Розділ 3. ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

Математична статистика розв'язує два види задач:
1) статистичне оцінювання (точкове, інтервальне) параметрів генеральної сукупності;
2) перевірка правдивості статистичних гіпотез про значення параметрів генеральної сукупності або про закон розподілу ознаки генеральної сукупності на підставі обробки результатів вибірки.

3. 1. Вибірковий метод

Статистичною сукупністю називають множину однорідних об'єктів, об'єднаних за деякою спільною ознакою.

Генеральною сукупністю називають сукупність всіх однорідних об'єктів, які вивчають.

Генеральна сукупність — це всеможливі результати спостережень при дослідженні певного явища. Явище має випадковий характер, тому воно задається за допомогою випадкової величини. Таким чином генеральна сукупність не що інше, як значення випадкової величини X.

Оскільки суцільна обробка всіх елементів генеральної сукупності практично неможлива, то застосовують вибірковий метод.

Вибірковою сукупністю або ***вибіркою*** називають сукупність об'єктів випадково вибраних для дослідження з генеральної сукупності.

Вважають, що елементи вибірки незалежні однаково розподілені випадкові величини, розподіл яких співпадає з розподілом генеральної сукупності.

Об'ємом сукупності називають кількість її об'єктів.

Як правило об'єм вибіркової сукупності набагато менший ніж об'єм генеральної сукупності, який у багатьох випадках може бути невідомим.

Вибірка буває ***повторною*** (із поверненням досліджуваного об'єкта в генеральну сукупність) і ***безповторною*** (без зазначеного повернення). На практиці частіше використовують безповторну вибірку. Вибірка повинна бути ***репрезентативною***, тобто такою, по якій можна дійти правильного висновку про ознаку генеральної сукупності, що вивчається. Вибірка буде репрезентативною, якщо її елементи вибрані випадково.

Нехай унаслідок n незалежних спостережень, які відбуваються в однакових умовах, одержано вибірку $x_1, x_2, \ldots, x_n$.

Послідовність спостережуваних значень, записаних у зростаючому порядку

$$x_1 \le x_2 \le \ldots \le x_n \tag{3.1}$$

називають ***дискретним варіаційним рядом (впорядкованою вибіркою)***, а самі x_i — ***варіантами.***

Кожна варіанта вибірки може бути спостереженою n_i $(n_i \geq 1)$ раз, число n_i називають ***частотою*** x_i, причому $\sum_{i=1}^{k} n_i = n.$

Відношення частоти n_i варіанти x_i до об'єму вибірки n називають ***відносною частотою*** x_i і позначають

$$W_i = \frac{n_i}{n}, \quad \sum_{i=1}^{k} W_i = 1. \tag{3.2}$$

Статистичним розподілом вибірки називають відповідність між варіантами та їхніми частотами. Його можна задати таблицею

x_1	x_2	. . .	x_k
n_1	n_2	. . .	n_k

Статистичний розподіл вибірки можна задати як відповідність між варіантами і відносними частотами

x_1	x_2	. . .	x_k
W_1	W_2	. . .	W_k

Статистичний розподіл вибірки можна зобразити графічно. Для цього по осі абсцис відкладають варіанти, а по осі ординат — частоти, будують точки $M_i(x_i, n_i)$ і послідовно з'єднують відрізками прямих. Одержану ламану називають ***полігоном частот.*** Аналогічно будують ***полігон відносних частот*** — ламану, вершинами якої є точки $M_i(x_i, W_i)$.

Якщо ознака генеральної сукупності, що вивчається, змінюється неперервно, то розглядають інтервальний варіаційний ряд.

Нехай відомі результати вимірювання неперервної випадкової величини X, для якої a та b — відповідно найменше і найбільше значення. Відрізок $[a, b]$ розіб'ємо на k елементарних проміжків (x_{i-1}, x_i), $(i = 1, 2, \ldots, k)$. Позначимо через n_i кількість значень випадкової величини X, які належать інтервалу (x_{i-1}, x_i).

