

## Лекція 2.

### МЕТОДИ ЗВЕДЕННЯ, ГРУПУВАННЯ І ЗОБРАЖЕННЯ СТАТИСТИЧНИХ ДАНИХ

#### План

1. Завдання статистичного зведення. Значення групування в аналізі інформації. Види групувань.
2. Принципи вибору груповальної ознаки. Способи утворення груп. Вторинне групування.
3. Види і функції статистичних таблиць та графіків. Загальні правила їх побудови, особливості аналізу.

1. Ефективність використання зібраних даних істотно залежить від якості їх оброблення. Систематизація одиничних фактів, яка дає змогу знайти узагальнюючі показники, що описують усю досліджувану сукупність, а також її окремі частини називається **статистичним зведенням**. Основне завдання зведення - виявлення і вивчення типових рис і закономірностей, властивих досліджуваному явищу в цілому.

За складністю побудови зведення може бути:

простим – загальний підсумок усіх одиниць сукупності (виконують без розбиття досліджуваної сукупності на частини за певними ознаками), наприклад, чисельність студентів Інституту журналістики у поточному навчальному році; застосовують, переважно, при оперативних дослідженнях;

складним (груповим) – розподіл одиниць сукупності на групи за декількома істотними ознаками методом групування, підрахунок групових та загальних підсумків, наприклад, чисельність студентів ІЖ по відділеннях, курсах, всього по курсу.

Метод групування, який покладено в основу зведення первинного статистичного матеріалу не тільки забезпечує систематизацію та узагальнення результатів спостереження, а є базою для застосування інших методів статистичного аналізу, таких як порівняльний аналіз, аналіз причин групових відмінностей, вивчення взаємозв'язків між факторами та визначення сили їх впливу на результативні показники (дисперсійний аналіз).

**Групування** – це процес утворення однорідних за найістотнішими ознаками груп. Групування може здійснюватись як за якісною, так і за кількісною ознакою, що має варіацію. Ознаки, за якими здійснюється групування називають груповальними (основою групування). На групування у статистичному аналізі покладаються певні функції:

визначення типів явищ, виокремлення однорідних груп і підгруп;

дослідження структури сукупності та структурних зрушень;

вивчення взаємозв'язків і залежностей між ознаками;

визначення меж і можливостей застосування інших статистичних методів (середніх, кореляційних, регресних).

Згідно з цими функціями групування поділяються на три види:

типологічні,  
структурні,  
аналітичні.

Типологічне – це групування, при якому складна сукупність, що містить неоднорідні одиниці, розкладається на ряд якісно однорідних груп за типологічною ознакою (соціальний склад населення, форма власності, рівень освіти тощо). Основне завдання такого групування – ідентифікація типів, класів.

Структурне групування використовується для характеристики співвідношень між окремими частинами якісно однорідної сукупності, вивчення їх зміни у різні періоди спостереження. Наприклад: структура населення за статтю, віком, місцем проживання,

освітою, сімейним складом; організацій – за кількістю працюючих, кількістю робочих місць тощо. Від типологічного відрізняється метою дослідження. Різновидом структурних групувань є ряди розподілу (ряд упорядкованих одиниць сукупності по групах за певною ознакою). Ці ряди є базою поглибленого аналізу закономірностей розподілу і будуть детальніше розглянуті у наступних розділах.

Аналітичне групування служить для дослідження взаємозв'язку і закономірностей між факторними (незалежними) та результатною (залежною) ознаками. Його проводять щонайменше за двома ознаками, одна з яких відображає причину (факторна), інша – наслідок (результатна). Групування здійснюється за факторною ознакою! Воно полягає в діленні ознак на однакову кількість груп, кожна з яких характеризується середнім значенням результатної ознаки, наприклад, колір очей (блакитні, сірі, зелені) – колір волосся (біле, русяве, чорне) чи стаж роботи (5, 10, 15 років) – кваліфікація (достатня, висока, профі). Умовою застосування аналітичного групування є якісна однорідність досліджуваної сукупності! Поділення групувань на види досить умовне і на практиці не завжди можна провести чітку межу між ними. Адже часто групування універсальні: одночасно виділяються типи, визначається склад сукупності та виявляється взаємозв'язок між ознаками.

