}^m f_j - S_{M_{e-1}}}{f_{M_e}}, \quad (3.15)$$

де x_0 – нижня границя медіанного інтервалу; h_e – значення медіанного інтервалу; f_{M_e} – частота медіанного інтервалу; $S_{M_{e-1}}$ – накопичена частота в інтервалі, що передує медіанному інтервалу.

Результат розрахунків медіани ряду наведено в таблиці 3.3.

Жирним позначено медіанний інтервал: серединою кількості робітників є 500, і він належить до накопиченої частки у третьому ряду.

$$M_e = 120 + 20 \cdot \frac{\frac{1000}{2} - 330}{320} = 120 + 7,5 = 127,5.$$

Таблиця 3.3 – Медіана ряду

Групування робітників за розміром зарплати (x)	Кількість робітників (f)	Середини інтервалу	Фонд заробітної плати	Наростаючий підсумок частот (накопичені частки)
До 100	80	90	7200	80
100 – 120	250	110	27500	330
120 – 140	320	130	41600	650
140 – 160	230	150	34500	880
Понад 160	120	170	20400	1000
Разом	1000	–	131200	–

Структурні величини мода і медіана застосовуються для вивчення внутрішньої будови рядів розподілу, тобто їх структури.

Контрольні запитання

1. Назвіть види абсолютних величин.
2. Чим відрізняються відносні величини від абсолютних?
3. Назвіть величини, які відносяться до відносних.
4. Вкажіть які величини належать до степеневих середніх величин.
5. Вкажіть які величини належать до структурних середніх величин.
6. Якими властивості характерні середній арифметичній?
7. Дайте характеристику моди ряду.
8. Дайте характеристику медіані ряду.
9. Назвіть відмінність застосування середньої гармонічною від середньої арифметичної.
10. Як знайти медіану в дискретному та інтервальному ряді?

ЛЕКЦІЯ 4. ПОКАЗНИКИ ВАРІАЦІЇ

4.1 *Поняття варіації та її основні показники.*

4.2 *Математичні властивості дисперсії та спрощені способи їх розрахунку.*

4.3 *Види дисперсій. Правило додавання дисперсій.*

4.1. Поняття варіації та її основні показники

Середні величини, як узагальнюючі показники, характеризують сукупності за варіюючою 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 навпаки, істотно відхиляються від середньої. В одних сукупностях індивідуальні значення ознаки щільно групуються навколо центру розподілу, в інших – значно відхиляються. Чим менші відхилення, тим однорідніша сукупність, а отже, тим більш надійні і типові середні характеристики розподілу.

Дослідження характеру та величини мінливості статистичних сукупностей дозволяє оцінити стійкість (однорідність або неоднорідність) статистичних сукупностей та інтенсивність структурних зрушень, виявити вплив різних чинників на коливання ознак та щільність взаємозв'язків соціально-економічних явищ, розробити конкретні заходи для підсилення позитивних коливань і зменшення негативних наслідків високої мінливості окремих чинників.

Колівання окремих значень ознаки характеризують *показники варіації*. Термін “варіація” походить від латинського *variatio* – зміна, коливання, відмінність. *Варіацією* називають кількісні зміни величини досліджуваної ознаки в межах однорідної сукупності, які зумовлені впливом дії різних факторів.

Неважко уявити дві сукупності, в яких середні величини варіюючої ознаки однакові, проте наближення індивідуальних значень у кожній сукупності до середньої є різним. Ось чому для характеристики сукупності велике практичне значення має вивчення відхилень досліджуваної

ознаки окремих одиниць сукупності від середньої величини. Треба брати до уваги не лише крайні відхилення, а й сукупність відхилень усіх варіантів. Від розміру і розподілу цих відхилень залежить типовість і надійність середніх величин.

Для вимірювання та оцінки варіації використовуються абсолютні й відносні характеристики. До абсолютних належать: розмах варіації, середнє лінійне та середнє квадратичне відхилення, дисперсії; відносні характеристики представлені низкою коефіцієнтів: варіації, локалізації, концентрації. Всі перелічені показники є іменованими величинами, крім коефіцієнтів, які обчислюються у відсотках.

Для характеристики розмірів коливань ознаки по відношенню до середньої величини обчислюють показники варіації.

У системі показників варіації найпростішим є показник *розмаху варіації* (амплітуда коливань). *Розмах варіації* характеризує межі, в яких зменшується значення ознаки, і обчислюється як різниця між максимальним і мінімальним значенням ознаки:

$$R = \chi_{\max} - \chi_{\min}.$$

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

Безумовною перевагою розмаху варіації, як міри коливання ознаки, є простота його обчислення, але надійність такої простої характеристики невисока, оскільки вона базується на двох крайніх значеннях ознаки, які часто не є типовими для сукупності, або мають випадковий характер. Тому розмах варіації використовують для попередньої оцінки варіації.

У статичних методах широко використовують характеристики варіації, що ґрунтуються на відхиленнях індивідуальних значень ознаки від середньої величини $\chi_i - \bar{\chi}$. Оскільки $\sum(\chi_i - \bar{\chi}) = 0$, то при розрахунку такого роду характеристик використовують або модулі, або квадрати відхилень. У результаті маємо такі характеристики варіації: середнє лінійне \bar{d} і середнє квадратичне відхилення σ в та дисперсію σ^2 . Методика обчислення цих показників залежить від характеру вихідних даних

(табл. 4.1).

Середнє квадратичне відхилення є мірилом надійності середньої. Чим менше середнє квадратичне відхилення, тим об'єктивніше середня арифметична відображує всю сукупність.

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

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

$$Q = \frac{Q_3 - Q_1}{3},$$

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

В нормально розподіленій сукупності існує співвідношення між квадратичним відхиленням, середнім лінійним відхиленням, квартильним відхиленням та розмахом варіації:

$$6\sigma \approx 7,5\bar{d} \approx 9Q \approx R.$$

В явищах, поряд з варіацією кількісних ознак має місце варіація якісних ознак. При цьому, якщо є два взаємно виключаючи варіанти, то таку варіацію називають *альтернативною*. При альтернативній мінливості одні одиниці сукупності володіють даною ознакою, а інші не володіють.

Таблиця 4.1 – Обчислення узагальнюючих показників варіації

Назва показників варіації	Формули показників варіації	
	Для незгрупованих даних	Для згрупованих даних
Середнє лінійне відхилення	$\bar{d} = \frac{\sum \chi_i - \bar{\chi} }{n}$	$\bar{d} = \frac{\sum \chi_i - \bar{\chi} \cdot f}{\sum f}$
Середній квадрат відхилень (дисперсія)	$\sigma^2 = \frac{\sum \chi_i - \bar{\chi} ^2}{n}$	$\sigma^2 = \frac{\sum \chi_i - \bar{\chi} ^2 \cdot f}{\sum f}$
Середнє квадратичне відхилення	$\sigma = \sqrt{\frac{\sum \chi_i - \bar{\chi} ^2}{n}}$	$\sigma = \sqrt{\frac{\sum \chi_i - \bar{\chi} ^2 \cdot f}{\sum f}}$

Усі розглянуті показники варіації – розмах варіації, середнє лінійне відхилення, середній квадрат відхилень та середнє квадратичне відхилення – завжди виражають у одиницях вихідних даних ряду та середньої величини. Всі вони є абсолютним виміром варіації. А це означає, що безпосередньо порівнювати абсолютні показники варіації у варіаційних рядах різних явищ не можна. Для того, щоб забезпечити їх порівняння потрібно обчислити показники, які характеризують варіацію, виражену в стандартних величинах, наприклад, у відсотках.

Відношення абсолютних характеристик варіації до середньої величини називаються *коефіцієнтами варіації*. Коефіцієнти варіації розраховують за формулами:

– лінійний: $V_{\bar{d}} = \frac{\bar{d} \cdot 100\%}{\bar{\chi}}$;

– квадратичний: $V_{\sigma} = \frac{\sigma \cdot 100\%}{\bar{\chi}}$;

– осциляції: $V_R = \frac{R \cdot 100\%}{\bar{\chi}}$.

Коефіцієнти варіації дозволяють порівнювати варіації різних ознак або варіацію однієї ознаки у різних сукупностях.

Для порівняння варіацій найчастіше використовують квадратичний коефіцієнт варіації. Цей показник вживається для оцінки однорідності сукупності, тобто надійності і типовості середньої величини. Роз-

різняють такі значення відносних коливань:

- $V_\sigma < 10\%$ – незначне коливання;
- $V_\sigma =$ від 10% до 30% – середнє коливання;
- $V_\sigma > 30\%$ – велике коливання.

Вважають, що сукупність є однорідною, а середня груповою, коли V_σ не перевищує 33% .

Лінійний коефіцієнт варіації або відносне лінійне відхилення характеризує частку середнього значення абсолютних відхилень від середньої величини.

Квадратичний коефіцієнт варіації характеризую зміну індивідуальних значень відносно середньої величини по сукупності.

Коефіцієнт осциляції відображує відносне коливання крайніх значень ознаки навколо середньої.

Розрахувавши коефіцієнти варіації можна зробити наступні висновки:

1. Середня величина (середня арифметична) має важливе пізнавальне значення, однак вона не завжди об'єктивно і не завжди з однаковим ступенем достовірності відображує внутрішній стан статистичної сукупності. При однаковому значенні середньої статистичні сукупності можуть бути досить нерівноцінні за рівнем коливання (варіації).

2. Чим менше відхилення, тим типовіша середня, тим більш однорідна сукупність.

3. Універсальним показником варіації є коефіцієнт варіації, цінність якого полягає в тому, що ним можна користуватись для характеристики і порівняння варіації різних сукупностей і різних явищ.

4.2 Математичні властивості дисперсії

Дисперсія (середній квадрат відхилень) має певні математичні властивості, урахування яких дає змогу суттєво спростити її обчислення.

1. Якщо всі значення варіант зменшити на будь-яке сталє число A , то середній квадрат відхилень від цього не зміниться:

$$\sigma_{\chi-A}^2 = \sigma_{\chi}^2.$$

2. Якщо всі значення варіант поділити на будь-яке стале число i , то дисперсія зменшиться внаслідок цього в i^2 разів, а середнє квадратичне відхилення – в i разів:

$$\sigma_{\frac{\chi}{i}}^2 = \sigma_{\chi}^2 : i^2.$$

3. Якщо обчислити квадрат відхилень від будь-якої величини A , що тією чи іншою мірою відмінна від середньої арифметичної $\bar{\chi}$, то він завжди буде більшим за середній квадрат відхилень (дисперсію) σ^2 , обчислений від середньої арифметичної $\sigma_A^2 > \sigma_{\chi}^2$, причому більший на певне значення – квадрат різниці між середньою і цією величиною, тобто на $(\bar{\chi} - A)^2$:

$$\sigma^2 = \sigma_A^2 + (\bar{\chi} - A)^2 \text{ або } \sigma^2 = \sigma_A^2 - (\bar{\chi} - A)^2.$$

Дисперсія від середньої величини має властивість мінімальності, тобто вона завжди менша від дисперсії, обчисленої від будь-яких інших величин.

У такому випадку, коли величину A прирівняти до нуля, то:

$$\sigma^2 = \frac{\sum \chi^2 \cdot f}{\sum f} - \left(\frac{\sum \chi \cdot f}{\sum f} \right)^2 \text{ або } \sigma^2 = \overline{\chi^2} - (\bar{\chi})^2.$$

Отже, дисперсія ознаки, або середній квадрат відхилень σ_{χ}^2 дорівнює різниці між середнім квадратом значень ознаки $\overline{\chi^2}$ і квадратом середнього значення ознаки $(\bar{\chi})^2$. Таким чином, не обчислюючи відхилень можна обчислити дисперсію.

