

Розділ 8

ЧИСЛОВІ ХАРАКТЕРИСТИКИ СТАТИСТИЧНОГО МАТЕРІАЛУ

8.1. Числові характеристики вибірки

Як доповнення до табличних та графічних методів представлення даних, ще одним найважливішим засобом обробки даних є обчислення їх числових характеристик. Найважливіші з них: *вибіркове середнє, дисперсія, середньоквадратичне відхилення*.

Ці характеристики можуть бути обчислені за даними, що знаходяться у вибірці, або за даними, що входять у кінцеву генеральну сукупність.

Числові характеристики, обчислені за вибіркою або ті, що використовуються для опису даних вибірки, називають **статистиками**.

Вибірковим середнім статистичного матеріалу називають суму всіх значень вибірки, поділену на обсяг вибірки n :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (8.1)$$

Якщо маємо частотну таблицю, то вибіркове середнє вираховуємо за формулою:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i x_i, \quad (8.2)$$

де n – об'єм вибірки, m – число різних варіант, $n_1, n_2, \dots, n_m$ – частоти варіант ($n = n_1 + n_2 + \dots + n_m$), x_i – значення i -ої варіанти.

Вибіркове середнє для інтервального статистичного ряду знаходиться за формулою

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s n_i z_i, \quad (8.3)$$

де n – об'єм вибірки, s – кількість інтервалів, z_i – середина i -ого інтервалу, $n_1, n_2, \dots, n_s$ – число елементів вибірки, які потрапили в i -ий інтервал.

Вибіркове середнє є аналогом математичного сподівання і використовується дуже часто. Воно може набувати різних числових значень при різних вибірках однакового об'єму.

Тому можна розглядати розподіли (теоретичний та емпіричний) вибіркового середнього та числові характеристики цього розподілу (цей розподіл називають *вибірковим*).

Основні властивості вибіркового середнього

1. При множенні усіх варіант вибірки на однаковий множник вибіркоче середнє також множиться на цей множник

$$\frac{1}{n} \sum_{i=1}^m n_i (cx_i) = \frac{c}{n} \sum_{i=1}^m n_i x_i = c\bar{x}.$$

2. Якщо додати (відняти) до всіх варіант вибірки однакове число, то вибіркоче середнє зростає (зменшується) на це число

$$\frac{1}{n} \sum_{i=1}^m n_i (x_i \pm c) = \frac{1}{n} \sum_{i=1}^m n_i x_i \pm \frac{1}{n} c \sum_{i=1}^m n_i = \bar{x} \pm c.$$

Степеневим середнім вибірки називають середнє значення, яке знаходять за формулою

$$\bar{x}_\alpha = \left(\sum_{i=1}^m \frac{n_i}{n} x_i^\alpha \right)^{\frac{1}{\alpha}}. \quad (8.4)$$

При $\alpha = 1$ одержимо вибіркоче середнє.

При $\alpha = 2$ одержимо середньоквадратичне значення вибірки

$$\bar{x}_2 = \left(\sum_{i=1}^m \frac{n_i}{n} x_i^2 \right)^{\frac{1}{2}}.$$

При $\alpha = -1$ одержимо середнє гармонічне

$$\bar{x}_{-1} = \bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^m \frac{n_i}{x_i}}.$$

Середнім геометричним вибірки називають середнє значення, яке обчислюється за формулою

$$\bar{x}_\Gamma = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_m^{n_m}}. \quad (8.5)$$

Це середнє обчислюється лише за умови, що всі значення варіанти є додатні, тобто $x_i > 0, i = 1, 2, \dots, m$.

Середнє геометричне застосовується у статистиці для визначення темпу зростання під час дослідження змін ознаки з часом.

Зауваження 1. Обрання того чи іншого середнього для характеристики розподілу пов'язане з якісним аналізом цього розподілу.

