Лекція 3.
УЗАГАЛЬНЮВАЛЬНІ СТАТИСТИЧНІ ПОКАЗНИКИ. ПОКАЗНИКИ ВАРІАЦІЇ

План

1. Абсолютні і відносні статистичні величини. Форми, види, способи обчислення.
2. Види середніх. Властивості та умови застосування.
3. Основні показники варіації.

1. Статистичне дослідження незалежно від його мети та масштабів завжди завершується розрахунком та аналізом різних за видами та формою вираження статистичних показників. За допомогою статистичних показників створюється, передається та зберігається інформація про розміри, пропорції, зміни в часі та інші закономірності досліджуваних явищ. Показники, які статистично характеризують досліджувану сукупність в цілому чи її окремі частини, називають узагальнювальними і розрізняють за способом обчислення (первинні та похідні), ознакою часу (інтервальні та моментні) та аналітичними функціями. Узагальнювальні показники можуть бути представлені:

- абсолютними,
- відносними,
- середніми величинами.

Важливою умовою статистичного аналізу є комплексне використання усіх видів узагальнювальних показників.

Абсолютні статистичні величини – початковий, первинний вид узагальнювальних показників, які характеризують чисельність сукупності, розмір (обсяг, рівень) досліджуваного явища і його чинників у конкретних межах часу і місця. Абсолютні величини завжди мають назву і можуть бути виражені в певних одиницях виміру – натуральних (км, кг, шт.), умовних (умовне паливо), трудових (чол.-дні), грошових (грн., $), комплексних (кВт/год., т/км) та часових (рік, сек.). Наприклад, на 1 січня 1997 р. населення Києва становило 2630,4 млн. чол.

За способом вираження абсолютні показники поділяються на:

- індивідуальні – отримують безпосередньо під час статистичного спостереження як результат виміру, підрахунку та оцінки розміру кількісної ознаки окремих одиниць сукупності, наприклад, обсяг статей, написаних конкретним журналістом за місяць;
- загальні – виражають розміри, обсяг, величину тієї чи іншої ознаки у всіх одиниць досліджуваної сукупності і отримуються в результаті підсумовування індивідуальних значень ознак окремих одиниць сукупності, наприклад, обсяг журналістських матеріалів у редакторському кошику певного видання у поточному місяці.

Абсолютні статистичні величини мають незаперечне значення в практичних і наукових дослідженнях – вони є основою для багатьох статистичних розрахунків. Але, щоб мати точне уявлення про величину зміни явищ, треба ці явища зіставити, адже все пізнається у порівнянні. І тоді абсолютні показники розглядаються не тільки самостійно, а й у порівнянні з іншим абсолютним показником, який приймається за базу порівняння.

Відносні статистичні величини –це похідні показники, які характеризують кількісне співставлення статистичних даних. Їх дістають як частку від ділення двох абсолютних величин. Чисельником є порівнювана (поточна) величина, що вивчається, а знаменником – база (основа) порівняння, величина з якою порівнюють і яка є своєрідним вимірювачем, еталоном, оптимальним рівнем. Співставляти можна одновимірні показники, які відносяться до різних періодів, різних об’єктів, територій, а також різновимірні.

В результаті співставлення одновимірних показників отримаємо ***неіменовані*** відносні величини:

- показники динаміки – характеризують зміни явища у часі; розраховують, зіставляючи значення показника за поточний період із його значенням за попередній або початковий період;
- показники структури – характеризують склад досліджуваної сукупності, зміни її руктури (порівнюють одержані частки за два періоди); розраховують як відношення частини до цілого;
- показники координації – оцінюють пропорції між окремими частинами одного цілого; в результаті отримують скільки одиниць однієї частини сукупності припадає на 1, 100, 1000 одиниць іншої частини, що взята за базу порівняння, наприклад, скільки студентів-редакторів припадає на 100 студентів-журналістів;
- показники порівнянь – відображують результати співставлення показників, які відносяться до одного періоду чи моменту часу, але належать до різних об’єктів чи територій; найчастіше це регіональні чи міжнародні порівняння;

Неіменовані відносні величини мають форму:

- коєфіцієнтів (разів), якщо базис приймають за 1.0 (одне ціле); застосовують переважно у випадках, коли порівнювана величина більша за базисну;
- процентів (%), якщо базис приймають за 100.0; проміле (‰) – за 1000.0; проценти, як правило, використовують у випадках, коли порівнюваний абсолютний показник менший за базисний, або більший лише у 2-3 рази, тобто відсотки більші ніж 200-300 % замінюють коефіцієнтом;
- процентних пунктів (п.п.) – у порівняннях структур, де п.п. дорівнює різниці між відповідними частками двох порівнюваних сукупностей.