На основі наведених математичних властивостей дисперсії базується спрощений спосіб розрахунку дисперсії та середнього квадратичного відхилення, який називається *способом моментів*, або *способ відліку від умовного нуля*. Він застосовується при умові рівних інтервалів.

Використання цього методу є доцільним при великих вихідних значеннях інтервального ряду розподілу:

$$\sigma_x^2 = \overline{x^2} - (\overline{x})^2 = \frac{\sum \left(\frac{x-A}{i} \right) \cdot f}{\sum f} \cdot i^2 + A^2 - \left(\frac{\sum \left(\frac{x-A}{i} \right) \cdot f}{\sum f} \cdot i + A \right)^2.$$

Після відповідних перетворень дістанемо:

$$\sigma_x^2 = i^2 \cdot \left[\frac{\sum \left(\frac{x-A}{i} \right)^2 \cdot f}{\sum f} - \left(\frac{\sum \left(\frac{x-A}{i} \right) \cdot f}{\sum f} \right)^2 \right],$$

де $\frac{\sum \left(\frac{x-A}{i} \right) \cdot f}{\sum f} = m_1$ – момент першого порядку; $\frac{\sum \left(\frac{x-A}{i} \right)^2 \cdot f}{\sum f} = m_2$ – момент другого порядку.

Отже, дисперсія, обчислена за способом моментів, дорівнює добутку квадрата інтервалу на різницю моменту другого порядку і квадрата моменту першого порядку:

$$\sigma_x^2 = i^2 \cdot (m_2 - m_1^2).$$

4.3 Види дисперсій. Правило додавання дисперсій

Варіація ознаки формується під впливом різних факторів. При вивченні дисперсії досліджуваної ознаки в межах даної сукупності можна визначити три показники коливань ознаки: загальну дисперсію, міжгрупову дисперсію і середню з групових дисперсій.

Загальна дисперсія (σ^2) характеризує загальну варіацію ознаки під впливом усіх умов і причин, що зумовили варіацію.

Групова дисперсія дорівнює середньому квадрату відхилень окремих значень ознаки всередині групи від середньоарифметичної відповідної групи:

$$\sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2}{n} \text{ – проста,}$$

$$\sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2 \cdot f}{\sum f} \text{ – зважена.}$$

Групову дисперсію також можна розрахувати за спрощеною формулою:

$$\sigma_i^2 = \overline{x^2} - (\overline{x})^2.$$

Ця дисперсія відображає варіацію лише за рахунок умов і причин, що діють в середині групи.

Середня з групових дисперсій ($\overline{\sigma^2}$) – це середня арифметична зв'язана з групових дисперсій, розраховується за формулою:

$$\overline{\sigma^2} = \frac{\sum \sigma_i^2 \cdot f}{\sum f}.$$

Міжгрупова дисперсія (δ^2) – це середній квадрат відхилень групових середніх \overline{x}_i від загальної середньої (\overline{x}). Характеризує систематичну варіацію, тобто відмінності у величині досліджуваного (результативного) ознаки за рахунок ознаки-фактора, покладеного в основу групування:

$$\delta^2 = \frac{\sum (\overline{x}_i - \overline{x})^2 \cdot f_i}{\sum f_i}.$$

Між зазначеними видами дисперсій існує співвідношення: загальна дисперсія дорівнює сумі середньої з групових дисперсій та міжгрупової дисперсії:

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

Це співвідношення називають *правилом додавання дисперсій*.

Контрольні запитання

1. Дайте визначення варіації. Наведіть формулу для розрахунку розмаху варіації.
2. Вкажіть як розрахувати середній квадрат відхилень (дисперсію).
3. Вкажіть як розрахувати середнє квадратичне відхилення.
- 4 В чому полягає залежність між величиною середнього квадратичного відхилення і кількістю спостережень в умовах нормального ро-

зподілу?

5. Наведіть формули для розрахунку коефіцієнтів варіації.
6. Наведіть математичні властивості дисперсії.
7. Що характеризує загальна дисперсія?
8. Як розраховується групова дисперсія?
9. Наведіть формулу для розрахунку середньої з групових дисперсій.
10. Наведіть формулу для розрахунку міжгрупової дисперсії.

ЛЕКЦІЯ 5. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

5.1 Сутність та різновиди вибіркового спостереження.

5.2 Способи відбору одиниць у вибірку сукупності.

5.3 Визначення середньої, граничної помилки та необхідної кількості вибірки.

5.1 Сутність та різновиди вибіркового спостереження

Вибіркове спостереження – це вид несучільного спостереження, при якому обстежуються не всі елементи сукупності, а лише певним чином відібрана її частина.

Сукупність, з якої вибирають елементи для обстеження, називається *генеральною сукупністю*, а сукупність, яку безпосередньо обстежують, – *вибірковою сукупністю*.

Кінцева мета вибіркового спостереження – поширення його статистичних характеристик на генеральну сукупність. Іншими словами вибірка сукупність представляє – *репрезентує* генеральну сукупність.

Вибірковий метод дозволяє через вивчення частини спеціально відібраних одиниць охарактеризувати масове явище в цілому.

Вибіркове спостереження забезпечує, у порівнянні з суцільним спостереженням, економію трудових, фінансових ресурсів та часу. При обстеженні частини генеральної сукупності зменшуються помилки реєстрації, але можлива помилка репрезентативності. При вивченні деяких соціально-економічних явищ єдине можливе спостереження – це вибіркоче. Наприклад, перевірка якості, яка пов'язана із руйнуванням продукції або наукове вивчення біологічних ресурсів.

Якщо генеральна сукупність містить N елементів, а для обстеження треба вибрати n елементів, то число можливих вибірок можна обчислити:

$$C_N^n = \frac{N!}{n!(N-n)!}.$$

Усі вони мають однакову імовірність $\frac{1}{C_N^n}$, а кожна з них несе певну похибку, що відображає факт випадкового вибору.

За причинами виникнення похибки поділяються на:

– *тенденційні (систематичні) помилки*, що виникають внаслідок порушення принципу випадковості: упереджений вибір елементів, недосконала основа вибірки тощо, які небезпечні внаслідок систематичного викривлення результатів.

– *випадкові помилки*, що виникають унаслідок випадковості вибору елементів для дослідження, які мають назву *помилки репрезентативності* і їх принципово неможливо уникнути, але завжди є можливість знайти ймовірність того, що випадкова помилка вибірки не перевищить заданого (припустимого) рівня, або знайти обсяг вибірки, який забезпечить потрібну точність результатів для взятої ймовірності.

5.2 Способи відбору одиниць у вибіркочну сукупність

Простий випадковий відбір – проводиться жеребкуванням чи за допомогою таблиць випадкових чисел. Для жеребкування на кожен одиницю генеральної сукупності необхідно заготувати відповідну фішку. Для використання таблиць випадкових чисел усі елементи генеральної сукупності мають бути пронумеровані. Якщо сукупність велика чи розповсюджена на території та у часі, так використовувати простий випадковий метод вибірки важко.

Систематичний (механічний) відбір використовують, якщо чисельність елементів сукупності впорядкована. Добір інтервалів здійснюється через рівні інтервали. Механічна вибірка порівняно із простою випадковою ефективніша, її простіше здійснити. Однак при наявності циклічних коливань значень ознаки, цикл яких збігається з інтервалом добірки, можливий зсув вибіркових оцінок.

Іноді, вивчаючи безпосередні в часі процеси, проводять моменти спостереження – фіксація стану процесу на певні моменти часу, які вибирають за схемою випадкової або механічної вибірки.

Розшиарований відбір – це спосіб формування вибірки з урахуванням структури генеральної сукупності. Незалежний добір здійснюється у кожній складовій частині генеральної сукупності, яка попередньо структурується.

Серійна вибірка ґрунтується на розподілі генеральної сукупності на частини – серії. Серії відбираються за схемою простої випадкової, або механічної вибірки. Вибрані серії підлягають суцільному обстеженню, а залишок – серій не обстежується зовсім.

Поєднання різних видів вибірки можливе у рамках *багатоступеневої вибірки*.

Вибірка елементів для вибіркового спостереження може здійснюватися способом *повторного і без повторного відбору*.

При повторному методі кожна вибрана одиниця сукупності повертається після обстеження до сукупності і може бути вибрана знову. При безповторному методі кожна обстежена одиниця не повертається до сукупності і принципово не може бути обстежена повторно.

5.3 Визначення середньої, граничної помилки та необхідної чисельності вибірки

Вибіркова сукупність має пізнавальне значення, оскільки дає уявлення (з певною імовірністю) про показники генеральної сукупності. Але при цьому виникають помилки спостереження.

У разі несучільного спостереження, зокрема вибіркового окрім помилок реєстрації, виникають помилки вибірки, або репрезентативності (відповідності).

Помилки реєстрації – розходження між записаними даними і реальними.

Помилки репрезентативності – це розходження між середніми величинами або частками ознаки вибіркової і генеральної сукупності. Бувають: *систематичні і випадкові*.

Достовірність вибіркового спостереження забезпечується розрахунками його помилок для середньої величини і для частки (питомої ваги) ознаки, що вивчається.

Обумовимо основні позначення статистичних характеристик, які використовуються при визначенні помилок вибіркового спостереження.

Умовні позначення статистичних характеристик генеральної та вибіркової сукупності показані в табл. 5.1.

Таблиця 5.1 – Умовні позначення статистичних характеристик генеральної та вибіркової сукупності

Характеристика	Сукупність	
	генеральна	вибіркова
Обсяг сукупності	N	n
Середнє значення ознаки	\bar{X}	\bar{x}
Загальна дисперсія	σ_r^2	σ^2
Середня з групових дисперсій	$\overline{\sigma_r^2}$	$\overline{\sigma^2}$
Міжгрупова дисперсія	$\overline{\delta_r^2}$	$\overline{\delta^2}$
Частка елементів сукупності, які мають певні значення ознаки	W	w
Частка вибіркової сукупності в генеральній	x	D
Кількість серій	R	r
Дисперсія альтернативної ознаки	$\sigma_r^2 = p \cdot g$	$\sigma^2 = w \cdot (1 - w)$

Для альтернативної ознаки:

p – частка вибіркової сукупності, якій властива ознака; $g = 1 - p$ – частка вибіркової сукупності, якій не властива ознака;

$$\sigma^2 = p \cdot g = p \cdot (1 - p)$$

Помилки вибірки (позначаються Δ):

- $\Delta x = \bar{x} - \bar{X}$ – помилка вибірки для середньої величини,
- $\Delta w = \bar{w} - \bar{W}$ – помилка вибірки для частки.

Величини помилок вибірки (репрезентативності) в основному залежать:

- від обсягу вибірки,

- від варіації досліджуваної ознаки,
- від способу і виду відбору вибіркової сукупності.

Для узагальнюючої характеристики помилок репрезентативності розраховують середню помилку вибірки μ . Способи розрахунку середньої помилки вибірки μ при різних способах відбору сукупності наведені в табл. 5.2.