Зауваження 2. Крім вказаних степеневих середніх, у статистиці застосовують ще структурні середні, які не залежать від значень варіант, які розташовані на краях розподілу, а пов'язані з рядом частот.

До структурних середніх відносять моду та медіану.

Мода M_o – це елемент, який найчастіше трапляється у вибірці.

Для дискретних статистичних рядів

$$M_o = x_i, \text{ якщо } n_i = \max_i n_i$$

Для інтервальних статистичних рядів

$$M_o = x_i + h \cdot \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}, \quad (8.6)$$

де x_i – початок інтервалу з найбільшою частотою, n_i – частота i -го інтервалу.

Медіаною M_e називають варіанту, яка ділить варіаційний ряд на дві частини, які рівні за числом варіант.

Для дискретних статистичних рядів

$$M_e = \begin{cases} x_m, & \text{при } n = 2m - 1, \\ \frac{x_m + x_{m+1}}{2}, & \text{при } n = 2m, \end{cases} \quad (8.7)$$

де x_m – середина варіаційного ряду.

Для інтервальних статистичних рядів

$$M_e = x_i + h \cdot \frac{\frac{n}{2} - \sum_{j=1}^{i-1} n_j}{n_i}, \quad (8.8)$$

де x_i – початок медіанного інтервалу, тобто такого, якому відповідає перша з нагромаджених частот, що перевищує половину всіх спостережень, h – довжина i -ого інтервалу, n_i – частота медіанного інтервалу.

Вибірковою дисперсією D_B називають середнє значення квадратів відхилення варіант від вибіркового середнього з урахуванням відповідних частот.

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i (x_i - \bar{x})^2. \quad (8.9)$$

Зауваження. Обчислення вибіркової дисперсії спрощується, якщо її знаходити за формулою:

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - \left(\frac{1}{n} \sum_{i=1}^m n_i x_i \right)^2 = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - (\bar{x})^2. \quad (8.10)$$

Виправленою вибірковою дисперсією (варіансою) називають суму квадратів відхилень елементів від вибіркового середнього, поділену на $(n-1)$:

$$S^2 = \frac{n}{n-1} D_B = \frac{1}{n-1} \sum_{i=1}^m n_i (x_i - \bar{x})^2. \quad (8.11)$$

Вибіркову дисперсію та виправлену вибірку дисперсію для інтервального розподілу знаходять, відповідно, за формулами:

$$D_B = \frac{1}{n} \sum_{i=1}^s n_i (z_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^s n_i z_i^2 - (\bar{x})^2, \quad (8.12)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^s n_i (z_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^s n_i z_i^2 - (\bar{x})^2, \quad (8.13)$$

$$\text{або } S^2 = \frac{n}{n-1} D_B. \quad (8.14)$$

Вибірковим середньоквадратичним відхиленням називають квадратний корінь із вибіркової дисперсії:

$$\sigma_B = \sqrt{D_B}. \quad (8.15)$$

Стандарт – це арифметичний квадратний корінь із варіанси:

$$S = \sqrt{S^2}. \quad (8.16)$$

Розмах – це різниця між найбільшим і найменшим значенням варіаційного ряду (між крайніми елементами) і позначається

$$R = x_{max} - x_{min}. \quad (8.17)$$

Приклад 8.1. Обчислити вибіркове середнє, вибірккову дисперсію, вибірккове середньоквадратичне відхилення, варіанту, стандарт, моду, медіану та розмах вибірки, заданої статистичним розподілом частот:

x_i	-1	2	5	8	10
n_i	5	10	20	5	10

Розв'язання. Об'єм вибірки $n = 5 + 10 + 20 + 5 + 10 = 50$.