Відношення між різновимірними (якісно відмінними) абсолютними показниками називають ***іменованими*** відносними величинами – це показники інтенсивності, які показують ступінь інтенсивності поширення досліджуваного явища чи події в певному середовищі. Обчислюються як відношення величини досліджуваного явища до обсягу того середовища, у якому це явище розвивається. Наприклад, показники забезпечення студентів навчально-методичною літературою, доступності бібліотечної мережі тощо. Важливою особливістю відносних величин є їх здатність порівнювати такі явища, абсолютні розміри яких безпосередньо порівняти не можна. Наприклад, порівняння числа народжених та померлих за певний період у різних країнах ще не дає змоги зробити висновок щодо рівня народжуваності чи смертності. Для цього треба мати показники чисельності населення у цей період, обчислити відносні величини і їх результати порівняти. Уважними слід бути до бази порівняння! Її зміна у розрахунках відносних показників може призвести до помилкових висновків, або просто приховати дійсне становище. Наприклад, фірма зменшила зарплату менеджера, що отримував 400$ на 10%, а потім підвищила теж на 10%. Але ця платня становила вже не 400$, а лише 396$, оскільки за 100% у першому випадку приймалися 400$, а у другому–360$.

2. **Середня величина** – узагальнювальний показник, який характеризує типовий рівень варіюючої ознаки, в розрахунку на одиницю однорідної сукупності і може не збігатися з жодним з індивідуальних значень ознаки. Виражається у одиницях виміру ознаки. Середня буде надійною і характеризуватиме типовий рівень ознаки лише за умови, що сукупність якісно однорідна! У неоднорідній сукупності за висловом П.Самуельсона, осереднюються “тигри та кицьки”, що лише створює ілюзію “благоденствія” і не віддзеркалює реалій. У випадках неоднорідності метод середніх поєднують з методом групування і замість однієї загальної середньої обчислюють групові середні величини.

Важлива властивість середньої полягає в тому, що в ній взаємно компенсуються індивідуальні відмінності елементів, які зумовлені дією випадкових факторів і узагальнюються типові риси. Умовою правильного використання середніх величин є і обґрунтований вибір їх форми (виду). Залежно від особливостей досліджень та характеру вихідних даних, а також певної математичної дії над емпіричними значеннями ознаки (підсумовування, множення, коренювання, степенювання) застосовують відповідний вид середніх величин. Усі види середніх поділяють на два класи:

- степеневі середні – середня арифметична, середня гармонійна, середня геометрична, середня квадратична; кожна може бути проста і зважена;
- структурні (позиційні) середні–мода і медіана.

Одним з найпоширеніших видів степеневих середніх у прикладній статистиці є **середня арифметична**, оскільки для більшості явищ характерна адитивність (підсумовування) обсягів (зарплата, витрати, виробництво, тираж, продаж тощо). Її застосовують при вивченні закономірностей розподілу, коли обсяг ознаки для всієї сукупності є сумою індивідуальних значень її окремих елементів. ***Середня арифметична проста*** ($x_{сер}$) обчислюється для незгрупованих даних таким чином: потрібно скласти всі індивідуальні значення ознаки (x) і суму поділити на їх кількість (n):

$$x_{сер} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x}{n} .$$

Середня арифметична зважена розраховується для згрупованих даних, коли відома частота повторення однакових значень ознаки (варіантів) у сукупності. Виконують такі операції: множення кожного варіанта (x) на його частоту (f) – число, що вказує скільки разів цей варіант повторюється, підсумовування отриманих добутків і ділення суми на суму частот:

$$x_{сер} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum xf}{f} ,$$

де процес множення (xf) – називають зважуванням, що відображує факт рівноваговості окремих варіант, а частоту (f)–вагою варіанти.

Наприклад, маємо розподіл студентів групи за балами атестації з певної дисципліни: 5 балів отримали 10 чоловік, 4 бали – 3 чол., 3 бали – 2 чол. Середній бал успішності групи буде таким:

$$\frac{5 \cdot 10 + 4 \cdot 3 + 3 \cdot 2}{10 + 3 + 2} = \frac{68}{15} = 4{,}5 .$$

Важливо відмітити, що на значення середньої вливає коливання структури сукупності! Чим більшу вагу мають великі значення ознаки, тим більша середня, і навпаки, що видно з наведеного вище прикладу. На цю властивість середніх слід зважати при використанні їх у порівняльному аналізі.

Якщо варіанти ознаки подаються у вигляді інтервалу (від…до), то слід знайти серединне значення кожного з інтервалів як півсуму двох меж. При обчисленні середньої з відносних величин (середній процент, середня питома вага) необхідно брати за ваги знаменники співвідношень, за допомогою яких були обчислені індивідуальні відносні показники.