Таблиця 5.2 – Середня помилка вибірки

Спосіб відбору сукупності	Метод відбору	
	Повторний	Безповторний
Помилка вибірки для середньої величини		
Випадковий і механічний	$\mu = \sqrt{\frac{\sigma^2}{n}}$	$\mu = \sqrt{\frac{\sigma^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Типовий	$\mu = \sqrt{\frac{\sigma_i^2}{n}}$	$\mu = \sqrt{\frac{\sigma_i^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Серійний	$\mu = \sqrt{\frac{\delta_x^2}{r}}$	$\mu = \sqrt{\frac{\delta_x^2}{r} \cdot \left(1 - \frac{r}{R}\right)}$
Помилка вибірки для частки		
Випадковий і механічний	$\mu = \sqrt{\frac{w \cdot (1-w)}{n}}$	$\mu = \sqrt{\frac{w \cdot (1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Типовий	$\mu = \sqrt{\frac{w \cdot (1-w)}{n}}$	$\mu = \sqrt{\frac{w \cdot (1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Серійний	$\mu = \sqrt{\frac{\delta_w^2}{r}}$	$\mu = \sqrt{\frac{\delta_w^2}{r} \cdot \left(1 - \frac{r}{R}\right)}$

Особливість обчислення помилок репрезентативності для середньої величини при різних способах відбору полягає в тому, що для її обчислення в основу беруться різні показники дисперсій.

При випадковому і механічному відборі – використовується загальна дисперсія σ^2 , для серійної і для частки – $w \cdot (1-w)$.

Для серійної вибірки похибка залежить від кількості вибраних серій.

У статистичному аналізі часто постає потреба порівняти помилки вибірки різних ознак або однієї і тієї ознаки в різних сукупностях. Такі порівняння виконують за допомогою *відносної помилки*, яка показує на скільки відсотків вибіркова оцінка може відхилитись від параметра генеральної сукупності.

Відносна стандартна помилка середньої – це коефіцієнт варіації вибірових середніх:

$$V_{\mu} = \frac{\mu_{\chi}}{\chi} \cdot 100.$$

Для узагальнюючої характеристики розраховують і *граничну помилку вибірки*.

В випадку вибіркового спостереження гранична помилка репрезентативності Δ може бути більшою, чи дорівнювати, або меншою від середньої помилки репрезентативності μ .

З основних положень теорії вибіркового методу впливає формула граничної помилки вибірки. Метод сформульований у теоремах ймовірностей, які відображають *закон великих чисел*.

Згідно з дослідженнями академіка А.М. Ляпунова ймовірність відхилень вибіркової середньої від генеральної середньої при достатньо великому обсягу вибірки та обмеженій дисперсії генеральної сукупності підпорядковується *закону нормального розподілу*. Ймовірність цих відхилень при різних значеннях t визначається за формулою:

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

Значення отриманого інтегралу при різних значеннях t табульовані, наводяться як правило в спеціальних таблицях.

Граничну помилку репрезентативності обчислюють з певною ймовірністю P , якій відповідає t -разове значення μ :

$$\Delta = t \cdot \mu,$$

де Δ – гранична помилка вибірки; μ – середня помилка вибірки; t – коефіцієнт довіри, який залежить від ймовірності, з якою гарантується значення граничної помилки вибірки.

Гранична помилка вибірки розраховується за вибірковим способом по-різному, залежно від видів і способів відбору. Якщо в формули для розрахунку граничної помилки вибірки підставити формули для розрахунків середньої помилка вибірки μ (дивись табл. 5.2) то отримаємо інші формули для розрахунку граничної помилки для повторного і безповторного методів відбору різних способів, які наведено в табл. 5.3.

Таблиця 5.3 – Розрахунок граничної помилки вибірки

Спосіб відбору одиниць сукупності	Метод відбору	
	Повторний	Безповторний
Помилка вибірки для середньої величини		
Випадковий і механічний	$\Delta_{\chi} = t \cdot \sqrt{\frac{\sigma^2}{n}}$	$\Delta_{\chi} = t \cdot \sqrt{\frac{\sigma^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Типовий	$\Delta_{\chi} = t \cdot \sqrt{\frac{\sigma_i^2}{n}}$	$\Delta_{\chi} = t \cdot \sqrt{\frac{\sigma_i^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Серійний	$\Delta_{\chi} = t \cdot \sqrt{\frac{\delta^2}{r}}$	$\Delta_{\chi} = t \cdot \sqrt{\frac{\delta^2}{r} \cdot \left(1 - \frac{r}{R}\right)}$
Помилка вибірки для частки		
Випадковий і механічний	$\Delta_p = t \cdot \sqrt{\frac{w \cdot (1-w)}{n}}$	$\Delta_p = t \cdot \sqrt{\frac{w \cdot (1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Типовий	$\Delta_p = t \cdot \sqrt{\frac{w \cdot (1-w)}{n}}$	$\Delta_p = t \cdot \sqrt{\frac{w \cdot (1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)}$
Серійний	$\Delta_p = t \cdot \sqrt{\frac{\delta_w^2}{r}}$	$\Delta_p = t \cdot \sqrt{\frac{\delta_w^2}{r} \cdot \left(1 - \frac{r}{R}\right)}$

Гранична помилка вибірки дає можливість встановити в яких межах лежать значення генеральної середньої або частки.

За допомогою формули граничної помилки вибірки визначають:

- 1) довірчі межі генеральної середньої і частки з певною імовірністю;
- 2) імовірність того, що відхилення між вибілковими та генеральними характеристиками не перевищують визначену величину;

3) необхідну чисельність вибірки, яка із заданою ймовірністю забезпечує очікувану точність вибіркових показників.

Під час вибіркового спостереження важливо правильно визначити чисельність вибірки, яка з відповідною ймовірністю забезпечує встановлену точність результатів спостереження.

Формули для визначення необхідної чисельності вибірки при повторному і без повторному відборах наведені в табл. 5.4.

Таким чином, чисельність вибірки залежить від:

- 1) розміру граничної помилки;
- 2) від показників варіації ознаки та частки;
- 3) від ймовірності, з якою вимагається гарантувати результати вибірки.

Таблиця 5.4 – Формули для розрахунку чисельності вибірки

Спосіб відбору	Визначення середньої	Визначення частки
Повторний	$n = \frac{t^2 \cdot \sigma^2}{\Delta_x^2}$	$n = \frac{t^2 \cdot w \cdot (1-w)}{\Delta_w^2}$
Безповторний	$n = \frac{t^2 \cdot \sigma^2 \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \sigma^2}$	$n = \frac{t^2 \cdot w \cdot (1-w) \cdot N}{\Delta_w^2 \cdot N + t^2 \cdot w \cdot (1-w)}$

Необхідну чисельність вибірки в разі серійного відбору визначають як відбір певної кількості серій, які забезпечують з відповідною ймовірністю потрібну точність результатів дослідження.

Для повторного відбору – необхідна чисельність вибірки:

$$r = \frac{t^2 \cdot \delta^2}{\Delta^2},$$

а для безповторного

$$r = \frac{t^2 \cdot \delta^2 \cdot R}{\Delta^2 \cdot R + t^2 \cdot \delta^2}.$$

Вибіркове спостереження з великих масивів генеральної сукупності можна здійснювати у вигляді комбінованої, ступінчатої або кілька фазної вибірки.

Загальна помилка для комбінованої вибірки складається з помилок, які можливі на кожному ступені, і визначається як корінь квадратний з квадратів помилок відповідних вибірок.

Приклад. Якщо серійну вибірку скомпонувати з випадковою або механічною то гранична помилка вибірки буде визначатись:

$$\Delta_z = t \cdot \sqrt{\frac{\delta^2}{r} \cdot \left(1 - \frac{r}{R}\right) + \frac{\sigma^2}{n} \cdot \left(1 - \frac{n}{N}\right)}.$$

Кінцева мета вибіркового спостереження – це поширення його характеристик на генеральну сукупність. Це забезпечується різними способами: або способом прямого перерахунку або спосіб поправочних коефіцієнтів.

Контрольні запитання

1. Назвіть переваги та недоліки вибіркового методу.
2. Вкажіть як за причинами виникнення поділяються похибки при вибіркового методі.
3. В чому полягає простий випадковий відбір одиниць сукупності?
4. В чому полягає систематичний (механічний) відбір одиниць сукупності?
5. В чому полягає розшарований відбір одиниць сукупності та серійна вибірка?
6. Назвіть від чого залежить величина помилки вибірки (репрезентативності).
7. Наведіть формули для розрахунку помилок вибірки для середніх величин, для вибірок відібраних повторним і безповторним метод відбору.
8. Як розраховується відносна стандартна помилка середньої?
9. Наведіть формулу для розрахунку граничної помилки репрезентативності.
10. Вкажіть від яких факторів залежить чисельність вибірки при вибіркового методі.

ЛЕКЦІЯ 6. СТАТИСТИЧНІ МЕТОДИ ВИМІРЮВАННЯ ВЗАЄМОЗВ'ЯЗКІВ

6.1 Види взаємозв'язків між явищами.

6.2 Метод аналітичного групування.

6.3 Дисперсійний аналіз.

6.1 Види взаємозв'язків між явищами

Одним із найзагальніших законів об'єктивного світу є закон загального зв'язку і залежності між явищами суспільного життя. Усі явища суспільного життя існують не ізольовано, а у нерозривному взаємозв'язку, тобто залежать одне від одного, тому вивчення будь-якого явища буде неповним, якщо не досліджені його зв'язки з іншими явищами і процесами. Статистичне дослідження взаємозв'язків дає можливість виявити не тільки наявність і напрямок зв'язку, але дозволяє кількісно оцінити і виразити його аналітично.

Визначення зв'язків між явищами дає змогу перейти від констатації фактів до пояснення і використання їх на практиці.

Так, при вивченні травматизму на машинобудівних підприємствах можна визначити кількісні характеристики впливу багатьох факторів на травматизм. Це дозволяє в майбутньому розробити заходи спрямовані на зниження рівня травматизму на виробництві. Визначення взаємозв'язків дозволяє проводити науково-обґрунтовані прогнози.

Зв'язки між явищами, окремими їх ознаками досить різноманітні, однак у будь-якому випадку одні ознаки виступають як фактори, що впливають на інші зумовлюють їх зміну, інші – як результати дії цих факторів. Одні із них є причиною, інші наслідком.

Якщо перші прийнято називати *ознаками-факторами*, або факторними (*причинними*) *ознаками*, то другі – *результативними* (*наслідковими*) *ознаками*.

Різноманітність зв'язків, в яких перебувають явища, зумовлює необхідність їх класифікації, зведення зв'язків до певних типів, форм за їх істотними рисами і властивостями.

В основу класифікації зв'язків у статистиці покладено відмінність і подібність зв'язків за такими їх особливостями, як ступінь тісноти, спрямованість, аналітичне вираження, одиничність або множинність факторів. Відповідно до цього розрізняють зв'язки:

- *функціональні і кореляційні;*
- *прямі і обернені;*
- *прямолінійні і криволінійні;*
- *однофакторні і багатфакторні.*

За статистичною природою зв'язки поділяють на *функціональні і стохастичні*. При *функціональному зв'язку* кожному можливому значенню факторної ознаки x відповідає чітко визначене значення результативної ознаки – y , тобто функціональні зв'язки характеризуються повною відповідністю між причиною і наслідком, факторною і результативною ознаками. Така залежність притаманна фізичним, хімічним явищам тощо.

У суспільних процесах це найчастіше зв'язок складових елементів розрахункових формул відповідних показників, наприклад, залежність кількості нещасних випадків від трудового стажу працівників, віку та статі працівника.

