Лекция 5. Показатели вариации

Основные показатели вариации

Вариация значений признака представляет наибольший интерес при исследовании социально-экономических явлений и процессов.

Вариация – колеблемость, многообразие, изменяемость величины признака у отдельных единиц совокупности.

Она возникает в результате того, что индивидуальные значения признака складываются под влиянием разнообразных факторов (условий), которые по-разному сочетаются в каждом отдельном случае.

Используемые в статистическом анализе показатели вариации можно разделить на три группы:

- показатели размаха;
- показатели, характеризующие отклонения от среднего уровня;
- относительные показатели вариации.

К показателям размаха относят:

- вариационный размах;
- децильный размах;
- квартильный размах.

К показателям, характеризующим отклонения от среднего, относят:

- среднее линейное отклонение;
- среднее квадратическое отклонение;
- дисперсию.

К относительным показателям относят:

- относительный квартильный размах;
- линейный коэффициент вариации;
- коэффициент вариации.

Показатели размаха

Вариационный размах или **размах вариации** характеризует абсолютную разницу между максимальным и минимальным значениями признака в изучаемой совокупности:

$$R = x_{\max} - x_{\min} \quad (5.1)$$

Основным недостатком данного показателя является то обстоятельство, что максимальные и минимальные значения признака могут быть обусловлены случайными обстоятельствами и в этой связи могут искажать типичный для изучаемой совокупности размах вариации.

Децильный размах (D) характеризует абсолютную разницу между значениями девятой (верхней) и первой (нижней) децилями:

$$D = D_9 - D_1 \quad (5.2)$$

Таким образом, децильный размах характеризует разброс 80% данных и, является более предпочтительным по сравнению с вариационным размахом, так как практически не зависит от экстремальных значений.

Квартильный размах или интерквартильный разброс (IQR) характеризует абсолютную разницу между третьим (верхним) и первым (нижним) квартилями:

$$IQR = \Delta_q = Q_3 - Q_1 \tag{5.3}$$

Третий квартиль (Q3) показывает значение признака больше которого расположено 25% значений. Таким образом квартильный размах характеризует разброс 50% центральных значений.

Среди показателей разброса **наиболее часто в практическом анализе используют квартильный размах**.

Показатели отклонения от среднего

Среднее линейное отклонение. Для абсолютной количественной оценки различий между всеми без исключения значениями признака в изучаемой совокупности используется оценка отклонений фактических значений от их среднего уровня. Чем больше различия между вариантами признака, тем больше и их отклонения от среднего уровня. При этом **сумма отклонений фактических значений от средней всегда равна 0**.

Существует два основных подхода к усреднению отклонений фактических значений от средней. Первый состоит в том, что используют абсолютные значения отклонений и в результате получают показатель, который называется **среднее линейное отклонение**. Второй состоит в том, что отклонения возводят в квадрат и в результате получают **дисперсию** и **среднее квадратическое отклонение**.

Среднее линейное или **среднее абсолютное отклонение** представляет собой среднее арифметическое из абсолютных значений отклонений фактических вариантов признака от среднего значения. В зависимости от характера исходных данных для расчета используют простую или взвешенную форму:

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n} \quad \text{- простая форма;} \tag{5.4}$$

$$\bar{d} = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i} \quad \text{- взвешенная форма,} \tag{5.5}$$

Если данные не сгруппированы, то используют простую форму, если сгруппированы – то взвешенную.

Дисперсия представляет собой средний квадрат отклонений значений признака от средней величины.

В зависимости от характера исходных данных для расчета используют простую или взвешенную формулу:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$ - простая форма; (5.6)

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$ - взвешенная форма, (5.7)

Для расчета дисперсии в отдельных случаях удобнее использовать формулу, которая представляет собой алгебраическое преобразование выражений (5.6) и (5.7):

$$\sigma^2 = \overline{x^2} - (\bar{x})^2$$, где (5.8)

$\overline{x^2}$ - **средняя квадратическая**.

В зависимости от характера исходных данных для расчета средней квадратической используются простая или взвешенная формы:

$$\overline{x^2} = \frac{\sum x_i^2}{n}$$ - простая, (5.9)

$$\overline{x^2} = \frac{\sum x_i^2 \cdot f_i}{\sum f_i}$$ - взвешенная. (5.10)

Если данные не сгруппированы, то используют простую форму, если сгруппированы – то взвешенную.

