1. Понятие, причины и показатели вариации
2. Расчеты дисперсии сокращенными способами
3. Виды дисперсий. Правило сложений дисперсий
4. Дисперсия альтернативного признака

Тема 6. ПОКАЗАТЕЛИ ВАРИАЦИИ

1. Понятие, причины и показатели вариации

Вариация – различие значений какого-либо признака у разных единиц совокупности за один и тот же промежуток времени.

Причиной возникновения вариации являются различные условия существования разных единиц совокупности.

Определение вариации необходимо при выборочном наблюдении, статистическом моделировании, проведении экспертных опросов.

По степени вариации судят об однородности совокупности, устойчивости значений признака, типичности средней, о взаимосвязи между признаками.

Данные о заработной плате рабочих двух бригад за январь 2010 г.

Бригада	Заработная плата отдельных рабочих, руб.					Средняя заработная плата, руб.
1	5340	6200	9120	7894	10370	7803
2	7340	7450	7400	8460	8364	7803

При оценке социально-экономических явлений нельзя ограничиваться расчетом только средней величины, надо знать её устойчивость и масштабы отклонения от средней.

Для измерения степени колеблемости отдельных значений признака от средней исчисляются абсолютные и относительные показатели вариации.

Абсолютные показатели **вариации:**

- размах вариации;
- среднее линейное отклонение;
- дисперсия;
- среднее квадратическое отклонение.

Относительные показатели **вариации:**

- коэффициент вариации;
- коэффициент осцилляции и др.

Размах вариации (*R*) является наиболее простым способом измерения колеблемости и рассчитывается как разность между максимальным и минимальным значениями признака:

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

Среднее линейное отклонение ($\overline{d}$) **рассчитывается как средняя из модулей отклонений вариантов признака от средней. Для ранжированного ряда:**

$$\overline{d} = \frac{\sum_{i=1}^{n} | x_i - \overline{x} |}{n}$$

Для вариационного ряда распределения:

$$\bar{d} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| \times f_i}{\sum f_i}$$

где xi – индивидуальные значения признака;

$\bar{x}$ – среднее значение признака;

n – объем совокупности;

f – частота значений признака.

Дисперсия (σ^2) – это средняя арифметическая квадратов отклонений отдельных значений признака от их средней арифметической.

Дисперсия является основным показателем вариации.

Дисперсия для ранжированного ряда рассчитывается по формуле:

$$\sigma^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}{n}$$

Для вариационного ряда распределения:

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \times f_i}{\sum f_i}$$

Среднее квадратическое отклонение (σ) - **представляет собой квадратный корень из среднего квадрата отклонений отдельных значений признака от их средней, т. е. корень квадратный из дисперсии.**

Для ранжированного ряда показатель рассчитывается по формуле:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n}}$$

Для вариационного ряда распределения:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \times f_i}{\sum_{i=1}^{n} f_i}}$$

Коэффициент вариации (γ) **является относительным показателем вариации и представляет собой процентное отношение среднего квадратического отклонения к средней арифметической:**

$$\gamma = \frac{\sigma \times 100}{\bar{x}}, \%$$

По коэффициенту вариации судят о степени вариации признака, об однородности совокупности.

Совокупность считается *однородной*, а среднее значение признака – *типичным* для совокупности, если коэффициент вариации *не превышает 33-35 %.*

<u>Пример 1:</u> Данные о распределении работников банка по стажу работы:

Стаж, лет	Численность работников, чел.
до 3	8
3-5	50
5-7	30
7-9	8
свыше 9	4
Итого	100

Определить:

- ✓ Средний стаж работников
- ✓ Размах вариации
- ✓ Среднее линейное отклонение
- ✓ Дисперсию
- ✓ Среднее квадратическое отклонение
- ✓ Коэффициент вариации

Для расчета показателей вариации в интервальных рядах его приводят к дискретному ряду, т.е. находят *середины интервалов.*

Стаж, лет	Численность работников, чел. *fi*	Середина интервала *xi*
до 3	8	2
3-5	50	4
5-7	30	6
7-9	8	8
свыше 9	4	10
Итого	100	-

Стаж, лет	Численность работников, чел. *fi*	Середина интервала *xi*	*xf*
до 3	8	2	16
3-5	50	4	200
5-7	30	6	180
7-9	8	8	64
свыше 9	4	10	40
Итого	100	-	500

При определении среднего стажа работников используем формулу средней арифметической взвешенной.