На відміну від функціональних, *стохастичні* зв'язки неоднозначні. Стохастичні зв'язки проявляються як узгодженість варіації двох чи більше ознак. У ланці зв'язку “ x – y ” кожному значенню ознаки x відповідає певна множина значень ознаки y , які утворюють так званий умовний розподіл. Стохастичний зв'язок, відбиваючи множинність причині наслідків, виявляється в зміні умовних розподілів.

Якщо умовні розподіли замінюються одним параметром середньою \bar{y}_i , то такий зв'язок називають кореляційним. Отже, кореляційний зв'язок є різновидом стохастичного і виявляється в зміні середніх умовних розподілів, що схематично ілюструє таблиця 6.1.

За напрямком дії (спрямованістю) розрізняють зв'язок *прямий і обернений*. *Прямий* – це такий зв'язок, при якому зі збільшенням, або зменшенням значень факторної ознаки відповідно збільшується або зменшується значення результативної ознаки, тобто факторна і результативна ознаки змінюються в одному напрямку.

Таблиця 6.1 – Види взаємозв'язків між явищами

Факторна ознака, X_i	Результативна ознака у при наявності зв'язку:		
	функціонального	стохастичного	кореляційного
X_1	Y_1	Y_1, Y_2	\bar{Y}_1
X_2	Y_2	Y_1, Y_2, Y_3	\bar{Y}_2
X_3	Y_3	Y_2, Y_3, Y_4	\bar{Y}_3
...
X_n	Y_n	Y_{n-1}, Y_n	\bar{Y}_n

Оберненим зв'язком називають такий, при якому значення результативної ознаки змінюється в протилежному напрямку відносно зміни значення факторної ознаки. Прикладом такого зв'язку може бути продуктивність праці і собівартість продукції.

Слід мати на увазі, що прямий і обернений зв'язки можуть переходити один в інший. Так, прямий зв'язок після досягнення факторною ознакою певного рівня в деяких випадках себе вичерпує і потім робиться оберненим, тобто з наступним збільшенням значення факторної ознаки значення результативної ознаки зменшується.

За формою аналітичного вираження в загальній класифікації виділяють зв'язки *прямолінійні та криволінійні*. Якщо певний зв'язок явищ можна точно або наближено зобразити рівнянням будь-якої прямої лінії, то його називають *лінійним (прямолінійним) зв'язком*, а якщо рівнянням будь-якої кривої лінії (параболи, гіперболи і т.п.) *нелінійним (криволінійним)*.

Аналітичним рівнянням можна описувати лише функціональні зв'язки. Кореляційні зв'язки описуються рівнянням лише наближено. Однак, навіть наближене описування за допомогою аналітичних рівнянь кореляційних зв'язків суспільних явищ дає можливість отримати результати, цілком придатні для наукових і практичних потреб.

Для відповіді на питання про наявність або відсутність кореляційного зв'язку використовують ряд специфічних методів:

1) елементарні прийоми:

– паралельне порівняння рядів значень факторної і результатив-

ної ознак;

- балансовий метод;
 - графічне зображення
 - метод аналітичного групування.
- 2) дисперсійний аналіз.
 - 3) кореляційно-регресивний аналіз.

6.2 Метод аналітичного групування

Елементарним способом виявлення зв'язку є *паралельне порівняння* ряду значень факторної ознаки і відповідних значень результативної ознаки. Значення факторної ознаки розташовують у порядку зростання і потім відслідковують напрям зміни величини результативної ознаки. В тих випадках, коли збільшення величини факторної ознаки призводить до збільшення результативної ознаки, можна стверджувати про наявність прямого кореляційного зв'язку. Якщо ж із збільшенням факторної ознаки, величина результативної ознаки має тенденцію до зменшення, то можна передбачати обернений зв'язок між ознаками.

Однак, наявність великої кількості різних значень результативної ознаки, що відповідають одному і тому ж значенню факторної ознаки, ускладнює сприйняття таких паралельних рядів особливо при великій кількості одиниць досліджуваної сукупності. В таких випадках для встановлення факту наявності зв'язку доцільно скористатися *графічним* або *табличними методами* наочного зображення і аналізу статистичних даних.

Найбільш поширеним і простим методом дослідження взаємозв'язків у статистиці є *метод аналітичного групування і побудова кореляційних таблиць*.

Метод аналітичного групування полягає в тому, що всі елементи сукупності групують, як правило, за факторною ознакою і в кожній групі обчислюють середні значення результативної ознаки у, тобто лінія регресії оцінюється лише в окремих точках, які відповідають певному значенню x .

Основні етапи методу аналітичного групування

На першому етапі аналізу кореляційного зв'язку при обґрунтуванні моделі постають два питання: вибір факторних ознак і визначення числа груп і меж інтервалів. При визначенні числа груп і меж інтервалів слід зважати на той факт, що типовість та сталість групових середніх залежить від чисельності груп. На практиці аналітичне групування часто виконується за принципом рівних інтервалів, що значно спрощує подальший аналіз зв'язку.

Наприклад, розглянемо дані про обіг коштів в банках та прибуток в них (див. табл. 6.2).

Таблиця 6.2 – Показники діяльності банків у звітному періоді

№ банку	Обсяг коштів, млн. грн.	Прибуток, млн. грн.
1	7	3
2	11	4
3	12	4
4	16	6
5	22	8
6	27	7
7	30	11
8	38	14
9	45	18
10	46	17

Проведемо комбіноване групування банків за двома ознаками: за обігом коштів і розміром одержаного прибутку, утворивши по три групи з рівними інтервалами (табл. 6.3).

Таблицю, що побудовано таким чином, називають *кореляційної таблицею*. Якщо частоти у кореляційній таблиці розташовані по діагоналі з лівого верхнього кута у правий нижній кут, тобто більшим значенням фактора відповідають більші значення функції, то передбачається наявність прямого кореляційного зв'язку між ознаками. Якщо ж частоти розташовані по діагоналі з правого кута у лівий, то передбачають

наявність оберненого зв'язку між ознаками.

Таблиця 6.3 – Розподіл банків за обігом коштів та розміром прибутку

Грипу банків за обсягом коштів	Кількість банків за розміром прибутків, млн грн. (частоти)			Разом
	3-8	8-13	13-18	
7-20	4	–	–	4
20-33	2	1	–	3
33-46	–	–	3	3
Разом	6	1	3	10

У наведеному прикладі кожен рядок побудованої кореляційної таблиці (крім підсумку) містить частоти розподілу банків за рівнем прибутку при фіксованому обігу коштів, тобто частоти умовного розподілу. Підсумковий рядок містить частоти безумовного розподілу. У даному випадку зв'язок між ознаками стохастичний, оскільки кожному значенню (інтервалу значень) ознаки x відповідає декілька значень ознаки y . При відсутності стохастичного зв'язку між ознаками умовні розподіли були б однакові і збігалися б з безумовним розподілом. Таблиця показує, що частоти розподілу банків концентруються по Діагоналі, яка йде з верхнього лівого кута у правий нижній. Це свідчить про те, що існує прямий зв'язок між обігом коштів і розміром прибутку банків. Із зростанням розмірів обігу коштів збільшується прибуток банків.

На другому етапі проводиться оцінка лінії регресії. У кожній групі за факторною ознакою обчислюють середні значення результативної та факторної ознак. Групові середні обчислюємо за вихідними незгрупованими даними. У нашому прикладі:

Середній прибуток, млн. грн. Середній обіг коштів, млн. грн.

$$\bar{y}_1 = \frac{3+4+4+6}{4} = 4,3;$$

$$\bar{x}_1 = \frac{7+11+12+16}{4} = 11,5;$$

$$\bar{y}_2 = \frac{8+7+11}{3} = 8,7;$$

$$\bar{x}_2 = \frac{22+27+30}{3} = 26,3;$$

$$\bar{y}_3 = \frac{14+18+17}{3} = 16,3;$$

$$\bar{x}_3 = \frac{38+45+46}{3} = 43,0.$$

Результати розрахунків оформляємо у вигляді табл. 6.4.

Таблиця 6.4 – Залежність прибутку від обсягу коштів банку

№ з/п	Групи банків за обсягом коштів, млн. грн., (x)	Кількість банків, штук, (f)	Середній прибуток банків, млн. грн., (\bar{y}_i)	Середній обіг коштів, млн. грн., (\bar{x}_i)
1	7-20	4	4,3	11,5
2	20-33	3	8,7	26,3
3	33-46	3	16,3	43,0
	Разом	10	9,2	25,4

Зростання групових середніх рівня прибутку від групи до групи свідчить про наявність кореляційного зв'язку між обігом коштів і розміром прибутку.

Крім того аналітичне групування дає змогу встановити кількісне співвідношення між ознаками, що вивчаються.

Отже, можна не лише стверджувати, що існує кореляційний зв'язок між факторною (x) і результативною (y) ознаками, а й визначити, як у середньому змінюється (y) зі зміною (x) на одиницю.

Ефекти впливу x на y визначаються відношенням приростів середніх групових $\Delta y : \Delta x$. Обчислимо ефект впливу розміру обігу коштів на рівень прибутку банків за даними таблиці 6.4:

$$\frac{\Delta y}{\Delta x} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1} = \frac{8,7 - 4,3}{26,3 - 11,5} = 0,297 ;$$

$$\frac{\Delta y}{\Delta x} = \frac{\bar{y}_3 - \bar{y}_2}{\bar{x}_3 - \bar{x}_2} = \frac{16,3 - 8,7}{43,0 - 26,3} = 0,455 .$$

Отже, з підвищенням обігу коштів на 1 млн. грн. прибуток збільшується: в другій групі порівняно з першою на 0,297 млн. грн., в третій групі порівняно з другою на 0,455 млн. грн.

Третій етап аналітичного групування – вимірювання тісноти зв'язку за допомогою дисперсійного аналізу.

6.3. Дисперсійний аналіз

Основною метою дисперсійного аналізу є виявлення впливу окремих факторів чи умов, які визначають варіацію ознаки.

В основі дисперсійного аналізу лежить закон розкладання загальної дисперсії на складові, згідно з яким загальна дисперсія результативної ознаки у складається з двох частин: між групової (факторної) дисперсії та середньої з групових (залишкової).

Взаємозв'язок факторної та залишкової варіації описується правилом розкладання (додавання) дисперсій: $\sigma_y^2 = \delta_y^2 + \overline{\sigma_y^2}$.

Суть дисперсійного аналізу полягає у зіставленні (порівнянні) різних видів дисперсій:

- міжгрупової та загальної;
- міжгрупової та внутрішньогрупової;
- загальної та внутрішньогрупової.

Відношення міжгрупової (факторної) дисперсії до загальної дисперсії розглядається як міра щільності кореляційного зв'язку і називається *коефіцієнтом детермінації*, який можна розрахувати за формулою:

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

За статистичною структурою це відношення є часткою варіації результативної ознаки у, що пов'язана з варіацією х.

Здобувши квадратний корінь з η^2 , одержимо *емпіричне кореляційне відношення*, що дорівнює:

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

Кореляційне відношення може змінюватись від 0 до 1. Якщо $\eta = 0$, то між групова дисперсія дорівнює нулю. В свою чергу, це можливе лише за умови, коли групові середні однакові і кореляційний зв'язок між ознаками відсутній.

При $\eta = 1$ міжгрупова дисперсія дорівнює загальній, а середня з групових – нулю. В цьому випадку кожному значенню факторної озна-

ки відповідає одне єдине значення результативної, тобто зв'язок між ознаками функціональний.

