



Средние величины и показатели вариации

- ***Определение средних величин, их формы и виды***
- ***Понятие вариационного ряда, его виды и графическое изображение***
- ***Показатели вариации. Порядок их построения***

ФОРМЫ И ВИДЫ СРЕДНИХ ВЕЛИЧИН

Формы средних величин	Виды средних величин	
	Простая	Взвешенная
Средняя арифметическая	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum xf}{\sum f}$
Средняя гармоническая	$\bar{x} = \frac{n}{\sum \frac{1}{x}}$	$\bar{x} = \frac{\sum f}{\sum \frac{f}{x}}$
Средняя квадратическая	$\bar{x} = \sqrt{\frac{\sum x^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum x^2 f}{\sum f}}$
Средняя геометрическая	$\bar{x} = \sqrt[n]{x_1 x_2 \dots x_n}$	$\bar{x} = \sqrt[\sum f]{x_1 f_1 x_2 f_2 \dots x_n f_n}$



Рекомендации при использовании средних величин

- Совокупность, по которой производится обобщение, должна быть однородной
- Необходимо обеспечить исчерпывающий учет единиц совокупности
- При расчете средних необходимо учитывать своеобразие и взаимосвязь признаков и использовать их в совокупности с другими статистическими показателями
- Порядок расчета средних сохраняется независимо от уровня обобщения



Простые и взвешенные средние

- Для расчета средних первичных признаков используется ***простая средняя***
- Для расчета средних вторичных признаков используется ***взвешенная средняя***
- Взвешенная средняя может быть рассчитана для первичных признаков, если они представлены в сгруппированном виде
- Несгруппированные данные осредняются по простой средней



Простые и взвешенные средние

- Простые и взвешенные средние различаются:
 - по величине (не всегда)
 - по способу вычисления
 - по своей роли в решении различных статистических задач



- Взвешенная средняя равна простой в трех случаях:
 - если изучаемый признак не варьирует
 - если не варьирует признак-вес
 - если между осредняемым признаком и признаком-весом нет линейной зависимости



Задача

Район	Численность безработных, тыс. чел.	Уровень безработицы, %	Доля женщин среди безработных
1	12,60	4,5	0,73
2	19,22	6,2	0,68
3	20,80	6,4	0,75



Решение

$$\bar{A} = \frac{\sum A}{n} = \frac{52,62}{3} = 17,54$$

$$\bar{B} = \frac{\sum A}{\sum \frac{A}{B}} = \frac{52,62}{\frac{12,6}{0,045} + \frac{19,22}{0,062} + \frac{20,80}{0,064}} = 0,058$$

$$\bar{C} = \frac{\sum CA}{\sum A} = \frac{12,6 \cdot 0,73 + 19,22 \cdot 0,68 + 20,80 \cdot 0,75}{52,62} = 0,72$$



Правило мажорантности средних

$$\bar{x}_{\text{гарм}} \leq \bar{x}_{\text{геом}} \leq \bar{x}_{\text{арифм}} \leq \bar{x}_{\text{квадр}}$$



СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВАРИАЦИИ

- **Вариация** – это различие значений признака у отдельных единиц изучаемой совокупности в один и тот же период или момент времени
- **Вариация** отражает колеблемость индивидуальных значений признака
- **Вариация** отражает неравномерность развития единиц совокупности



- ***Ряд распределения*** – упорядоченное распределение единиц совокупности по возрастающим или убывающим значениям признака
- Ряд распределения, построенный по качественному признаку – ***атрибутивный ряд***
- Ряд распределения, построенный по количественному признаку – ***вариационный ряд***



Виды вариационных рядов

- ***Дискретные***, в которых значения варьирующего признака выражены в виде вполне определенных величин (обычно целых).
- ***Интервальные***, в которых значения варьирующего признака представлены в виде интервалов.



***Распределение частных домохозяйств РФ по размеру
(по данным переписи 2010 г.)***

Домохозяйства, состоящие:	Число домохозяйств, миллионов		В % к итогу	
	2002 г.	2010 г.	2002 г.	2010 г.
из 1 человека	11,8	14,0	22,3	25,7
из 2 человек	14,5	15,6	27,6	28,5
из 3 человек	12,5	12,3	23,8	22,5
из 4 человек	9,0	7,9	17,0	14,5
из 5 человек	3,0	2,9	5,7	5,4
из 6 и более человек	1,9	1,9	3,6	3,4
Всего домохозяйств	52,7	54,6	100	100



