Тема 2.6. Елементи теорії відносності та математичної статистики

2.6.1. Поняття про математичну статистику та її методи

Математична статистика — розділ математики та інформатики, в якому на основі дослідних даних вивчаються імовірнісні закономірності масових явищ. Основними задачами математичної статистики є статистична перевірка гіпотез, оцінка розподілу статистичних імовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок, якими є: вибіркове середнє, вибіркові дисперсії, стандартне відхилення. Прикладом перевірки таких гіпотез є з'ясування питання про те, змінюється чи не змінюється виробничий процес з часом. Прикладом оцінки параметрів є оцінка середнього значеннястатистичної змінної за дослідними даними. Для вивчення статистичної залежності використовують методи теорії кореляції. Загальні методи математичної статистики є основою теорії похибок.

Математична статистика широко використовує методи теорії ймовірностей для вибудови і перевірки математичних моделей. Її методи розширюють можливості наукового передбачення і раціонального прийняття рішення багатьох задач, де суттєві параметри не можуть бути з'ясовані чи контрольовані з достатньою точністю.

До основних **завдань** математичної статистики можна віднести наступні великі класи задач:
встановлення законів розподілу різних випадкових змінних, одержаних у результаті статистичного спостереження;перевірка статистичних гіпотез; оцінка невідомих параметрів різних розподілів.

Методи математичної статистики широко застосовують в організації виробництва, радіотехніці, військовій справі, теорії автоматичного керування, біології, економіці,статистичній фізиці, зоряній астрономії тощо.

Математичну статистику використовують також при розв'язанні теоретичних і практичних задач кібернетики. Порівняно новим напрямом розвитку математичної статистики є послідовний аналіз та загальна теорія статистичних рішень, яка тісно пов'язана з теорією ігор.

Першим етапом будь-якого дослідження є **збирання інформації**, а саме, статистичне спостереження.

Статистичне спостереження – це спланований, науково організований збір масових даних про соціально-економічні явища та процеси.

Прикладами статистичних спостережень можуть бути: щоденний облік відвідування; облік успішності за семестр; перепис населення; анкетування; перелік релігійних громад країни; обстеження фінансової діяльності інвестиційної компанії; реєстрація шлюбів у загсах; опитування окремих учасників презентації; облік числа зареєстрованих злочинів; реєстрація рівня цін на сільськогосподарські продукти; телефонне опитування та інші.

Найпоширенішим серед видів статистичних спостережень є вибіркове спостереження. У процесі вибіркового спостереження вивчається лише частина сукупності, відібрана спеціальним методом, яка називається вибіркою. Всю сукупність, з якої роблять вибірку називають генеральною сукупністю. Число об'єктів генеральної сукупності і вибірки називають відповідно обсягом генеральної сукупності і обсягом вибірки.

Приклад 2.6.1.

Якщо із 1000 деталей відібрано для обстеження 100 деталей, то обсяг генеральної сукупності $N = 1000$, а обсяг вибірки $n = 100$.

У результаті статистичного спостереження отримують матеріал, який характеризує окремі елементи сукупності. Після зведення і групування статистичних даних для найбільш раціонального та наукового вигляду їх результатів використовують статистичні таблиці.

Статистичні таблиці мають підмет і присудок. Статистичний підмет – це перелік окремих об’єктів або груп об’єктів. Статистичний присудок – це ті ознаки або показники, які характеризують статистичний підмет.

У таблицях інформація подається компактно, у зручній для порівняння та аналізу формі. Вид таблиці залежить від мети та особливостей об’єкта дослідження, обсягу наявної інформації.