Індекс кореляції (R), визначають зіставленням внутрішньогрупової дисперсії і загальної, і розраховують за формулою:

$$R = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{\frac{\sigma^2 - \overline{\sigma^2}}{\sigma^2}} = \sqrt{1 - \frac{\overline{\sigma^2}}{\sigma^2}}.$$

Необхідно зазначити, що значення $\eta^2 > 0$ не завжди є доказом наявності кореляційного зв'язку між ознаками. Відмінне від нуля кореляційне відношення може з'явитись при випадковому розподілі сукупності на групі.

Перевірку істотності відхилень групових середніх здійснюють за допомогою критеріїв математичної статистики. Це перевірка базується на порівнянні значення η^2 з так званим *критичним значенням*. Критичне значення є тим максимально можливим значенням кореляційного відношення, яке може виникнути випадково при відсутності кореляційного зв'язку.

Якщо фактичне значення η^2 більше критичного, то зв'язок між результативною і факторною ознаками вважається *істотним*.

Якщо фактичне значення η^2 менше критичного, то наявність кореляційного зв'язку між ознаками не доказана і зв'язок вважається *неістотним*.

Для оцінки надійності кореляційних характеристик використовують Критерій Фішера (F) та критерій Стюдента (t) за відповідними формулами.

Для вивчення зв'язку результативної ознаки з двома і більше факторними ознаками використовують *комбінаційні аналітичні групування* і дисперсійний аналіз. Вони дають можливість аналізувати залежність результативної ознаки від кожного з факторів при фіксованих значення інших. Але в даному випадку зазначені методи є громіздкими і потребують використання комп'ютерної техніки.

Контрольні запитання

1. Назвіть види взаємозв'язків між явищами.
2. Вкажіть чим відрізняється різні види взаємозв'язків між собою.
3. Що таке факторні та результативні ознаки?
4. Назвіть методи, які використовуються для отриманні відповіді на питання про наявність або відсутність кореляційного зв'язку.
5. В чому полягає суть методу аналітичного групування?
6. В чому полягає суть дисперсійного методу аналізу?
7. Вкажіть формулу для розрахунку коефіцієнта детермінації.
8. Наведіть формулу для розрахунку емпіричного кореляційного відношення.
9. В чому полягає перевірка істотності відхилень групових середніх?
10. Які критерії використовують для оцінки надійності отриманих кореляційних характеристик?

ЛЕКЦІЯ 7. КОРЕЛЯЦІЙНИЙ АНАЛІЗ

7.1 *Поняття про кореляційний аналіз.*

7.2 *Основні етапи проведенні кореляційного аналізу.*

7.3 *Проста (парна) лінійна кореляція.*

7.4 *Показники тісноти зв'язку.*

7.1 Поняття про кореляційний аналіз

Вивчення реальної дійсності показує, що практично кожне суспільне явище знаходиться в тісному зв'язку і взаємодії з іншими явищами, якими б випадковими вони не здавалися на перший погляд. Так, наприклад, рівень урожайності сільськогосподарських культур залежить від множини природних і економічних факторів, тісно пов'язаних між собою.

Для дослідження взаємозв'язків між явищами статистика використовує ряд методів і прийомів: статистичні групування (прості і комбінаційні), індексний, кореляційний і дисперсійний аналіз, балансовий, табличний, графічний та ін.

Поряд з уже розглянутими методами вивчення взаємозв'язків особливе місце займає метод кореляції, який є логічним продовженням таких методів як аналітичне групування, дисперсійний аналіз і зіставлення паралельних рядів. В поєднанні з цими методами він надає статистичному аналізу закінчений, завершений характер.

Засновниками теорії кореляції є англійські статистики Ф. Гальтон (1822–1911 рр.) і К. Пірсон (1857–1936 рр.).

Термін кореляція походить від англійського слова “корелейшен” співвідношення, відповідність (взаємозв'язок, взаємозалежність) між ознаками, що виявляється при масовому спостереженні зміни середньої величини однієї ознаки залежно від значення іншої. Ознаки, що пов'язані між собою кореляційним зв'язком, називають *корельованими*.

Кореляційний аналіз дає змогу виміряти ступінь впливу факторних ознак на результативні, встановити єдину міру тісноти зв'язку і

роль досліджуваного фактора (факторів) у загальній зміні результативної ознаки.

Кореляційний метод дозволяє одержати кількісні характеристики ступеня зв'язку між двома і більшим числом ознак, а тому на відміну від розглянутих вище методів, дає більш широке уявлення про зв'язок між ними.

Кореляційний зв'язок є неповним, він проявляється при великій кількості спостережень, при порівнянні середніх значень результативної і факторної ознак. У цьому відношенні виявлення кореляційних залежностей пов'язано з дією *закона великих чисел*.

Згідно з законом великих чисел: "Тільки при досить великій кількості спостережень індивідуальні особливості і другорядні фактори згладяться і залежність між результативною і факторною ознаками, якщо вона має місце, виявиться досить виразно".

За допомогою кореляційного аналізу вирішують такі основні завдання:

1) визначення середньої зміни результативної ознаки під впливом одного або кількох факторів (в абсолютному або відносному вимірі);

2) характеристика ступеня залежності результативної ознаки від одного з факторів при фіксованому значенні інших факторів, включених до кореляційної моделі;

3) визначення тісноти зв'язку між результативними і факторними ознаками (як з усіма факторами, так і з кожним фактором окремо при виключенні впливу інших);

4) визначення і розкладання загального обсягу варіації результативної ознаки на відповідні частини і встановлення ролі кожного окремого фактора в цій варіації;

5) статистична оцінка вибірових показників кореляційного зв'язку.

Кореляційний зв'язок виражається відповідними математичними рівняннями. За напрямом зв'язок між корелюючими ознаками може бути прямим і оберненим. При *прямому зв'язку* обидві ознаки змінюються в одному напрямі, тобто із збільшенням факторної ознаки зростає результативна і навпаки (наприклад, зв'язок між якістю ґрунту і врожай-

ністю, рівнем годівлі і продуктивністю тварин, стажем роботи і продуктивністю праці). При оберненому зв'язку обидві ознаки змінюються в різних напрямках (наприклад, зв'язок між урожайністю і собівартістю продукції, продуктивністю праці і собівартістю продукції).

За формою або аналітичним вираженням розрізняють зв'язки прямолінійні (або просто лінійні) і нелінійні (або криволінійні). Якщо зв'язок між ознаками виражається рівнянням прямої лінії, то його називають *лінійним зв'язком*, якщо ж він виражається рівнянням будь-якої кривої (параболи, гіперболи, показникової, степеневої і т.д.), то такий зв'язок називають *нелінійним або криволінійним*.

Залежно від кількості досліджуваних ознак розрізняють парну (просту) і множинну кореляцію. При *парній кореляції* вивчають зв'язок між двома ознаками (результативною і факторною), при *множинній кореляції* зв'язок між трьома і більшим числом ознак (результативною і двома і більшим числом факторів).

За допомогою методу кореляційного аналізу вирішується два головних завдання:

- 1) визначення форми і параметрів рівняння зв'язку;
- 2) вимірювання тісноти зв'язку.

Перше завдання вирішується знаходженням рівняння зв'язку і визначенням його параметрів. Друге – за допомогою розрахунку різних показників тісноти зв'язку (коефіцієнта кореляції, кореляційного відношення, індексу кореляції та ін.).

7.2 Основні етапи проведенні кореляційного аналізу

Схематично кореляційний аналіз можна поділити на 5 етапів:

1 етап – постановка завдання, встановлення наявності зв'язку між досліджуваними ознаками;

2 етап – відбір найістотніших факторів для аналізу;

3 етап – визначення характеру зв'язку, його напрямку і форми, вибір математичного рівняння для вираження існуючих зв'язків;

4 етап – розрахунок числових характеристик кореляційного зв'язку (визначення параметрів рівняння і показників тісноти зв'язку);

5 етап – статистична оцінка вибірових показників зв'язку.

Науково обґрунтоване застосування кореляційного методу потребує перед усім глибокого розуміння суті взаємозв'язків соціально-економічних явищ. Сам метод не встановлює наявності і причин виникнення зв'язків між досліджуваними явищами, його призначення полягає в їх кількісному вимірюванні.

Розглянемо детальніше проведення цих етапів.

1 етап. На цьому етапі кореляційного аналізу здійснюється загальне ознайомлення з досліджуваним об'єктом і явищами, уточнюються мета і завдання дослідження, встановлюється теоретична можливість причинно-наслідкового зв'язку між ознаками.

Встановлення причинних залежностей в досліджуваному явищі передуює власне кореляційному аналізу. Тому застосуванню методів кореляції повинен передувати глибокий теоретичний аналіз, який охарактеризує основний процес, що протікає в досліджуваному явищі, визначить суттєві зв'язки між окремими його сторонами і характер їх взаємодії.

Попередній аналіз даних створює основу для формулювання конкретного завдання дослідження зв'язків, відбору найважливіших факторів, встановлення можливої форми взаємозв'язку ознак і тим самим приводить до математичної формалізації - до вибору математичного рівняння, яке найбільш повно відтворить існуючі зв'язки.

2 етап. Одним із найважливіших питань кореляційного аналізу є відбір результативної і факторної (факторних) ознак. Факторні і результативні ознаки, що відбираються для кореляційного аналізу, повинні бути суттєвими, перші повинні безпосередньо впливати на другі. Відбір факторів для включення їх в кореляційну модель повинен базуватися передусім на теоретичних основах і практичному досвіді аналізу досліджуваного соціально-економічного явища. Велику допомогу в розв'язанні цього завдання можуть надати такі статистичні прийоми і методи, як зіставлення паралельних рядів, побудова таблиць розподілу чисельностей за двома ознаками (кореляційних таблиць), побудова статистичних групувань як за результативною ознакою з аналізом взаємо-

пов'язаних з нею факторів, так і за факторною ознакою (або комбінацією факторних ознак) з аналізом їх впливу на результативну ознаку.

Відбір факторів для парних кореляційних моделей не складний: з множини факторів, що впливають на результативну ознаку, відбирається один із найважливіших факторів, який в основному визначає варіацію результативної ознаки або ж фактор, істотність впливу якого на результативну ознаку передбачається вивчити або перевірити. Відбір факторів для множинних кореляційних моделей має ряд особливостей і обмежень.

3 етап. Одною з головних проблем побудови кореляційної моделі є визначення форми зв'язку і на цій основі встановлення типу аналітичної функції, що відображає механізм зв'язку результативної ознаки з факторною (факторними). Під *формою кореляційного зв'язку* розуміють тип аналітичного рівняння, що виражає залежність між досліджуваними ознаками.

Вибір того або іншого рівняння для дослідження зв'язків між ознаками є найбільш важким і відповідальним завданням, від якого залежать результати кореляційного аналізу. Всі подальші найретельніші розрахунки можуть бути знецінені, якщо форма зв'язку вибрана невірно. Важливість цього етапу полягає в тому, що правильно встановлена форма зв'язку дає змогу підібрати і побудувати найбільш адекватну модель і на основі її розв'язання одержати статистично вірогідні і надійні характеристики.

Встановлення форми зв'язку між ознаками в більшості випадків обґрунтовується теорією або практичним досвідом попередніх досліджень.

Якщо форма зв'язку невідома, то при парній кореляції математичне рівняння може бути встановлено за допомогою складання кореляційних таблиць, побудови статистичних групувань, перегляду різних функцій на ПЕОМ і вибір такого рівняння, яке дає найменшу суму квадратів відхилень фактичних даних від вирівняних (теоретичних) значень та ін.

