

# ВСТУП ДО СТАТИСТИКИ

*Математична статистика, яка виникла спочатку для цілей демографії і страхування, перетворюється на один із методів кількісного дослідження явищ природи, технічних процесів, економіки і лінгвістики.*

Б. В. Гнеденко

## § 48. Про статистику та її методи

Термін «статистика» походить від латинського слова *status* — стан, становище.

Статистика — наука, що збирає, обробляє і вивчає різні дані, пов'язані з масовими явищами, процесами і подіями.

Предметом вивчення статистики є, зокрема, кількісний бік масових суспільних явищ і процесів у зв'язку з їх якісним боком.

Іноді неточно визначають статистику як «науку збирання даних». Це галузь прикладної математики.

Математична статистика — розділ математики, присвячений математичним методам систематизації, обробки й використання статистичних даних для наукових і практичних висновків.

Статистика виникла з практичних потреб людей, їхньої господарської діяльності, необхідності обліку земельних угідь, майна, кількості населення, вивчення його занять, вікового складу тощо. Цікаво, що в Англії у XVII ст., коли статистичне вивчення поширилося на явища суспільного життя, людей, які займалися цими питаннями, називали «політичними арифмети-ками». Одним з головних представників «політичних арифметиків» Англії був **В. Петті** (1625—1687).

Статистику розділяють на описову і пояснювальну.

**Описова статистика** займається добором кількісної інформації, необхідної (або цікавої) для різних людей. Такою є

спортивна інформація, відомості про середній рівень заробітної плати в державі, середньорічну температуру в певному регіоні тощо.

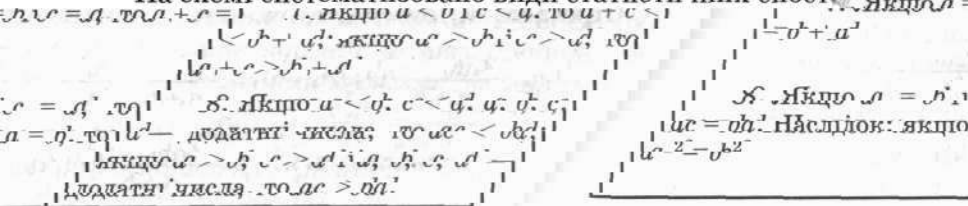
Великі масиви даних, перш ніж вони вивчатимуться людиною, мають узагальнюватися або згортатися. Саме це робить описова статистика, яка описує, узагальнює або зводить до бажаного виду властивості масивів даних.

За допомогою **пояснювальної статистики** з добутих статистичних результатів роблять певні висновки, складають прогнози. Предметом вивчення статистики є такі об'єкти, як кількість і склад населення, трудові ресурси суспільства (їх розподіл і використання), національне багатство, виробництво і розподіл суспільного продукту і національного прибутку, матеріальний достаток населення, освіта, культура, охорона здоров'я, показники статистики органів державного управління і громадських організацій.

У процесі статистичного дослідження застосовують особливі прийоми вивчення, які в сукупності утворюють **статистичний метод**. Складовими елементами статистичного методу є масове спостереження, статистичне зведення, групування, обчислення середніх величин та індексів, побудова графіків.

**Статистичне спостереження** — перший етап статистичного дослідження.

На схемі систематизовано види статистичних спостережень:



## *Спостереження за часовою ознакою*

**Поточне** спостереження передбачає систематичне вивчення змін, що відбуваються в певній сукупності в міру їх надходження. Наприклад, щоденний облік відвідування в певному класі.

**Періодичне** спостереження проводиться через строго визначені інтервали часу — місяць, квартал, рік. Наприклад, облік успішності в школі за чверть, навчальний рік. У сільському господарстві — щорічний перепис худоби тощо.

**Одиничне** спостереження, як правило, проводиться в разі потреби в якийсь певний момент за особливим завданням. Наприклад, перепис житлового фонду в певному районі міста.

## *Спостереження за способом організації*

**Звітне** спостереження — вивчення певних явищ і процесів на основі статистичних відомостей, які містяться у різноманітній звітності.

Під час **експедиційного** спостереження спеціальні люди — обліковці — обходять закріплені за ними ділянки території і здійснюють там реєстрацію. Прикладом є перепис населення.