**Распределение населения РФ в 2013 г. по размеру
среднедушевого денежного дохода**

Все население	100
в том числе со среднедушевыми денежными доходами, руб. в месяц:	
до 5000,0	4,2
5000,1–7000,0	5,6
7000,1–10000,0	10,4
10000,1–14000,0	14,2
14000,1–19000,0	15,2
19000,1–27000,0	17,5
27000,1–45000,0	19,3
свыше 45000,0	13,6



Элементы вариационного ряда

- ***Варианты*** – значения, которые принимает исследуемый признак
- ***Частоты*** – абсолютная численность отдельных групп с различными значениями признака
- ***Частоты*** – удельные веса (доли) отдельных групп в общей численности совокупности



Порядок построения интервальных вариационных рядов

- Определяется число групп (число интервалов)
- При определении числа групп учитывается объем совокупности с тем, чтобы обеспечить заполненность интервалов
- Определяется величина интервала



Расчет числа интервалов

- Формула Стерджесса: $k \approx 1 + 3,32 \lg n$
- или $k \approx 1 + 1,44 \ln n$
- где: k - число интервалов; n - объем совокупности.



Расчет величины интервалов

$$i = \frac{x_{\max} - x_{\min}}{k}$$

- Где: $x_{\max}$, $x_{\min}$ - максимальное и минимальное значения признака в вариационном ряду.



Плотность распределения

- Если в интервальном вариационном ряду ширина интервала отлична от единицы, то определяют абсолютную и относительную плотности распределения



- Отношение частоты интервала к ширине этого интервала называют **абсолютной плотностью распределения** для i -го интервала
-
- **Относительной плотностью распределения** для i -го интервала называют отношение относительной частоты интервала к его ширине



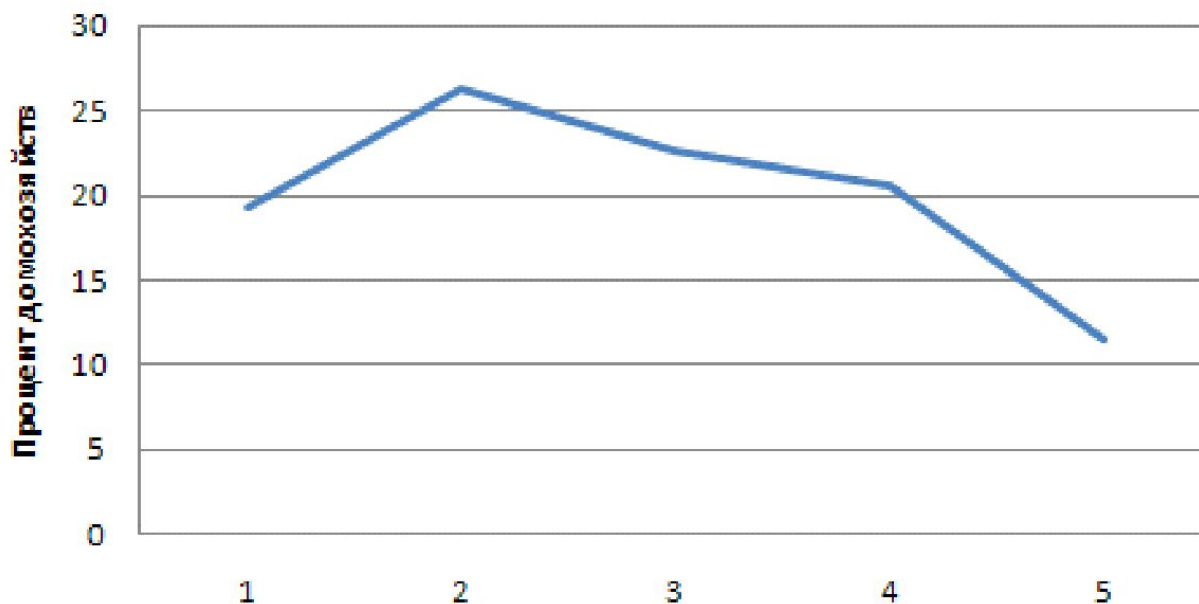
Графическое изображение вариационного ряда

- ***Полигон распределения***
- ***Гистограмма*** – столбиковая диаграмма, для построения которой на оси абсцисс откладывают отрезки, равные величине интервалов вариационного ряда
- ***Кумулята распределения*** строится по накопленным частотам (частостям). Накопленные частоты (частости) определяются последовательным суммированием частот (частостей)