У структурованій сукупності (складається з декількох груп) при розрахунку середньої зваженої варіантами є групові середні, кожна з яких має відповідну вагу у вигляді групових частот. Обчислену таким способом середню називають ***загальною***.

Основні ***властивості*** середньої арифметичної:

- алгебраїчна сума відхилень усіх варіантів від середньої дорівнює нулю:

$\sum (x - x_{сер}) = 0$,

отже в середній взаємно компенсуються додатні та від'ємні відхилення окремих варіант;

- сума квадратів відхилень варіантів від середньої менша за будь-яку іншу величину: $\sum (x - x_{сер})^2 \Rightarrow \min$;

- якщо частоти всіх варіантів поділити чи помножити на будь-яке число, то середня від цього не зміниться:

$$x_{сер} = \frac{\sum x\,(f \cdot a)}{\sum (f \cdot a)};$$

тому величина середньої залежить не від абсолютних значень ваг, а від пропорцій між ними, від питомої ваги варіанта в сукупності; Згідно з цією властивістю замість абсолютних ваг – частот (f) – можна використати відносні ваги у вигляді часток (d).

Інші степеневі середні використовують рідше. ***Середню гармонійну*** – як обернену середній арифметичній, застосовують тоді, коли чисельність сукупності невідома. ***Середню геометричну*** – для аналізу рядів динаміки. ***Середню квадратичну*** – при розрахунках абсолютних і відносних показників варіації ознаки.

Структурним середнім надають перевагу в умовах недостатньої кількості вихідних даних (наприклад, обмеження інформації у зв'язку з "комерційною таємницею"), а також для більш детального розкриття властивостей розподілу і характеристики структури досліджуваної сукупності.

У середніх узагальнювальних показниках, якими є степеневі середні, не видно ні найбільш визначних досягнень, ні відставань, оскільки середня нівелює (стирає, ігнорує) усі індивідуальні особливості. Тому їх доповнюють особливими показниками – модою та медіаною, які є конкретними описовими характеристиками статистичного ряду. Ці характеристики завжди відповідають повному варіанту.

Мода (Мо) – значення ознаки, яке найчастіше зустрічається в сукупності, тобто має найбільшу вагу (частоту, частку). Наприклад: маємо 5 контрольних робіт у яких 1-на помилка, 10 – з 2-ма помилками, 3 – без помилок. Модальним є варіант із значенням 2-ві помилки, оскільки має найбільшу частоту – 10. Мода має практичне значення у біологічних, соціологічних, маркетингових та інших дослідженнях, оскільки вказує, яке значення ознаки є найбільш ймовірним, масовим – найбільш поширена думка, рейтинг популярності, споживчий попит тощо. Зверніть увагу, іноді трапляються сукупності, в яких два, або навіть декілька варіантів повторюються однакову кількість разів. Наявність декількох мод свідчить про неоднорідність сукупності за досліджуваною ознакою – про об'єднання в одній сукупності різноякісних одиниць.

Медіана (Ме) – значення варіюючої ознаки, яке поділяє впорядкований ряд даних на дві рівні частини: 50% одиниць досліджуваної сукупності матимуть значення ознаки менше, ніж медіана, а 50% – більше. Це серединне (центральне) значення. За незгрупованими даними, попередньо розташувавши їх за зростанням, можна визначити “позицію” медіани за формулою:

$$N_{Me} = \frac{n+1}{2},$$

де n – число одиниць сукупності.

Сума відхилень варіант від медіани є найменшою величиною. Ця властивість медіани зумовлює її практичне використання, наприклад, у маркетингу, будівництві –при плануванні розміщення торговельного закладу, зупинок міського транспорту тощо.

3. Показники варіації надають можливість оцінити залежність зміни ознаки від суттєвих для неї причин. На основі цих показників оцінюється інтенсивність структурних зрушень, щільність взаємозв′язків, точність результатів вибіркового обстеження. Виражають показники варіації в абсолютних і відносних величинах. До абсолютних показників належать:

- розмах варіації;
- середнє арифметичне відхилення;
- дисперсія (середній квадрат відхилень);
- середнє квадратичне відхилення.

Функції цих показників полягають у встановленні середньої величини з відхилень індивідуальних значень ознаки від їх середнього значення. Чим меншою є величина відхилень, тим краще, надійніше середній узагальнювальний показник характеризує сукупність.

Розмах варіації (R) є найпростішим з показників варіації і використовується для встановлення амплітуди варіаційної ознаки, тобто різниці між найбільшим і найменшим значенням ознаки:

$R = x_{max} – x_{min}$.

