Характеристики варіації

Якщо індивідуальні значення ознаки істотно різняться між собою, то середня арифметична не є надійною характеристикою сукупності. Чим меншим є відхилення значень ознаки від середньої, тим краще вона характеризує загальний рівень ряду.

До абсолютних показників, з допомогою яких вимірюють відхилення від середньої, відносяться розмах варіації, середнє арифметичне відхилення та середнє квадратичне відхилення.

Розмах варіації обчислюють як різницю між найбільшим та найменшим значеннями ознаки:

$$R = x_{\max} - x_{\min}. \tag{1}$$

Розмах варіації є найпростішим показником варіації ознаки і використовується у випадках, коли важливо встановити її амплітуду, наприклад, оцінити варіацію цін на товар за певний проміжок часу, оцінити річну варіацію попиту на певний товар.

Точнішою характеристикою варіації ознаки є *середнє арифметичне* (*лінійне*) *відхилення*. Воно визначається за формулою:

$$\overline{d} = \frac{\sum\left|x-\overline{x}\right|}{n} = \frac{\sum\left|x-\overline{x}\right|\cdot f}{\sum f}. \qquad (2)$$

Середнє квадратичне відхилення обчислюють за формулою:

$$\sigma = \sqrt{\frac{\sum\left(x-\overline{x}\right)^2}{n}} = \sqrt{\frac{\sum\left(x-\overline{x}\right)^2\cdot f}{\sum f}}. \qquad (3)$$

Розмах варіації, середнє лінійне та середнє квадратичне відхилення вимірюють у тих же одиницях, що і ознаку. У статистиці для вимірювання варіації ознаки найбільшого поширення набуло середнє квадратичне відхилення. Квадрат середнього квадратичного відхилення σ^2 називають *дисперсією*.

При обчисленні дисперсії та, відповідно, середнього квадратичного відхилення використовують наступні властивості дисперсії.

1. Дисперсія сталої ознаки дорівнює нулю.
2. Зменшення (збільшення) всіх значень ознаки на одне й те ж число не змінює значення дисперсії ознаки.
3. Зменшення (збільшення) значень ознаки у A разів призводить до збільшення (зменшення) значення її дисперсії у A^2 разів.
4. Зменшення (збільшення) всіх частот у A разів не впливає на значення дисперсії.
5. Дисперсію можна обчислити за формулою:

$$\sigma^2 = \overline{x^2} - \left(\overline{x}\right)^2. \tag{4}$$

Формула (4) свідчить, що дисперсія ознаки дорівнює різниці між середньою арифметичною квадрату ознаки та квадратом її середньої арифметичної.

Для обчислення дисперсії та середнього квадратичного відхилення ознаки у випадку інтервального варіаційного ряду замість значень ознаки x використовують значення середин інтервалів, замість частот ознаки – частоти відповідних інтервалів. При такій заміні може виникнути систематична похибка обчислення дисперсії та середнього квадратичного відхилення.

Порівнюючи варіацію ознаки у різних сукупностях або варіацію різних ознак у одній сукупності, недостатньо виявити абсолютну величину варіації, оскільки вона залежить і від розміру варіації і від рівня ознаки. Тому для дослідження варіаційного ряду використовують також *відносні показники варіації*.

Розрізняють такі відносні показники варіації:

1) *коефіцієнт осциляції* $V_R = \frac{R}{\overline{x}} \cdot 100\%$ (5)

2) *лінійний коефіцієнт варіації* $V_{\overline{d}} = \frac{\overline{d}}{\overline{x}} \cdot 100\%$; (6)

3) *коефіцієнт варіації* $V = \frac{\sigma}{\overline{x}} \cdot 100\%$. (7)

Коефіцієнт варіації V використовують як показник однорідності сукупності. Вважається, що при значенні цього показника, яке не перевищує 33%, сукупність є однорідною, отже, середня арифметична $\overline{x}$ є типовою характеристикою сукупності.

Приклад 1. У таблиці 1 наведено дані щодо кваліфікаційного розряду робітників підприємства. Визначити абсолютні та відносні показники варіації цієї ознаки.

Таблиця 1. Дані щодо кваліфікаційного розряду робітників підприємства.