Знайдемо, відповідно, вибірккове середнє, вибірккову дисперсію, вибірккове середньоквадратичне відхилення, варіансу, стандарт, моду та медіану вибірки:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_i x_i = \frac{1}{50} (-1 \cdot 5 + 2 \cdot 10 + 5 \cdot 20 + 8 \cdot 5 + 10 \cdot 10) = \frac{255}{50} = 5,1;$$

$$D_B = \frac{1}{n} \sum_{i=1}^m n_i x_i^2 - (\bar{x})^2 = \frac{1}{50} (1 \cdot 5 + 4 \cdot 10 + 25 \cdot 20 + 64 \cdot 5 + 100 \cdot 10) - 5,1^2 = \frac{1}{50} \cdot 1865 - 26,01 = 37,3 - 26,01 = 11,29;$$

$$\sigma_B = \sqrt{D_B} = \sqrt{11,29} \approx 3,36;$$

$$S^2 = \frac{n}{n-1} D_B = \frac{50}{49} \cdot 11,29 = 11,52;$$

$$S = \sqrt{S^2} = \sqrt{11,52} \approx 3,39;$$

$$M_o = 5;$$

$$M_e = 5;$$

$$R = x_{max} - x_{min} = 10 - (-1) = 11.$$

Приклад 8.2. Обчислити вибірккове середнє, вибірккову дисперсію, вибірккове середньоквадратичне відхилення, варіанту, стандарт, моду, медіану та розмах вибірки, заданої інтервальним розподілом:

інтервал	[2; 4)	[4; 6)	[6; 8)	[8; 10)	[10; 12)
n_i	2	8	35	40	15

Розв'язання. Об'єм вибірки $n = 2 + 8 + 35 + 40 + 15 = 100$.

Перетворимо цей інтервальний статистичний розподіл вибірки на точковий, знайшовши середини частинних інтервалів:

$$z_1 = \frac{x_1+x_2}{2} = \frac{2+4}{2} = 3; \quad z_2 = \frac{x_2+x_3}{2} = \frac{4+6}{2} = 5;$$

$$z_3 = \frac{x_3+x_4}{2} = \frac{6+8}{2} = 7; \quad z_4 = \frac{x_4+x_5}{2} = \frac{8+10}{2} = 9;$$

$$z_5 = \frac{x_5+x_6}{2} = \frac{10+12}{2} = 11.$$

Отже, ми отримали такий статистичний розподіл вибірки:

z_i	3	5	7	9	11
n_i	2	8	35	40	15

Знайдемо, відповідно, вибіркове середнє, вибіркиму дисперсію, вибіркиму середньоквадратичне відхилення, варіансу, стандарт, моду та медіану вибірки:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^s n_i z_i = \frac{1}{100} (3 \cdot 2 + 5 \cdot 8 + 7 \cdot 35 + 9 \cdot 40 + 11 \cdot 15) = \\ &= \frac{1}{100} \cdot 816 = 8,16; \end{aligned}$$

$$\begin{aligned} D_B &= \frac{1}{n} \sum_{i=1}^s n_i z_i^2 - (\bar{x})^2 = \frac{1}{100} (9 \cdot 2 + 25 \cdot 8 + 49 \cdot 35 + 81 \cdot 40 + 121 \cdot 15) - 8,16^2 = \\ &= \frac{1}{100} \cdot 6988 - 66,5856 = 69,88 - 66,59 = 3,29; \end{aligned}$$

$$\sigma_B = \sqrt{D_B} = \sqrt{3,29} \approx 1,81;$$

$$S^2 = \frac{n}{n-1} D_B = \frac{100}{99} \cdot 3,29 = 3,32;$$

$$S = \sqrt{S^2} = \sqrt{3,32} \approx 1,82;$$

$$M_o = 8 + 2 \cdot \frac{40-35}{2 \cdot 40 - 35 - 15} = 8 + 2 \cdot \frac{5}{30} = 8,33;$$

$$M_e = 8 + 2 \cdot \frac{\frac{100}{2} - (8+2+35)}{40} = 8 + 2 \cdot \frac{50-45}{40} = 8,25;$$

$$R = x_{max} - x_{min} = 12 - 2 = 10.$$

8.2. Метод добутків обчислення вибіркового середнього та вибіркової дисперсії

Нехай вибірку задано у вигляді розподілу рівновіддалених варіант і відповідних їм частот. У цьому разі вибіркове середнє та вибіркиму дисперсію зручно знаходити методом добутків.

