ЗМІСТОВИЙ МОДУЛЬ 2.
МЕТОДОЛОГІЧНІ ОСНОВИ СТАТИСТИЧНОГО ОЦІНЮВАННЯ ЗАКОНОМІРНОСТЕЙ РОЗВИТКУ. МЕТОДОЛОГІЯ ВИБІРКОВОГО СПОСТЕРЕЖЕННЯ

ТЕМА 5. СТАТИСТИЧНЕ ВИВЧЕННЯ ВАРІАЦІЇ І ФОРМИ РОЗПОДІЛУ

1. Суть варіації та завдання її статистичного аналізу.

Варіація ознаки

в статистиці називається відмінність індивідуальних значень ознаки всередині сукупності, що вивчається.

Термін „варіація” пішов від латинського слова variatio – зміна, коливання, відмінність. Під варіацією розуміють мінливість, коливання значень ознаки у сукупності. Варіацію можна вивчати як на основі рядів розподілу, так і за індивідуальними даними. Основні причини формують центр розподілу, а сукупна їх дія – форму розподілу.

Головні завдання вивчення варіації

три групи показників

– характеристики центру розподілу (середня, мода і медіана);

– характеристики розміру та ступеня варіації;

– характеристики виду та типу розподілу.

Абсолютні показники варіації: економічний зміст та способи обчислення.

Для оцінювання розміру варіації використовується система абсолютних показників, які розглядаються як абсолютна міра варіації.

Розмах варіації (R) → максимальна амплітуда коливань значень ознаки в сукупності: $R = x_{max} - x_{min}$.

В інтервальних рядах розподілу розмах варіації визначається як різниця між верхньою межею останнього та нижньою межею першого інтервалу.

Проте, коли частоти крайніх варіант надто малі, варіаційний розмах неадекватно характеризує варіацію. У таких випадках використовують квартильні або децильні розмахи.

Квартильний розмах $R_Q = Q_3 - Q_1$ охоплює 50% обсягу сукупності,
децильний $R_{D_2} = D_8 - D_2$ — 60% або $R_{D_1} = D_9 - D_1$ — 80%.

Узагальнюючою характеристикою варіації є ***середнє*** відхилення:

Лінійне відхилення

$$\bar{l} = \frac{\sum\left|X - \overline{X}\right|}{n}$$

де X – індивідуальні значення ознаки;

$\bar{X}$ – середнє значення ознаки;

n– кількість одиниць у сукупності.

Лінійне відхилення інтервального ряду

$$\bar{l} = \frac{\sum\left|X - \overline{X}\right| f}{\sum f}$$

де X– варіанти; f– частоти.

Середнє квадратичне відхилення (σ) — показує середній розмір відхилень значень ознаки від середнього рівня. Залежно від вихідних даних використовують середнє квадратичне відхилення просте і зважене:

Середне
Квадратичне
Відхилення
(стандартне)

$$\sigma = \sqrt{\frac{\sum (X - \overline{X})^2}{n}} = \sqrt{\sigma^2} - \text{просте};$$

$$\sigma = \sqrt{\frac{\sum (X - \overline{X})^2 f}{\sum f}} = \sqrt{\sigma^2} - \text{зважене}.$$

Середнє квадратичне відхилення найчастіше використовують у статистичному аналізі, тому його також називають стандартним відхиленням. Слід мати на увазі, що при симетричному розподілі одиниць сукупності .

$$\sigma = 1{,}25l$$

Відносні показники варіації використовуються:

– для оцінки ступеня варіації;

– для порівняння варіації різних ознак;

– для порівняння варіації однієї ознаки у різних сукупностях.

Дисперсія (середній квадрат відхилень)

$$\sigma^2 = \frac{\sum (X - \overline{X})^2}{n}$$

проста

$$\sigma^2 = \frac{\sum (X - \overline{X})^2 f}{\sum f}$$

зважена

де X – індивідуальні значення ознаки;

$\bar{X}$ – середнє значення ознаки;

n– кількість одиниць у сукупності.