Поэтому наиболее удобным и широко распространенным на практике показателем вариации является среднее квадратическое отклонение, которое определяется как квадратный корень из дисперсии и имеет ту же размеренность, что и изучаемый признак.

Среднее квадратическое отклонение характеризует среднее отклонение фактических значений признака в статистической совокупности от их среднего значения и рассчитывается на основе следующих формул:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$ - простая форма, (5.11)

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}$$ - взвешенная форма (5.12)

$$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2}$$ (5.13)

Среднее квадратическое отклонение также называют **стандартным отклонением**.

Среднее квадратическое отклонение и среднее линейное отклонение близки друг другу по экономическому смыслу и между ними есть определенная связь. **Для симметричных** или **умеренно ассиметричных распределений** $\bar{d} \approx \frac{4}{5}\sigma$.

Среднее квадратическое отклонение более широко применяется в статистическом анализе по сравнению со средним линейным отклонением благодаря своим математических свойствам. Так среднее квадратическое отклонение является одним из параметров многих распределений и в первую очередь нормального распределения.

В нормальном распределении примерно 2/3 всех значений отклоняются от среднего уровня не больше, чем на одну величину среднего квадратического отклонения. Приблизительно 95% всех значений отклоняются от среднего уровня не более чем на две величины среднего квадратического отклонения. И, наконец, **около 99,7% всех значений лежат в пределах трех средних квадратических отклонений** (правило 3-х сигм).

Относительные показатели вариации

Чтобы оценить масштабы вариации используют **относительные показатели вариации**, которые измеряют изменчивость значений признака в относительном выражении по сравнению со средним уровнем, что во многих случаях является более предпочтительным.

Для оценки относительных размеров вариации используют **коэффициент осцилляции**, **линейный коэффициент вариации** и **квадратический коэффициент вариации,** который называют также просто **коэффициентом вариации**. Относительные показатели вариации, как правило, рассчитывают в процентах.

Коэффициент осцилляции – процентное отношение размаха вариации к средней:

$$V_{\mathrm{R}} = \frac{\mathrm{R}}{\bar{x}} \times 100\% \qquad (5.14)$$

Линейный коэффициент вариации измеряют через соотношение среднего линейного отклонения и средней:

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \times 100\% \qquad (5.15)$$

Коэффициент вариации измеряют через соотношение среднего квадратического отклонения и средней:

$$V_{\sigma} = \frac{\sigma}{\bar{x}} \times 100\% \qquad (5.16)$$

По величине коэффициента вариации можно, в частности, судить о степени однородности признаков совокупности. Чем больше его величина, тем больше разброс значений признаков вокруг средней, тем менее однородна совокупность по своему составу и тем менее представительна средняя.

Под однородными данными понимается некоторый уровень их рассеяния, при котором рассчитываемые статистические показатели (средняя, дисперсия и др.) будут давать надежную и качественную характеристику анализируемой совокупности.

В статистике принято считать, что, **если значение коэффициента вариации менее 33%, то совокупность данных является однородной, если более 33%, то – неоднородной.**

Пример. По данным выборочного обследования произведена группировка вкладчиков по размеру вклада в Сбербанке города:

Размер вклада, руб.	До 400	400 - 600	600 - 800	800 - 1000	Свыше 1000
Число вкладчиков	32	56	120	104	88

Определить:

1) размах вариации;
2) средний размер вклада;
3) среднее линейное отклонение;
4) дисперсию;
5) среднее квадратическое отклонение;
6) коэффициент вариации вкладов.

Решение

Данный ряд распределения содержит открытые интервалы. В таких рядах условно принимается величина интервала первой группы равной величине интервала последующей, а величина интервала последней группы равной величине интервала предыдущей.

Величина интервала второй группы равна 200, следовательно, и величина первой группы также равна 200. Величина интервала предпоследней группы равна 200, значит и последний интервал будет иметь величину, равную 200.

Получаем ряд:

Размер вклада, руб.	200 - 400	400 - 600	600 - 800	800 - 1000	1000 - 1200
Число вкладчиков	32	56	120	104	88

1) Определим **размах вариации** как разность между наибольшим и наименьшим значением признака:

$$R = x_{max} - x_{min} = 1200 - 200 = 1000$$

Размах вариации размера вклада равен **1000** руб.