Залежно від вихідних даних теоретичною лінією регресії можуть бути різні типи кривих або пряма лінія. Так, якщо зміна результативної

ознаки під впливом фактора характеризується постійними приростами, то це вказує на лінійний характер зв'язку, якщо ж зміни результативної ознаки під впливом фактора характеризується постійними коефіцієнтами зростання, то є підстава припустити криволінійний зв'язок.

Особливе місце в обґрунтуванні форми зв'язку при проведенні кореляційного аналізу належить графікам, побудованих у системі прямокутних координат на основі емпіричних даних. Графічне зображення фактичних даних дає наочне уявлення про наявність і форму зв'язку між досліджуваними ознаками.

Згідно з правилами математики при побудові графіка на осі абсцис відкладають значення факторної ознаки, а на осі ординат – значення результативної ознаки. Відклавши на перетині відповідних значень двох ознак точки, одержимо точковий графік, який називають *кореляційним полем*. За характером розміщення точок на кореляційному полі роблять висновок про напрям і форму зв'язку. Достатньо поглянути на графік, щоб прийти до висновку про наявність і форму зв'язку між ознаками. Якщо точки концентруються навколо уявної осі напрямленої зліва, знизу, направо, вгору, то зв'язок прямий, якщо ж навпаки зліва, зверху, направо, вниз – зв'язок обернений. Якщо точки розкидані по всьому полю, то це свідчить про те, що зв'язок між ознаками відсутній або дуже слабкий. Характер розміщення точок на кореляційному полі вказує також і на наявність прямолінійного або криволінійного зв'язку між досліджуваними ознаками.

За допомогою графіка добирають відповідне математичне рівняння для кількісної оцінки зв'язку між результативною і факторною ознаками.

Рівняння, що відображає зв'язок між ознаками, *називають рівнянням регресії або кореляційним рівнянням*. Якщо рівняння регресії зв'язує лише дві ознаки, то воно називається *рівнянням парної регресії*. Якщо рівняння зв'язку відображає залежність результативної ознаки від двох і більше факторних ознак, воно називається *рівнянням множинної регресії*. Криві, побудовані на основі рівнянь регресії, називають *кривими регресії або лініями регресії*.

Розрізняють емпіричну і теоретичну лінії регресії. Якщо на кореляційному полі з'єднати точки відрізками прямої лінії, то одержимо ламану лінію з деякою тенденцією, яка називається емпіричною лінією регресії. *Теоретичною лінією регресії* називається та лінія, навколо якої концентруються точки кореляційного поля і яка вказує основний напрям, основну тенденцію зв'язку. Теоретична лінія регресії повинна відображати зміну середніх величин результативної ознаки в міру зміни величин факторної ознаки при умові повного взаємопогашення всіх інших випадкових по відношенню до фактора - причин. Отже, ця лінія має бути проведена так, щоб сума відхилень точок кореляційного поля від відповідних точок теоретичної лінії дорівнювала нулю, а сума квадратів відхилень була б мінімальною величиною. Пошук, побудова, аналіз і практичне застосування теоретичної лінії регресії називають *регресійним аналізом*.

За емпіричною лінією регресії не завжди вдається встановити форму зв'язку і добрати рівняння регресії. В таких випадках будують і розв'язують різні рівняння регресії. Потім оцінюють їх адекватність і добирають таке рівняння, яке забезпечує найкращу апроксимацію (наближення) фактичних даних до теоретичних і достатню статистичну вірогідність і надійність.

Регресійний аналіз вирішує питання побудови, розв'язання і оцінки рівнянь регресії, а при кореляційному аналізі до цих питань приєднується ще коло питань пов'язаних із визначенням тісноти зв'язку між результативною і факторною (факторними) ознаками. Зазвичай регресійно-кореляційний аналіз дуже часто розглядається як єдине ціле і називається просто кореляційний аналіз.

Щоб результати кореляційного аналізу знайшли практичне застосування і дали науково обґрунтовані результати, повинні виконуватись певні вимоги відносно об'єкта дослідження і якості вихідної статистичної інформації. Основні з цих вимог такі:

– якісна однорідність досліджуваної сукупності, що передбачає близькість формування результативних і факторних ознак. Необхідність виконання цієї умови впливає із змісту параметрів рівняння зв'язку. З математичної статистики відомо, що параметри є середніми величина-

ми. В якісно однорідній сукупності вони будуть типовими характеристиками, в якісно різнорідній - спотвореними, що перекручують характер зв'язку.

Кількісна однорідність сукупності полягає у відсутності одиниць спостереження, які за своїми числовими характеристиками суттєво відрізняються від основної маси даних. Такі одиниці спостереження слід виключати із сукупності і вивчати окремо;

– досить велике число спостережень, оскільки зв'язки між ознаками виявляються тільки внаслідок дії закону великих чисел. Кількість одиниць спостереження повинна в 6 – 8 разів перевищувати кількість включених у модель факторів;

– випадковість і незалежність окремих одиниць сукупності одна від одної. Це означає, що значення ознак у одних одиниць сукупності не повинні залежати від значень у інших одиниць даної сукупності;

– стійкість і незалежність дії окремих факторів;

– сталість дисперсії результативної ознаки при зміні факторних ознак;

– нормальний розподіл ознак.

7.3 Проста (парна) лінійна кореляція

Найпростішим видом кореляційного зв'язку є зв'язок між двома ознаками: результативною і факторною. Такий зв'язок називають *парною кореляцією або простою кореляцією*.

Дуже часто в дослідженнях взаємозв'язку двох факторів серед множини функцій часто розглядається прямолінійна форма зв'язку, яка виражається рівнянням прямої лінії:

$$\overline{y_x} = a + b \cdot x,$$

де $\overline{y_x}$ – вирівняне значення результативної ознаки (залежна змінна); x – значення факторної ознаки (незалежна змінна); a – початок відліку, або $\overline{y_x}$; b – коефіцієнт регресії, який показує середню змінну залежної змінної при зміні незалежної змінної на одиницю (одне своє значення).

Коефіцієнти регресії є величинами іменованими і мають одиниці

вимірювання, що відповідають змінним, між якими вони характеризують зв'язок.

Якщо $b > 0$, то зв'язок прямий, якщо $b < 0$, то зв'язок обернений, якщо $b = 0$, то зв'язок відсутній.

Параметри рівняння a і b визначають способом найменших квадратів. Він дає можливість знайти ту криву, яка порівняно з іншими кривими проходить найближче до точок кореляційного поля, що відображають фактичні дані, тобто дає найменшу суму квадратів відхилень фактичних значень результативної ознаки від вирівняних (теоретичних) значень:

$$\sum (y_i - \bar{y}_x)^2 \rightarrow \min .$$

Порядок одержання системи нормальних рівнянь при парній кореляції

Для одержання першого рівняння системи необхідно всі члени вихідного рівняння кореляційного зв'язку помножити на коефіцієнти при першому невідомому (a) і одержані добутки підсумувати.

Потім для отримання другого рівняння необхідно всі члени вихідного рівняння помножити на коефіцієнт при другому невідомому (b) і також всі добутки підсумувати.

Техніка одержання системи нормальних рівнянь залишається аналогічною і для побудови системи рівнянь з більшим числом змінних.

Так, для парного лінійного зв'язку система нормальних рівнянь має вигляд:

$$\begin{cases} \sum y = an + b \sum x \\ \sum yx = a \sum x + b \sum x^2 \end{cases}$$

де n – число спостережень.

Параметри a і b рівняння прямої лінії можна визначити за іншими робочими формулами:

$$a = \frac{\sum y \sum x^2 - \sum yx \sum x}{n \sum x^2 - \sum x \sum x} \text{ і } b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \sum x \sum x} \text{ або}$$

$$a = \bar{y} - b\bar{x} \text{ та } b = \frac{\overline{xy} - \bar{x}\bar{y}}{x^2 - (\bar{x})^2}.$$

Рівняння кореляційного зв'язку мають як пізнавальне, так і практичне значення, їх використовують для обчислення теоретичної лінії регресії, очікуваних (теоретичних, вирівняних) і прогнозованих значень залежної змінної при тих або інших значеннях фактора (факторів). При цьому слід мати на увазі, що рівняння дає середнє співвідношення між результативною і факторною ознаками, тому найбільшу точність збігання мають розрахункові значення результативної ознаки при величині фактора, близького до середнього його рівня.

Ступінь наближення розрахункових значень результативної ознаки до її фактичного значення залежить від того, наскільки досконала кореляційна модель. Якщо вона включає всі основні фактори, що визначають варіацію результативної ознаки, то точність буде досить високою.

7.4 Показники тісноти зв'язку

При кореляційному зв'язку разом з досліджуваним фактором або кількома факторами при множинній кореляції на результативну ознаку впливають і інші фактори, які не враховуються або не можуть бути точно враховані. При цьому дія їх може бути направлена як в сторону підвищення результативної ознаки, так і в сторону її зниження. Отже, дослідження зв'язку відбувається в умовах, коли цей зв'язок більшою або меншою мірою затушовується суперечливою дією інших причин. Тому одне із завдань кореляційного аналізу полягає у визначенні тісноти зв'язку між ознаками, у визначенні сили дії досліджуваного фактора (факторів) на результативну ознаку.

Тіснота зв'язку у кореляційному аналізі характеризується за допомогою спеціального відносного показника, який отримав назву *коефіцієнта кореляції*.

При парній лінійній залежності тіснота зв'язку визначається за допомогою *лінійного коефіцієнта кореляції*:

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

$$\text{де } \bar{x} = \frac{\sum x}{n}, \bar{y} = \frac{\sum y}{n}, \overline{xy} = \frac{\sum xy}{n}, \sigma_x = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}, \sigma_y = \sqrt{\frac{\sum y^2}{n} - (\bar{y})^2}.$$

Він може бути обчислений і за допомогою інших формул:

$$r = \frac{\sum (y - \bar{y})(x - \bar{x})}{n\sigma_x \cdot \sigma_y} \text{ або } r = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sqrt{\sum (y - \bar{y})^2 \cdot \sum (x - \bar{x})^2}} \text{ або } r = b \frac{\sigma_x}{\sigma_y}.$$

Коефіцієнт кореляції знаходиться в межах від 0 до ± 1 . Якщо коефіцієнт кореляції дорівнює нулю, то зв'язок відсутній, а якщо одиниці, то зв'язок функціональний. Знак при коефіцієнті кореляції вказує на напрям зв'язку "+" – прямий, "-" – зв'язок обернений). Чим ближче коефіцієнт кореляції до одиниці, тим зв'язок між ознаками тісніший.

Квадрат коефіцієнта кореляції називається *коефіцієнтом детермінації* (r^2). Він показує, яка частка загальної варіації результативної ознаки визначається досліджуваним фактором. Якщо коефіцієнт детермінації виражений в процентах, то його слід читати так: варіація (коливання) залежної змінної на стільки-то процентів зумовлена варіацією фактора.

Між лінійним коефіцієнтом кореляції (r) і коефіцієнтом повної регресії (b) є такий зв'язок: $r = b \frac{\sigma_x}{\sigma_y}$. Звідки $b = r \frac{\sigma_y}{\sigma_x}$.