Полигон распределения

Размер домохозяйств, чел

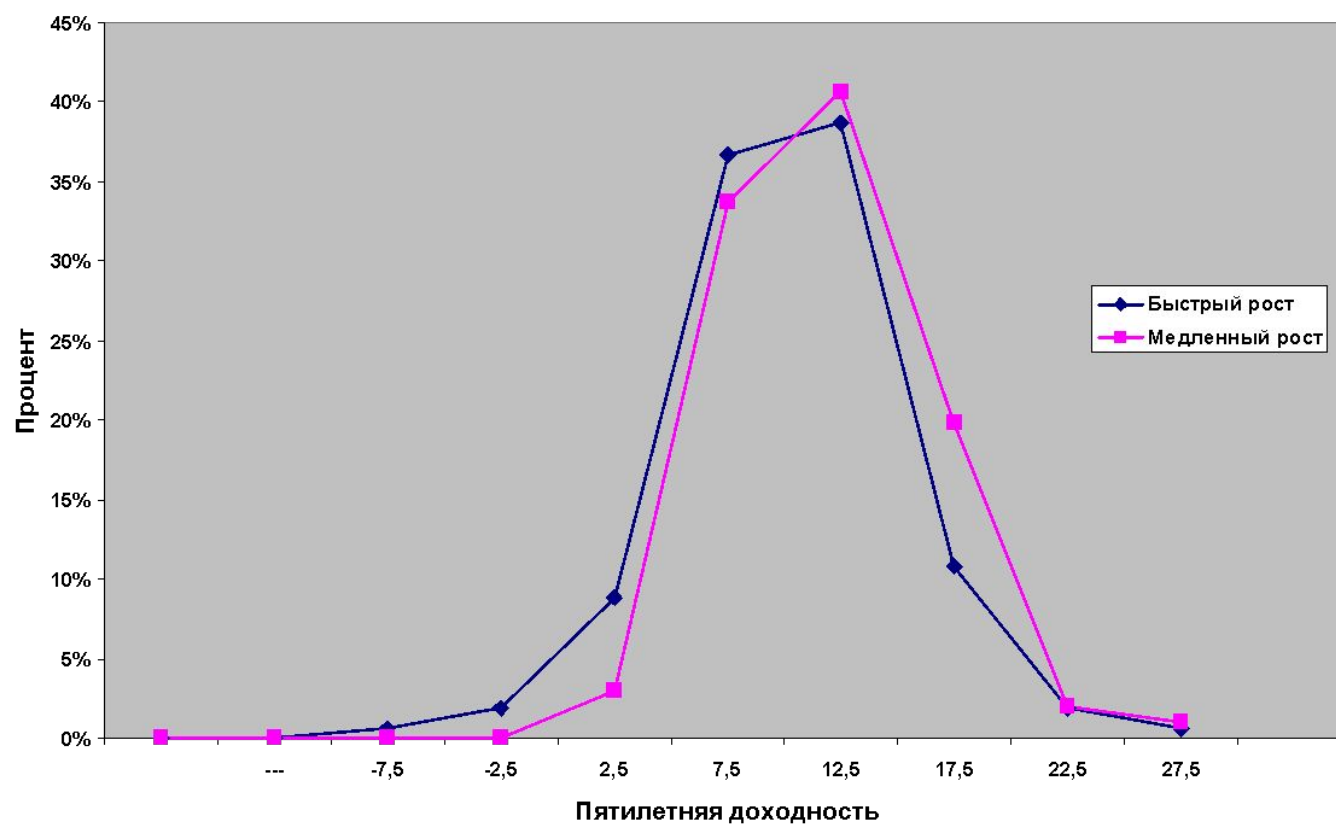


Распределение домохозяйств по размеру



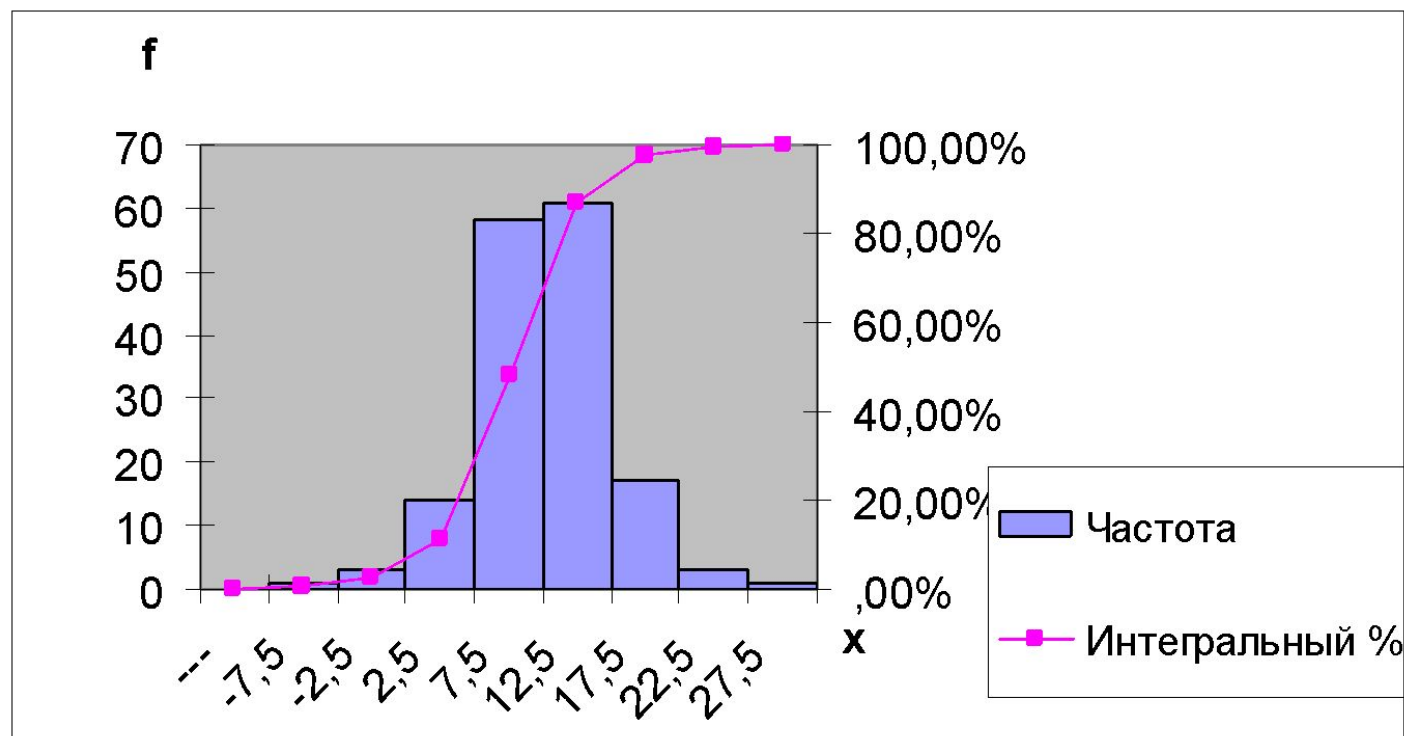
Полигон распределения

Процентный полигон для пятилетней доходности





Гистограмма, кумулята





Показатели вариации

- ***Размах вариации*** $R = x_{\max} - x_{\min}$
- ***Среднее линейное отклонение***
 - а) для несгруппированных данных

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

- б) для сгруппированных данных

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



Показатели вариации

Дисперсия:

- а) для несгруппированных данных

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

- б) для сгруппированных данных

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



Показатели вариации

Среднее квадратическое отклонение

а) для несгруппированных данных

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

б) для сгруппированных данных

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



Показатели вариации

Коэффициент вариации

$$v = \frac{\sigma}{x} \cdot 100$$

Среднее значение признака

- для несгруппированных данных

$$\bar{x} = \frac{\sum x}{n}$$

- для сгруппированных данных

$$\bar{x} = \frac{\sum xf}{\sum f}$$



Шкала значений коэффициента вариации

Коэффициент вариации, %	Степень однородности совокупности
До 30	однородная
30-60	средняя
60 более	неоднородная



Показатели вариации

Мода – наиболее часто встречающееся в данной совокупности значение признака



- **В дискретном ряду** мода – вариант с наибольшей частотой
- **В интервальном ряду** мода определяется по формуле:

$$Mo = x_0 + i \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}$$



Показатели вариации

- ***Медиана*** – то значение признака, которое находится в середине упорядоченного ряда и делит совокупность на две равные части



- ***В дискретном*** ряду медиана определяется по сумме накопленных частот, которая должна превышать половину всей численности совокупности



Показатели вариации

- ***В интервальном*** ряду медиана определяется по формуле:

$$Me = x_0 + i \frac{\frac{1}{2} \Sigma f - S_{me-1}}{f_{me}}$$



Пример

- Имеются данные о численности работников по 6 магазинам: 7 4 9 7 3 12
- Необходимо рассчитать показатели вариации и описать форму распределения этих данных.



