ТЕМА 6. ПОКАЗНИКИ ВАРІАЦІЇ

6.1 Поняття варіації та її основні показники.

6.2 Дисперсія, її різновиди, властивості та взаємозв'язок дисперсій.

Автор Олександр Маценко, PhD

6.1 ПОНЯТТЯ ВАРІАЦІЇ ТА ЇЇ ОСНОВНІ ПОКАЗНИКИ

Для повноти аналізу і кращого обґрунтування його результатів використовують ***показники варіації***.

Варіація (від лат. variatio) – зміна значень ознаки у одиниць сукупності в один і той же момент або період часу.

У статистико-економічному аналізі застосовуються такі показники варіації:

1) *розмах варіації* – (R);
2) *середнє лінійне відхилення* – ($\overline{d}$);
3) *дисперсія* (середній квадрат відхилень) – (σ^2);
4) *середнє квадратичне відхилення* – (σ);
5) *коефіцієнт варіації* – ($\overline{V}$).

Розмах варіації

Варіаційний розмах (амплітуда коливань) є найпростішим показником варіації ознаки. Обчислюється як різниця між найбільшим і найменшим значеннями кількісної ознаки в деякій сукупності.

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

де R – розмах варіації;
x_{max} – найбільше значення ознаки;
x_{min} – найменше значення ознаки.

Розмах варіації має **істотний недолік**: він не дає уявлення про ступінь коливання ознаки всередині сукупності, оскільки обчислюється на основі тільки двох крайніх значень ознаки. Ці ознаки можуть мати випадкові характеристики, не типові для даної сукупності. Тому цей показник використовується *для попередньої, орієнтовної* оцінки варіації.

Розмах варіації

ПРИКЛАД. Є дані про виробіток робітниками 2-х бригад, у складі 5-и чоловік. Яку із бригад доцільно найняти для виконання робіт?

№ з/п	І бр.	ІІ бр.
1	8	4
2	2	5
3	3	4
4	5	3
5	2	4
Всього	20	20

Середній виробіток по бригадах:

$$\bar{x}_1 = \bar{x}_2 = \frac{20}{5} = 4$$

Обчислимо розмах варіації:

$$R_1 = x_{max} - x_{min} = 8 - 2 = 6 \qquad R_2 = x_{max} - x_{min} = 5 - 3 = 2$$

Висновок: одержані дані свідчать, що ІІ бригада більш однорідна за виробітком і кваліфікацією.

Середнє лінійне відхилення

Більш досконалим показником, який враховує відхилення значень кожної ознаки від середньої є **середнє лінійне відхилення (СЛВ)**.

***СЛВ** є середньою арифметичною з відхилень окремих абсолютних значень ознаки від їх середнього значення. Може бути простим і зваженим.*

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n}$$ – *просте СЛВ*

$$\bar{d} = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i}$$ – *зважене СЛВ*

де $\bar{d}$ – середнє лінійне відхилення;

x_i – конкретне значення варіанти ознаки;

$\bar{x}$ – середнє значення ознаки;

n – число варіантів;

f_i – частоти.

Середнє лінійне відхилення

ПРИКЛАД:

Табельний номер робітника	I бр.		II бр.	
	x_i	$\left\|x_i - \overline{x}\right\|$	x_i	$\left\|x_i - \overline{x}\right\|$
1	8	4	4	0
2	2	2	5	1
3	3	1	4	0
4	5	1	3	1
5	2	2	4	0
Всього:	20	10	20	2

$$\overline{d}_1 = \frac{\sum \left|x_i - \overline{x}\right|}{n} = \frac{10}{5} = 2 \ шт \qquad \overline{d}_2 = \frac{\sum \left|x_i - \overline{x}\right|}{n} = \frac{2}{5} = 0{,}4 \ шт$$

Висновок: у середньому виробіток робітників I бригади відхиляється від середньої на 2 шт. Тобто, II бригада більш однорідніша за виробітком.