Для спрощення розрахунків використовують формули:.

$$\sigma^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2 \text{ або } \sigma^2 = \frac{\sum X^2 f}{\sum f} - \left(\frac{\sum Xf}{\sum f}\right)^2.$$

В інтервальних рядах розподілу з рівними інтервалами дисперсію можна визначити методом «моментів» за формулою:

$$\sigma^2 = i^2\left(m_2 - m_1^2\right),$$

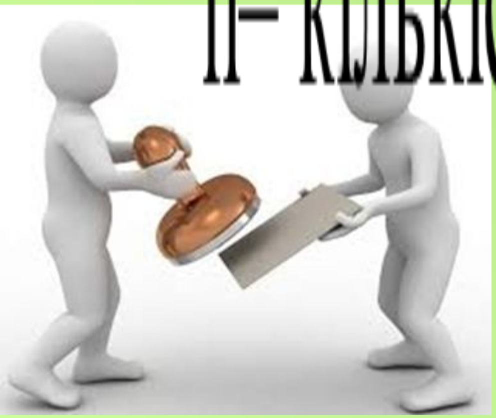
$$m_1 = \frac{\sum \left(\frac{X-A}{i}\right)^2 f}{\sum f} - \text{момент 1 порядку} \qquad m_2 = \frac{\sum \left(\frac{X-A}{i}\right)^2 f}{\sum f} - \text{момент 2 порядку},$$

де X – індивідуальні значення ознаки;

$\sigma = 1{,}25\bar{l}$ $R = 6\sigma$

$\bar{X}$ – середнє значення ознаки;

n– кількість одиниць у сукупності.



На основі взаємозв’язку між варіаційним розмахом *R*, середнім квадратичним відхиленням і чисельністю сукупності *n* Пірсон обчислив коефіцієнти *k*, за допомогою яких орієнтовно можна визначити середнє квадратичне відхилення за варіаційним розмахом:

$$\sigma \approx kR$$

КОЕФІЦІЄНТИ *k* ДЛЯ РІЗНОГО ОБСЯГУ СУКУПНОСТІ

n	10	20	30	40	50	100	200
k	0,32	0,27	0,24	0,23	0,22	0,20	0,18

Очевидний взаємозв’язок середнього квадратичного відхилення та дисперсії: $\sigma = \sqrt{\sigma^2}$. Дисперсія входить до більшості теорем теорії ймовірностей, які є фундаментом математичної статистики, і широко використовується для вимірювання зв’язку й перевірки статистичних гіпотез.

3.Відносні показники варіації.

Відносні показники варіації використовуються:

– для оцінки ступеня варіації;

– для порівняння варіації різних ознак;

– для порівняння варіації однієї ознаки у різних сукупностях.

У загальному вигляді відносні показники варіації визначаються за формулою:

$$K_в = \frac{Абсолютний\ показник\ варіації}{Середня\ величина}.$$

Існує 12 варіантів обчислення відносного показника варіації:

коефіцієнт осциляції

$$K_o = \frac{R}{\overline{X}}$$

лінійний коефіцієнт варіації

$$K_l = \frac{l}{\overline{X}}$$

квадратичний коефіцієнт варіації

$$K_V = \frac{\sigma}{\overline{X}}$$

На практиці переважно використовують коефіцієнт варіації такого виду:

$$V = \frac{\sigma}{\overline{X}} x100.$$

Вважається, що сукупність є однорідною, якщо $V \leq 33\%$.

Крім цього, наведений коефіцієнт варіації застосовують для оцінки ступеня варіації:

$V < 15\%$ – варіація слабка;

$15 \leq V \leq 25\%$ – середня;

$V > 25\%$ – сильна.

Загальну варіацію ознаки можна розкласти на дві складові – систематичну та випадкову.

Для цього необхідно виконати аналітичне групування, у якому досліджувана ознака є результативною, а групувальна ознака розглядається як систематичний фактор.

