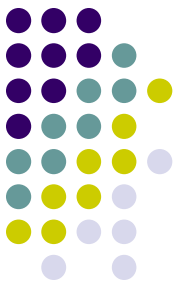


Показатели вариации

Необходимость измерения вариации

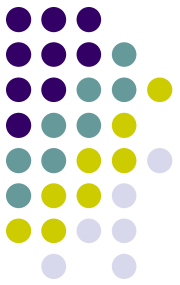


- Средняя величина характеризует совокупность по изучаемому признаку, такой характеристики совокупности будет достаточно, если разброс индивидуальных значений невелик
- Когда ряд характеризуется значительным рассеиванием индивидуальных значений, то применение средней величины ограничено

Необходимость измерения вариации

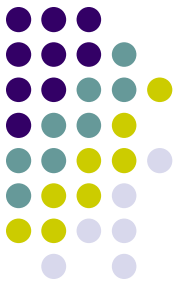


- При значительном рассеивании индивидуальных значений необходимо рассчитать специальную систему показателей, характеризующих средний размер отклонений индивидуальных значений от средней величины и степень колеблемости признака в совокупности, т. е. показателей вариации



Показатели вариации

- Используются две группы показателей вариации:
 - абсолютные: размах вариации, среднее линейное отклонение, дисперсия, среднеквадратическое отклонение
 - относительные: коэффициент осцилляции, линейный коэффициент и коэффициент вариации



1. Размах вариации

- РВ – разность между экстремальными значениями признака в совокупности. РВ имеет единицу измерения, совпадающую с единицей измерения признака у единиц совокупности

Размах вариации

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

Недостаток РВ: он учитывает только крайние значения и не учитывает промежуточные значения

2. Среднее линейное отклонение

Недостаток РВ устраняет показатель СЛО. Он рассчитывается по двум формулам:
а) для несгруппированных данных (по формуле средней арифметической простой)

б) для сгруппированных данных (по формуле средней арифметической взвешенной)

Среднее линейное отклонение

а) для несгруппированных данных

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

б) для сгруппированных данных

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

Среднее линейное отклонение

У СЛО есть единица измерения.

Он обладает серьезным **недостатком**: в числителе нет минуса, а сам показатель – положительное число. Эта проблема решается третьим и четвертым показателями вариации – дисперсией и среднеквадратическим отклонением

3. Дисперсия -

Это средний квадрат отклонений индивидуальных значений от средней величины. Она рассчитывается по простой и взвешенной формулам. Для ее обозначения используется греческая буква сигма.

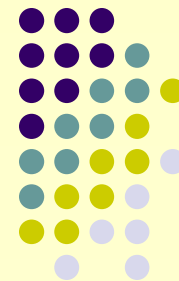
Дисперсия

а) для несгруппированных данных

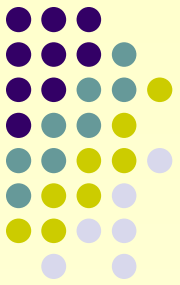
$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

б) для сгруппированных данных

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



Расчет дисперсии для вариационного ряда



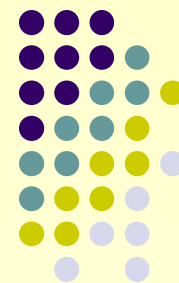
**Осуществляется при помощи
взвешенной формулы:**

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

Выработка, м.	Число рабочих, f	x	xf	x'	$x' \cdot f$	S	$x - \bar{x}$	$(x - \bar{x})^2 \cdot f$	$x'^2 \cdot f$
до 200	3	190	570	-3	-9	3	-63,9	12249,63	27
200-220	12	210	2520	-2	-24	15	-43,9	23126,52	48
220-240	50	230	11500	-1	-50	65	-23,9	28560,50	50
240-260	56	250	14000	0	0	121	-3,9	851,76	0
260-280	47	270	12690	1	47	168	16,1	12182,87	47
280-300	23	290	6670	2	46	191	36,1	29973,83	92
300-320	7	310	2170	3	21	198	56,1	22030,47	63
320 и более	2	330	660	4	8	200	76,1	11582,42	32
Итого:	200		50780		39			140558	359

$$\sigma^2 = \frac{140558}{200} = 702,79$$

$$\sigma = \sqrt{702,79} \approx 26,5(\mathcal{M})$$



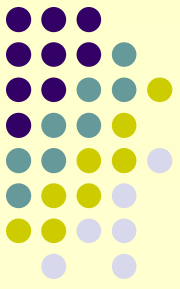
Свойства дисперсии

1. Если из всех вариантов вычесть какую-либо константу, то дисперсия от этого не изменится:

$$\begin{aligned}\sigma_{(x-A)}^2 &= \frac{\sum [(x-A) - (\bar{x}-A)]^2 \cdot f}{\sum f} = \\ &= \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f} = \sigma^2\end{aligned}$$