Отже, знаючи коефіцієнт кореляції (r) і значення середніх квадратичних відхилень по x і y , можна визначити коефіцієнт регресії (b) і навпаки, знаючи коефіцієнт регресії (b) і відповідні середні квадратичні відхилення можна обчислити коефіцієнт кореляції (r).

При парній лінійній залежності коефіцієнт кореляції і коефіцієнт повної регресії мають однакові знаки (плюс, мінус).

Перевірку істотності зв'язку в кореляційному аналізі здійснюють за допомогою таких самокритеріїв і за тими саме процедурами, що і при дисперсному аналізі та аналітичному групуванні.

Контрольні запитання

1. Назвіть переваги та недоліки кореляційного аналізу.
2. Зазначте види кореляційного зв'язку та різновиди кореляційно-

го аналізу.

3. Вкажіть основні етапи проведення кореляційного аналізу.
4. Що таке кореляційне поле, як воно будується?
5. В чому полягає відмінність між теоретичною на емпіричною лінією регресії?
6. Назвіть задач які вирішує кореляційний аналіз.
7. Наведіть рівняння парної лінійної кореляції.
8. Вкажіть показник, який характеризує тісноту зв'язку при кореляційному аналізі. Які значення він може приймати?
9. Наведіть формули для розрахунку коефіцієнта кореляції.
10. Що показує коефіцієнт регресії?

ЛЕКЦІЯ 8. СТАТИСТИЧНЕ ВИВЧЕННЯ ДИНАМІКИ

8.1 *Поняття про ряди динаміки. Види рядів динаміки.*

8.2 *Аналітичні показники ряду динаміки. Середні показники динаміки.*

8.3 *Розрахунок тенденції.*

8.4 *Коефіцієнт випередження. Екстраполяція та інтерполяція.*

8.1 Поняття про ряди динаміки. Види рядів динаміки

Динамікою (від грецького *динаміс* – "сила, розвиток") називається процес розвитку явища в часі і просторі. Для того, щоб відобразити ці процеси динаміки будують ряди динаміки (інша назва – динамічні ряди).

Динамічним рядом (рядом динаміки) називають ряд статистичних показників, що розташовані в хронологічній послідовності і характеризують зміну явища в часі.

Динамічний ряд складається з двох елементів:

1) статистичний показник (інша назва – рівень ряду) – характеризує величину явища, його розмір і найчастіше позначається через "у";

2) момент часу, ряд періодів – показник, який характеризує певний час, у який дійсний відповідний статистичний показник.

Приклад ряду динаміки наведено в табл. 8.1.

Таблиця 8.1 – Ряд динаміки процесу

Момент часу (ряд періодів)	Показник
1990	54,2
1991	54,1
1992	53,9
1993	53,8
1994	53,2

Види рядів динаміки

Ряд динаміки може бути в залежності від показників, які утворюють дану сукупність: *абсолютним, відносним і середнім*.

В залежності від часу, який визначений в динамічних рядах вони поділяються на *інтервальні і моментні*.

Залежно від відстані між рівнями ряду динаміки, ряди можуть бути *рівні і нерівні* (тобто з *рівними і нерівними інтервалами*).

Залежно від кількості статистичних показників: *одномірний і багатомірний*.

8.2 Аналітичні показники ряду динаміки. Середні показники динаміки

Наведемо приклад аналітичних показників ряду динаміки. Аналітичні показники ряду динаміки наведено в таблиці 8.2.

Таблиця 8.2 – Аналітичні показники ряду динаміки

Роки	Всього побудовано ЖБК, млн м ²	Абсолютний приріст Δu , млн м ²		Коефіцієнти або темпи зростання		Темпи приросту, %		Абсолютне значення одного відсотку приросту, тис. м ²	Пункти росту, пункто-відсотки
		Порівняно з 1990 р	Порівняно з попереднім роком	Порівняно з 1990 р	Порівняно з попереднім роком	Порівняно з 1990 р	Порівняно з попереднім роком		
1	2	3	4	5	6	7	8	9	10
1990	2,9	–	–	–	–	–	–	–	–
1991	2,4	-0,5	-0,5	0,8276	0,8276	-17,24%	-17,24%	290	-17,24
1992	2,1	-0,8	-0,3	0,7241	0,8750	-27,59%	-12,5%	240	-10,35
1993	1,9	-1	-0,2	0,6552	0,9048	-34,48%	-9,52%	210	-6,89
1994	1,8	-1,1	-0,1	0,6207	0,9474	-38,93%	-5,26%	190	-3,45

В залежності від того, яка база взята для порівняння, розрізняють характеристики *базисні і ланцюгові*. Якщо база порівняння постійна, то характеристики динаміки називають базовими. Якщо база порівняння

змінюється, то характеристики динаміки будуть називатися ланцюговими.

Одним із показників аналітичного дослідження динаміки є *абсолютний приріст (зменшення)*. Це різниця між двома рівнями ряду динаміки. Він показує, наскільки даний рівень ряду перевищує рівень ряду, прийнятий за базу порівняння.

Для ланцюгових показників: $\Delta y = y_i - y_{i-1}$,

для базисних показників: $\Delta y = y_i - y_0$,

де Δy – абсолютний приріст ряду, y_i – рівень періоду, що порівнюється, y_{i-1} – рівень попереднього періоду, y_0 – рівень базисного періоду.

Середній абсолютний приріст (середня швидкість росту) розраховується як середня арифметична з показників швидкості росту за певний період або за окремі проміжки часу.

Для ланцюгового ряду:

$$\bar{\Delta} = \frac{\sum \Delta}{n}, \quad (8.1)$$

де Δ – абсолютний приріст, n – кількість ланцюгових темпів зростання.

Для базисного ряду:

$$\bar{\Delta} = \frac{y_n - y_1}{n - 1}, \quad (8.2)$$

де n – кількість періодів.

Коефіцієнти або темпи зростання показує, у скільки разів збільшився або зменшився рівень ряду відносно базового. Коефіцієнт виражається лише в частках, а темп – частіше виражається у відсотках.

Для базового ряду:

$$K = \frac{y_i}{y_0} \quad (8.3)$$

Для ланцюгового ряду:

$$K = \frac{y_i}{y_{i-1}}, \quad (8.4)$$

де Δy – абсолютний приріст ряду, y_i – рівень періоду, що порівнюється, y_{i-1} – рівень попереднього періоду, y_0 – рівень базисного періоду.

Середній темп росту – обраховується за формулою середньої геометричної.

Для ланцюгового ряду:

$$\overline{T}_{zp} = \sqrt[n]{T_1 \cdot T_2 \cdot T_3 \cdot T_4 \cdot \dots \cdot T_n}, \quad (8.5)$$

де n – кількість ланцюгових темпів зростання

Для базисного ряду:

$$\overline{T}_{zp} = \sqrt[n-1]{\frac{y_n}{y_1}}, \quad (8.6)$$

де n – кількість періодів

Темп приросту показує, наскільки рівень ряду більший від того, з яким ми порівнюємо. Темп приросту обчислюється відношенням абсолютного приросту до базисного рівня.

$$T_{np} = \frac{\Delta y}{y_1} = T_{zp} - 1 \text{ або } 100 \% \quad (8.7)$$

Середньорічний темп приросту: $\overline{T}_{np} = \overline{T}_{zp} - 1$ або 100 %.

Середній рівень ряду. Обрахування середнього рівня ряду залежить від того, який це ряд (інтервальний чи моментний), а також які інтервали він утримує (рівні чи нерівні):

– для *інтервального ряду з рівними інтервалами* середній рівень ряду обраховується через середню арифметичну просту.

– для *інтервального ряду з нерівними інтервалами* середній рівень ряду розраховується як середня арифметична зважена:

$$x = \frac{\sum x \cdot t}{\sum t}, \quad (8.8)$$

де t – число періодів часу, протягом яких рівень не змінюється.

– для *моментного ряду з рівними інтервалами* середній рівень ряду обраховується як середня хронологічна проста:

$$\overline{x}_{xp} = \frac{\frac{1}{2}x_1 + x_2 + x_3 + \dots + x_{n-1} + \frac{1}{2}x_n}{n-1}; \quad (8.9)$$

– якщо ми маємо *моментний ряд але нерівні інтервали*, то використовується середня хронологічна зважена:

$$\overline{x}_{xp.zв.} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \dots + (y_{n-1} + y_n)t_{n-1}}{2 \sum t_{n-1}}. \quad (8.10)$$

Частіше використовується середня арифметична зважена:

$$\bar{x} = \frac{\sum \bar{y}_i \cdot t_i}{\sum t_i}, \quad (8.11)$$

де $\bar{y}_i = \frac{y_i + y_{i-1}}{2}$.

8.3 Розрахунок тенденції (тренду)

Тенденція (або тренд) – це основний напрям розвитку того явища, яке ми досліджуємо. Графічне зображення лінії тренду приведено на рис. 8.1.

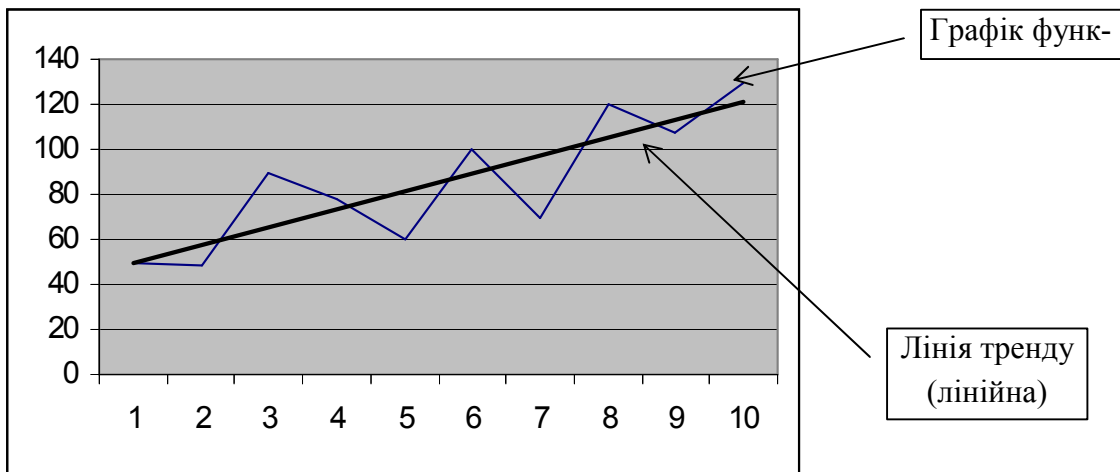


Рисунок 8.1 – Лінія тренду на графіку

Існує декілька аналітичних методів обчислення тренду.

1. *Метод укрупнення інтервалів.* Принцип цього прийому полягає в тому, що дані динамічного ряду об'єднують в групи по періодам, і для них розраховують середній показник на період 3, 5, 10 і більше років.

Приклад 1

Інтервал	Значення ознаки
1991	50
1992	48
1993	55
1994	57
1995	60
1996	58

Інтервал	Значення ознаки
1991-1993	$\frac{50 + 48 + 55}{3} = 51$
1994-1996	$\frac{57 + 60 + 58}{3} = 58,3$

За результатами маємо дві точки для графічної побудови лінії тренду: 51 та 58,3.

2. *Метод ковзної середньої.* 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 арифметичними за визначені періоди (тобто 3, 5, 10 років).