Спостереження **самообчисленням** полягає в тому, що представники статистичних органів роздають населенню або установам статистичні формуляри, які періодично, через певні інтервали часу, збирають і обробляють для отримання узагальнених даних.

## *Спостереження за ступенем повноти охоплення одиниць*

Цей вид статистичного спостереження безпосередньо пов'язаний з математикою, а тому після загального огляду методів, якими користується статистика, ми повернемося до нього.

**Суцільним** є спостереження, в якому реєструється ознака всіх без винятку одиниць, що входять у сукупність, яка вивчається. Суцільне спостереження застосовується, наприклад, під час перепису населення.

**Несуцільним** називають такий вид спостереження, під час якого реєструють ознаки лише частини одиниць досліджуваної сукупності, і за частиною роблять висновок про всю сукупність.

Видами несуцільного спостереження є: вибіркове спостереження, спостереження основного масиву, анкетне спостереження і монографічний опис.

Найпоширенішим з видів несуцільного спостереження є **вибіркове** спостереження. Всю сукупність, з якої роблять відбір одиниць спостереження, називають **генеральною**. Сукупність одиниць, відібраних для вибіркового спостереження, називається **вибірковою**.

Під час вибіркового спостереження обстеженню підлягає відібрана певним чином частина одиниць усієї її сукупності, а результати обчислення цієї частини сукупності поширюються на всю сукупність у цілому.

Для того щоб за вибіркою можна було досить впевнено судити про властивості генеральної сукупності, вибірка має бути представницькою (репрезентативною).

**Репрезентативність вибірки означає, що об'єкти вибірки досить добре представляють генеральну сукупність.**

Репрезентативність вибірки забезпечується випадковістю відбору. Останнє означає, що будь-який об'єкт вибірки відібраний випадково, при цьому всі об'єкти мають однакову імовірність потрапити до вибірки. Існує кілька способів відбору, які забезпечують репрезентативність вибірки. Розглянемо основні з них.

Коли об'єкти генеральної сукупності невеликі і знаходяться, наприклад, в ящику, то, перемішавши їх, з ящика беруть по одному об'єкту доти, доки не утвориться вибірка. Але такий відбір неможливий, коли генеральна сукупність складається з досить великих за розмірами об'єктів, наприклад з потужних електродвигунів, або з таких об'єктів, які під час перемішування псуються, наприклад з електроламп. У такому разі всі об'єкти генеральної сукупності нумерують, кожний номер записують на окрему картку, а картки ретельно перемішують. З отриманої сукупності карток вибирають одну навмання. Об'єкт, номер якого відповідає номеру картки, вважається таким, що потрапив до вибірки. Таке виїмання карток продовжують доти, доки не утворять потрібну вибірку. При цьому можуть бути два варіанти вибірки: випадкова вибірка з поверненням відібраних об'єктів у генеральну сукупність і без такого повернення. Якщо обсяг генеральної сукупності великий, то різниця між вибірками з поверненням і без повернення незначна, якщо обсяг невеликий, то різниця буде істотною.

Для генеральних сукупностей великого обсягу застосування карток для утворення вибірок також ускладнюється через написання великої кількості карток з номерами і труднощі перемішування їх великої кількості. У такому разі використовують таблицю випадкових чисел (табл. 16).

Нехай, наприклад, треба утворити вибірку з генеральної сукупності великого обсягу, якою є виготовлені заводом протягом кварталу електродвигуни. Кожний двигун має чотиризначний номер. Якщо у вибірці має бути 20 двигунів, то з таблиці навмання вибирають 20 чисел (можна й підряд) і дви-

гуни з відповідними номерами відправляють на контроль. Якщо не звертати увагу на те, що деякі номери можуть повторюватися і, отже, деякі двигуни мають обстежуватися двічі, то вибірка буде з поверненням. Якщо треба утворити випадкову вибірку без повернення, то при відборі випадкових чисел з таблиці треба пропустити число, яке випало другий раз.

Після того як утворена вибірка, всі її об'єкти обстежуються щодо властивості, яка цікавить, і в результаті дістають дані, що спостерігаються.