Якщо центр розподілу поданий медіаною, то за відносну міру варіації беруть квартильний коефіцієнт варіації

$$V_Q = \frac{Q_3 - Q_1}{2\mathrm{Me}}$$

Для оцінювання ступеня варіації застосовують також співвідношення децилів. Так, коефіцієнт децильної диференціації показує кратність співвідношення дев’ятого та першого децилів:

$$V_D = \frac{D_9}{D_1}$$

Необхідні для розрахунку узагальнюючих характеристик варіації величини подано в табл.2 на прикладі розподілу домогосподарств за рівнем забезпеченості житлом. Середня розподілу становить 9 м^2.

ДО РОЗРАХУНКУ УЗАГАЛЬНЮЮЧИХ ХАРАКТЕРИСТИК ВАРІАЦІЇ

x_j	f_j	$x_j - \bar{x}$ $(\bar{x} = 9)$	$\lvert x_j - \bar{x} \rvert f_j$	$(x_j - \bar{x})^2$	$(x_j - \bar{x})^2 f_j$
4	17	–5	85	25	425
6	39	–3	117	9	351
8	51	–1	51	1	51
10	42	1	42	1	42
12	29	3	87	9	261
14	15	5	75	25	375
16	7	7	49	49	343
Разом	200	x	506	x	1848

$$\bar{l} = \frac{\sum_{1}^{m} |x_j - \bar{x}| f_j}{\sum_{1}^{m} f_j} = \frac{506}{200} =$$ 2,53 m2

$$\sigma^2 = \frac{\sum_{1}^{m} (x_j - \bar{x})^2 f_j}{\sum_{1}^{m} f_j} = \frac{1848}{200} = 9{,}24$$

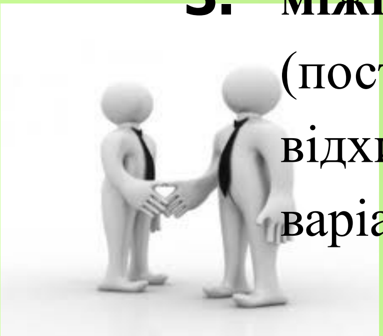
$$\sigma = \sqrt{9{,}24} = 3{,}04$$ m2

$$V_\sigma = \frac{\sigma}{\bar{x}} = \frac{3{,}04}{9} = 0{,}34$$

Децильний коефіцієнт $V_D = 13{,}3 : 5{,}2 = 2{,}5$ показує, що нижня межа 10% відносно забезпечених житлом домогосподарств в 2,5 рази перевищує верхню межу 10% малозабезпечених домогосподарств

4.Міжгрупова та внутрішньогрупова дисперсії. Правило додавання дисперсій. Види дисперсій

1. **загальна (χρ2)** – є результатом впливу усіх факторів, що спричинили варіацію ознаки, як постійних (систематичних), так і випадкових; характеризує варіацію ознаки навколо загальної середньої;
2. **групова (часткова) (χρі)** – є результатом впливу випадкових факторів (усіх крім фактора, покладеного в основу групування), характеризує варіацію ознаки у межах групи навколо групової середньої; узагальнюючою мірою внутрішньогрупової варіації є **середня з групових дисперсій (χρі)сер**;
3. **міжгрупова дисперсія (δ2)** – є результатом впливу фактора (постійного), який покладено в основу групування; характеризує відхилення групових середніх від загальної, тобто систематичну варіацію.



Дисперсія має певні математичні властивості.

1. Якщо всі значення варіант xj зменшити на сталу величину А, то дисперсія не зміниться: $\sigma^2_{(x-A)} = \sigma^2_x$

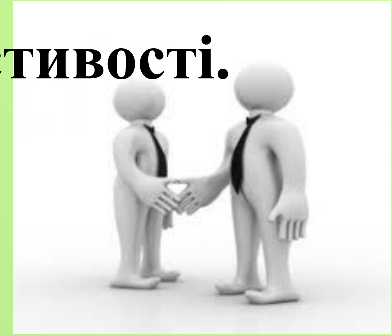
2. Якщо всі значення варіант xj змінити в А раз, то дисперсія зміниться в А2 раз: $\sigma^2_{xA} = \sigma^2_x A^2$

3. Якщо частоти замінити частками, дисперсія не зміниться.