2. Если все варианты разделить на константу A , то дисперсия уменьшится от этого в A^2 раз:


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




$$= \frac{1}{A^2} \cdot \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f} = \frac{\sigma^2}{A^2}$$

3. Дисперсия равна разности среднего квадрата вариант и квадрата их средней:

$$\begin{aligned}\sigma^2 &= \frac{\sum (x - \bar{x})^2}{n} = \frac{\sum (x^2 - 2x \cdot \bar{x} + \bar{x}^2)}{n} = \\ &= \frac{\sum x^2}{n} - \frac{\sum 2x \cdot \bar{x}}{n} + \frac{\sum \bar{x}^2}{n} =\end{aligned}$$

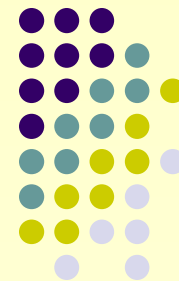

$$= \overline{x^2} - 2 \cdot \bar{x} \cdot \frac{\sum x}{n} + \frac{n \cdot \bar{x}^2}{n}$$

$$= \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 = \overline{x^2} - \bar{x}^2$$


4. Если рассчитать среднее квадратическое отклонение от любой константы A , отличной от средней арифметической, то оно всегда будет больше дисперсии на квадрат разности между средней и данной константой A :

$$\sigma_A^2 = \sigma^2 + (\bar{X} - A)^2 \quad , \text{ где}$$

$$\sigma_A^2 = \frac{\sum (x - A)^2 \cdot f}{\sum f}$$



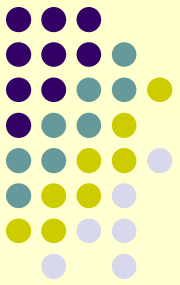
Расчет дисперсии упрощенным способом

Расчет дисперсии упрощенным способом осуществляется на основе перечисленных свойств по формуле:


$$\sigma^2 = \left[\frac{\sum x'^2 \cdot f'}{\sum f'} - \left(\frac{\sum x' \cdot f'}{\sum f'} \right)^2 \right] \cdot h^2 \quad , \text{ где}$$

$$x' = \frac{x - C}{h}; \quad f' = \frac{f}{A}$$

Выработка, м.	Число рабочих, f	x	xf	x'	$x' \cdot f$	S	$x - \bar{x}$	$(x - \bar{x})^2 \cdot f$	$x'^2 \cdot f$
до 200	3	190	570	-3	-9	3	-63,9	12249,63	27
200-220	12	210	2520	-2	-24	15	-43,9	23126,52	48
220-240	50	230	11500	-1	-50	65	-23,9	28560,50	50
240-260	56	250	14000	0	0	121	-3,9	851,76	0
260-280	47	270	12690	1	47	168	16,1	12182,87	47
280-300	23	290	6670	2	46	191	36,1	29973,83	92
300-320	7	310	2170	3	21	198	56,1	22030,47	63
320 и более	2	330	660	4	8	200	76,1	11582,42	32
Итого:	200		50780		39			140558	359




$$\sigma^2 = \left[\frac{359}{200} - \left(\frac{39}{200} \right)^2 \right] \cdot 20^2 = 702,79$$



Недостаток дисперсии состоит в том, что она имеет размерность вариант, возведенную в квадрат (рублей в квадрате, человек в квадрате)

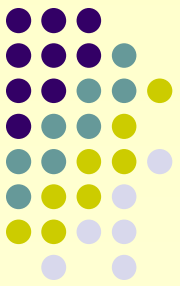
Чтобы устранить этот недостаток, используется *среднее квадратическое отклонение*



4. Среднее квадратическое отклонение

a) для несгруппированных данных

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



б) для сгруппированных данных

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

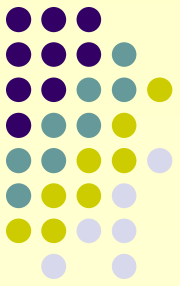
σ представляет собой среднее квадратическое отклонение вариант ряда от средней величины

Среднее квадратическое отклонение

имеет единицы измерения, а также может принимать положительные и отрицательные значения, поскольку получается в результате извлечения квадратного корня.