Решение

№ п/п	x	$ x - \bar{x} $	$(x - \bar{x})^2$
1	3	4	16
2	4	3	9
3	7	0	0
4	7	0	0
5	9	2	4
6	12	5	25
Итого	42	14	54

$$R = 12 - 3 = 9$$

$$\bar{x} = \frac{42}{6} = 7$$

$$\bar{d} = \frac{14}{6} = 2,3$$

$$\sigma^2 = \frac{54}{6} = 9$$

$$\sigma = \sqrt{9} = 3$$

$$v = \frac{3}{7} \cdot 100 = 43\%$$

$$Mo = 7$$

$$Me = 7$$



Пример

Возраст, лет	Численность сотрудников
20-30	8
30-40	17
40-50	11
50-60	8
60-70	2



Решение

Возраст, лет	f	x	xf					
20-30	8	25	200					
30-40	17	35	595					
40-50	11	45	495					
50-60	8	55	440					
60-70	2	65	130					
Итого	46	X	1860					



$$\bar{x} = \frac{1860}{46} = 40,4 \text{ лет};$$



Возраст, лет	f	x	xf	$ x - \bar{x} $	$ x - \bar{x} f$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$	Накопленные частоты
20-30	8	25	200	15,4	123,2	237,16	1897,28	8
30-40	17	35	595	5,4	91,8	29,16	495,72	25
40-50	11	45	495	4,6	50,6	21,16	232,76	36
50-60	8	55	440	14,6	116,8	213,16	1705,28	44
60-70	2	65	130	24,6	49,2	605,16	1210,32	46
Итого	46	X	1860	X	431,6	X	5541,36	X



$$\bar{d} = \frac{431,6}{46} = 9,38 \text{ лет};$$

$$\sigma^2 = \frac{5541,36}{46} = 120,46;$$

$$\sigma = \sqrt{120,46} = 11,0 \text{ лет};$$

$$v = \frac{11,0}{40,4} \cdot 100 = 27,2\%;$$



$$Mo = 30 + 10 \frac{17 - 8}{(17 - 8) + (17 - 11)} = 36 \text{ лет};$$

$$Me = 30 + 10 \frac{23 - 8}{17} = 38,8 \text{ лет.}$$



- **Квартили** – делят совокупность на 4 равные части

$$Q_k = x_0 + i \frac{k \frac{\Sigma f}{4} - S_{Q_k-1}}{f_{Q_k}}$$



Децили – делят совокупность на 10
равных частей

$$D_k = x_0 + i \frac{k \frac{\Sigma f}{10} - S_{D_k-1}}{f_{D_k}}$$



Показатели асимметрии

Коэффициент асимметрии

$$As = \frac{M_3}{\sigma^3}$$

M_3 - центральный момент 3-го порядка.

$$M_3 = \frac{\sum (x - \bar{x})^3 f}{\sum f}$$



Показатели асимметрии

Для нормального распределения $As = 0$

Если $As < 0$, асимметрия – *левосторонняя*

Если $As > 0$, асимметрия – *правосторонняя*.

Если $|As| < 0,25$, асимметрия –
незначительная.

Если $|As| > 0,5$, асимметрия – *значительная*.



Средняя квадратическая ошибка коэффициента асимметрии

$$\sigma_{As} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}$$

Если отношение $|As| : \sigma_{As} > 3$, то асимметрия
существенная.

Если отношение $|As| : \sigma_{As} < 3$, то асимметрия не
существенная, вызванная влиянием случайных факторов



Структурный коэффициент асимметрии (формула Пирсона)

$$As = \frac{\bar{x} - Mo}{\sigma}$$

Если $\bar{x} < Mo$, то асимметрия
левосторонняя

Если $\bar{x} > Mo$, то асимметрия –
правосторонняя.



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

M_4 - центральный момент 4-го порядка.

$$M_4 = \frac{\sum (x - \bar{x})^4 f}{\sum f}$$



ЭКССЕСС

Для нормального распределения $M_4 : \sigma^4 = 3$, следовательно $Ex = 0$.

При $Ex > 0$ (положительный эксцесс) распределение является более островершинным, чем нормальное распределение.

При $Ex < 0$ (отрицательный эксцесс) распределение является более пологим, чем нормальное распределение.



Средняя квадратическая ошибка коэффициента эксцесса

$$\sigma_{Ex} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}$$

Если отношение $|Ex| : \sigma_{Ex} > 3$, то отклонение от нормального распределения можно считать существенным.



ЭКСЦЕСС

- ***Положительный эксцесс***
свидетельствует о том, что в совокупности есть слабо варьирующее по данному признаку «ядро»
- ***Чем круче распределение, тем ярче проявляется закономерность в формировании значений показателей***



ЭКСЦЕСС

- ***В плосковершинном распределении*** единицы рассеяны по всем значениям признака более равномерно.
- При существенном отрицательном эксцессе результаты анализа не надежны.
- Значительный отрицательный эксцесс может указывать на качественную неоднородность совокупности.