

Предмет та задачі математичної статистики

Означення. **Статистика** – наука, яка займається збиранням, обробкою та вивченням різноманітних даних, пов'язаних з масовими явищами, процесами та подіями.

Статистика використовується в економіці, політиці та різних експериментальних дослідженнях, зокрема психолого-педагогічних.

Математична статистика виникла у XVII ст. і розвивалася паралельно з теорією ймовірностей. У XX ст. великий внесок зробили В.І. Романовський, А.Н. Колмогоров, Стьюдент, Е. Пірсон, А.В. Скороходов.

Предметом вивчення статистики можуть бути такі об'єкти: кількість і склад населення, національні багатства їх виробництво та розподіл, освіта, культура, охорона здоров'я.

У процесі статистичного дослідження використовуються особливі прийоми, які в сукупності і утворюють **статистичний метод**. Найчастіше статистичну інформацію збирають за допомогою спостереження: перепис, опитування, облік. Часто статистичні відомості про конкретні об'єкти, що входять у велику сукупність, досліджують в результаті аналізу тільки її частини – **вибірки**.

Означення. **Вибірка** – сукупність об'єктів, які випадково відібрані із усієї сукупності об'єктів, що називається **генеральною сукупністю**.

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

Наприклад, щоб дізнатися про найпоширеніші розміри чоловічого взуття досить опитати кілька десятків чоловік. Припустимо, що опитали 60 осіб. Одержаний результат в табл. 1.

Таблиця 1

Розмір взуття	24	25	26	27	28	29	30	31	32	33	34	35
К-ть осіб	1	2	3	7	10	9	8	8	6	4	1	1

+Проаналізувавши таку таблицю, роблять загальні висновки, наприклад, що найбільше потрібно виготовити взуття 28 розміру. Це наближені відомості, але на практиці таких наближень буває достатньо.

Дослідження в будь-якій області, в тому числі і в психології, передбачає отримання результатів – зазвичай у вигляді числових масивів. Але цього недостатньо. Навіть об'єктивно і коректно зібрані дані ніпрощо не говорять. Досліднику необхідне вміння організувати їх, обробити і проінтерпретувати, що неможливо зробити без застосування математичної статистики. Саме математичним аналізом різних вибірок і займається математична статистика. Основною задачею математичної статистики є розробка ефективних методів вивчення великої сукупності об'єктів на основі порівняно невеликих вибірок.

Однак для того, щоб за вибіркою можна було досить впевнено судити вона повинна бути **представницькою** або **репрезентативною**.

Означення. **Репрезентативність вибірки** означає, що об'єкти вибірки досить добре представляють генеральну сукупність, забезпечується випадковість їх вибору.

Це означає, що будь-який об'єкт вибірки відібраний випадково, при цьому всі об'єкти мають однакову ймовірність потрапити до вибірки.

Види та способи відбору

1. Відбір, який не потребує розділення генеральної сукупності на частини називається **простим**. Розрізняють *простий випадковий безповторний відбір* (відібрані об'єкти після дослідження не повертаються до генеральної сукупності) та *простий випадковий повторний відбір* (відібрані об'єкти після дослідження повертаються до генеральної сукупності).
2. **Механічний відбір**. Відбір, при якому генеральна сукупність механічно поділяється на стільки частин скільки має бути об'єктів у вибірці. Розрізняють *типовий відбір* (коли об'єкти вибираються не з усієї генеральної сукупності, а лише з її типових частин) та *серійний відбір* (при якому об'єкти генеральної сукупності відбираються не по-одному, а серіями які досліджують).

Після того, як утворилася вибірка всі її елементи обстежують щодо властивості, яка цікавить і в результаті дістають дані, що спостерігаються. Поширюючи дані вибіркового спостереження на всю генеральну сукупність застосовують два способи:

1. Спосіб прямого перерахунку, який полягає в тому, що результат вибіркового спостереження приймають і для генеральної сукупності.
2. Спосіб поправочних коефіцієнтів. Суть якого в тому, що дані вибіркового спостереження зіставляють з даними спеціального спостереження і визначають коефіцієнт розходження.

Означення. Число об'єктів усієї сукупності називають **обсягом сукупності**.

Означення. Число об'єктів із сукупності, які потрапили у вибірку називають **об'ємом вибірки**.

Означення. Кожний елемент вибірки називають **варіантою**.

Наприклад, дано вибірку 4,3,7,9,6,8,2,6,1,7,7,3,2,5, тоді об'єм вибірки дорівнює 14.

Якщо значення варіант розташувати у порядку зростання або скоріше неспадання (1,2,2,3,3,4,5,6,6,7,7,7,8,9), то здійснимо ранжування досліджуваних даних.