С помощью СКО можно утверждать, что i -тое значение признака в совокупности находится в пределах:

$$\bar{x} - \sigma \leq x_i \leq \bar{x} + \sigma$$



Относительные показатели вариации

Относительные показатели вариации применяются для решения следующих задач:


- сравнение степени вариации различных вариационных рядов**
- характеристика степени однородности совокупности**

Коэффициент осцилляции


$$V_r = \frac{R}{\bar{x}} \cdot 100\%, \quad \text{где}$$

R - размах вариации

\bar{x} - среднее значение



**Коэффициент осцилляции отражает
относительную колеблемость крайних
значений признака относительно среднего
значения**



Линейный коэффициент вариации

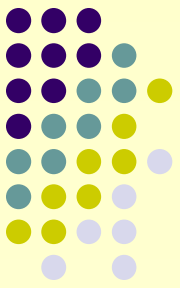
$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\%, \quad \text{где}$$

\bar{d} - среднее линейное отклонение


Коэффициент вариации

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


Характеризует долю усредненного значения отклонений от средней величины. При этом совокупность считается однородной, если V не превышает 33%



$$V = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{26,5}{253,9} = 10,44\%$$



Правило трех сигм




В условиях нормального распределения существует зависимость между величиной σ и количеством наблюдений:

в пределах $\bar{X} \pm \sigma$ располагается 68,3 % наблюдений;

в пределах $\bar{X} \pm 2\sigma$ располагается 94,5 % наблюдений;

в пределах $\bar{X} \pm 3\sigma$ располагается 99,7 % наблюдений



На практике почти не встречаются отклонения, которые превышают 3σ . Отклонение в 3σ может считаться максимальным

При помощи этого правила можно получить примерную оценку σ :


$$\sigma \approx \frac{X_{\max} - X_{\min}}{6}$$







Дисперсия альтернативного признака






Признаки, которыми обладают одни единицы совокупности и не обладают другие, называются ***альтернативными***. Количественно вариация альтернативного признака проявляется в значении 0 у единиц, которые им не обладают, или в значении 1 у единиц, которые им обладают






x	f
0	q
1	p

где q - доля единиц, не обладающих признаком
 p - доля единиц, обладающих признаком

$$p + q = 1$$




Среднее значение альтернативного признака

$$\bar{x} = \frac{\sum x \cdot f}{\sum f} = \frac{p \cdot 1 + q \cdot 0}{p + q} = p$$



Дисперсия альтернативного признака :

$$\begin{aligned}\sigma^2 &= \frac{\sum(x - \bar{x}) \cdot f}{\sum f} = \frac{(0 - p)^2 \cdot q + (1 - p)^2 \cdot p}{p + q} = \\ &= p^2 q + q^2 p = (p + q) \cdot pq = pq\end{aligned}$$


Максимальное значение дисперсии альтернативного признака **0,25**







Правило сложения дисперсий





Выделяют дисперсии:

- 1) общую
 - 2) межгрупповую
 - 3) внутригрупповую
- 



Величина общей дисперсии характеризует вариацию признака под воздействием всех факторов, вызывающих эту вариацию:

$$\sigma^2 = \frac{\sum (x_j - \bar{x})^2 \cdot f_j}{\sum f_j}$$


где j – номер варианты








Межгрупповая дисперсия (дисперсия групповых средних или факторная дисперсия) характеризует систематическую вариацию, т. е. различия в величине изучаемого признака, возникающие под влиянием одного фактора, положенного в основание группировки





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


где \bar{X}_i – среднее значение изучаемого признака для i
– i – й группы
 \bar{X} – общая средняя для всей
совокупности
 i – номер группы
 n_i – количество единиц в i – й группе





$$\sum_j f_j = \sum_i n_i$$






Внутригрупповая (средняя из групповых или остаточная) дисперсия характеризует случайную вариацию, т. е. ту часть вариации, которая вызвана действием других неучтённых факторов, и не зависящую от фактора, положенного в основании группировки:




$$\bar{\sigma}_i^2 = \frac{\sum (\sigma_i^2 \cdot n_i)}{\sum n_i},$$

где σ_i^2 - групповая дисперсия

$$\sigma_i^2 = \frac{\sum (x_j - \bar{x}_i)^2 \cdot f_{ij}}{\sum f_{ij}}$$




Общая дисперсия равна сумме межгрупповой и внутригрупповой дисперсий:

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






Эмпирический коэффициент детерминации:

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

Эмпирический коэффициент детерминации показывает долю общей вариации изучаемого признака, обусловленную вариацией группировочного признака (факторного)