**2. В застосуванні методу групування** первісним є розуміння природи явищ, чітке визначення істотних ознак та інтервалів групування таким чином, щоб утворені групи складались з подібних одиниць сукупності, а окремі групи суттєво відрізнялись одна від одної. Тому у кожному конкретному дослідженні під час розподілу сукупності на групи необхідно вирішити такі питання:

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

Вибір ознаки групування – одне з найсерйозніших питань теорії групування. Основні принципи вибору такі:

- в основу групування необхідно покласти найбільш суттєві ознаки, що відповідають задачам дослідження;
- слід урахувати конкретні історичні і територіальні умови (умови місця і часу), в яких знаходиться явище, зміна цих умов може призвести і до зміни групувальної ознаки;
- якщо вивчається явище, на яке впливає декілька різних факторів, необхідно проводити групування за кількома ознаками у комбінації.

При цьому слід зважати на те, що групування з великою кількістю груп стає ненаочним. Тому на практиці використовують складні групування не більш ніж за трьома ознаками. Число груп залежить від задач дослідження і виду ознаки, покладеної в основу групування, обсягу сукупності, ступеня варіації групувальної ознаки. Якщо групування проводять за атрибутивною ознакою, то кількість груп, на які поділяється сукупність, визначається кількістю різновидів (варіант) цих ознак, наприклад, групування населення за статтю (альтернативна ознака) передбачає 2 групи, групування населення Києва по районах – 10 груп тощо).

Для групування за кількісною ознакою необхідно встановити число відокремлених груп і розмір інтервалу. Важливою вимогою тут є вибір такої кількості груп і значення інтервалу, які б давали змогу рівномірно розподілити всі одиниці сукупності в розрізі груп та забезпечили їх представництво і якісну однорідність. Кількість груп і величина інтервалу пов'язані між собою: чим більше створено груп, тим менше інтервал, і навпаки. Зверніть увагу, якщо брати занадто широкі інтервали, то групи будуть складатись з одиниць, що якісно відрізняються, вони будуть неоднорідними. Також не можна поділяти невелику за обсягом сукупність на значну кількість груп. Утворення малочисельних груп розпоршує

матеріал і характеристики цих груп можуть бути недостатньо типовими. У цих випадках не можна буде виявити масові закономірності та отримати адекватну характеристику досліджуваного явища. Тому головну роль при визначенні числа груп мають відігравати розуміння дослідником природи явища, певний досвід та інтуїція. Інтервали бувають рівні (однакові), нерівні (неоднакові), відкриті (від 10 осіб; понад 51 особу) і замкнені ([2-7], [8-14], [14-20]).

При типологічному групуванні найчастіше вживають нерівні *інтервали*. Межа інтервалу розглядається як умовна межа переходу кількості в нову якість. Наприклад, при групуванні населення за ознакою працездатності використовуються вікові групи: до 16 років – особи до працездатного віку; 16-59 років – працездатного; 60 і більше – пенсійний вік. Такі інтервали формування груп називають спеціалізованими.

У структурних, аналітичних групуваннях та при обмеженій кількості одиниць сукупності дані розміщують впорядковано за зростанням ознаки, а кожна група вміщує їх однакову кількість. Це виключає утворення малочисельних груп, особливо в аналітичних групуваннях, для яких коректний вибір інтервалів має важливе значення, оскільки невдалий або упереджений підхід може спотворити дійсний характер зв'язку між явищами. Іноді виникає потреба у перегрупуванні вже побудованого групування для забезпечення порівнянності структур двох сукупностей за однією і тією самою ознакою. Тобто коли зведені дані не можна безпосередньо зіставити або вони не задовольняють дослідників щодо кількості груп. Постає необхідність утворення нових груп на основі тих, що є.

Результат перегрупування називають вторинним групуванням. Розрізняють два способи вторинного групування:

- об'єднання (збільшення) інтервалів – найбільш простий і поширений, використовують коли межі нових і старих інтервалів збігаються, показники об'єднувальних інтервалів просто підсумовуються;
- розбиття інтервалів первинного групування – застосовують при необхідності визначення долі (частки) одиниць сукупності, яка перейде зі старих груп у нові.

**3. Статистичні дані за формою зображення** можуть подаватися у вигляді числових масивів, таблиць чи графіків. Найефективнішою формою викладення результатів спостережуваних даних є статистичні таблиці. Вони є невіддільним елементом зведення та групування. Статистичну таблицю від інших табличних форм відрізняє:

- вона вміщує результати підрахунку емпіричних (спостережуваних) даних (статистичні показники);
- є підсумком зведення первинної інформації.