Кваліфікацій-ний розряд, x	Частота, f	$\lvert x-\overline{x}\rvert \cdot f$	$\left(x-\overline{x}\right)^2 \cdot f$
1	3	9	27
2	6	12	24
3	8	8	8
4	13	0	0
5	22	22	22
6	6	12	24
Разом	58	72	105

Середнє значення кваліфікаційного розряду робітників:

$$\overline{x}=\frac{1\cdot 3+2\cdot 6+3\cdot 8+4\cdot 13+5\cdot 22+6\cdot 6}{58}=4,09\approx 4$$

Визначимо абсолютні показники варіації ознаки.
Розмах варіації дорівнює:

$$R = x_{\max} - x_{\min} = 6 - 1 = 5 \text{ (розрядів)}.$$

Обчислимо лінійне та середнє квадратичне відхилення. За даними таблиці 1 отримуємо:

$$\overline{d} = \frac{\sum \left| x - \overline{x} \right| \cdot f}{\sum f} = \frac{72}{58} = 1,24 \approx 1 \text{ (розряд)};$$

$$\sigma = \sqrt{\frac{\sum \left(x - \overline{x} \right)^2 \cdot f}{\sum f}} = \sqrt{\frac{108}{58}} = 1,35 \approx 1 \text{ (розряд)}.$$

Знайдемо відносні показники варіації.

$$V_R = \frac{R}{\overline{x}} \cdot 100\% = \frac{5}{4,09} \cdot 100\% = 125,2\%;$$

$$V_{\overline{d}} = \frac{\overline{d}}{\overline{x}} \cdot 100\% = \frac{1,24}{4,09} \cdot 100\% = 30,3\%;$$

$$V = \frac{\sigma}{\overline{x}} \cdot 100\% = \frac{1,35}{4,09} \cdot 100\% = 33\%.$$

Оскільки коефіцієнт варіації ознаки становить 33%, то сукупність можна вважати однорідною, а середню величину $\overline{x} = 4$ – типовою характеристикою сукупності.

Приклад 2. За даними таблиці 2 знайти абсолютні та відносні показники варіації заробітної платні працівників підприємства.

Таблиця 2. Розподіл працівників за розміром заробітної платні

Заробітна платня, г.о.	Серединне значення заробітної платні, $\overline{x}_i$, г.о.	Частота, Кількість працівників, f чол.	Частка, d, %	Накопичена	
				частота, чол.	частка, %
До 120	104,5	8	12,3	8	12,3
120-150	134,5	10	15,4	18	27,7
150-180	164,5	15	23,1	33	50,8
180-210	194,5	18	27,7	51	78,5
210-240	224,5	9	13,8	60	92,3

Понад 240	254,5	5	7,7	65	100,0
Разом	–	65	100	–	–

Розмах варіації заробітної платні становить

$$R = x_{\max} - x_{\min} = 269 - 90 = 179 \text{ (г.о.)}$$

Ознака сукупності подана у вигляді інтервального варіаційного ряду. Для розрахунку середнього лінійного та середнього квадратичного відхилення побудуємо таблицю 3. Середнє значення заробітної платні $\overline{x} = 176$ г.о. отримуємо з таблиці 2.

Таблиця 3. Дані для розрахунку середнього лінійного та середнього квадратичного відхилення заробітної платні працівників

Середня в інтервалі заробітна платня, г.о., $\overline{x}_i$	Кількість працівників, f	$\lvert\overline{x}_i - \overline{x}\rvert \cdot f$	$\left(\overline{x}_i - \overline{x}\right)^2 \cdot f$
104,5	8	572,32	40943,77

134,5	10	415,40	17255,72
164,5	15	173,10	1997,57
194,5	18	332,28	6133,89
224,5	9	436,14	21135,34
254,5	9	392,30	30779,86
Разом	65	2321,54	118246,15

За даними цієї таблиці отримуємо наступні значення абсолютних показників варіації:

$$\bar{d} = \frac{\sum |x - \bar{x}| \cdot f}{\sum f} = \frac{2321,54}{65} = 35,72 \text{ (г.о.)};$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}} = \sqrt{\frac{118246,15}{65}} = 42,65.$$

Обчислимо відносні показники варіації.

$$V_R = \frac{R}{\bar{x}} \cdot 100\% = \frac{179}{176,04} \cdot 100\% = 101,7\%;$$

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\% = \frac{35,72}{176,04} \cdot 100\% = 20,3\%;$$

$$V = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{42,65}{176,04} \cdot 100\% = 24,2\%.$$

Оскільки коефіцієнт варіації $V = 24,2\% < 33\%$, то можна вважати, що структура є однорідною, а середня величина заробітної платні є типовою та надійною характеристикою сукупності.