Перелік часткових інтервалів і відповідних їм частот або відносних частот називають ***інтервальним статистичним розподілом вибірки.*** У табличній формі цей розподіл має такий вигляд:

(x_0, x_1)	(x_1, x_2)	. . .	(x_{k-1}, x_k)
n_1	n_2	. . .	n_k
W_1	W_2	. . .	W_k

Різниці $x_i - x_{i-1}$ називають ***інтервальними різницями***, $b - a$ — ***розмахом варіації.***

Частку $\dfrac{n_i}{x_i - x_{i-1}}$ називають ***щільністю розподілу частот*** на інтервалі (x_{i-1}, x_i).

Гістограмою частот називають функцію визначену на відрізку [a, b], яка на кожному частинному інтервалі (x_{i-1}, x_i) набуває сталого значення $\dfrac{n_i}{x_i - x_{i-1}}$.

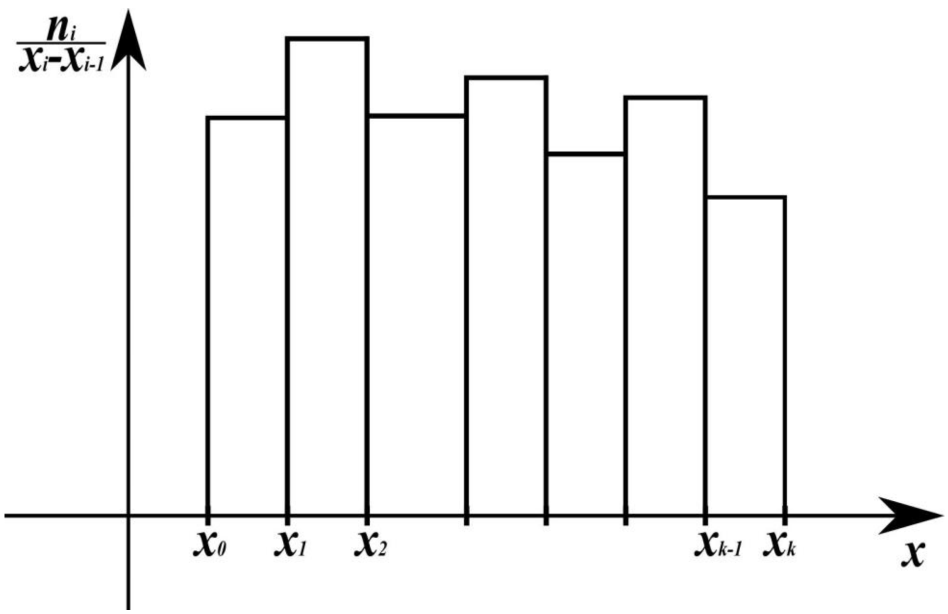

Рис. 26. Гістограма частот

Сума площ прямокутників з основами $x_i - x_{i-1}$ і висотами $\dfrac{n_i}{x_i - x_{i-1}}$ дорівнює об'єму вибірки.

Гістограма відносних частот елементів вибірки — це функція задана на відрізку [a, b], яка на кожному частинному інтервалі (x_{i-1}, x_i) набуває сталого значення $\dfrac{W_i}{x_i - x_{i-1}}$.

Площа гістограми відносних частот дорівнює сумі відносних частот, тобто 1.

Зауваження. У переважній більшості випадків відрізок [a, b] розбивають на k рівних проміжків, довжину i – го проміжку $h = x_i - x_{i-1}$ називають кроком.

Емпіричною функцією розподілу або функцією розподілу вибірки називають функцію, яка визначає для кожного дійсного значення x визначає відносну частоту події $x < X$. Якщо n_x — число варіант менших за x, n — об'єм вибірки, то

$$F_e(x) = \frac{n_x}{n} \qquad (3.3)$$

Цю функцію називають ще ***функцією нагромадження відносних частот.***

Властивості емпіричної функції розподілу

1) $0 \le F_e(x) \le 1$;

$\lhd\ 0 \le n_x \le n \ \rhd$

2) $F_e(x)$ — неспадна функція, тобто, $x_1 < x_2 \Rightarrow F_e(x_1) \le F_e(x_2)$;

$\lhd\ x_1 < x_2 \Rightarrow n_{x_1} \le n_{x_2} \Rightarrow F_e(x_1) \le F_e(x_2) \rhd$

3) якщо a — найменша варіанта, b — найбільша варіанта, то $F_e(x) = 0$ при $x \le a$, $F_e(x) = 1$, при $x \ge b$.

