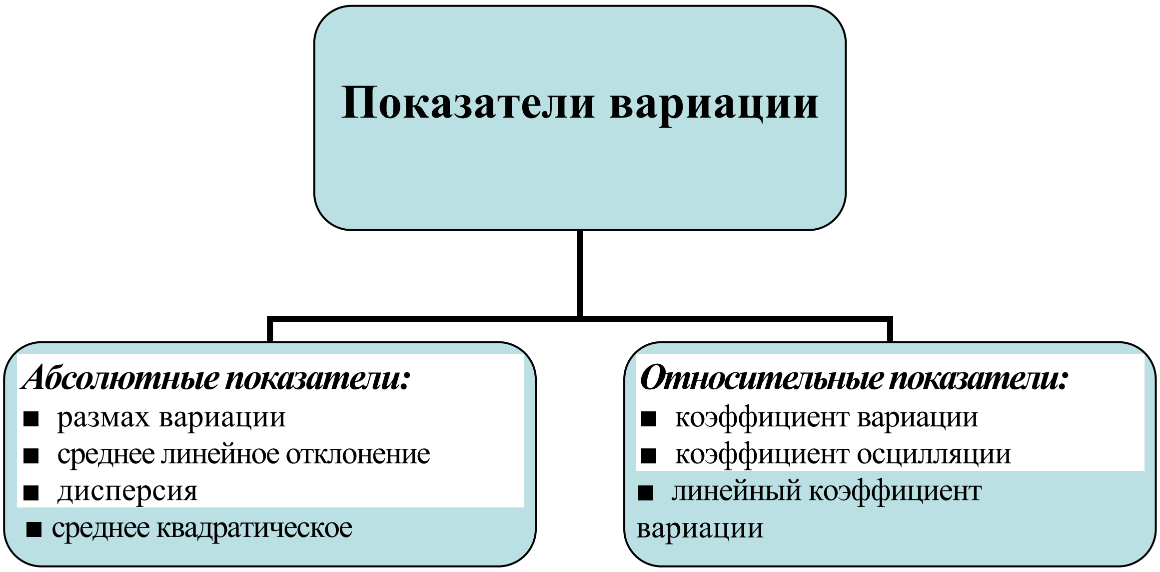
Лекция № 6

Показатели вариации ряда

Помимо показателей среднего значения, для описания распределений широко применяют показатели вариации, характеризующие разброс или отклонение значений ряда от среднего. Показатели вариации дополняют анализ статистических совокупностей.

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



Размах вариации

Самым простым (с точки зрения вычислений) показателем вариации является ***размах вариации***, равный разности наибольшего и наименьшего значения:

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

Размах вариации — это абсолютное отклонение, сохраняет размерность изучаемого признака.

Размах вариации применяется в ряде специальных случаев (разбиение на интервалы, оценка дисперсии), однако в целом этот показатель мало информативен: он полностью определяется двумя крайними значениями ряда и никак не зависит от остальных значений.

Однако крайние значения признака могут быть аномальными для данной совокупности, обусловленными какими-то аномальными или критическими обстоятельствами. Тогда размах вариации будет служить характеристикой только этих двух аномальных единиц совокупности. В этом случае с целью дальнейшего изучения вариации единиц совокупности аномальные единицы следует убрать из совокупности. Но не для того, чтобы забыть о них, а для того, чтобы изучать их отдельно. Аномальные значения могут отражать необычные, яркие явления, которые могут представлять интерес для исследователя как раз своей нетипичностью.

Для получения средних характеристик вариации рассчитывают средние отклонения вариантов признака x_i от средней величины признака $\bar{x} = m$, которые являются наиболее важными и часто употребляемыми показателями вариации: среднее линейное отклонение d, среднее квадратическое отклонения σ и дисперсия D.

Среднее линейное отклонение (d) — средний модуль отклонения вариантов признака от средней арифметической величины признака. Сохраняет размерность изучаемого признака.

Для расчета среднего линейного отклонения d используется формула средней арифметической

простой: $d = \dfrac{\sum_{i=1}^{n} |x_i - m|}{n}$ - для несгруппированных данных;

взвешенной: $d_{взв.} = \dfrac{\sum_{j=1}^{k} |x_j - m| f_i}{\sum_{j=1}^{k} f_j}$ - для вариационного ряда распределения.