Алгоритм методу добутків

1. У перший стовпчик таблиці записують рівновіддалені варіанти x_i вибірки, розміщуючи їх у порядку зростання.

2. У другий стовпчик таблиці записують відповідні частоти n_i відповідних варіант. Суму елементів цього стовпчика (об'єм вибірки n) записують в останню його клітинку.

3. Третій стовпчик містить умовні варіанти u_i вибірки. Для знаходження умовних варіант вибірки треба:

а) значення варіанти вибірки, яка знаходиться приблизно посередині варіаційного ряду C обрати за умовний нуль;

б) знайти різницю h між будь-якими двома сусідніми варіантами;

в) обчислити умовні варіанти $u_i = \frac{x_i - C}{h}$. Зауважимо, що умовні варіанти завжди будуть цілими числами.

4. У четвертий стовпчик записують добутки частот та відповідних їм умовних варіант $n_i \cdot u_i$. Суму елементів стовпчика $\sum_{i=1}^m n_i \cdot u_i$ записують в його останню клітинку.

5. Знаходять добутки частот та квадратів умовних варіант $n_i \cdot u_i^2$ і записують їх у п'ятий стовпчик. Суму елементів стовпчика $\sum_{i=1}^m n_i \cdot u_i^2$ записують в його останню клітинку.

6. Знаходять добутки частот та квадратів умовних варіант, збільшених на одиницю, $n_i \cdot (u_i + 1)^2$ і записують їх у шостий стовпчик. Суму елементів стовпчика $\sum_{i=1}^m n_i \cdot (u_i + 1)^2$ записують в його останню клітинку.

7. Перевіряють обчислення так: сума елементів шостого стовпчика повинна задовольняти тотожність:

$$\sum_{i=1}^m n_i \cdot (u_i + 1)^2 = \sum_{i=1}^m n_i \cdot u_i^2 + 2 \cdot \sum_{i=1}^m n_i \cdot u_i + n.$$

8. Обчислюють умовні моменти за формулами:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i; \quad M_2^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i^2.$$

9. Обчислюють вибіркове середнє та вибіркву дисперсію за формулами:

$$\bar{x} = M_1^* \cdot h + C, \quad D_B = (M_2^* - (M_1^*)^2) \cdot h^2.$$

Зауваження. Якщо варіанти вибірки не є рівновіддаленими, то інтервал, у якому розміщені всі варіанти вибірки, ділять на кілька однакових за довжиною частинних інтервалів, кожен з яких має містити не менше 8–10 варіант. Потім знаходять середини частинних інтервалів, які й утворюють послідовність рівновіддалених варіант. За частоту кожної середини частинного інтервалу беруть суму частот варіант, які потрапили у відповідний частинний інтервал.

Під час обчислення вибіркової дисперсії з метою зменшення похибки, зумовленої групуванням (особливо при малій кількості інтервалів), роблять поправку Шеппарда, а саме: обчислюють дисперсію за формулою:

$$D'_B = D_B - \frac{1}{12} \cdot h^2.$$

Приклад 8.3. Знайти методом добутків вибіркове середнє та вибіркву дисперсію для статистичного розподілу вибірки об'єму $n = 100$.