Розглянемо приклад, де сукупність спостережень поділяється на кілька груп.

Приклад 3. У таблиці 4 наведено дані щодо рівня місячного сукупного доходу двох груп населення чисельністю 100 та 300 осіб. Групувальною ознакою є наявність вищої освіти. Порівняти варіацію сукупного місячного доходу у цих групах.

Таблиця 4. Розподіл населення за рівнем сукупного місячного доходу

Рівень доходу x, г.о.	Серединне значення ознаки, $\overline{x}_i$, г.о.	Частота, осіб		
		Загалом, f_i	Особи з вищою освітою, f_{i1}	Особи без вищої освіти, f_{i2}
До 50	25,5	13	1	12
51-100	75,5	124	37	87

101-150	125,5	134	18	116
151-200	175,5	50	6	44
201-250	225,5	49	21	28
251-300	275,5	18	10	8
301-350	325,5	6	3	3
Понад 350	375,5	6	4	2
Разом	–	400	100	300

Виконаємо розрахунок середнього рівня та коефіцієнта варіації доходу у обох групах. Результати розрахунку зводимо у таблицю 5. Як видно з цієї таблиці, середній рівень доходу для осіб з вищою освітою є більшим за середній рівень доходу осіб без вищої освіти на 29,7 г.о.

Оскільки коефіцієнти варіації ознаки набули значення, що перевищують 33% ($V_1 = 55,4\%$, $V_2 = 47\%$), то обидві виділені групи населення є неоднорідними, тобто середні значення для них не є надійними характеристиками. У таких випадках як характеристики центру розподілу використовують моду та медіану.

Таблиця 5. Розрахунок коефіцієнтів варіації

$\overline{x}_i \cdot f_i$	$\overline{x}_i \cdot f_{i1}$	$\overline{x}_i \cdot f_{i2}$	$\left(x_i - \overline{x}\right)^2 f_i$	$\left(x_i - \overline{x}_1\right)^2 f_i$	$\left(x_i - \overline{x}_2\right)^2 f_i$
331,5	25,5	306	166732,3	18360,25	134323,7
9362	2793,5	6568,5	496069,7	270479,25	270886,7
16817	2259	14558	23525,4	22684,5	3902,2
8775	1053	7722	67528,1	1261,5	85960,2
11049,5	4735,5	6314	368752,6	87365,25	248461,9
4959	2755	2204	336610,1	131102,5	166349,1
1953	976,5	976,5	209253,4	81180,75	113140,9
2253	1502	751	336303,4	184041	119267,3
55500	16100	39400	2004775	796475	1142292
$\overline{x} = 138,75$	$\overline{x}_1 = 161$	$\overline{x}_2 = 131,3$	$\sigma^2 = 5011,9$	$\sigma_1^2 = 7964,75$	$\sigma_2^2 = 3807,6$
			$\sigma = 70,8$	$\sigma_1 = 89,25$	$\sigma_2 = 61,7$
			$V = 51\%$	$V_1 = 55,4\%$	$V_2 = 47\%$

Знайдемо значення моди та медіани для всієї сукупності та кожної з груп за формулами Орженцького

$$M_0 = \underline{x}_{M_o} + h \cdot \frac{f_{M_o} - f_{M_o - 1}}{\left(f_{M_o} - f_{M_o-1}\right)\left(f_{M_o} - f_{M_o+1}\right)}.$$

та Фехнера

$$M_e = \underline{x}_{M_e} + h_{M_e} \cdot \frac{\frac{1}{2}\sum f - S_{M_e - 1}}{f_{M_e}}.$$

$$M_o^1 = 51 + 49 \cdot \frac{37-1}{(37-1)+(37-18)} =$$

$$= 133,07(г.о.);$$

$$M_o^2 = 101 + 49 \cdot \frac{116-87}{(116-87)+(116-4)} = ;$$

$$= 115,07(г.о.);$$

$$M_e^1 = 101 + 49 \cdot \frac{50-38}{18} = 133,67(г.о.);$$

$$M_e^2 = 101 + 49 \cdot \frac{150-99}{116} = 122,54(г.о.).$$

Отже, найбільш поширеним рівнем доходу для першої групи є значення близько 133 г.о., для другої – близько 115 г.о. Для половини осіб з першої групи величина місячного доходу не перевищує 133,67 г.о., для половини осіб з другої групи – 122, 54 г.о.