1. Дисперсія — це різниця квадратів $\overline{x^2} - \bar{x}^2$

Де $\bar{x}^2$ — квадрат середньої величини;
$\overline{x^2}$ — середній квадрат значень ознаки.

Дисперсія має певні математичні властивості.



1. **Дисперсія альтернативної ознаки обчислюється як добуток часток:**

$$\sigma^2 = d_1 d_0$$

2. **d1 — частка елементів сукупності, яким властива ознака,**
3. **d0 — частка решти елементів .** $(d_0 = 1 - d_1)$

Якщо сукупність розбито на групи за певною ознакою х, то для будь-якої іншої ознаки у можна обчислити дисперсію як у цілому по сукупності, так і в кожній групі.

Центром розподілу сукупності в цілому є загальна середня $\bar{y} = \frac{\sum_1^n y}{n}$

центром розподілу в j-й групі — групова середня $\bar{y}_j = \frac{\sum_1^{f_i} y}{f_j}$

Відхилення індивідуальних значень ознаки у від загальної середньої $\bar{y}$ можна подати як дві складові:

$$(y - \bar{y}) = (y - \bar{y}_j) + (\bar{y}_j - \bar{y})$$

Узагальнюючими характеристиками цих відхилень є дисперсії: загальна, групова та міжгрупова.

Загальна дисперсія характеризує варіацію ознаки y навколо загальної середньої

$$\sigma^2 = \frac{\sum_1^n (y-\bar{y})^2}{n}$$

Групова дисперсія характеризує варіацію відносно групової середньої

$$\sigma_j^2 = \frac{\sum_1^{f_i} (y-\bar{y}_j)^2}{f_j}$$

Узагальнюючою мірою внутрішньогрупової варіації є **середня з групових дисперсій**

$$\overline{\sigma^2} = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j}$$

Мірою варіації їх навколо загальної середньої є міжгрупова дисперсія

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j-\bar{y})^2 f_j}{\sum_1^m f_j}$$

Дисперсія складається з двох частин.
Перша характеризує внутрішньогрупову,
друга — міжгрупову варіацію

Взаємозв'язок дисперсій називається правилом розкладання (декомпозиції) варіації:

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

«Правило додавання дисперсій» використовується для того, щоб розкласти загальну варіацію результативної ознаки на систематичну та випадкову.
При цьому мірою систематичної варіації є міжгрупова дисперсія δ , а випадкової – середня внутрішньогрупова дисперсія ($\overline{\sigma_i^2}$).

Приклад розрахунку абсолютних і відносних показників варіації за індивідуальними значеннями показника.

Маємо дані про загальну площу п'ятнадцяти обстежених двокімнатних квартир (перша і друга графи таблиці):

Номер з/п	Загальна площа, м^2	$\lvert X - \overline{X} \rvert$	$(X - \overline{X})^2$	X^2
1	76,3	22,7	515,29	5821,69
2	54,2	0,6	0,36	2937,64
3	41,7	11,9	141,61	1738,89
4	51,6	2,0	4,00	2662,56
5	49,3	4,3	18,49	2430,49
6	60,4	6,8	46,24	3648,16
7	52,4	1,2	1,44	2745,76
8	48,2	5,4	29,16	2323,24
9	40,3	13,3	176,89	1624,09
10	64,0	10,4	108,16	4096,00
11	54,5	0,9	0,81	2970,25
12	48,7	4,9	24,01	2371,69
13	62,2	8,6	73,96	3868,84
14	51,8	1,8	3,24	2683,24
15	48,9	4,7	22,09	2391,21
Разом	**804,5**	**99,5**	**1165,75**	**44313,75**

Для розрахунку розмаху варіації знайдемо максимальне і мінімальне значення ознаки: $X_{max} = 76{,}3$; $X_{min} = 40{,}3$.

