Тема: ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ ТА СТАТИСТИЧНИЙ РОЗПОДІЛ

Мета: ознайомлення з основними поняттями математичної статистики, дискретним та інтервальним статистичними розподілами, числовими характерстиками вибірки

Термін "**статистика**" походить від латинського слова **status** (статус), що означає "визначене положення речей". У даний час термін "*статистика*" вживається в трьох значеннях:

1) як *галузь* практичної діяльності зі збору, обробки та аналізу даних соціально-економічного та іншого масового характеру;

2) як *наука*, що містить теоретичні положення і методи розв'зання практичних задач статистики (**математична статистика**);

3) як *підсумкові* статистичні дані, тобто як результати застосування статистичних методів до початкової статистичної інформації.

Математична статистика – це наука про методи реєстрації, опису й аналізу статистичних експериментальних даних, які одержуються у результаті спостереження масових випадкових явищ.

Задачами математичної статистики, які найчастіше зустрічаються на практиці, є:

1) визначення закону розподілу випадкової величини (або системи випадкових величин) за статистичними початковими даними;

2) перевірка правдоподібності гіпотез, наприклад – чи узгодяться результати експерименту з гіпотезою про те, що дана випадкова величина підпорядкована закону *розподілу* $F(x)$;

3) відшукання невідомих параметрів розподілу випадкової величини або системи випадкових величин (або оцінок цих параметрів).

У математичній статистиці використовується властива цій предметній галузі система категорій і понять. Розглянемо основні з цих понять.

1. Статистичний розподіл

Припустимо, що виникла ситуація, у якій потрібно вивчити сукупність об'єктів (наприклад покупців продукції фірми), оцінюючи деяку ознаку, яка властива кожному з цих об'єктів (наприклад, вік). Якщо аналізована ознака має кількісне значення, то частіше за все цю ознаку ототожнюють із випадковою величиною, а конкретне значення ознаки сприймається як значення випадкової величини.

Вибірковою сукупністю (синоніми: вибірка, простий статистичний ряд) називається сукупність значень x_i тієї самої ознаки X у випадково відібраних n об'єктів.

Звичайно вибіркова сукупність оформляється у вигляді таблиці в два рядки (див. табл. 1). У першому рядку вказують номер об'єкта (або спроби), у другому – значення ознаки цього об'єкта або в цій спробі (значення випадкової величини ***X***). Така вибіркова сукупність далі підлягає обробці, наприклад – побудові статистичної функції розподілу. Вибіркова сукупність є елементом більш загальної – генеральної сукупності.

Таблиця 1

i	1	2	…	k	…	$\boldsymbol{n}$
X	x_1	x_2	…	x_k	…	x_n

Генеральною сукупністю називається сукупність значень тієї самої ознаки X в ***усіх*** об'єктах, зі складу яких проводиться вибірка.

Обсягом сукупності (вибіркової або генеральної) називають ***кількість*** (***n***) об'єктів цієї сукупності.

Вибірки бувають повторні і безповторні.

Повторної називають вибірку, при якій відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність. Безповторною називають вибірку, при якій відібраний об'єкт в генеральну сукупність не повертається.

Зазвичай використовуються безповторні вибірки.

Вибірка повинна бути репрезентативною або представницької, що означає, що дані вибірки повинні правильно відбивати ознака генеральної сукупності.

Способи відбору:

1. Відбір, що не вимагає поділу генеральної сукупності на частини.

Сюди відносяться: 1) простий випадковий безповторний відбір; 2) простий випадковий повторний відбір.

Простим випадковим називають такий відбір, при якому об'єкти витягують по одному з усієї генеральної сукупності.

2. Відбір, при якому генеральна сукупність ділиться на частини. Сюди відносять: типический відбір, механічний відбір, серійний відбір.

Типовим називається відбір, при якому об'єкти відбираються не з усієї генеральної сукупності, а з кожної її «типової» частини.

Механічним називають відбір, при якому генеральну сукупність «механічно» ділять на стільки груп, скільки об'єктів повинно увійти у вибірку, а з кожної групи випадковим чином вибирається один об'єкт. Щоб механічний відбір був репрезентативним, необхідно враховувати специфіку технологічного процесу.

Серійним називається відбір, при якому об'єкти відбирають з генеральної сукупності не по одному, а «серіями», які піддаються суцільного обстеження. Використовується тоді, коли досліджуваний ознака коливається в різних серіях незначно.

В економічних дослідженнях іноді використовують комбінований відбір.