Звичайно, одержані дані є множиною розташованих як завгодно чисел. У такій множині важко помітити яку-небудь закономірність їх варіювання. Щоб помітити таку закономірність (якщо вона є), дослідні дані піддають обробці. Наприклад, обробка результатів спостережень полягає в тому, що отримані наслідки спостережень розташовують у порядку неспадання. Така обробка називається **ранжуванням** дослідних даних.

Після ранжування дослідні дані легко об'єднати в групи так, щоб у кожній окремій групі дані були однаковими.

**Приклад.** На телефонній станції проводилися спостереження над числом  $X$  неправильних з'єднань за хвилину. Спостереження протягом години дали такі результати: 3; 1; 3; 1; 4; 2; 2; 4; 0; 3; 0; 2; 2; 0; 2; 1; 4; 3; 3; 1; 4; 2; 2; 1; 1; 2; 1; 0; 3; 4; 1; 3; 2; 7; 2; 0; 0; 1; 3; 3; 1; 2; 4; 2; 0; 2; 3; 1; 2; 5; 1; 1; 0; 1; 1; 2; 2; 1; 1; 5.

Розташуємо ці дані в порядку неспадання і згрупуємо їх. Дістанемо такий *ранжирований* ряд даних спостережень: 0; 0; 0; 0; 0; 0; 0; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 4; 4; 4; 4; 4; 5; 5; 7.

Отримали 7 груп різних спостережень. Кожне таке значення, що належить різним групам, називають **варіантом**, а зміну цього значення — **варіюванням**.

Таблиця 16

1534	7106	2836	7873	5574	7545
6128	8993	4102	2551	0330	2358
6047	8556	8644	9343	9297	6751
0806	5201	5705	7355	1448	9562
9915	8274	4525	5695	5752	9630
2882	7158	4341	3463	1178	5786
9213	1223	4388	9760	6691	6861
8410	9836	3899	3683	1253	1683
9974	2362	2103	4326	3825	9079
3402	8162	8226	0782	3364	7871

Чисельність окремої групи згрупованого ряду даних називають **частотою**, або **вагою** відповідного варіанта і позначають  $m_i$ , де  $i$  — індекс варіанта. Наприклад, для варіанта  $x_4$  частота  $m_4 = 10$  (див. попередній приклад).

Відношення частоти даного варіанта до загальної суми частот усіх варіантів називається **частотою** варіанта і позначається  $P_i$ , де  $i$  — індекс варіанта.

**Дискретним варіаційним рядом** розподілу називається ранжирована сукупність варіантів  $x_i$  з відповідними їм частотами  $m_i$  або частотами  $P_i^*$ .

Результати спостережень для наведеного прикладу (робота телефонної станції) зручно подати у вигляді таблиці (табл. 17).

У практиці статистично розрізняють три способи відбору одиниць сукупності, яка вивчається.

1) **Випадковий відбір** — усі одиниці сукупності мають однакову можливість потрапити до вибірки. Відбір здійснюється з усієї сукупності жеребкуванням.

2) **Механічний відбір** — одиниці спостереження відбирають у певному порядку. Наприклад, під час механічного відбору при вивченні якості продукції береться на вибір кожна десята або двадцята деталь.

3) **Типовий відбір** — усю масу одиниць, що вивчаються, розчленовують на дрібніші однорідні групи і здійснюють наступний відбір одиниць — «представників» кожної групи у випадковому або механічному порядку. Наприклад, під час вивчення бюджету сімей їх попередньо поділяють на групи за соціальним станом і рівнем прибутків.

Поширюючи дані вибіркового спостереження на всю генеральну сукупність, застосовують два способи поширених даних: *спосіб прямого перерахунку і спосіб поправочних коефіцієнтів*. Перший спосіб полягає в тому, що результати вибіркового спостереження приймають і для генеральної сукупності.

Другий спосіб застосовують під час уточнення результатів суцільного спостереження. Суть його полягає в тому, що дані вибіркового обстеження зіставляють з даними суцільного спостереження і визначають коефіцієнт розходження.