Розмах варіації становить: $R = X_{max} - X_{min} = 76{,}3 - 40{,}3 = 36{,}0\,м^2$

Розрахунок середнього лінійного відхилення розпочнемо з визначення середнього значення, яке становить:

$$\overline{X} = \frac{\sum X}{n} = \frac{804{,}5}{15} = 53{,}6\,м^2$$

Результати розрахунку модулів відхилень варіант від середньої занесено у третю графу таблиці. Середнє лінійне відхилення дорівнює:

$$\bar{l} = \frac{\sum \left| X - \overline{X} \right|}{n} = \frac{99{,}5}{15} = 6{,}6\,м^2$$

Дисперсію обчислимо двома способами, для чого визначимо квадрати відхилень значень ознаки від середньої $(X - \overline{X})^2$ та квадрати значень ознаки (X2), які занесемо у четверту і п'яту графи таблиці.

Дисперсія дорівнює:

$$\sigma^2 = \frac{\sum (X - \overline{X})^2}{n} = \frac{1165{,}75}{15} = 77{,}7$$

$$\sigma^2 = \frac{\sum X^2}{n} - \overline{X}^2 = \frac{44313{,}75}{15} - 53{,}6^2 = 2954{,}25 - 2876{,}5 = 77{,}7$$

Середнє квадратичне відхилення становить: $\sigma = \sqrt{\sigma^2} = \sqrt{77{,}7} = 8{,}8\ м^2$

Для оцінки ступеня варіації обчислимо відносні показники:

– коефіцієнт осциляції: $Ko = \frac{R}{\overline{\overline{X}}} = \frac{36{,}0}{53{,}6} * 100 = 67{,}1\%$

– лінійний коефіцієнт варіації: $K_l = \frac{l}{\overline{\overline{X}}} = \frac{6{,}6}{53{,}6} * 100 = 12{,}4\%$

– квадратичний коефіцієнт варіації: . $K_v = \frac{\sigma}{\overline{\overline{X}}} = \frac{8{,}8}{53{,}6} * 100 = 16{,}4\%$

Розглянемо розрахунок дисперсій на прикладі варіації якості твердого сиру *у* залежно від терміну його зберігання *х*. Результати вибіркового обстеження якості 20 партій сиру, розподіл їх за терміном зберігання (1, 2, 3 місяці), розрахунки середніх та дисперсій наведено в табл.4 Згідно з даними таблиці маємо:

1) середній бал якості сиру (за 10-бальною шкалою)

$$\bar{y} = \frac{\sum_{1}^{n} y}{n} = \frac{148}{20} = 7{,}4$$

2)загальна дисперсія балів якості

$$\sigma^2 = \frac{\sum_{1}^{n} (y - \bar{y})^2}{n} = \frac{26{,}26}{20} = 1{,}313$$

3) групові середні бали якості та групові дисперсії:

$$\bar{y}_1 = \frac{60{,}9}{7} = 8{,}7; \qquad \sigma_1^2 = \frac{0{,}36}{7} = 0{,}051;$$

$$\bar{y}_2 = \frac{57{,}6}{8} = 7{,}2; \qquad \sigma_2^2 = \frac{1{,}4}{8} = 0{,}175;$$

$$\bar{y}_3 = \frac{29{,}5}{5} = 5{,}9; \qquad \sigma_3^2 = \frac{1{,}1}{5} = 0{,}220.$$