Серед отриманих $\boldsymbol{n}$ значень ознаки можуть бути однакові значення. Так, значення x_i може зустрічатися $\boldsymbol{n_i}$ разів. У результаті, ***різні*** значення ознаки $\boldsymbol{X}$ можуть виявитися рівними $x_1, x_2, \ldots, x_m$, а кількість таких *різних* значень ознаки Х у вибірці може дорівнювати деякій величині $\boldsymbol{m}$ ($\boldsymbol{m} \le \boldsymbol{n}$). Для опису зазначеної ситуації використовують поняття “варіант”.

Варіантом називають конкретне, відмінне від інших, отримане в спробі значення $\boldsymbol{x_i}$ ознаки $\boldsymbol{X}$.

Частотою варіанту $\boldsymbol{x_i}$ у вибірці називають кількість $\boldsymbol{n_i}$ *однакових* значень ($\boldsymbol{x_i}$) ознаки у вибірці.

Сума частот усіх варіантів дорівнює обсягу вибірки:

$$\sum_{i=1}^{m} n_i = n .$$

Відносною частотою ($\boldsymbol{n_i^*}$) значення $\boldsymbol{x_i}$ ознаки $\boldsymbol{X}$ називається відношення частоти $\boldsymbol{n_i}$ варіанта $\boldsymbol{x_i}$ до обсягу вибірки $\boldsymbol{n}$:

$$n_i^* = \frac{n_i}{n} \le 1; \qquad \sum_{i=1}^{m} n_i^* = 1 .$$

Дискретним варіаційним рядом (*варіацією*) називають упорядковану за зростанням значень x_i послідовність варіантів із вказівкою їх абсолютних або відносних частот і подану у вигляді таблиці (див., наприклад, табл. 2).

Таблиця.2

Дискретний варіаційний ряд

Варіант $\boldsymbol{x_i}$	x_1	x_2	x_3	$\ldots$	x_m
Частота $\boldsymbol{n_i}$	$\boldsymbol{n_1}$	$\boldsymbol{n_2}$	$\boldsymbol{n_3}$		$\boldsymbol{n_m}$
Відносна частота $\boldsymbol{n_i^*}$	$\boldsymbol{n_1^*}$	$\boldsymbol{n_2^*}$	$\boldsymbol{n_3^*}$		$\boldsymbol{n_m^*}$

Полігоном частот (див. рис. 2) називають ламану лінію, яка з'єднує точки з координатами (x_i, n_i).

Полігоном відносних частот (див. рис. 1) називають ламану лінію, яка з'єднує точки з координатами (x_i, n_i^*).

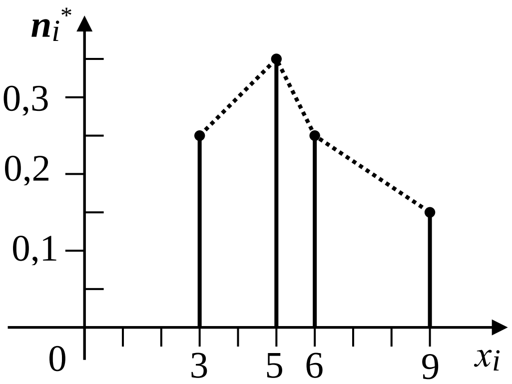

Рис. 1. Полігон відносних частот

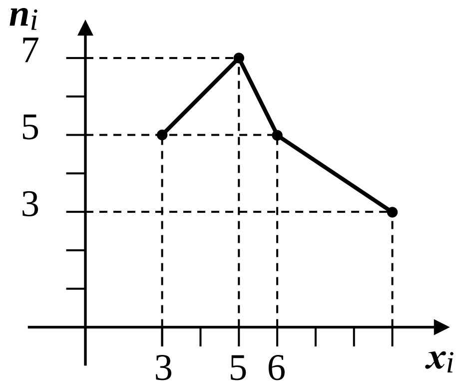

Рис. 2. Полігон частот

Розмахом варіації (***R***) називають *різницю* між максимальним (x_{max}) і мінімальним (x_{min}) значеннями варіантів у вибірці:

$$R = x_{max} - x_{min}.$$

Приклад. Записати у вигляді варіаційного ряду вибірку:

3; 9; 9; 5; 3; 6; 9; 5; 5; 5; 6; 3; 6; 5; 6; 3; 3; 5; 6; 5.

Визначити розмах вибірки, побудувати полігон частот і полігон відносних частот.

Розв'язання. Обсяг вибірки (число її елементів) дорівнює $n = 20$. Упорядкуємо варіанти за зростанням і підрахуємо кількість повторень значень x_i у кожному варіанті, одержимо:

$$x_1 = 3; \quad x_2 = 5; \quad x_3 = 6; \quad x_4 = 9;$$
$$n_1 = 5; \quad n_2 = 7; \quad n_3 = 5; \quad n_4 = 3.$$

Контроль обчислень виконаємо шляхом підсумовування частот варіантів:

$$\Sigma n_i = 20.$$

Для кожного варіанту x_i знаходимо відносну частоту: $n_i^* = n_i / n$.