x_i	2	8	9	13	15	18	20	21	24	27
n_i	8	1	15	13	9	3	4	10	22	15

Розв'язання. Розіб'ємо інтервал $[2; 27]$ на п'ять частинних інтервалів довжиною $h = 5$:

$$[2; 7], \quad (7; 12], \quad (12; 17], \quad (17; 22], \quad (22; 27].$$

Новими варіантами будуть середини цих частинних інтервалів:

$$y_1 = \frac{2+7}{2} = 4,5; \quad y_2 = \frac{7+12}{2} = 9,5; \quad y_3 = \frac{12+17}{2} = 14,5;$$

$$y_4 = \frac{17+22}{2} = 19,5; \quad y_5 = \frac{22+27}{2} = 24,5.$$

За частоти m_i варіант y_i візьмемо суму частот варіант, які потрапили у відповідний i -ий інтервал:

$$m_1 = 8; \quad m_2 = 1 + 15 = 16; \quad m_3 = 13 + 9 = 22;$$

$$m_4 = 3 + 4 + 10 = 17; \quad m_5 = 22 + 15 = 37.$$

Отже, ми отримали такий статистичний розподіл рівновіддалених варіант:

y_i	4,5	9,5	14,5	19,5	24,5
m_i	8	16	22	17	37

Складемо розрахункову таблицю:

y_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$	$m_i (u_i + 1)^2$
4,5	8	-2	-16	32	8
9,5	16	-1	-16	16	0
14,5	22	0	0	0	22
19,5	17	1	17	17	68
24,5	37	2	74	148	333
	$n = 100$		$\sum_{i=1}^5 m_i u_i = 59$	$\sum_{i=1}^5 m_i u_i^2 = 213$	$\sum_{i=1}^5 m_i (u_i + 1)^2 = 431$

Перевірка:

$$\sum_{i=1}^5 m_i (u_i + 1)^2 = 431;$$

$$\sum_{i=1}^5 m_i u_i^2 + 2 \cdot \sum_{i=1}^5 m_i u_i + n = 213 + 2 \cdot 59 + 100 = 431.$$

Обчислимо умовні моменти першого і другого порядків:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i = \frac{59}{100} = 0,59; \quad M_2^* = \frac{1}{n} \sum_{i=1}^m n_i \cdot u_i^2 = \frac{213}{100} = 2,13.$$

Методом добутків знайдемо вибіркове середнє й дисперсію, враховуючи, що хибний нуль $C = 14,5$:

$$\bar{x} = M_1^* \cdot h + C = 0,59 \cdot 5 + 14,5 = 17,45,$$

$$D_B = (M_2^* - (M_1^*)^2) \cdot h^2 = (2,13 - 0,59^2) \cdot 5^2 = 44,5475.$$

Оскільки кількість частинних інтервалів мала, то скористаємося поправкою Шеппарда:

$$D'_B = D_B - \frac{1}{12} \cdot h^2 = 44,5475 - \frac{1}{12} \cdot 5^2 \approx 42,46.$$

8.3. Метод сум обчислення вибіркового середнього та вибіркової дисперсії

Нехай вибірку задано у вигляді розподілу рівновіддалених варіант і відповідних їм частот. Тоді вибіркове середнє та вибірккову дисперсію зручно знаходити методом сум.

Алгоритм методу сум

1. У перший стовпчик таблиці записують рівновіддалені варіанти x_i вибірки, розміщуючи їх у порядку зростання.

2. У другий стовпчик таблиці записують відповідні частоти n_i варіант. Суму елементів цього стовпчика (об'єм вибірки n) записують в його останню клітинку.

3. Третій стовпчик містить умовні варіанти u_i вибірки. Для знаходження умовних варіант вибірки треба:

а) значення варіанти вибірки з найбільшою частотою C обрати за умовний нуль, це значення називають модою;

б) у порожніх клітинках третього стовпця, які розміщені над нулем (крім верхньої), зверху вниз записують послідовно нагромаджені частоти. Додавши всі нагромаджені частоти зверху до нуля, отримують число b_1 , яке записують у верхній клітинці;

в) у порожніх клітинках третього стовпця, які розміщені під нулем (крім нижньої), знизу вверх записують послідовно нагромаджені частоти. Додавши всі нагромаджені частоти знизу до нуля, отримують число a_1 , яке записують у його нижню клітинку.