РОЗРАХУНОК ЗАГАЛЬНОЇ ТА ГРУПОВИХ ДИСПЕРСІЙ ЯКОСТІ СИРУ

№ з/п	Термін зберігання x, міс	Бал якості, y	Розрахунок дисперсій якості						
			загальної	групових σ_i^2					
				1-ша група		2-га група		3-тя група	
			$(y-\bar{y})^2$	y	$(y-\bar{y}_1)^2$	y	$(y-\bar{y}_2)^2$	y	$(y-\bar{y}_3)^2$
1	2	7,3	0,01			7,3	0,01		
2	1	8,8	1,96	8,8	0,01				
3	1	8,4	1,00	8,4	0,09				
4	3	6,5	0,81					6,5	0,36
5	2	7,5	0,01			7,5	0,09		
6	3	6,4	1,00					6,4	0,25
7	1	9,1	2,89	9,1	0,16				
8	1	8,6	1,44	8,6	0,01				
9	3	5,7	2,89					5,7	0,04
10	2	6,8	0,36			6,8	0,16		
11	2	7,7	0,09			7,7	0,25		
12	3	5,6	3,24					5,6	0,09
13	1	8,9	2,25	8,9	0,04				
14	2	7,8	0,16			7,8	0,36		
15	3	5,3	4,41					5,3	0,36
16	1	8,5	1,21	8,5	0,04				
17	2	6,8	0,36			6,8	0,16		
18	2	7,1	0,09			7,1	0,01		
19	1	8,6	1,44	8,6	0,01				
20	2	6,6	0,64			6,6	0,36		
Разом	—	148,0	26,26	60,9	0,36	57,6	1,4	29,5	1,1
Середня	—	7,4	—	8,7	—	7,2	—	5,9	—
Дисперсія	—	—	1,313	—	0,051	—	0,175	—	0,220

Значення групових середніх підтверджують залежність якості сиру від терміну його зберігання.

У 1-й групі середній бал якості становить 8,7, у 2-й групі якість сиру знижується на 1,5 бала, а в 3-й зниження якості порівняно з першою групою становить 2,8 бала.

Водночас зростає варіація балів у групах, що відбиває посилення впливу інших чинників на якість сиру.

Необхідні величини для розрахунку середньої з групових і міжгрупової дисперсій наведено в табл.5.

ДО РОЗРАХУНКУ МІЖГРУПОВОЇ ТА СЕРЕДНЬОЇ З ГРУПОВИХ ДИСПЕРСІЙ

Групи за терміном зберігання, міс.	Число партій f_j	Середній бал якості $\bar{y}_j$	Групова дисперсія σ_j^2	Розрахунок дисперсій		
				середньої з групових	міжгрупової	
				$\sigma_j^2 f_j$	$\bar{y}_j - \bar{y}$	$(\bar{y}_j - \bar{y})^2 f_j$
1	7	8,7	0,051	0,36	1,3	11,83
2	8	7,2	0,175	1,40	– 0,2	0,32
3	5	5,9	0,220	1,10	– 1,5	11,25
Разом	20	7,4	×	2,86	×	23,4

За даними таблиці міжгрупова дисперсія становить

$$\delta^2 = \frac{\sum_1^m \left(\bar{y}_j - \bar{y}\right)^2 f_j}{\sum_1^m f_j} =$$

$$= \frac{(8,7-7,4)^2 7 + (7,2-7,4)^2 8 + (5,9-7,4)^2 5}{20} = \frac{23,4}{20} = 1,170;$$

середня з групових дисперсій

$$\overline{\sigma^2} = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j} = \frac{0,051 \cdot 7 + 0,175 \cdot 8 + 0,220 \cdot 5}{20} = \frac{2,86}{20} = 0,143$$

Сума їх дорівнює загальній дисперсії: 0,143 + 1,170 = 1,313.

Міжгрупова варіація — це результат впливу фактора, який покладено в основу групування, внутрішньогрупова — інших факторів, окрім групувального. Відношення міжгрупової дисперсії до загальної характеризує частку варіації результативної ознаки *y*, яка пов'язана з варіацією групувальної ознаки. Це відношення називають ***кореляційним*** і позначають символом η^2:

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

У розглянутому прикладі кореляційне відношення становить $\eta^2 = 1{,}17 : 1{,}313 = 0{,}842$, тобто 84,2% варіації якості сиру пов'язані з терміном зберігання. На інші фактори припадає 100–84,2=15,8% варіації.

Правило декомпозиції варіації є основою вимірювання щільності зв'язку.

Дякую за увагу!