Отриманий варіаційний ряд набуде вигляду (див. табл. 3).

Таблиця 3

Дискретний варіаційний ряд

x_i	3	5	6	9
$\boldsymbol{n_i}$	5	7	5	3
$\boldsymbol{n_i^*}$	0,25	0,35	0,25	0,15

Контроль обчислень виконаємо шляхом підсумовування відносних частот варіантів:

$$\Sigma\, \boldsymbol{n_i^*} = 1.$$

Розмах вибірки одержимо:

$$\boldsymbol{R} = \boldsymbol{x}_{max} - \boldsymbol{x}_{min} = 9 - 3 = 6.$$

Для побудови полігону частот нанесемо отримані точки ($\boldsymbol{x_i}$, $\boldsymbol{n_i}$) із табл. 3 на графік (див. рис. 4).

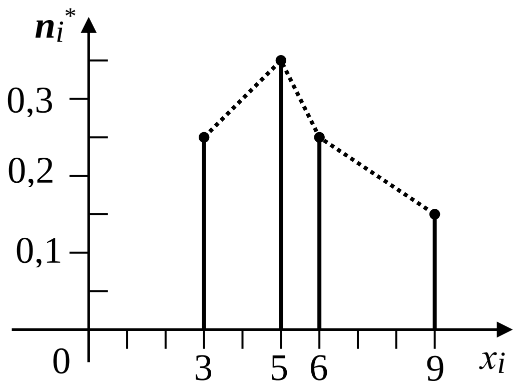

Рис. 3. Полігон відносних частот (див. табл.3)

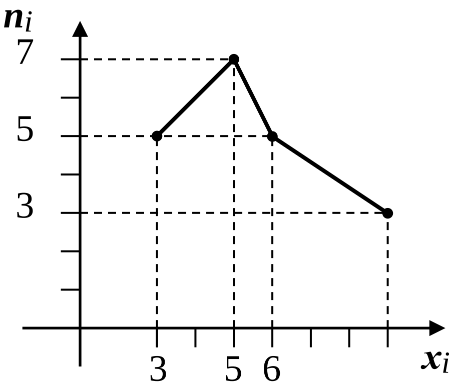

Рис. 4. Полігон частот (див. табл. 3)

Якщо обсяг вибірки великий (сотні варіантів), то дискретний варіаційний ряд стає незручною формою запису. Тоді увесь діапазон значень ознаки $\boldsymbol{X}$ розбивають на $\boldsymbol{k}$ часткових інтервалів, які не перетинаються, (синонім – розрядів) $[a_{i-1}; a_i)$, $(\boldsymbol{i} = 1, 2, \ldots, \boldsymbol{k})$.

Для кожного $\boldsymbol{i}$-го інтервалу підраховують кількість значень x_i ознаки $\boldsymbol{X}$, що потрапили в цей інтервал, – частоту інтервалу ($\boldsymbol{n_i}$). Елемент x_i, який *співпав* із границею інтервалу, відносять до наступного інтервалу, а не до попереднього.

Якщо до точності розрахунків немає дуже високих вимог, то частіше використовують розряди рівної довжини. Результати групування значень ознаки $\boldsymbol{X}$ за інтервалами записують у вигляді таблиці.

Інтервальним варіаційним рядом називається таблиця відповідності інтервалів значень (розрядів) випадкової величини ознаки $\boldsymbol{X}$ і частот $\boldsymbol{n_i}$ або/і відносних частот $\boldsymbol{n_i^*}$ входження значень ознаки в ці розряди, яка (таблиця) отримана за результатами спостережень або спроб (див. табл. 4). Такий ряд іноді називають *безперервним статистичним* рядом.

Таблиця 4

Інтервальний варіаційний ряд

Інтервал (розряд)	$[a_1; a_2)$	$[a_2; a_3)$	$[a_3; a_4)$	…	$[a_{m-1}; a_m]$
Частота $\boldsymbol{n}_i$	$\boldsymbol{n}_1$	$\boldsymbol{n}_2$	$\boldsymbol{n}_3$		$\boldsymbol{n}_m$
Відносна частота $\boldsymbol{n}_i^*$	$\boldsymbol{n}_1^*$	$\boldsymbol{n}_2^*$	$\boldsymbol{n}_3^*$		$\boldsymbol{n}_m^*$

Якщо в кожному інтервалі як представницьке значення ознаки $\boldsymbol{X}$ узяти середнє значення інтервалу: $x_i = \dfrac{a_{i-1} + a_i}{2}$, ($i$=1, 2, …, k),

то *інтервальний варіаційний* ряд можна умовно подати вже розглянутим *дискретним варіаційним* рядом. У цьому випадку в першому рядку табл. 4 указуються *середні* значення ознак для кожного інтервалу.