Таким чином *статистична таблиця* – це таблиця, що вміщує зведену (загальну, підсумкову) числову характеристику досліджуваної сукупності по одній чи декілька істотним ознакам, які взаємопов'язані логікою аналізу. Статистичну таблицю тлумачать як форму логічного статистичного речення, що містить підмет і присудок.

*Підметом* таблиці є об'єкт дослідження (те, про що йдеться в таблиці): перелік елементів сукупності, їх групи, окремі територіальні одиниці або часові інтервали тощо. Звичайно підмет розміщують зліва, у назві рядків.

*Присудок* таблиці – це система показників, що характеризують підмет в кількісній формі. Присудок формує верхні заголовки і складає зміст граф з логічно послідовним розміщенням показників зліва направо.

**Розміщення підмета і присудка може мінятися місцями**, що залежить від досягнення кожним дослідником окремо найбільш повного й кращого способу читання та аналізу вихідної інформації про досліджувану сукупність.

Складена, але не заповнена цифрами таблиця є макетом статистичної таблиці:

**Назва таблиці\*** (загальний заголовок)

<u>Підмет</u> Зміст рядків	<u>Присудок</u> Назви граф (верхні заголовки)			
A	1	2	...	n
Назви рядків (бічні заголовки)				
Підсумковий рядок				Підсумкова графа

\*Примітка до таблиці.

Види статистичних таблиць розрізняють за будовою підмета і присудка:

- прості (монографічні, перелікові, хронологічні) – підмет не ділиться на групи, наприклад, динаміка видання книжок видавництвами України; мають інформативну функцію;
- групові – підмет поділяється на групи за однією ознакою, наприклад, розподіл періодичних видань за мовами (використовують з метою зіставлення узагальнюючих показників по групах);
- комбінаційні – підмет ділиться на групи за декількома ознаками, наприклад, розподіл населення водночас за місцем проживання та статтю (дозволяють характеризувати типові групи, зв'язки і відмінності між ознаками);
- залежно від структури присудка розрізняють таблиці простої або складної розробки.

Складаючи статистичну таблицю слід дотримуватись таких загальних правил:

- таблиця повинна бути невеликою за розміром, містити лише ті дані, що безпосередньо характеризують об'єкт дослідження; таку таблицю простіше читати і аналізувати;
- цифровий матеріал необхідно викладати таким чином, щоб під час аналізу таблиці суть явища розкривалася читанням рядків зліва направо та зверху вниз;
- назва таблиці, заголовки рядків і граф мають бути чіткими, короткими, лаконічними, зрозумілими, без скорочень; у назві вказується зміст, місце, час, одиниці виміру(якщо вони єдині для усієї сукупності);

У групових і комбінаційних таблицях завжди необхідно давати підсумкові графи та рядки! Не обов'язково це лише для аналітичних групувань;

В статистичній таблиці не повинно бути пустих клітин; відсутність значень ознаки позначають знаком тире, відсутність відомостей – трьома крапками.

Читають та аналізують статистичні таблиці у певній послідовності. Починають з загальних підсумків, потім переходять до вивчення даних окремих рядків та граф – від абсолютних до пов'язаних з ними відносних величин. В аналізі даних поряд із статистичними таблицями використовуються й інші види таблиць, одним з яких є таблиці співзалежності. Найпоширенішими є таблиці 2x2, наприклад, пари щеплений /не щеплений – захворів/не захворів. Цей вид таблиць найчастіше використовують при вивченні соціальних явищ, у медичних обстеженнях. Статистичні дослідження не обмежуються лише таблицями. Широкого застосування набув і графічний спосіб зображення статистичних величин.

**Статистичний графік** – це умовне зображення кількісних характеристик та їх співвідношень за допомогою геометричних образів (крапок, ліній, знакових систем, геометричних фігур, малюнків) або географічних картосхем.

**За допомогою графіків** легко виявити закономірності розвитку явища, встановити існуючі взаємозв'язки. Графіки широко використовуються для вивчення структури, змін у часі і розміщення у просторі. В них більш виразно проявляються порівнювані характеристики (тобто швидкість зміни різних показників) і чітко видно основні тенденції розвитку явищ. Графіки також є ефективним засобом для наукового узагальнення статистичної інформації та її популяризації!

Для побудови статистичного графіка потрібно знати, з якою метою складається графік та оволодіти методикою графічних зображень.