6.2 ДИСПЕРСІЯ, ЇЇ РІЗНОВИДИ, ВЛАСТИВОСТІ ТА ВЗАЄМОЗВ'ЯЗОК ДИСПЕРСІЙ

На цей час, в міру розвитку математичної статистики, замість СЛВ частіше використовують середній квадрат відхилень, так звану **дисперсію** (σ^2) і середнє квадратичне відхилення (σ).

Дисперсія є середньою величиною з квадратів відхилень окремих значень ознаки від їх середньої арифметичної.

Дисперсія визначається за такими формулами:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

проста

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$$

зважена

Дисперсія

ПРИКЛАД:

Табельний номер робітника	І бр.			ІІ бр.		
	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1	8	4	16	4	0	0
2	2	-2	4	5	1	1
3	3	-1	1	4	0	0
4	5	1	1	3	-1	1
5	2	-2	4	4	0	0
Всього:	20	10	26	20	2	2

$$\sigma_1^2 = \frac{26}{5} = 5{,}2; \qquad \sigma_2^2 = \frac{2}{5} = 0{,}4;$$

Отже дисперсія виражає *міру ступеня коливання ознаки*, що вивчається. Вона враховує знак відхилення і на відміну від СЛВ не має розмірності.

Середнє квадратичне відхилення

Середнє квадратичне відхилення (СКВ) є коренем квадратним з дисперсії.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$ просте СКВ

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f}{\sum f}}$$ зважене СКВ

Для нашого прикладу: $\sigma_1^2 = \sqrt{5{,}2} = 2{,}3;$ $\sigma_2^2 = \sqrt{0{,}4} = 0{,}6;$

Смисловий зміст СКВ такий самий, як і СЛВ – чим менше σ, тим сукупність однорідніша, тим типовіша середня і стабільніше явище чи процес.

СКВ завжди більше СЛВ, а в симетричних або помірно асиметричних розподілах:

$$\sigma \approx 1{,}25\bar{d} \qquad \sigma \approx \frac{1}{3}\bar{x}$$

Особливості показників варіації

СЛВ вимірює і узагальнює відхилення від середньої, не вносячи того, що не пов'язане з абсолютними розмірами відхилень.

СКВ і дисперсія, підносячи відхилення до квадрату, зменшують питому вагу малих відхилень і збільшують питому вагу великих відхилень в загальній сумі відхилень.

Тобто *структура лінійного відхилення краще відображає реальні відхилення*, ніж структура квадратичного відхилення.

Варіаційний розмах, середнє лінійне відхилення і середнє квадратичне відхилення завжди виражаються в іменованих одиницях, тобто в тих самих, що і ознака, що вивчається (гривнях, тонах, літрах і т.д.).

Властивості дисперсії

1) Якщо всі значення варіюючої ознаки зменшити на будь-яку сталу величину *k*, то дисперсія при цьому не зміниться:

$$\sigma^2_{x_i - k} = \sigma^2_x$$

2) Якщо всі значення варіант поділити на будь-яку сталу величину *k*, то дисперсія зменшиться в k^2 разів.

$$\sigma^2_{\frac{x_i}{k}} = \frac{\sigma^2}{k^2}$$

3) Дисперсія дорівнює різниці середньої з квадратів варіант і квадрату їх середньої:

$$\sigma^2 = \overline{x^2} - (\bar{x})^2 = \left[\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right]$$

4) Сума квадратів відхилень індивідуальних значень ознаки від середньої менша суми квадратів їх відхилень від будь-якого значення числа *k*:

$$\sigma^2 = \sum (x_i - \bar{x})^2 < \sum (x_i - k)^2$$

Різновиди дисперсій

На варіацію ознаки впливають різні причини і фактори. Одні з них можуть бути випадкові, інші – систематичні. Тому і варіація може бути випадковою і систематичною.