Приклад. Використати 6 інтервалів рівної довжини і побудувати інтервальний варіаційний ряд розподілу за даною вибіркою:

15	18	22	26	15	10	23	27	19	18	14	15	28	6	29	7	11
26	24	19	14	15	7	27	14	19	8	20	5	29	16	10	16	5
18	20	12	16	22	23	20	21	11	16	22	22	6	18	14	11	

Потім перейти до дискретного варіаційного ряду.

Розв'язання. Обсяг вибірки $\boldsymbol{n} = 50$.

Знайдемо розмах вибірки: $\boldsymbol{R} = x_{max} - x_{min} = 29 - 5 = 24$.

За умовою кількість інтервалів $\boldsymbol{m} = 6$, тому довжина кожного часткового інтервалу $\boldsymbol{h} = \boldsymbol{R} / \boldsymbol{m} = 24 / 6 = 4$. Перший інтервал починається з $x_{min} = 5$ і закінчується точкою $x_{min} + \boldsymbol{h} = 5 + 4 = 9$, значення якої до складу інтервалу *не входить*. Ця ж точка є початком другого інтервалу, до складу якого вона *входить*. Виконуючи послідовність таких же розрахунків, одержимо границі 6 інтервалів: [5;9), [9;13), [13;17), [17; 21), [21; 25) [25; 29).

Підраховуємо кількість елементів вибірки, що потрапили в кожний із знайдених інтервалів. Елемент вибірки 21 є границею інтервалів [17; 21) і [21; 25). При підрахунку частоти відносимо його до інтервалу [21; 25). У результаті одержимо такий інтервальний варіаційний ряд (див. табл. 5).

Таблиця 5

Інтервальний варіаційний ряд

$a_{i-1} \div a_i$	[5÷9)	[9÷13)	[13÷17)	[17÷21)	[21÷25)	[25÷29]
$\boldsymbol{n_i}$	7	6	12	10	7	8
$\boldsymbol{n_i^*}$	0,14	0,12	0,24	0,20	0,14	0,16
$n_i^{(n)}$	7	13	25	35	42	50
$F_n(x)$	0,00	0,14	0,26	0,50	0,70	0,84 ($F_n(x$>29)=1

Контроль обчислень виконаємо шляхом підсумовування частот варіантів, тоді одержимо: $\Sigma\, \boldsymbol{n_i} = 50$.

Для кожного часткового інтервалу обчисляємо відносну частоту за формулою: $\boldsymbol{n_i^*} = \boldsymbol{n_i} / \boldsymbol{n}$ і результат заносимо в нижній рядок таблиці.

Контроль обчислень виконаємо шляхом підсумовування відносних частот варіантів, одержимо: $\Sigma\, \boldsymbol{n_i^*} = 1$.

Для переходу від інтервального варіаційного ряду до дискретного варіаційного ряду візьмемо як варіанти середнє значення в кожному інтервалі: $x_i = (a_{i-1} + a_i)/2$. Одержимо наступний дискретний варіаційний ряд:

x_i	7	11	15	19	23	27
$\boldsymbol{n_i}$	7	6	12	10	8	7
$\boldsymbol{n_i^*}$	0,14	0,12	0,24	0,20	0,16	0,14

2. Числові характеристики вибірки

Вибірковим середнім (середнім арифметичним, позначається $\bar{x}$) називається середнє арифметичне значення ознаки ξ у вибірці x_i.

Коли всі значення ознаки *різні*, розрахункова формула має вигляд:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i\ .$$

Коли у вибірці є *частоти* значень ознаки або обчислення виконуються для інтервального варіаційного ряду, тоді варто визначити: x_i – середнє значення ознаки ξ в i-му інтервалі, $\boldsymbol{n_i}$ – частоту значень ознаки в i-му інтервалі. Тоді розрахункова формула набуде вигляду:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_m n_m}{n} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i \,.$$

Вибірковою дисперсією (дисперсією варіаційного ряду) називається середнє арифметичне квадратів відхилення варіантів x_i від свого вибіркового середнього значення, при цьому зберігаються відомі з теорії ймовірностей співвідношення:

$$D_в = S_x^2 = \frac{1}{n}\cdot\sum_{i=1}^{m}\left(x_i - \bar{x}\right)^2 \cdot n_i \,;$$

Тут S_x – вибіркове значення середнього квадратичного відхилення варіантів x_i.