Неоднорідність виділених груп свідчить про те, що у межах відокремлення груп за ознакою «наявність вищої освіти» на рівень місячного доходу суттєво впливають інші ознаки (стать, кваліфікація тощо).

Якщо одиниці сукупності характеризуються ознакою, яка властива або невластива їм, то цю ознаку називають *альтернативною*. Альтернативна ознака може набувати одне з двох кількісних значень: $x_1 = 1,\ x_2 = 0$.

Для дослідження альтернативної ознаки використовують середню арифметичну, дисперсію та середнє квадратичне відхилення. Вони обчислюються за формулами:

$$\overline{x} = \frac{x_1 \cdot p + x_2 \cdot q}{p+q} = \frac{1 \cdot p + 0 \cdot q}{1} = p; \tag{8}$$

$$\sigma^2 = \frac{\left(x_1 - p\right)^2 \cdot p + \left(x_2 - p\right)^2 \cdot q}{p+q} =$$

$$= \frac{\left(1 - p\right)^2 \cdot p + \left(0 - q\right)^2 \cdot q}{1} = \tag{9}$$

$$= q^2 p + p^2 q = qp\left(q + p\right) = qp;$$

$$\sigma = \sqrt{qp}\,. \qquad (10)$$

Тут p – частка одиниць сукупності, для яких значення ознаки $x_1 = 1$, q– частка одиниць сукупності, для яких $x_2 = 0$, $p + q = 1$.

Показники варіації альтернативної ознаки використовують під час обробки даних соціологічних досліджень, організації вибіркових спостережень, статистичного контролю якості продукції тощо.

Нехай статистичну сукупність можна розділити на декілька груп за певними ознаками. Для характеристики впливу групувальних ознак на загальну варіацію ознаки сукупності використовують наступні показники: *загальну дисперсію*, *групову дисперсію*, *середню з групових дисперсію*, *міжгрупову дисперсію*. Для інтервального варіаційного ряду вони обчислюються за формулами:

$$\sigma^2 = \frac{\sum\left(\overline{x}_i - \overline{x}\right)^2 f_i}{\sum f_i}, \qquad (11)$$

$$\sigma_j^2 = \frac{\sum\left(\overline{x}_i - \overline{x}_j\right)^2 f_{ij}}{\sum f_{ij}}, \tag{12}$$

$$\overline{\sigma}^2 = \frac{\sum \sigma_j^2 f_j}{\sum f_j}; \quad f_j = \sum_i f_{ij}; \tag{13}$$

$$\delta^2 = \frac{\sum\left(\overline{x} - \overline{x}_j\right)^2 f_j}{\sum f_j}. \tag{14}$$

У формулах (11)-(14) $\overline{x}_j$ – середня для j-ої групи, $\overline{x}_i$ – серединне значення ознаки x у i-му інтервалі, f_j – обсяг j-ої групи, f_i – частота появи значення ознаки $\overline{x}_i$ у загальній сукупності (частота i-го інтервалу), f_{ij} частота появи значення ознаки $\overline{x}_i$ у j-ій групі, $\overline{x}$ – середня для сукупності, σ^2 – загальна дисперсія, σ_j^2 – внутрішньогрупова дисперсія j-ої групи, $\overline{\sigma}^2$ – середня з внутрішньогрупових дисперсій, δ^2 – міжгрупова дисперсія.

Якщо групові дисперсії не дорівнюють нулю, то у межах виділених груп залишилась варіація, зумовлена впливом ознак, відмінних від ознаки групування. Загальною мірою впливу цих ознак (факторів) є середня з групових дисперсій $\bar{\sigma}^2$.