Приклад 2

Інтервал	Значення ознаки
1991	50
1992	48
1993	55
1994	57
1995	60
1996	58

Значення ознаки (середня)
$(50 + 48) / 2 = 49$
$(48 + 55) / 2 = 51,5$
$(55 + 57) / 2 = 56$
$(57 + 60) / 2 = 58,5$
$(60 + 58) / 2 = 59$

3. *Метод зімкнення рядів* – об'єднання двох і більше рядів, що характеризують зміну одного і того ж явища, використовується тоді, коли показники динамічних рядів не можуть бути співставлені. Змикання рядів проводять наступним чином: рахують 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.

Приклад 3

Нехай маємо два ряди.

1990	1991	1992
100	139	153

1992	1993	1994	1995
100	120	154	176

Коефіцієнт буде дорівнювати: $K = \frac{153}{100} = 1,53$. Отримане значення – базове значення.

З'єднаймо ці ряди, помножуючи значення у другому ряду на цей

коефіцієнт. Отримали третій, зімкнений ряд.

1990	1991	1992	1993	1994	1995
100	139	153	184	236	269

4. *Метод аналітичного вирівнювання.* Для розрахунку тенденції за допомогою цього методу виходять з того, що зміна рівнів ряду динаміки може бути тим, чи іншим ступенем точності наближення виражена математичними функціями. Вирівнювання визначається характером динаміки розвитку конкретного явища.

8.4 Коефіцієнт випередження. Екстраполяція та інтерполяція

Коефіцієнт випередження – це показник інтенсивності зміни одного ряду динаміки порівняно з іншим за однакові проміжки часу.

$$k_k = \frac{k'}{k''},$$

де k' – темп зростання першого ряду, k'' – темп зростання другого ряду, обчислені на базовій основі.

Екстраполяція та інтерполяція

Інтерполяція – це знаходження відсутнього показника всередині ряду.

Екстраполяція – знаходження наступних рівнів ознаки (у кінці або на початку) при умові, що попередні відомі.

І екстраполяція і інтерполяція базуються на одній умові – існує тенденція, яка характерна для всього ряду, і з її допомогою можна обрахувати невістачаючі дані. Однак для обрахування невістачаючих даних потрібно знати функцію (закон) за яким змінюються дані всередині ряду. Дуже часто інтерполяція використовується для знаходження необхідних табличних даних.

Наведемо ряд у вигляді табл. 8.1.

Таблиця 8.1 – Застосування інтерполяції та екстраполяції

№ елемента	x	y
1	1	3,5
2	2	5
3	3	–
4	4	8
5	5	–

Ми бачимо, що для $x = 3$ відсутнє значення “ y ”. Знайдемо “ y ” використовуючи інтерполяцію.

Якщо розташувати “ x ” у вигляді ряду, то отримаємо пряму лінію. Отже ми маємо тренд у вигляді прямої лінії.

Позначимо наступним чином інтервал і дані в якому вони знаходяться: $x_1 = 2$; $x_2 = 4$ $x_x = 3$ і відповідно $y_1 = 5$; $y_2 = 8$; y_y – знайдемо.

Запишімо для y_y рівняння, використовуючи рівняння прямої лінії:

$$y_y = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)} \cdot (x_x - x_1)$$

і підставимо необхідні числа

$$y_y = 5 + \frac{(8 - 5)}{(4 - 2)} \cdot (3 - 2) = 6,5.$$

Окрім наведеного рівняння необхідний y_y можна знайти іншим способом:

$$y_y = y_2 - \frac{(y_2 - y_1)}{(x_2 - x_1)} \cdot (x_2 - x_x)$$

і також підставимо необхідні числа

$$y_y = 8 - \frac{(8 - 5)}{(4 - 2)} \cdot (4 - 3) = 6,5.$$

Приклад використання екстраполяції. Користуючись наведеною вище таблицею знайдемо для останнього “ x ” відсутнє у таблиці значення “ y ”.

Для вирішення цієї задачі знайдемо наскільки одиниць збільшиться “ y ” при збільшенні на одиницю “ x ”.

$$\Delta x_1 = x_2 - x_1 = 2 - 1 = 1 \text{ при цьому } \Delta y_1 = y_2 - y_1 = 5 - 3,5 = 1,5,$$

$$\Delta x_2 = x_3 - x_2 = 3 - 2 = 1 \text{ при цьому } \Delta y_2 = y_3 - y_2 = 6,5 - 5 = 1,5,$$

$$\Delta x_3 = x_4 - x_3 = 4 - 3 = 1 \text{ при цьому } \Delta y_3 = y_4 - y_3 = 8 - 6,5 = 1,5 .$$

Отже в середині ряду при збільшенні “ x ” на 1 “ y ” збільшується на 1,5.

$$\text{Для } x_5 \text{ знайдемо відсутнє } y_5: y_5 = y_4 + \Delta y = 8 + 1,5 = 9,5 .$$

Контрольні запитання

1. Дайте визначення динамічного ряду. З яких елементів він складається?
2. Зазначте види рядів динаміки які існують.
3. Вкажіть аналітичні показники рядів динаміки.
4. Дайте визначення тенденції (тренду). Назвіть методи обчислення тренду.
5. В чому полягає суть методу укрупнення інтервалів, як метода обчислення тренду?
6. В чому полягає суть методу ковзної середньої, як метода обчислення тренду?
7. В чому полягає суть методу зімкнення рядів, як метода обчислення тренду?
8. В чому полягає суть методу аналітичного вирівнювання, як метода обчислення тренду?
9. Наведіть приклад застосування інтерполяції.
10. Наведіть приклад застосування екстраполяції.

СПИСОК ДЖЕРЕЛ ІНФОРМАЦІЇ

1. Бек В. Л. Теорія статистики: курс лекцій : навчальний посібник / В.Л.Бек. – Київ : ЦУЛ, 2003. – 288 с.
2. Мармоза А. Т. Теорія статистики : підручник \ А. Т. Мармоза.– Київ : Центр учбової літератури, 2013. – 592 с.
3. Костюк В. О. Статистика: навчальний посібник \ В.О. Костюк, І.В.Мількін. – Харків : ХНУМГ ім. О. М. Бекетова, 2015. – 166 с.
4. Ковалевський Г. В. Статистика: учебник \ Г.В.Ковалевський. – Харьков : ХНАГХ, 2012. – 445 с.
5. Кушнір Н.Б. Статистика : навчально-методичний посібник для самостійного вивчення дисципліни / Н. Б. Кушнір [та ін.]. – Київ : Центр учбової літератури, 2009. – 208 с.
6. Опря А. Т. Статистика (модульний варіант з програмованою формою контролю знань) : навчальний посібник \ А. Т. Опря. – Київ : Центр учбової літератури, 2012. – 448 с.
7. Єріна А. М. Статистичне моделювання та прогнозування: навчальний посібник \ А. М. Єріна. – Київ : КНЕУ, 2001. – 170 с.
8. Герасименко С.С. Статистика : підручник / С. С. Герасименко [та ін.]; за наук. ред. д-ра екон. наук С. С. Герасименка. 2-ге вид., перероб. і доп. – Київ : КНЕУ, 2000. – 467 с.
9. Минашкин В.Г. Теория статистики : учебно-методический комплекс / В. Г. Минашкин [и др.]. – Москва : Издательство Центр ЕАОИ. 2008. – 296 с.
10. Казарезов А. Я. Задачі з теорії статистики : навчальний посібник \ А. Я. Казарезов , І. В. Прядко, Г. О. Бурдельна. – Миколаїв: ЧДУ ім. П. Могили, 2012. – 248 с.
11. Єріна А. М. Теорія статистики : практикум \ А. М. Єріна, З. О. Пальян. – Київ : Знання, 2004. – 255 с.
12. Громько Г. Л. Общая теория статистики: практикум \ Г.Л.Громько. – Москва : Инфра-М, 2008. – 240 с.
13. Практикум по теории статистики: учебное пособие / под ред. проф. Р. А. Шмойловой. – Москва : Финансы и статистика, 2003. – 416 с.

ЗМІСТ

Вступ	3
ЛЕКЦІЯ 1. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ	5
1.1 Сутність статистичного спостереження	5
1.2 План статистичного спостереження. Основні визначення	6
1.3 Форми, види і способи статистичного спостереження	8
ЛЕКЦІЯ 2. ПЕРВИННА ОБРОБКА ТА ПРЕДСТАВЛЕННЯ ЇЇ РЕЗУЛЬТАТІВ	12
2.1 Суть статистичного зведення	12
2.2 Статистичні групування. Види групувань	13
2.3 Ряди розподілу	16
2.4 Статистичні таблиці	18
2.5 Графічні способи зображення статистичних даних. Види графіків	20
ЛЕКЦІЯ 3. АБСОЛЮТНІ, ВІДНОСНІ ТА СЕРЕДНІ ВЕЛИЧИНИ	22
3.1 Абсолютні та відносні величини. Основні поняття і категорії	22
3.2 Середні величини: поняття, види та способи їх обчислення	24
ЛЕКЦІЯ 4. ПОКАЗНИКИ ВАРІАЦІЇ	32
4.1 Поняття варіації та її основні показники	32
4.2 Математичні властивості дисперсії та спрощені способи їх розрахунку	36
4.3 Види дисперсій. Правило додавання дисперсій	38
ЛЕКЦІЯ 5. ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ	41
5.1 Сутність та різновиди вибіркового спостереження	41
5.2 Способи відбору одиниць у вибірку сукупність	42
5.3 Визначення середньої, граничної похибки та необхідної чисельності вибірки	43

ЛЕКЦІЯ 6. СТАТИСТИЧНІ МЕТОДИ ВИМІРЮВАННЯ ВЗАЄ- МОЗВ'ЯЗКІВ.....	50
6.1 Види взаємозв'язків між явищами	50
6.2 Метод аналітичного групування	53
6.3 Дисперсійний аналіз	57
ЛЕКЦІЯ 7. КОРЕЛЯЦІЙНИЙ АНАЛІЗ.....	60
7.1 Поняття про кореляційний аналіз	60
7.2 Основні етапи проведенні кореляційного аналізу	62
7.3 Проста (парна) лінійна кореляція	67
7.4 Показники тісноти зв'язку	69
ЛЕКЦІЯ 8. СТАТИСТИЧНЕ ВИВЧЕННЯ ДИНАМІКИ.....	72
8.1 Поняття про ряди динаміки. Види рядів динаміки	72
8.2 Аналітичні показники ряду динаміки. Середні показники динаміки.....	73
8.3 Розрахунок тенденції.....	76
8.4 Коефіцієнт випередження. Екстраполяція та інтерполяція	78
Список джерел інформації.....	81

Навчальне видання

СЕМЕНОВ Євгеній Олександрович
РАЙКО Валентина Федорівна
ЇЛЬІНСЬКА Ольга Ігорівна

ЧИСЛОВІ МЕТОДИ АНАЛІЗУ З ОХОРОНИ ПРАЦІ

Текст лекцій для студентів
освітньо-кваліфікаційного рівня бакалавр,
спеціальність 263 «Цивільна безпека»

Відповідальний за випуск проф. Березуцький В. В.
Роботу до видання рекомендувала проф. Самойленко Н. М.
В авторській редакції

План 2020 р., поз.7

Підп. до друку 23.06.2020 р . Формат 60×84 1/12. Папір офсет.
Друк — різнографія. Гарнітура Minion Pro. Ум. друк. арк. 2,7.
Наклад 50 прим. Зам. № . Ціна договірна.

Видавничий центр НТУ «ХП».

Свідоцтво про державну реєстрацію ДК № 5478 від 21.08.2017 р.
61002, Харків, вул. Кирпичова, 2

Друкарня