Таблиця 17

Індекс	$i$	1	2	3	4	5	6	7
Число неправильних з'єднань за хвилину	$x_i$	0	1	2	3	4	5	7
Частота	$m_i$	8	17	16	10	6	2	1
Частість	$P_i^*$	$\frac{8}{60}$	$\frac{17}{60}$	$\frac{16}{60}$	$\frac{10}{60}$	$\frac{6}{60}$	$\frac{2}{60}$	$\frac{1}{60}$

*Спостереження основного масиву* передбачає облік лише частини одиниць певної сукупності, яка має переважну питому вагу в обсязі досліджуваного об'єкта. Наприклад, вивчення цін на ринках, які мають найбільшу питому вагу в оборотах торгівлі.

*Анкетне спостереження* не надійне (частина анкет не повертається). Воно використовується переважно транспортними організаціями і органами зв'язку для вивчення ефективності обслуговування населення.

*Монографічний опис* полягає в тому, що для обстеження береться один об'єкт, який докладно вивчають (здебільшого це має місце під час вивчення і поширення передового досвіду).

Важливу частину статистичних методів становлять планування і аналіз експериментів, спрямованих на виявлення і перевірку причинних зв'язків між змінними. Планування експериментів спирається в основному на поєднання теорії імовірностей з елементарною логікою.

Статистичні дослідження проводяться за таким планом:

1) формулюється завдання дослідження і визначаються обсяг, місце і час потрібної вибірки;

2) здійснюються збирання необхідних даних та їх наочне подання;

3) проводиться обробка зібраного статистичного матеріалу.

На першому етапі важливо чітко визначити мету дослідження, встановити, які об'єкти вивчатимуться і в якій кількості (обсяг вибірки). Необхідно встановити, які ознаки при цьому братимуться до уваги, які кількісні та якісні характеристики об'єктів слід оцінити.

На другому етапі використовують різні методи збирання даних: спостереження, порівняння, усне і письмове анкетування, систематизація даних.

На третьому етапі (частково на другому) результати статистичних досліджень піддають обробці і оформляють у вигляді таблиць, діаграм, графіків. За результатами виконаної роботи роблять певні висновки.

Статистичні висновки роблять від окремих властивостей вибірок до часткових властивостей сукупності; опис властивостей як вибірок, так і сукупностей здійснюється за допомогою методів описової статистики.

Описова статистика включає в себе табулювання (складання таблиць), подання і опис сукупності даних. Ці дані можуть бути або кількісними, як наприклад, вимірювання зросту і маси людини, або якісними, наприклад вивчення певних явищ, в яких принципове значення має стать.

## ЗАПИТАННЯ І ЗАВДАННЯ ДЛЯ ПОВТОРЕННЯ

1. Чим займається наука «статистика»?
2. У чому полягає відмінність між описовою і пояснювальною статистикою?
3. Які ви знаєте елементи статистичного методу дослідження?
4. Перелічити види статистичних спостережень. Охарактеризувати кожний із них.
5. Пояснити, що таке вибіркове спостереження.
6. Що таке генеральна сукупність? Що таке вибірка?
7. Які ви знаєте способи відбору одиниці сукупності, що вивчається?
8. За яким планом здійснюються статистичні дослідження?

## § 49. Статистичні таблиці

Статистичні таблиці мають підмет і присудок.

Статистичний підмет — це та сукупність, про яку йдеться в таблиці.

Як правило, підмет розміщується у лівій частині таблиці.

Статистичний присудок — це ті ознаки або показники, які характеризують статистичний підмет.

Підмет розміщується в заголовках граф — стовпців.

За структурою підмета статистичні таблиці поділяються на прості, групові і комбінаційні.

Прості — підмет задається переліком окремих об'єктів (назви підприємств, міст, країн і т. п.).

Групові — в підметі одиниці сукупності групуються за однією якоюсь ознакою.

Комбінаційні — в підметі одиниці групуються за двома і більше ознаками, пов'язаними між собою.

Наведемо приклади таблиць (табл. 18—21).

Якщо групування здійснено за інтервалами зміни ознаки, то таке групування називають **інтервальним**.

Наприклад, вибіркове вимірювання врожайності жита на площі 1200 га дало результати, які ми подаємо за допомогою інтервального групування (табл. 22).

Подавши результат групування рядом варіант або інтервалів варіації, розміщених у зростаючій послідовності, і низкою відповідних частот, дістанемо **варіаційний ряд** (відповідно дискретний або інтервальний).