2) **Средний размер вклада** определим по формуле средней арифметической взвешенной.
Предварительно определим дискретную величину признака в каждом интервале. Для этого найдём середины интервалов. Среднее значение для первого интервала будет равно: (200+400)/2=300, второго - 500 и т. д.
Занесём результаты вычислений в таблицу:

Размер вклада, руб.	**Число вкладчиков, f**	**Середина интервала, x**	**xf**
200-400	32	300	9600
400-600	56	500	28000
600-800	120	700	84000
800-1000	104	900	93600
1000-1200	88	1100	96800
Итого	400	-	312000

Средний размер вклада будет равен:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{300 * 32 + 500 * 56 + 700 * 120 + 900 * 104 + 1100 * 88}{32 + 56 + 120 + 104 + 88}$$
$$= \frac{312000}{400} = 780$$

3) **Среднее линейное отклонение** есть средняя арифметическая из абсолютных отклонений отдельных значений признака от общей средней:

$$\bar{d} = \frac{\sum |x - \bar{x}| f}{\sum f}$$

Порядок расчёта **среднего линейного отклонения в интервальном ряду** распределения следующий:

1. Вычисляется средняя арифметическая взвешенная. Мы в пункте 2 получили это значение равным 780.

2. Определяются абсолютные отклонения вариант от средней:

$$|x - \bar{x}|$$

3. Полученные отклонения умножаются на частоты:

$$|x - \bar{x}| f$$

4. Находится сумма взвешенных отклонений без учёта знака:

$$\sum |x - \bar{x}| f$$

5. Сумма взвешенных отклонений делится на сумму частот:

$$\frac{\sum |x - \bar{x}| f}{\sum f}$$

Удобно пользоваться расчётной таблицей:

Размер вклада, руб.	**Число вкладчиков, f**	**Середина интервала, x**	$x - \bar{x}$	$\lvert x - \bar{x}\rvert$	$\lvert x - \bar{x}\rvert f$
200-400	32	300	-480	480	15360
400-600	56	500	-280	280	15680
600-800	120	700	-80	80	9600
800-1000	104	900	120	120	12480
1000-1200	88	1100	320	320	28160
Итого	400	-	-	-	81280

$$\bar{d} = \frac{\sum|x - \bar{x}|f}{\sum f} = \frac{81280}{400} = 203{,}2$$

Среднее линейное отклонение размера вклада клиентов Сбербанка составляет **203,2** руб.

4) **Дисперсия** - это средняя арифметическая квадратов отклонений каждого значения признака от средней арифметической.

Расчёт дисперсии в интервальных рядах распределения производится по формуле:

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

Порядок расчёта дисперсии в этом случае следующий:

1. Определяется средняя арифметическая взвешенная, как показано в п. 2).
2. Вычисляются отклонения вариант от средней:

$$x - \bar{x}$$

3. Возводятся в квадрат отклонения каждой варианты от средней:

$$(x - \bar{x})^2$$

4. Умножаются квадраты отклонений на веса (частоты):

$$(x - \bar{x})^2 * f$$

5. Суммируются полученные произведения:

$$\sum(x - \bar{x})^2 * f$$

6. Полученная сумма делится на сумму весов (частот):

$$\frac{\sum(x - \bar{x})^2 * f}{\sum f}$$

Расчёты сведём в таблицу:

Размер вклада, руб.	**Число вкладчиков, f**	**Середина интервала, x**	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 * f$
200-400	32	300	-480	230400	7372800
400-600	56	500	-280	78400	4390400
600-800	120	700	-80	6400	768000
800-1000	104	900	120	14400	1497600
1000-1200	88	1100	320	102400	9011200
Итого	400	-	-	-	23040000

$$\sigma^2 = \frac{\sum(x-\bar{x})^2 * f}{\sum f} = \frac{23040000}{400} = 57600$$

5) **Среднее квадратическое отклонение** размера вклада определяется как корень квадратный из дисперсии:

$$\sigma = \sqrt{\sigma^2} = \sqrt{57600} = 240$$

6) **Коэффициент вариации** - это отношение среднего квадратического отклонения к средней арифметической:

$$V = \frac{\sigma * 100\%}{\bar{x}} = \frac{240 * 100\%}{780} = 30{,}77\%$$

Используем полученное значение коэффициента вариации для анализа степени однородности значений исследуемого признака. Так как V=30,77%<33%, то можно считать, что рассматриваемая совокупность значений признака является однородной