Формулы для расчета дисперсии для данных различного типа имеют вид:

не сгруппированные данные	$D = \left((x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_N - m)^2 \right) \Big/ N = \left(\sum_{i=1}^{N} (x_i - m)^2 \right) \Big/ N$
дискретное частотное распределение	$D = \left(\sum_{i=1}^{N} f_i (x_i - m)^2 \right) \Big/ N = \left(\sum_{i=1}^{N} (f_i x_i^2) - m^2 N \right) \Big/ N = \left(\sum_{i=1}^{N} (f_i x_i^2) \right) \Big/ N - m^2$

интервальное частотное распределение	$D=\left(\sum_{i=1}^{N} f_i\left(x_i-m\right)^2\right)\Big/N=\left(\sum_{i=1}^{N}\left(f_i x_i^2\right)-m^2 N\right)\Big/N=\left(\sum_{i=1}^{N}\left(f_i x_i^2\right)\right)\Big/N-m^2$

где K число групп. Среднее квадратическое отклонение равняется корню квадратному из дисперсии, $\sigma=\sqrt{D}$.

Математические свойства дисперсии:

1) Дисперсия постоянной величины равна нулю.

Доказательство: Пусть $C=const$, тогда $M[C]=0$,

$D(C)=M\left[C-M(C)\right]^2=M\left[C-C\right]^2=M[0]=0$

2) Уменьшение всех значений признака на одну и ту же величину A не меняет величины дисперсии.

3) Уменьшение всех значений признака в k раз уменьшает дисперсию в k^2 раз, а среднее квадратическое отклонение — в k раз.

4) Средний квадрат отклонений от любой величины A всегда будет больше среднего квадрата отклонений, вычисленного от средней арифметической, на определенную величину — на квадрат разности средней и этой условно взятой величины, т.е. на $(m-A)^2$.

Пример. Проведено испытание на стойкость двух видов краски (А и В) для наружних работ - из каждой партии случайным образом отобрано по 6 банок, краской
из них сделаны надписи и определено, через сколько времени (в месяцах) надписи начали блекнуть. Получены следующие результаты:

A	10	60	50	30	40	20
B	35	45	30	35	40	25

Рассчитать величины среднего арифметического, дисперсии, среднего квадратического отклонения и сравнить качество краски.

Решение:

$$m_A = (10+60+50+30+40+20)/6 = 210/6 = 35 \text{ месяцев.}$$

$$m_A = (35+45+30+35+40+25)/6 = 210/6 = 35 \text{ месяцев}$$

Средние значения для двух видов краски одинаковы.

$$D_A = \left((10-35)^2+(60-35)^2+(50-35)^2+(30-35)^2+(40-35)^2+(20-35)^2\right)/6 =$$

$= 1950/6 = 325$ месяцев^2; $\sigma_A \approx 18$ месяцев.

$$D_B = \left((35-35)^2+(45-35)^2+(30-35)^2+(35-35)^2+(40-35)^2+(22-35)^2\right)/6 =$$

$= 250/6 \approx 42$ месяцев^2; $\sigma_B \approx 6.5$ месяцев.

Разброс значений для второй краски оказался почти в три раза меньше! Это означает, что предсказуемость ее свойств будет намного выше, а именно это ценится и составляет основу качества любого изделия. В ряде случаев (отклонение в положительную сторону) первая краска может прослужить и дольше второй. В другом же случае, например, для отклонения на $1.5\sigma_A$ в отрицательную сторону, краска вместо почти трех лет прослужит менее 1 года. Итак, качество второй краски выше.

Есть и другие примеры, когда качество изделий напрямую связано с величиной отклонения. Также, дисперсия и среднее квадратическое отклонение широко используются в задачах статистического анализа, например, при проверке гипотез и построении статистических оценок.

Коэффициент вариации

Дисперсия и среднее квадратическое отклонение описывают разброс значений признака в размерных единицах.

$$V = \frac{\sigma}{m}$$

Коэффициент вариации используется для оценки интенсивности вариации.

Как относительный показатель интенсивности, он имеет размерность; показывает, сколько единиц среднего квадратического отклонения приходится на единицу среднего значения изучаемого признака. Совокупность считается однородной, если коэффициент вариации не превышает 0,33. В этом случае средняя величина исследуемого признака может считаться типичной, надежной характеристикой статистической совокупности.