Міру впливу групувальної ознаки на утворення загальної дисперсії характеризує міжгрупова дисперсія δ^2.

Можна довести, що виконується рівність:

$$\sigma^2 = \bar{\sigma}^2 + \delta^2, \tag{15}$$

тобто загальна дисперсія дорівнює сумі середньої з групових дисперсій $\bar{\sigma}^2$ та міжгрупової дисперсії δ^2

Останню рівність можна звести до рівності:

$$1 = \frac{\bar{\sigma}^2}{\sigma^2} + \frac{\delta^2}{\sigma^2}.$$

Перший доданок у цій рівності показує, яка частка загальної варіації зумовлена дією ознак, відмінних від ознаки групування, а другий доданок – частку, зумовлену дією ознаки групування.

Відношення $\eta^2 = \dfrac{\delta^2}{\sigma^2}$ називають <u>*коефіцієнтом детермінації*</u>, а квадратний корінь з коефіцієнта детермінації – *емпіричним кореляційним відношенням*:

$$\eta = \sqrt{\eta^2} = \sqrt{\frac{\delta^2}{\sigma^2}}, \eta \in [0; 1]. \qquad (16)$$

Коефіцієнт детермінації характеризує частку варіації ознаки, зумовлену впливом ознаки, покладеної у основу групування, тобто він показує, яку частину загальної дисперсії ознаки становить міжгрупова дисперсія.

Емпіричне кореляційне відношення використовується для кількісної оцінки щільності зв’язку між ознакою сукупності та ознакою, за якою здійснювалося групування. У таблиці 6 наведено залежність між значенням емпіричного кореляційного відношення та щільністю цього зв’язку.

Таблиця 6. Залежність між числовим значенням η та щільністю зв’язку між ознакою сукупності та групувальною ознакою

Емпіричне кореляційне відношення	Оцінка щільності зв'язку
0-0,1	відсутній
0,1-0,3	слабкий
0,3-0,5	помірний
0,5-0,7	відчутний
0,7-0,9	щільний
0,9-1,0	дуже щільний

Приклад 4. За даними прикладу 3 встановити, чи впливає наявність вищої освіти на рівень сукупного місячного доходу.

Розрахуємо загальну, групові та міжгрупову дисперсії за формулами (11)-(14):

$$\sigma^2 = 5011,9 \text{ (розраховано у прикладі 3);}$$

$$\bar{\sigma}^2 = \frac{\sigma_1^2 f_1 + \sigma_2^2 f_2}{f_1 + f_2} =$$

$$= \frac{7964,75 \cdot 100 + 3807,6 \cdot 300}{100 + 300} = 4847;$$

$$\delta^2 = \sigma^2 - \bar{\sigma}^2 = 5011.9 - 4847 = 164,9.$$

За цими даними отримуємо таке значення коефіцієнта детермінації:

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{164,9}{5011,9} = 0,033.$$

Знайдемо значення емпіричного кореляційного відношення $\eta = \sqrt{\eta^2} = \sqrt{0,033} = 0,18$. Воно вказує на слабкість зв'язку між рівнем доходу та наявністю вищої освіти.

При вивченні варіації альтернативної ознаки у декількох групах загальну, групову та міжгрупову дисперсії розраховують за формулами:

$$\sigma^2 = \overline{p} \cdot \overline{q}, \quad \overline{p} = \frac{\sum p_j \cdot f_j}{\sum f_j}, q = 1 - p. \qquad (17)$$

Тут p_j – частка одиниць сукупності, яким властива ознака у j-ій групі, f_j – чисельність j-ої групи, $\overline{p}$ – частка одиниць сукупності, яким властива ознака у всій сукупності, $\overline{q}$ – частка одиниць сукупності, яким не властива ознака у всій сукупності.

Середня з групових дисперсій обчислюється за формулою:

$$\sigma^2 = \frac{\sum p_j q_j f_j}{\sum f_j}, \tag{18}$$

де q_j – частка одиниць сукупності, яким не властива ознака у j-ій групі.

Міжгрупову дисперсію обчислюють за формулою:

$$\delta^2 = \frac{\sum \left(p_j - \overline{p}\right)^2 \cdot f_j}{\sum f_j}. \tag{19}$$