$\lhd\ n_x = 0$ при $x \le a$; $n_x = n$ при $x \ge b$. $\rhd$

Функцію розподілу генеральної сукупності на відміну від емпіричної функції розподілу називають теоретичною функцією розподілу. $F_e(x)$ випадкова величина і змінюється від вибірки до вибірки, вона визначає відносну частоту події $x < X$, а теоретична функція розподілу визначає ймовірність цієї ж події.

Приклад 1. Заданий статистичний розподіл вибірки

x_i	-3	-1	2	5
n_i	2	5	7	6

Побудувати полігон частот; знайти емпіричну функцію розподілу та побудувати її графік.

$\lhd$Знайдемо об'єм вибірки $n = 2 + 5 + 7 + 6 = 20$.

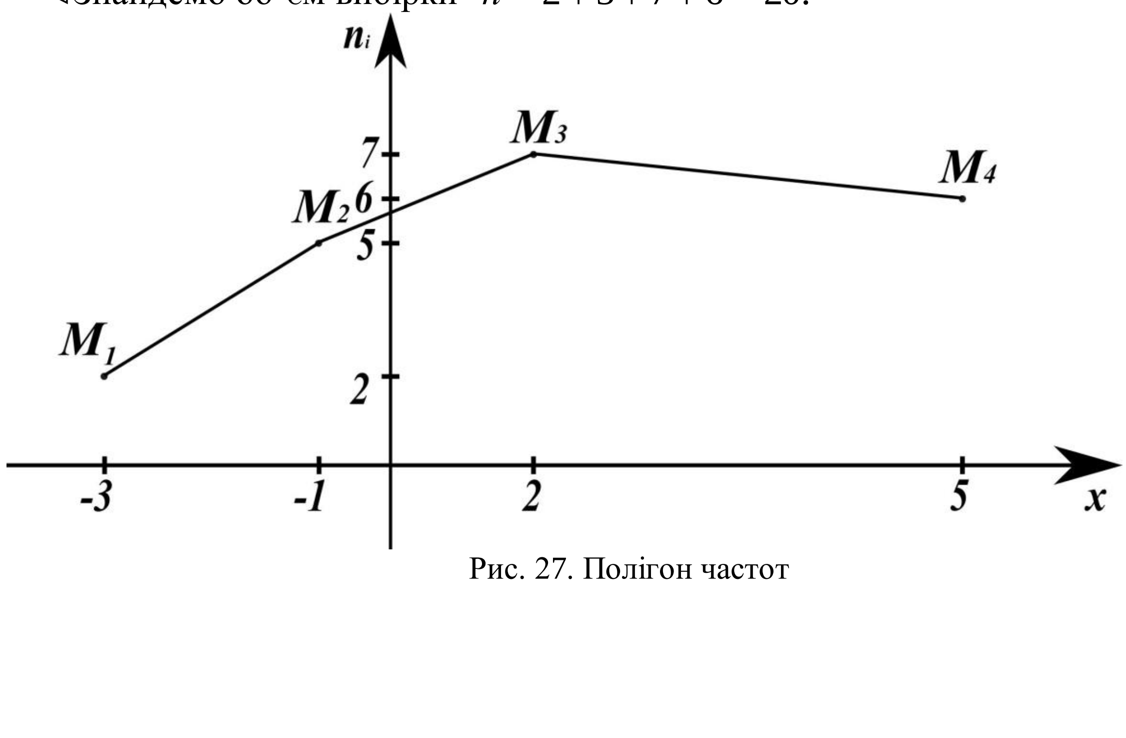

Рис. 27. Полігон частот

Найменша варіанта $x_1 = -3,$ тому $F_e(x) = 0,$ якщо $x \le -3.$

Значення $X < -1$ (тобто $x_1 = -3$) спостерігалось два рази, тому $F_e(x) = \frac{2}{20} = 0{,}1$, якщо $-3 < x \le -1.$