Эмпирическое корреляционное отношение

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

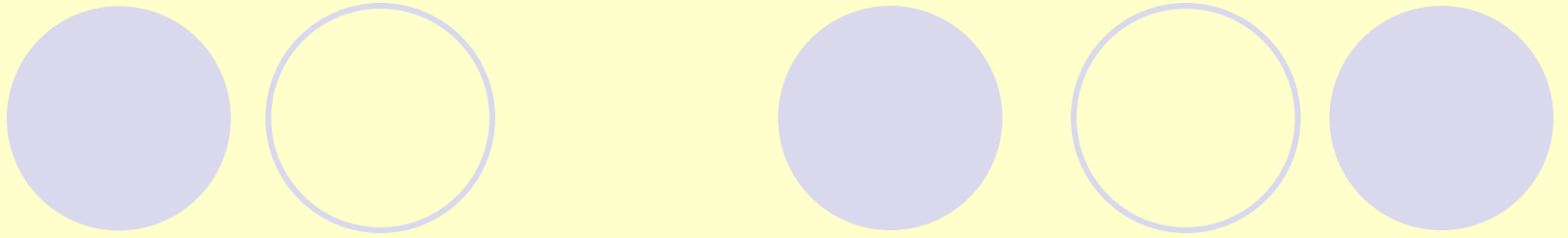
Эмпирическое корреляционное отношение характеризует степень влияния группировочного признака на результативный показатель. Эмпирическое корреляционное отношение изменяется в пределах от 0 до 1. Чем ближе η к единице, тем степень влияния больше

$$0 \leq \eta \leq 1$$

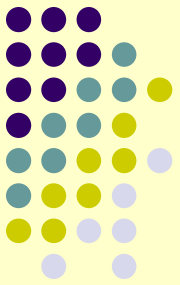



Моменты распределения






Обобщающие характеристики вариационного ряда могут быть представлены системой величин, носящих название моментов распределения



Формула момента k -го порядка:

$$M_k = \frac{\sum (x - A)^k \cdot f}{\sum f},$$

где: x – варианты
 k – показатель степени
 f – частоты
 A – const




1. При $A = 0$ получаем систему начальных моментов. Начальный момент k -го порядка выражается формулой:

$$M_k' = \frac{\sum x^k \cdot f}{\sum f}$$

Начальный момент первого порядка равен \bar{x}





2. При $A = \bar{X}$ получаем систему центральных моментов.

Центральный момент k -го порядка выражается формулой:

$$M_k^* = \frac{\sum (x - \bar{x})^k \cdot f}{\sum f}$$

Центральный момент первого порядка равен 0

Центральный момент второго порядка равен σ^2







При $A = x_0$ получаем **систему условных моментов**:


$$M_k'' = \frac{\sum (x - x_0)^k \cdot f}{\sum f},$$

где: x_0 – некоторый вариант ряда, обычно близкий к его середине





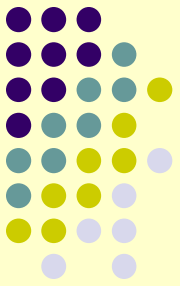
Нормированный момент представляет собой отношение центрального момента k -го порядка к k -ой степени среднего квадратического отклонения:

$$r_k = \frac{M_k^*}{\sigma^k}$$





Нормированный момент

- первого порядка равен 0
- второго порядка равен 1
- третьего и четвертого порядков используется для характеристики асимметрии и эксцессов



Показатели асимметрии и эксцесса

Симметричным называется такое распределение, при котором варианты, равноотстоящие от средней, имеют равные частоты. Если распределение асимметрично, частоты вариантов, равноотстоящих от средней, не равны между собой



Для характеристики асимметрии используется нормированный момент третьего порядка:

$$A = r_3 = \frac{M_3^*}{\sigma^3}$$

Если $A = 0$ распределение симметрично

Если $A > 1$ имеет место правосторонняя асимметрия

Если $A < 1$ имеет место левосторонняя асимметрия





Под эксцессом понимается степень островершинности распределения, при этом в качестве эталона берется нормальное распределение. Характеристикой эксцесса является нормированный момент четвертого порядка



Формула коэффициента эксцесса:

$$E = \frac{M_4^*}{\sigma^4} - 3$$



Для нормального распределения $E = 0$. Для более островершинных распределений, чем нормальное, $E > 0$, для более плосковершинных $E < 0$



Выработка, метры	Число рабочих	x	$\bar{x} - x$	$(\bar{x} - x)^2 f$	$x'f$	$x'^2 f$
до 200	3	190	-64	12249,63	-9	27
200-220	12	210	-44	23126,52	-24	48
220-240	50	230	-24	28560,5	-50	50
240-260	56	250	-3,9	851,76	0	0
260-280	47	270	16,1	12182,87	47	47
280-300	23	290	36,1	29973,83	46	92
300-320	7	310	56,1	22080,47	21	63
свыше 320	2	330	76,1	11582,42	8	32
ИТОГО:				140558	39	359