Наприклад, встановити зміну попиту на книжкову продукцію, оцінити варіацію цін тощо. Але слід зважати на те, що крайні значення ознаки, на яких базується цей показник, часто бувають нетиповими (дуже низькі чи дуже високі), або мають випадковий характер. Це дає викривлене значення амплітуди і вимагає уважного ставлення до даних спостереження.

Середнє арифметичне (лінійне) відхилення ($d_{лін}$) служить точнішою характеристикою варіації, оскільки враховує усі відхилення ознаки від її середнього значення. Обчислюється як частка від ділення суми всіх відхилень на їх число. Відхилення беруть за модулем, враховуючи властивість середньої, що $\sum(x – x_{сер})=0$. Отже:

$$d_{лін} = \frac{\sum |x - x_{сер}|}{n} = \frac{\sum |x - x_{сер}| \cdot f}{f},$$

відповідно для незгрупованих та згрупованих даних.

Для більш об’єктивного оцінювання ступеня варіації при якому не порушуються закони алгебри щодо знаку відхилень (“+” чи “–“) прийнято використовувати **дисперсію** (σ^2) – показник середнього квадрата відхилень, міра розсіювання варіантів. Обчислюють аналогічно лінійному відхиленню з різницею у тому, що усі відхилення варіантів від середньої підносять у квадрат:

$$\sigma^2 = \frac{\sum (x - x_{cep})^2}{n} = \frac{\sum (x - x_{cep})^2 f}{\sum f}.$$

Види дисперсій:

 загальна (σ^2) – є результатом впливу усіх факторів, що спричинили варіацію ознаки, як постійних (систематичних), так і випадкових; характеризує варіацію ознаки навколо загальної середньої;

 групова (часткова) (σ_i^2) – є результатом впливу випадкових факторів (усіх крім фактора, покладеного в основу групування), характеризує варіацію ознаки у межах групи навколо групової середньої; узагальнюючою мірою внутрішньогрупової варіації єсередня з груповихдисперсій $(\sigma_i^2)_{cep}$;

 міжгрупова дисперсія (δ^2) – є результатом впливу фактора (постійного), який покладено в основу групування; характеризує відхилення групових середніх від загальної, тобто систематичну варіацію. Між видами дисперсій існує співвідношення – правило складання дисперсій – загальна дисперсія дорівнює сумі середньої з групових дисперсій і міжгрупової дисперсії:

$(\sigma^2) = (\sigma_i^2)_{cep} + \delta^2$,

отже, чим більший внесок однієї з складових у загальну дисперсію, тим сильніший вплив відповідних їм факторів.

Це правило широко використовується при обчислюванні щільності зв′язку, у дисперсійному аналізі та в ряді інших випадків як для кількісних, так і для якісних ознак (дисперсія альтернативної ознаки дорівнює добутку частки одиниць, які мають цю ознаку, на частку одиниць, що її не мають). Наприклад, можемо дослідити, як впливає наявність спеціальної освіти на плинність кадрів (стаж), якість виконання роботи залежно від кваліфікації та інших умов, якість зберігання продуктів харчування залежно від терміну їх зберігання тощо. Тобто визначити який з факторів, що впливають на варіацію ознаки має більш (або менш) суттєве значення.

Середнє квадратичне (стандартне) відхилення (σ) – це квадратний корінь з дисперсії:

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

Чим менше стандартне відхилення, тим повніше середня арифметична характеризує усю досліджувану сукупність, тим більш однорідною вона є.

Абсолютні показники варіації завжди виражаються у одиницях виміру ознаки.

Порівнюючи варіацію різних ознак в одній сукупності чи варіацію однієї ознаки в різних сукупностях, недостатньо виявити абсолютні величини варіації, оскільки вони залежать і від розміру варіації, і від рівня ознаки. Щоб забезпечити порівняння обчислюють відносні показники варіації, значення яких залежить від того, який саме абсолютний показник варіації використовується.

Базою порівняння служить середня арифметична величина ознаки (іноді медіана).

До **відносних** показників належать:

 коефіцієнт осциляції (V_R) – характеризує відносне коливання крайніх значень ознаки навколо середньої:

$$V_R = \frac{R}{x_{cep}} \cdot 100\%\ ;$$

- **відносне лінійне відхилення** (V_d):

$$V_d = \frac{d}{x_{сер}} \cdot 100\% = \frac{d}{Me} \cdot 100\% ;$$

- **коефіцієнт варіації** (V) – найчастіше застосовується:

$$V = \frac{\sigma}{x_{сер}} \cdot 100\% ,$$

при $V \leq 33\%$ сукупність є однорідною, а середня є типовою та надійною її характеристикою.