Упорядкована таким чином вибірка називається **варіаційним рядом**.

Для зручності складають частотну таблицю

1	2	3	4	5	6	7	8	9
1	2	2	1	1	2	3	1	1

Означення. Різниця між крайніми значеннями варіаційного ряду називається **розмахом вибірки** $R = 9 - 1 = 8$

Первинна обробка даних

Статистичне дослідження проводять за таким планом:

1. Формулюється завдання дослідження і визначається об'єм, місце і час потрібної вибірки.
2. Збираються необхідні дані та наочно подаються (аналітично, таблично, графічно).
3. Проводиться обробка зібраних статистичних даних та формулюється висновок.

Нехай маємо вибірку: $x_1, x_2, x_3, \dots, x_n$. Об'єм вибірки дорівнює n .

Наприклад, ці числа одержані в результаті підрахунку числа неправильних з'єднань за 1 хв. на телефонній станції чи в результаті вимірювання довжин деталей тощо.

Наступні наші дії залежать від кількості у цій вибірці різних чисел.

I випадок. Якщо різних чисел небагато, то маємо справу з дискретною величиною.

II випадок. Якщо багато – з неперервною.

Відповідно до цього розглядаємо два випадки: дискретний і неперервний.

Дискретний випадок

Перший етап обробки вибірки – це складання варіаційного ряду. Його одержують таким чином: серед усіх значень x_i ($i=1,2,3,\dots,n$) відбирають всі різні і розміщують у порядку зростання, отримаємо

$a_1, a_2, a_3, \dots, a_m$ де $m \leq n$, $m=1,2,3, \dots$

$a_1 < a_2 < a_3 < \dots < a_m$

Другий етап обробки вибірки – це складання дискретної таблиці частоти та відносної частоти.

a_i	a_1	a_2	a_3	...	a_m
k_i	k_1	k_2	k_3	...	k_m
$\frac{k_i}{n_i = n}$	$\frac{k_1}{n_1 = n}$	$\frac{k_2}{n_2 = n}$	$\frac{k_3}{n_3 = n}$	...	$\frac{k_m}{n_m = n}$

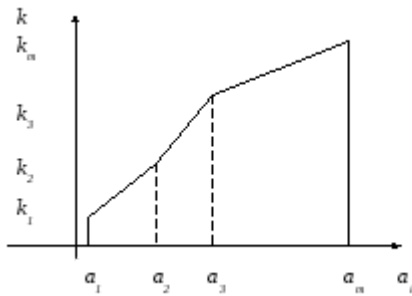
де $\sum k_i = n$, $\sum n_i = 1$,

k_i – число вимірів в яких спостерігалася ознака a_i або **частота**,

$n_i = \frac{k_i}{n}$ – **відносні частоти**.

Третій етап – графічне зображення дискретної таблиці. Це можна зробити за допомогою стовпцевої діаграми та полігону частот.

Для зображення діаграми будують систему координат, на вісі абсцис відкладають значення a_i , а на вісі ординат – відповідні частоти. Потім будують точки з координатами (a_i, k_i) та з них опускають перпендикуляри на вісь абсцис. Отримаємо:



Означення. **Полігоном частот** (відносних частот) називають ламану, відрізки якої з'єднують точки, абсцисами яких є значення варіант a_i , а ординатами – відповідні їм частоти k_i (відносні частоти n_i).

Приклад 1. Результати оцінювання 56 студентів на екзамені з психології представлені такою вибіркою:

3,4,5,4,3,3,5,4,3,5,5,2,3,5,3,5,3,5,4,4,3,3,4,3,4,3,3,5,3,3,4,3,4,3,5,3,4,4,3,5,3,3,5,4,2,5,3,4,2,3,5,4,3,5,3,5.

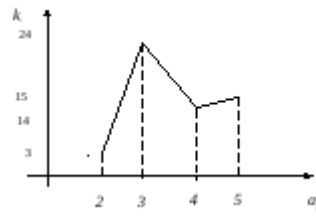
Значення, що складають вибірку є реалізацією випадкової величини – оцінки на екзамені.

Складемо дискретну таблицю частоти

a_i	2	3	4	5
k_i	3	24	14	15
$n_i = \frac{k_i}{n}$	$\frac{3}{56}$	$\frac{24}{56}$	$\frac{14}{56}$	$\frac{15}{56}$

$$\sum k_i = 56, \quad \sum n_i = 1$$

Стовпцева діаграма частоти матиме вигляд



Тренувальні тести по темі
(перейдіть за посиланням нище)

<https://www.iznotest.info/pochatki-teoriyi-ymovirnosti/>