Значення $X < 2$ (тобто $x_1 = -3,\ x_2 = -1$) спостерігалось 2 + 5 = 7 разів, тому $F_e(x) = \frac{7}{20} = 0{,}35$, якщо $-1 < x \le 2.$

Значення $X < 5$ (тобто $x_1 = -3,\ x_2 = -1,\ x_3 = 2$) спостерігалось 2 + 5 + 7 =14 разів, тому $F_e(x) = \frac{14}{20} = 0{,}7$, якщо $2 < x \le 5.$

$x_4 = 5$ — найбільша варіанта, тому $F_e(x) = 1$ при $X > 5.$

Отож, матимемо

$$F_e(x) = \begin{cases} 0, & x \le -3; \\ 0{,}1, & -3 < x \le -1; \\ 0{,}35, & -1 < x \le 2; \\ 0{,}7, & 2 < x \le 5; \\ 1, & x > 5. \end{cases}$$

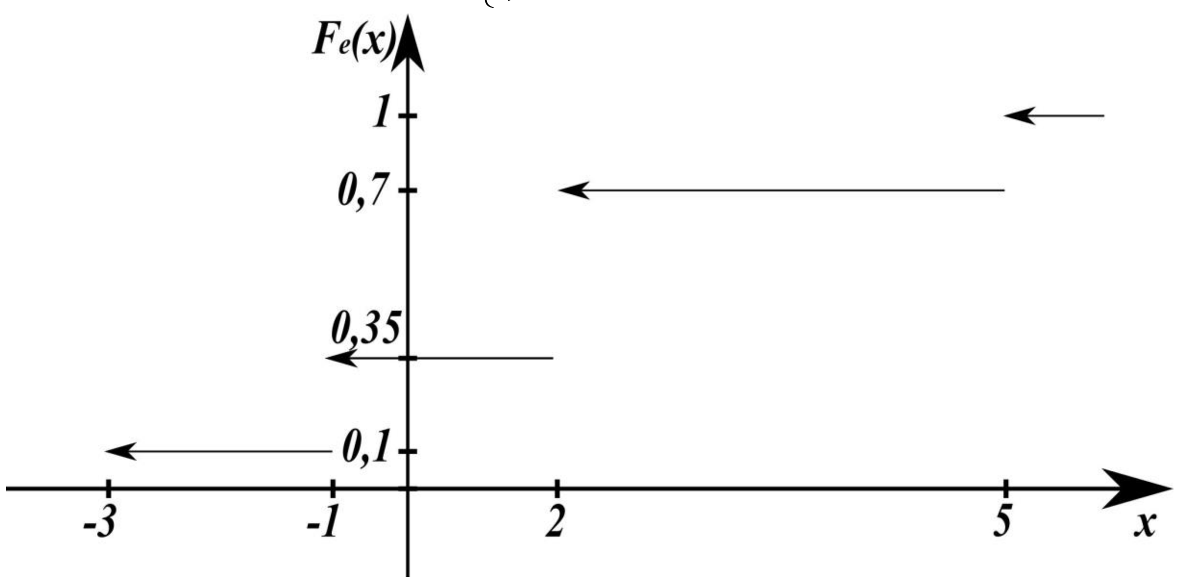

Рис. 28. Графік функції розподілу

Приклад 2. Відсоток виконання плану підприємства за рік та кількість підприємств, що виконують план, наведено у вигляді інтервального статистичного розподілу:

10;20	20;30	30;40	40;50	50;60	60;70	70;80	80;90	90;100	100;110
2	6	13	16	25	12	10	8	5	3

Побудувати гістограму відносних частот і обчислити вибіркову середню і вибіркове середнє квадратичне відхилення.

◁ Об'єм вибірки $n = 2+6+13+16+25+12+10+8+5+3 = 100$.

Знаходимо відносні частоти $W_i = \dfrac{n_i}{100}$ та щільності розподілу відносних частот $\dfrac{W_i}{h} = \dfrac{W_i}{10}$ заданого інтервального статистичного розподілу.