Если же коэффициент вариации больше 0,33, то это означает, что вариация исследуемого признака велика. Найденная средняя плохо представляет всю статистическую совокупность и не может считаться ее типичной характеристикой. Статистическая совокупность является неоднородной по рассматриваемому признаку.

В ряде случаев оказывается удобнее (необходимо) использовать безразмерный показатель вариации - ***коэффициент вариации:***

$$V = \frac{\sigma}{m} \times 100\%$$

Аналогично коэффициенту вариации рассчитывают другие относительные показатели вариации, которые в практике статистики применяются реже:

показатель осцилляции $V_R = \frac{R}{m}$ или $V_R = \frac{R}{m} \times 100\%$,

линейный коэффициент вариации $V_d = \frac{d}{m}$ или $V_d = \frac{d}{m} \times 100\%$.

Пример. В январе 1995 года средняя стоимость квадратного метра муниципального жилья в одном из городов составила 3.88 млн рублей при среднем квадратическом отклонении 0.76 млн, а в мае - 5.82 млн при. отклонении 0.975 млн. В каком месяце реальный разброс цен был больше?

Решение. На первый взгляд кажется, что разброс цен в мае был больше. Однако нужно учитывать, что в тот период в силу инфляции быстро менялся масштаб цен, что проявляется в быстром росте среднего значения. В этом случае более адекватный способом сравнения будет сравнение коэффициентов вариации: $V_1 = \frac{0.76}{3.88} \times 100\% = 19.6\%$; $V_2 = \frac{0.975}{5.82} \times 100\% = 16.8\%$.

Относительный разброс цен уменьшился.

Коэффициент вариации может использоваться также для сравнения разброса величин
разной размерности.

Пример. Данные наблюдений крупного автоцентра показали, что в среднем за день в нем продается 87 автомобилей при среднем квадратическом отклонении 8. Размер дневной прибыли в среднем составлял 15675 У.Е. при среднем квадратическом отклонении 2319 У.Е. Какая из величин имеет больший разброс?

Для сравнения разброса разноименных величин применим коэффициент вариации:

$$V_{avto} = \frac{8}{87} \times 100\% \approx 9.2\%; \quad V_{profit} = \frac{2319}{15675} \times 100\% \approx 14.8\%$$

Разброс значений прибыли выше.

Задача.

Имеется следующее распределение 60 рабочих по тарифному разряду (ряд дискретный):

Тарифный разряд x_i	2	3	4	5	6
Число рабочих f_i	8	16	17	12	7

Рассчитать основные показатели вариации.

Решение:

1. Определяем средний тарифный разряд рабочих, по формуле средней арифметической взвешенной:

$$\overline{x} = m = \frac{\sum_{i=1}^{5} x_i f_i}{\sum_{i=1}^{5} f_i} = \frac{2\cdot 8 + 3\cdot 16 + 4\cdot 17 + 5\cdot 12 + 6\cdot 7}{8+16+17+12+7} = \frac{234}{60} = 3,9\,.$$

2. Рассчитываем таблицу:

x_i	f_i	$\lvert x_i - m \rvert$	$\lvert x_i - m \rvert f_i$	$(x_i - m)^2 f_i$
2	8	1,9	15,2	28,88
3	16	0,9	14,4	12,96
4	17	0,1	1,7	0,17
5	12	1,1	13,2	14,52
6	7	2,1	14,7	30,87
Итого	60	-	59,2	87,40

3. Размах вариации $R = x_{\max} - x_{\min} = 6 - 2 = 4$

4. Среднее линейное отклонение

$$d = \frac{\sum_{i=1}^{5} \lvert x_i - m \rvert f_i}{\sum_{i=1}^{5} f_i} = \frac{15,2+14,4+1,7+13,2+14,7}{8+16+17+12+7} = \frac{59,2}{60} = 0,987$$

5. Дисперсия $D = \frac{\sum_{i=1}^{5} (x_i - m)^2 f_i}{\sum_{i=1}^{5} f_i} = \frac{28,88+12,96+0,17+14,52+30,87}{8+16+17+12+7} = \frac{87,4}{60} = 1,46$

6. Среднее квадратическое отклонение $\sigma = \sqrt{D} = \sqrt{1,46} = 1,21$

7. Коэффициент вариации $V = \frac{\sigma}{m} \times 100\% = \frac{1,21}{3,9} \times 100\% = 31,0\%$ или $V = \frac{\sigma}{m} = \frac{1,21}{3,9} = 0,31$