Визначити роль кожної з них можна шляхом розрахунку і аналізу різних видів дисперсії:

- загальної;
- міжгрупової;
- середньої із групових (залишкової).

Загальна (σ^2) – характеризує загальну варіацію ознаки під впливом усіх факторів.

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n};$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f}{\sum f}$$

Різновиди дисперсій

<u>Міжгрупова</u> (δ^2) – характеризує систематичну варіацію результатної ознаки під впливом **факторної (<u>групувальної</u>) ознаки.** Тобто вона характеризує коливання групових середніх навколо загальної середньої

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f_i}{\sum f_i}$$

<u>Середня із групових</u> ($\overline{\sigma}^2$) – розраховується як середня із групових дисперсій (σ_i^2)тобто із дисперсій всередині кожної групи.

Групові дисперсії дисперсії характеризують відхилення окремих значень ознаки в групі від середньої арифметичної даної групи. Групова дисперсія розраховуються як загальна дисперсія, або спрощеним способом:

$$\sigma_i^2 = \frac{\sum \left(x_i - \bar{x}_j\right)}{n} = \overline{x^2} - (\bar{x})^2$$

Різновиди дисперсій

Тобто внутрішньогрупова дисперсія (σ_i^2) характеризує випадкову варіацію, що виникає під впливом усіх інших факторів, окрім фактора-ознаки, за яким здійснюється групування.

Якщо відомі значення групових дисперсій, то *середня із групових* розраховується за формулою:

$$\overline{\sigma}^2 = \frac{\sum \sigma_i^2 f_i}{\sum f_i};$$

Взаємозв'язок дисперсій

Між дисперсіями існує співвідношення:

$$\sigma^2 = \delta^2 + \overline{\sigma}^2,$$

яке називається <u>**правилом додавання дисперсій**</u>

Емпіричний коефіцієнт детермінації

В статистико-економічному аналізі часто застосовуються показник, який характеризує **частку міжгрупової дисперсії у загальній**. Він називається **<u>емпіричним коефіцієнтом детермінації.</u>**

$$\eta^2 = \frac{\delta^2}{\sigma^2};$$

Цей показник, а також корінь із нього

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$$

- емпіричне кореляційне відношення

використовують для оцінки щільності зв'язку між групувальною ознакою (фактором) та результатною ознакою.

Альтернативні ознаки

Серед варіаційних ознак є ознаки, які мають одні із одиниць сукупності, але не мають інші. Такі ознаки називають **альтернативними** (науковий ступінь у викладача, наявність дітей чи авто в сім'ї і т. ін.).

Альтернативні ознаки можуть приймати всього 2 значення: **0 або 1**. При цьому, якщо частку одиниць, які володіють даною ознакою (мають дану ознаку) позначити ***p***, а тих, що не мають – ***q***, то ***p* + *q* = 1**, тоді дисперсія альтернативної ознаки = добутку частки одиниць, що мають дану ознаку на частку одиниць, які не мають даної ознаки:

$$\sigma^2 = pq \qquad \sigma = \sqrt{pq}$$

Граничне значення дисперсії альтернативної ознаки = 0,25=$p \cdot q$=0,5·0,5. Її ще позначають так: $\sigma^2 = \omega(1-\omega)$

Відносні показники варіації

Коли ознаки виражені в різних одиницях виміру і порівняти варіацію неможливо для вимірювання варіації часто розраховують відносні показники, виражені у %.

$$v_R = \frac{R}{\bar{x}} \cdot 100\%$$

– коефіцієнт осциляції характеризує коливання крайніх значень ознаки навколо середньої.

$$v_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\%$$

– лінійний коефіцієнт варіації, характеризує відсоток середнього відхилення ознак від середньої.

$$v_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%$$

– квадратичний коефіцієнт варіації. Якщо $v \leq 33\%$ - *сукупність однорідна, середня типова*; якщо $v \geq 33\%$ - *сукупність неоднорідна, середня нетипова*.

Дякую за увагу!