10;20	20;30	30;40	40;50	50;60	60;70	70;80	80;90	90;100	100;110
0,02	0,06	0,13	0,16	0,25	0,12	0,10	0,08	0,05	0,03
0,002	0,006	0,013	0,016	0,025	0,012	0,010	0,008	0,005	0,003

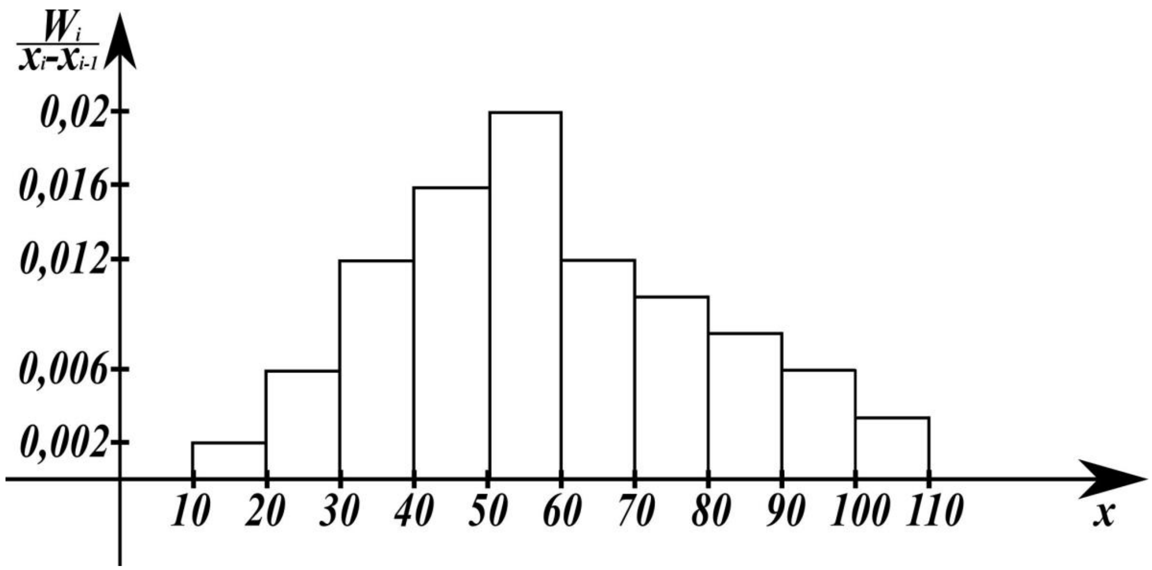

Рис. 29. Гістограма відносних частот

Для визначення точкових оцінок від інтервального статистичного розподілу переходимо до дискретного, варіантами якого є середини частинних інтервалів.

x_i^*	15	25	35	45	55	65	75	85	95	105
n_i	2	6	13	16	25	12	10	8	5	3

$$x_B = \frac{1}{100}(15\cdot 2 + 25\cdot 6 + 35\cdot 13 + 45\cdot 16 + 55\cdot 25 + 65\cdot 12 + 75\cdot 10 + 85\cdot 8 + 95\cdot 5 +$$

$$+105\cdot 3) = \frac{1}{100}\left(30 + 150 + 455 + 720 + 1375 + 780 + 750 + 680 + 475 + 315\right) = 57{,}3;$$

$$x_B^2 = \frac{1}{100}(15^2\cdot 2 + 25^2\cdot 6 + 35^2\cdot 13 + 45^2\cdot 16 + 55^2\cdot 25 + 65^2\cdot 12 + 75^2\cdot 10 + 85^2\cdot 8 +$$

$$+95^2 \cdot 5 + 105^2 \cdot 3) = \frac{1}{100}(450 + 3750 + 15925 + 32400 + 75625 + 50700 + 56250 +$$

$$+ 57800 + 45125 + 33075) = 3711;$$

$$D_B = x_B^2 - (x_B)^2 = 3711 - 57{,}3^2 = 427{,}11;$$

$$\sigma_B = \sqrt{D_B} = \sqrt{427{,}11} = 20{,}68. \triangleright$$