4. У четвертому стовпці знову записують хибний нуль, над і під ним записують ще по одному нулю. Аналогічно до третього, заповнюють четвертий стовбець, причому нагромаджуються частоти третього стовпця. Суму нагромаджених частот, розміщених над нулем, позначають b_2 і записують у верхній клітинці четвертого стовпця. Суму нагромаджених частот, розміщених під нулем, позначають a_2 і записують у нижню клітинку четвертого стовпця.

5. Обчислюють d_1, s_1, s_2 :

$$d_1 = a_1 - b_1; \quad s_1 = a_1 + b_1; \quad s_2 = a_2 + b_2.$$

6. Обчислюють умовні моменти за формулами:

$$M_1^* = \frac{d_1}{n}; \quad M_2^* = \frac{s_1 + 2s_2}{n}.$$

7. Обчислюють вибіркве середнє та вибіркву дисперсію за формулами:

$$\bar{x} = M_1^* \cdot h + C,$$

$$D_B = (M_2^* - (M_1^*)^2) \cdot h^2.$$

Приклад 8.4. Знайти методом сум вибіркве середнє та вибіркву дисперсію для статистичного розподілу вибірки об'єму $n = 100$.

x_i	52	58	64	70	76	82	88	94	100	106
n_i	4	8	10	13	18	21	14	9	2	1

Розв'язання. Складемо розрахункову таблицю

1	2	3	4
x_i	n_i	$b_1 = 126$	$b_2 = 131$
52	4	4	4
58	8	12	16
64	10	22	38
70	13	35	73
76	18	53	0
82	21	0	0
88	14	26	0
94	9	12	16
100	2	3	4
106	1	1	1
		$a_1 = 42$	$a_2 = 21$

Знайдемо d_1, s_1, s_2 :

$$d_1 = a_1 - b_1 = 42 - 126 = -84;$$

$$s_1 = a_1 + b_1 = 42 + 126 = 168;$$

$$s_2 = a_2 + b_2 = 21 + 131 = 152.$$

Обчислимо умовні моменти першого і другого порядку:

$$M_1^* = \frac{d_1}{n} = \frac{-84}{100} = -0,84;$$

$$M_2^* = \frac{s_1 + 2s_2}{n} = \frac{168 + 2 \cdot 152}{100} = 4,72.$$

Знайдемо вибіркове середнє та дисперсію, врахувавши, що хибний нуль $C = 82$, а крок (відстань між двома сусідніми варіантами) $h = 6$, отримаємо:

$$\bar{x} = M_1^* \cdot h + C = -0,84 \cdot 6 + 82 = 76,96;$$

$$D_B = (M_2^* - (M_1^*)^2) \cdot h^2 = (4,72 - (-0,84)^2) \cdot 6^2 = 144,5184.$$

Запитання для самоконтролю

1. Що відноситься до числових характеристики вибірки?
2. Які числові характеристики називаються статистичними?
3. Як визначається вибіркова середня для статистичного розподілу вибірки?
4. Як визначається вибіркова середня для інтервального розподілу вибірки?
5. Які властивості має вибіркова середня?
6. Що таке степеневе середнє; середнє геометричне?
7. Як визначається дисперсія для статистичного розподілу вибірки?
8. Як визначається дисперсія для інтервального розподілу вибірки?
9. Як визначається середньоквадратичне відхилення?
10. Що таке мода і як її знайти?
11. Що таке медіана і як її знайти?
12. Що таке варіанса, стандарт, розмах і як їх обчислити?
13. Метод добутків для знаходження вибіркового середнього та вибіркової дисперсії, його алгоритм.
14. Метод сум для знаходження вибіркового середнього та вибіркової дисперсії, його алгоритм.
15. У яких випадках використовують виправлену вибірккову дисперсію і як вона пов'язана із вибірковою дисперсією?
16. В яких випадках обчислюють характеристики вибірки методом добутків; методом сум?