Ряди розподілу

Нехай із генеральної сукупності зроблено вибірку, причому x_1 спостерігалось n_1 раз, x_2 – n_2 раз, x_3 – n_3 раз, ... , x_m – n_m і $n_1+n_2+\ldots+n_m=N$ – обсяг вибірки. Значення x_1, x_2, ... , x_m називаються варіантами, послідовність варіант записаних в зростаючому (спадному) порядку, – варіаційним рядом. Числа спостережень n_1, n_2, …, n_m називають частотами, а їх відношення до обсягу вибірки $\frac{n_1}{N}=p_1$; $\frac{n_2}{N}=p_2$; ..., $\frac{n_m}{N}=p_m$- відносними частотами. Відзначимо, що сума відносних частот дорівнює одиниці:

$$p_1+p_2+...+p_m=\frac{n_1}{N}+\frac{n_2}{N}+...+\frac{n_m}{N}=\frac{n_1+n_2+...+n_m}{N}=\frac{N}{N}=1.$$

Таблиця яка встановлює зв’язок між рядом варіант і відповідними частотами називається частотною (таблицею), або **статистичним розподілом**.

Таблиця 2.6.1

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

Статистичний розподіл можна задати у вигляді послідовності інтервалів і відповідних їм частот.

Приклад 2.6.2. Перейти від частот до відносних частот у такому розподілі вибірки обсягом N = 20.

Варіанта x_i	2	6	12
Частота n_i	3	10	7

Розв’язання

Знайдемо відносні частоти:

$$p_1 = \frac{3}{20} = 0{,}15; \quad p_2 = \frac{10}{20} = 0{,}50; \quad p_3 = \frac{7}{20} = 0{,}35.$$

Тому одержимо такий розподіл:

Варіанта x_i	2	6	12
Відносна частота p_i	0,15	0,50	0,35

Для графічного зображення статистичного розподілу використовуються полігони і гістограми.

Для побудови полігона на осі OX відкладають значення варіант x_i, на осі ординат – значення частот n_i.

Точки (x_i; n_i) з’єднують відрізками прямих і одержують полігон частот.

Для прикладу 2.6.2 на рис. 2.6.1 побудовано полігон частот, а на рис. 2.6.2 побудовано полігон відносних частот.

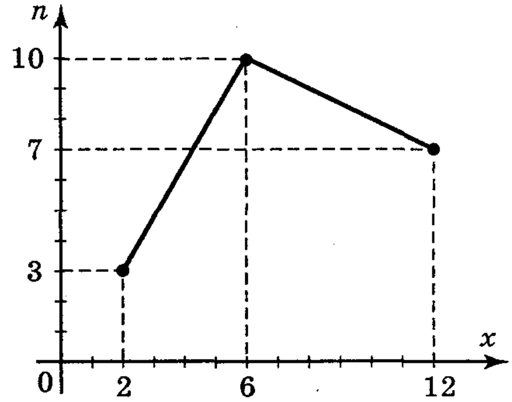

Рис. 2.6.1

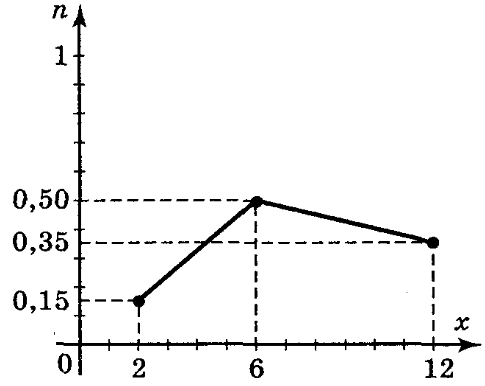

Рис. 2.6.2

Вибірка характеризується центральними тенденціями: **середнім значенням, модою і медіаною.** Дамо означення кожній з них. Середнім значенням вибірки називається середнє арифметичне всіх її значень:

$$\boxed{\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}, \text{ або } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

($\sum$ – знак суми – “сигма” велика).

Мода вибірки – те її значення, яке трапляється найчастіше. Позначається *Mo*. **Медіана вибірки** – це число, яке “поділяє” “навпіл” упорядковану сукупність усіх значень вибірки, тобто середня величина змінюваної ознаки, яка міститься в середині ряду, розміщеного в порядку зростання або спадання ознаки. Позначається *Me*.

Приклад 2.6.3

Нехай дано вибірку 2, 3, 4, 4, 6, 6, 6, 7, 7, 8. Знайдемо центральні тенденції вибірки.

Розв’язання

Мода даної вибірки *Mo* = 6, бо число 6 зустрічається найчастіше.