$$\bar{x} = \frac{\sum x \times f}{\sum f} = \frac{500}{100} = 5\ \text{лет}$$

Размах вариации определяем по формуле:

$$R = x_{max} - x_{min} = 10 - 2 = 8 \text{ (лет).}$$

Стаж, лет	Численность работников, чел. f_i	Середина интервала x_i	xf	$\|x-\tilde{x}\|$	$\|x-\tilde{x}\|f$
до 3	8	2	16	3	24
3-5	50	4	200	1	50
5-7	30	6	180	1	30
7-9	8	8	64	3	24
свыше 9	4	10	40	5	20
Итого	100	-	500	-	148

Среднее линейное отклонение:

$$\overline{d} = \frac{\sum \left| x - \overline{x} \right| f}{\sum f} = \frac{148}{100} = 1{,}48$$

Стаж, лет	Числ-ть работ-в, чел. fi	Середина интервала xi	xf	$\lvert x-\tilde{x}\rvert$	$\lvert x-\tilde{x}\rvert f$	$(x-\tilde{x})^2$	$(x-\tilde{x})^2 f$
до 3	8	2	16	3	24	9	72
3-5	50	4	200	1	50	1	50
5-7	30	6	180	1	30	1	30
7-9	8	8	64	3	24	9	72
свыше 9	4	10	40	5	20	25	100
Итого	100	-	500	-	148	-	324

Дисперсия:

$$\sigma^2 = \frac{\sum\left(x-\overline{x}\right)^2 f}{\sum f} = \frac{324}{100} = 3{,}24$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\sigma^2} = \sqrt{3{,}24} = 1{,}8 \quad \text{лет}$$

Коэффициент вариации:

$$V = \frac{\sigma \cdot 100}{\overline{x}} = \frac{1{,}8 \cdot 100}{5} = 36\%$$

2. Расчеты дисперсии сокращенными способами

Дисперсия сокращенным способом определяется по формуле:

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(\frac{xi - A}{k}\right)^2 \times f_i}{\sum f_i} \times k^2 - \left(\bar{x} - A\right)^2$$

где k – величина интервала; A – условный нуль, в качестве которого удобно использовать середину интервала с наибольшей частотой.

Формула расчёта дисперсии по «методу моментов»:

$$\sigma^2 = k^2 \times (m_2 - m_1^2),$$

где:

$$m_1 = \frac{\sum_{i=1}^{n}\left(\frac{xi - A}{k}\right)\times f_i}{\sum f_i}$$

– *момент первого порядка;*

$$m_2 = \frac{\sum_{i=1}^{n} \left(\frac{x_i - A}{k} \right)^2 \times f_i}{\sum f_i}$$

– момент второго порядка.

В случае, когда $A = 0$ и, следовательно, не вычисляются отклонения, формула расчета дисперсии «от условного нуля» принимает вид:

$$\sigma^2 = \square x^2 - \left(\bar{x}\right)^2$$

$$\sigma^2 = \frac{\sum x_i^2 \times f_i}{\sum f_i} - \left[\frac{\sum_{i=1}^{n} \left(x_i \times f_i \right)}{\sum f_i} \right]^2$$

Пример 2: По данным о стаже работников банка определить дисперсию по способу «от условного нуля» и «методом моментов»:

Стаж работы, лет	**Численность работников, чел.**
до 3	**10**
3-5	**48**
5-7	**28**
7-9	**10**
свыше 9	**4**
Итого	***100***

1. Средний стаж работы:

$$\overline{x'} = \frac{\sum x'_i \times f_i}{\sum f_i}$$

$$\overline{x} = \overline{x'} \times k + A$$

Упростим варианту и найдем суммарное значение $x'f$

Стаж работы, лет	Численность работников, чел.	Середина интервала *xi*	*xi-A* *A=6*	$\frac{xi-A}{k}$ *k=2* x'	$\frac{(xi-A)\,fi}{k}$ $x'f$
до 3	10	2	-4	-2	-20
3-5	48	4	-2	-1	-48
5-7	28	6 - *A*	0	0	0
7-9	10	8	2	1	10
свыше 9	4	10	4	2	8
Итого	*100*	-	-	-	*-50*

$$\bar{x}=\left[\frac{(-50)}{100}\right]\times 2+6\ \textbf{\textit{лет}}\,(1)+6=5$$

2. Дисперсия по способу расчёта «от условного нуля»:

$$\sigma^2 = \frac{\sum_{i=1}^{n} \left(\frac{x_i - A}{k} \right)^2 \times f_i}{\sum f_i} \times k^2 - \left(\bar{x} - A \right)^2$$