#### **Основні елементи графіка:**

- поле графіка – частина площини, де зображено графік, має визначені розміри, які залежать від призначення графіка;
- графічний образ (основа графіка) – вигляд графіка, сукупність геометричних знаків (точок, ліній, фігур), які відображують статистичні показники; кожній зміні показника відповідає пропорційна зміна графічного образу;
- просторові орієнтири задають у вигляді системи координатних сіток, найбільш поширеною є система прямокутних координат (вісі X та Y), але застосовують також і полярні координати, які необхідні для наочного зображення циклічного руху в часі;
- масштабні орієнтири визначаються масштабом та системою масштабних шкал (звичайних арифметичних і логарифмічних); найкраще співвідношення масштабу по осі абсцис і ординат становить 1,41:1 (“золотий перетин”);
- експлікація графіка – словесний опис його змісту, складається з заголовка над чи під графіком, підписів уздовж масштабних шкал та пояснень до окремих частин графіка.

Статистичні графіки класифікують за функціональним і загальним призначенням, видами, формами і типами основних елементів. Найпоширенішими є діаграми (лінійні, площинні, об'ємні і фігурні). Тому термін “діаграма” часто ототожнюють з терміном “статистичний графік”.

#### **Діаграми, які застосовують для порівняння статистичних величин:**

- стовпчикові та стрічкові – основи стовпчиків розташовують відповідно на вісі абсцис чи ординат; зверніть увагу! – шкала, за якою встановлюють розмір стовпчика (стрічки), повинна починатися з нуля, бути неперервною, тобто охоплювати усі числа статистичного ряду; розрив шкали і відповідно стовпчиків (стрічок) не допускається, це приводить до викривленого графічного представлення статистичного матеріалу;
- квадратні, кругові, прямокутні (знаки Варзара) – виражають величину зображуваного явища розміром своєї площі;
- фігурні – досліджувана статистична величина зображується окремою кількістю однакових за розмірами фігур-знаків, послідовно розташованих на малюнку; використовуються для популяризації статистичних даних та реклами. Для графічного зображення структури використовують діаграми:
- стовпчикові, стрічкові, кругові, секторні (найпоширеніші) – відображують співвідношення різних частин сукупності пропорційно до відповідних абсолютних чи відносних показників; якщо сукупність поділяється більше ніж на 5-6 частин, або між

окремими частинами велика різниця, секторні діаграми втрачають свою виразність, тоді доцільно використовувати стовпчикові чи стрічкові діаграми.

***Діаграми динаміки розвитку явищ:***

- стовпчикові, стрічкові, квадратні, кругові та фігурні – зображують обсяг явища на певну дату чи за відповідний період, добре запам'ятовуються, але не зручні для великого числа рівнів, тому, що громіздкі;
- лінійні - відображують зміни явищ у часі; вісь абсцис визначає періоди часу (дні, місяці, роки), ордината – рівні (показники) динаміки; шкала ординати починається з нуля, може перериватися у випадку, коли мінімальне значення рівня динаміки значно перевищує нуль; необхідним є забезпечення рівноваги масштабних шкал між вісями координат для усунення викривлення зображення (розтягненні чи різкі коливання) і його невідповідності фактам;
- радіальні (спіральні та замкнені) – будують в полярних координатах; відображують ритмічний (циклічний) рух у часі; найчастіше ці діаграми застосовують для ілюстрації сезонних коливань, що притаманні процесам, які пов'язані з с/г, туризмом, транспортом тощо.

Для характеристики розташування і поширення явищ у просторі використовують ***статистичні карти***: картограму (схематична статистична географічна карта, в якій на визначеній території за допомогою штрихування, зафарбування, нанесення точок, зображується порівнювана інтенсивність будь-якого явища та закономірність його розподілу), картодіаграму (поєднання діаграм з географічною картою), центрограму (контурна карта з цифровими таблицями-списками, у яких вміщується інформація про історико-географічний розвиток досліджуваного явища; застосовується при вивченні міграцій населення, переміщення центрів виробництва різних галузей тощо).

Статистичні графіки надзвичайно різноманітні. Крім розглянутих існують спеціальні графіки зображення варіаційних рядів (гістограма, полігон, огіва, кумулята), взаємозв'язку і взаємозалежності (діаграми розсіювання, матричні графіки, обличчя Чернова тощо). Ефективними у статистичних дослідженнях є комп'ютерні засоби побудови графічних зображень у пакетах STATISTICA, SPSS, STATGRAPHICS, EXCEL та інших.