Середнє значення вибірки:

$$\bar{x} = \frac{2+3+4+4+6+6+6+7+7+8}{10} = \frac{53}{10} = 5{,}3.$$

Медіана даної вибірки *Me* = 6, бо вибірка має парне число значень і її медіана дорівнює півсумі двох її середніх значень:

$$Me = \frac{6+6}{2} = 6.$$

Приклад 2.6.4.

Розглянемо вибірку 0, 0, 1, 1, 3, 3, 3, 5; n = 8, $\bar{x}$ =2.

Знайдемо відхилення x_i-$\bar{x}$ кожного значення x_i від середнього значення $\bar{x}$. Результати занесемо в таблицю 2.6.2.

Таблиця 2.6.2

Значення x	Середнє арифметичне $\bar{x}$	Відхилення x_i-$\bar{x}$
0	2	-2
0	2	-2
1	2	-1
1	2	-1
3	2	1
3	2	1
3	2	1
5	2	3
$\sum_{i=1}^{8} x_i = 16$	$\sum_{i=1}^{8} \bar{x} = 16$	$\sum_{i=1}^{8} (x_i - \bar{x}) = 0$

Сума всіх відхилень дорівнює 0.

Для будь-якої вибірки $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$, тому в статистиці користуються іншим показником – середнім квадратичним відхиленням, який знаходиться так: усі відхилення підносяться до квадрата; знаходять середнє арифметичне цих квадратів, із знайденого середнього арифметичного добувають квадратний корінь. Середнє квадратичне відхилення позначається грецькою буквою σ (“сигма” мала):

$$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}.}$$

σ^2 в статистиці називають дисперсією.

Приклад 2.6.5.

Знайдемо середнє квадратичне відхилення значень вибірки: 5, 8, 10, 12, 17, 20.

Розв’язання

Знаходження середнього квадратичного подано в таблиці 2.6.3

Таблиця 2.6.3

Значення x	Середнє арифметичне $\bar{x}$	Відхилення $x_i - \bar{x}$	Квадрат відхилення $(x_i - \bar{x})^2$	Квадратичне не відхилення σ
5 8 10 12 17 20		-7 -4 -2 0 5 8	49 16 4 0 25 64	
$\sum_{i=1}^{6} x_i = 72$	$\bar{x} = \frac{72}{6} = 12$	$\sum_{i=1}^{6}(x_i - \bar{x}) = 0$	$\sum_{i=1}^{6}(x_i - \bar{x})^2 = 158$	$\sigma = \sqrt{\frac{158}{6}} \approx 5{,}13$

Якщо вибірку задано статистичним рядом, то

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

$$\sigma = \sqrt{\frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \ldots + n_m(x_m - \bar{x})^2}{n}} \text{ або } \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \bar{x})^2} .$$

Приклад 2.6.6.

Для статистичного ряду знайти $\bar{x}$ та σ.

x_i	-1	0	3	5	8
n_i	2	1	4	2	1

Розв’язання

Обсяг вибірки $n = 10$.

Середнє значення вибірки: $\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{2\cdot(-1) + 4\cdot 3 + 2\cdot 5 + 1\cdot 8}{10} = 2{,}8$.

Середнє квадратичне відхилення значень:

$$\sigma = \sqrt{\frac{(-1-2{,}8)^2\cdot 2 + (0-2{,}8)^2\cdot 1 + (3-2{,}8)^2\cdot 4 + (5-2{,}8)^2\cdot 2 + (8-2{,}8)^2\cdot 1}{10}} =$$

$$\sqrt{\frac{14{,}44\cdot 2 + 7{,}84\cdot 1 + 0{,}04\cdot 4 + 4{,}84\cdot 2 + 27{,}04\cdot 1}{10}} = \sqrt{\frac{73{,}6}{10}} = \sqrt{7{,}36} \approx 2{,}71.$$

Контрольні запитання

1. Чим займається наука «Статистика»?
2. Які засоби наочного зображення інформації в статистиці ви знаєте?
3. Як побудувати полігон?
4. У чому полягає основне завдання математичної статистики?