Достроим таблицу двумя графами:

$\left(\frac{x_i - A}{k}\right)^2$	$\left(\frac{x_i - A}{k}\right)^2 \times f_i$
4	**40**
1	**48**
0	**0**
1	**10**
4	**16**
-	**114**

Рассчитаем дисперсию «от условного нуля»:

$$\sigma^2 = \frac{114}{100} \times 2^2 - (5-6)^2 = 3,56$$

По способу "моментов" получаем:

$$\sigma^2 = k^2 \times (m_2 - m_1^2),$$

$$\sigma^2 = 2^2 \times \left[\frac{114}{100} - \left(\frac{50}{100} \right)^2 \right] = 3{,}56$$

3. Виды дисперсий. Правило сложений дисперсий

Если совокупность разбивается на группы по изучаемому признаку, то для этой совокупности вычисляются **дисперсии:**

- общая (σ^2);
- внутригрупповые (частные) (σ^2_i);
- средняя из внутригрупповых ($\overline{\sigma_i^2}$);
- межгрупповая (δ^2).

«правило сложения дисперсий»

Общая дисперсия равна сумме средней из внутригрупповых и межгрупповой дисперсий:

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

Внутригрупповая дисперсия **отражает случайную вариацию, происходящую под влиянием неучтённых факторов.**

$$\sigma_i^2 = \frac{\sum_{i=1}^{n} \left(x_i - \tilde{x}\right)^2 \times f_i}{\sum_{i=1}^{n} f_i}$$

Средняя из внутригрупповых дисперсий:

$$\overline{\sigma}_i^{\,2} = \frac{\sum_{i=1}^{n} \sigma_i^2 \times f_i}{\sum_{i=1}^{n} f_i}$$

Межгрупповая дисперсия является объясненной дисперсией.

Она вызывается влиянием фактора, положенного в основу группировки.

$$\delta^2 = \frac{\sum \left(\tilde{x}_i - \bar{x}_{общ} \right)^2 \times f_i}{\sum f_i}$$

Пример 3: Определить общую дисперсию стажа рабочих двух бригад по правилу сложения дисперсий

Распределение рабочих по стажу работы

1 бригада		2 бригада	
Стаж, лет x	Число рабочих, чел fi	Стаж, лет x	Число рабочих, чел fi
1-4	3	1-4	4
4-10	8	4-10	7
10-15	3	10-15	6
15-20	2	15-20	3
Итого	16	Итого	20

Введем в таблицу 1 значения, необходимые для расчета дисперсии по *1* бригаде:

Стаж, лет *x*	Число рабочих, чел *fi*	Середина интервала, *xi*	*xifi*
1-4	3	2,5	7,5
4-10	8	7	56
10-15	3	12,5	37,5
15-20	2	17,5	35
Итого	16	-	136

Средний стаж у рабочих *1 бригады*:

$$\widetilde{x}_i = \frac{\sum x_i f_i}{\sum f_i} = \frac{136}{16} = 8{,}5$$

лет

Введем в таблицу 2 значения, необходимые для расчета дисперсии по 2 бригаде:

Стаж, лет x	Число рабочих, чел fi	Середина интервала, xi	$xifi$
1-4	4	2,5	10
4-10	7	7	49
10-15	6	12,5	75
15-20	3	17,5	52,5
Итого	20	-	186,5

Средний стаж у рабочих <u>2 бригады</u>:

$$\tilde{x}_i = \frac{\sum x_i f_i}{\sum f_i} = \frac{186{,}5}{20} = 9{,}3$$

лет

Введем данные для расчета внутригрупповой дисперсии по 1 бригаде:

Стаж, лет x	Число рабочих, чел fi	Середина интервала, xi	$xifi$	$(x_i - \tilde{x}_i)^2 \times f_i$
1-4	3	2,5	7,5	108
4-10	8	7	56	18
10-15	3	12,5	37,5	48
15-20	2	17,5	35	162
Итого	16		136	336

Внутригрупповая дисперсия для рабочих:

$$\sigma i^{2} = \frac{\sum_{i=1}^{n} \left(x_i - \widetilde{x}\right)^{2} \times f_i}{\sum_{i=1}^{n} f_i}$$

Внутригрупповая дисперсия по 1 бригаде:

$$\sigma_i^2 = \frac{336}{16} = 21$$

Введем данные для расчета внутригрупповой дисперсии по 2 бригаде:

Стаж, лет x	Число рабочих, чел fi	Середина интервала, xi	$xifi$	$\left(x_i - \tilde{x}_i\right)^2 \times f_i$
1-4	4	2,5	10	184,96
4-10	7	7	49	37,3
10-15	6	12,5	75	61,44
15-20	3	17,5	52,5	201,72
Итого	20	-	186,5	485,15

Внутригрупповая дисперсия по 2 бригаде:

$$\sigma_i^2 = \frac{485{,}15}{20} = 24{,}26$$

Средняя из внутригрупповых дисперсий:

$$\overline{\sigma}_i^{\,2} = \frac{\sum_{i=1}^{n} \sigma_i^2 \times f_i}{\sum_{i=1}^{n} f_i}$$

$$\overline{\sigma}_i^2 = \frac{21 \times 16 + 24{,}26 \times 20}{16 + 20} = 22{,}8$$

Для межгрупповой дисперсии δ^2 рассчитаем средний стаж для рабочих двух бригад вместе:

$$\bar{x} = \frac{8{,}5 \times 16 + 9{,}3 \times 20}{16 + 20} = 8{,}9(\textit{лет})$$

Определим межгрупповую дисперсию δ2:

$$\delta^2 = \frac{\sum \left(\tilde{x}_i - \bar{x}_{общ}\right)^2 \times f_i}{\sum f_i}$$

$$\delta^2 = \frac{(8{,}5 - 8{,}9)^2 \times 16 + (9{,}3 - 8{,}9)^2 \times 20}{16 + 20} = 0{,}16$$

Используя правило сложений дисперсий, определяем общую дисперсию:

$$\sigma^2 = 22{,}8 + 0{,}16 = 22{,}96.$$

Для характеристики оценки тесноты связи между факторным и результативным признаками <u>используются показатели</u>:

- **эмпирического коэффициента детерминации (η^2);**
- **эмпирического корреляционного отношения (η).**

Эмпирический коэффициент детерминации :

$$\eta^2 = \frac{\delta^2}{\sigma^2} \times 100,\%$$

Если $\eta^2 = 100$ %, то вариация результативного признака целиком определяется изменением факторного;

если $\eta^2 = 0$, то группировочный фактор не влияет на изменение результативного;

если $0< \eta^2 <100$, то на измерение результативного влияет не только группировочный, но и другие признаки.

Эмпирическое корреляционное отношение :

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

Значения эмпирического корреляционного отношения находятся в пределах от 0 до 1 и характеризуют тесноту связи между факторным и результативным признаком:

0,0-0,3 –связь слабая

0,3-0,5 – умеренная

0,5-0,7 – заметная

0,7-0,9 – высокая

0,9-0,99 – очень высокая.

4. Дисперсия альтернативного признака

Существуют качественные признаки, которые имеют лишь два противоположных значения: имеется этот признак или нет (наличие высшего образования, наличие стипендии, бракованность продукции и т. д.).

Таким качественным признакам можно придать условные количественные значения:

1 – наличие признака;

0 – отсутствие признака

Число единиц совокупности, имеющих признак, обозначим *p*, не имеющих его – *q*.

Признак, *xi*	Частота признака, *fi*
1	**p**
0	**q**

Среднее значение признака во всей совокупности (по формуле средней арифметической взвешенной) будет равно:

$$\bar{x} = \frac{1 \times p + 0 \times q}{p + q} = \frac{p}{p + q} = \frac{p}{1} = p$$

Находим дисперсию альтернативного признака (по обычной формуле):

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

$$\sigma^2 = \frac{\left(1 - p\right)^2 \times p + \left(0 - p\right)^2 \times q}{p + q} =$$

$$= \frac{q^2 \times p + p^2 \times q}{p + q} = \frac{qp\left(q + p\right)}{q + p} = qp$$

Значение дисперсии альтернативного признака <u>не может </u>превышать 0,25, так как:

0,1 х 0,9=0,09

0,2 х 0,8=0,16

0,3 х 0,7=0,21

0,4 х 0,6=0,24

0,5 х 0,5=0,25.

Пример 4: Определить дисперсию альтернативного признака, если известно, что доля преподавателей, имеющих ученые степени, в ВУЗах города составляет 63 %.

~~Есть ученая степень~~ – 63 % – 0,63

~~Нет ученой степени~~ – 37 % – 0,37.

~~Дисперсия равна~~: