Тема: Средние величины

Средняя величина

Обобщающий показатель, который дает количественную характеристику признака в статистической совокупности в условиях

Условия правильного применения средней величины:

Средняя величина должна исчисляться лишь для совокупностей, состоящих из однородных единиц

Совокупность, неоднородную в качественном отношении, необходимо разделять на однородные группы и вычислять для них групповые типичные средние, характеризующие каждую из этих групп. В этом проявляется связь между методами группировок и средних величин

Средняя величина сглаживает индивидуальные значения и тем самым может элиминировать различные тенденции в развитии, скрыть передовое и отстающее, поэтому кроме средней величины следует исчислять другие показатели

Среднюю величину целесообразно исчислять не для отдельных единичных фактов, взятых изолированно друг от друга, а для совокупности фактов


Виды средних величин
Степенные
Гармоническая
Геометрическая
Арифметическая
Квадратическая
Кубическая
Биквадратическая
Структурные
Мода
Медиана
Квартили
Децили
Квинтили
Перцентили

Средняя степенная простая

где К – показатель степени.
Применяется в случае, если каждая варианта Х встречается в совокупности один или

Средняя степенная взвешенная

$$\overline{X} = \sqrt{\frac{\sum X^K \cdot f_i}{\sum f_i}}$$

где f_i - показатель повторяемости вариант (веса, частоты).

Применяется в случае, если каждая варианта X встречается в совокупности не одинаковое число раз, т.е. по сгруппированным данным

Вид средней	Методика расчета и содержание показателя
Средняя гармоническая	К = - 1; $$\overline{X} = \frac{n}{\sum (1 : x_i)}, \text{ или } \overline{X} = \frac{\sum \omega}{\sum (\omega : x_i)},$$ где $\omega = x_i \cdot f_i$ Средняя гармоническая применяется в случае, если известны варьирующие обратные значения признака
Средняя геометрическая	К = 0; $$\overline{X} = \sqrt[n]{\prod x_i}, \text{ или } \overline{X} = \sqrt[f_i]{\prod x_i^{f_i}},$$ где П – знак умножения. Наиболее широкое применение средняя геометрическая получила для определения средних темпов изменения в рядах динамики, а также в рядах распределения
Средняя арифметическая	К = 1; $$\overline{X} = \frac{\sum X_i}{n}, \text{ или } \overline{X} = \frac{\sum X_i \cdot f_i}{\sum f_i}$$ Средняя арифметическая применяется в тех случаях, когда объем варьирующего признака для всей совокупности образуется как сумма значений признака отдельных ее единиц
Вид средней	**Методика расчета и содержание показателя**
Средняя квадратическая	К = 2; $$\overline{X} = \sqrt[2]{\frac{\sum X_i^2}{n}}, \text{ или } \overline{X} = \sqrt[2]{\frac{\sum X_i^2 \cdot f_i}{\sum f_i}}$$
Средняя кубичес-кая	К = 3; $$\overline{X} = \sqrt[3]{\frac{\sum X_i^3}{n}}, \text{ или } \overline{X} = \sqrt[3]{\frac{\sum X_i^3 \cdot f_i}{\sum f_i}}$$
Средняя биквадратическая	К = 4; $$\overline{X} = \sqrt[4]{\frac{\sum X_i^4}{n}}, \text{ или } \overline{X} = \sqrt[4]{\frac{\sum X_i^4 \cdot f_i}{\sum f_i}}$$ и др.

Основные свойства средней арифметической

Свойство	Формула расчета
Произведение средней на сумму частот всегда равно сумме произведений вариант частот	$\overline{X} \cdot \sum f_i = \sum X_i \cdot f_i$
Если от каждой варианты отнять какое-либо произвольное число, то средняя увеличится на это же число	$\frac{\sum (X_i - A) \cdot f_i}{\sum f_i} = \overline{X} - A$
Если к каждой варианте прибавить какое-либо произвольное число, то средняя увеличится на это же число	$\frac{\sum (X_i + A) \cdot f_i}{\sum f_i} = \overline{X} + A$
Если каждую варианту разделить на какое-либо произвольное число, то средняя арифметическая уменьшится во столько же раз	$\frac{\sum \frac{X_i}{A} \cdot f_i}{\sum f_i} = \frac{\overline{X}}{A}$
Если каждую варианту умножить на какое-либо произвольное число, то средняя арифметическая увеличится во столько же раз	$\frac{\sum (X_i \cdot A) \cdot f_i}{\sum f_i} = \overline{X} \cdot A$
Если все частоты (веса) разделить или умножить на какое-либо число, то средняя арифметическая от этого не изменится	$\frac{\sum X_i \cdot (f_i \cdot A)}{\sum (f_i \cdot A)} = \overline{X}$
Сумма отклонений вариант от средней арифметической всегда равна нулю	$\sum (X - \overline{X}) \cdot f_i = 0$

Мода

Величина признака (варианта), которая чаще всего встречается в данной совокупности. В вариационном дискретном ряду модой выступает варианта, имеющая наибольшую частоту

Медиана

варианта, которая находится в середине вариационного ряда. Медиана делит ряд пополам, по обе стороны от нее (вверх и вниз) находится одинаковое количество единиц совокупности

Виды структурных (непараметрических) средних

Вид Средней	Методика расчета показателя
Мода	В интервальных рядах с равными интервалами мода вычисляется по формуле $$Mo = X_0 + i \cdot \frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m-1})},$$ Где X_0 – минимальная граница модального интервала; i – величина модального интервала; f_m – частота модального интервала; f_{m-1} – частота интервала, предшествующего модальному интервалу; f_{m+1} – частота интервала, следующего за модальным. Модальный интервал в интервальном ряду определяется по наибольшей частоте

Медиана	В дискретном вариационном ряду определение медианного значения признака сводится к определению номера медианной единицы ряда по формуле $$N_{ME} = \frac{n+1}{2},$$ где n – объем совокупности. Полученное значение показывает, где точно находится номер медианной единицы (номер середины ряда). Медианное значение характеризуется тем, что его кумулятивная частота (сумма накопленных частот по группам) равна половине суммы всех частот или превышает ее. В интервальном ряду с равными интервалами медиана рассчитывается по формуле $$Me = X_0 + i \cdot \frac{\frac{1}{2}\sum f - S_{m-1}}{f_m},$$ где X_0 – начальное значение медианного интервала; i - величина медианного интервала; $\sum f$ – сумма частот ряда; S_{m-1} - сумма накопленных частот в интервалах, предшествующих медианному; f_m – частота медианного интервала. Для определения медианного интервала необходимо рассчитать суммы накопленных частот. Медианный интервал характерен тем, что его кумулятивная частота равна полусумме всех частот ряда или превышает ее

Квартили	Значения признака, делящие ранжированную совокупность на четыре равные части. Различают нижний квартиль (Q_1), отделяющий ¼ часть совокупности с наименьшими значениями признака, и верхний квартиль (Q_3), отсекающий ¼ часть с наибольшими значениями признака. Средний квартиль (Q_2) совпадает с медианой (Ме). Для расчета квартилей по интервальному вариационному ряду используют формулы $$Q_1 = X_{Q_1} + i \cdot \frac{\frac{1}{4}\sum f - S_{Q_1 - 1}}{f_{Q_1}};$$ $$Q_3 = X_{Q_3} + i \cdot \frac{\frac{3}{4}\sum f - S_{Q_3 - 1}}{f_{Q_3}},$$ где X_{Q_1} (X_{Q_3}) – нижняя граница интервала, содержащего нижний квартиль (верхний) квартиль; i - величина интервала; $S_{Q_1\text{-}1}$ ($S_{Q_3\text{-}1}$) – накопленная частота интервала, предшествующего интервалу, содержащему нижний квартиль (верхний) квартиль; f_{Q_1} (f_{Q_3}) – частота интервала, содержащего нижний квартиль (верхний) квартиль

Децили	Варианты, делящие ранжированный ряд на десять равных частей; они вычисляются по той же схеме, что и квартили: $$d_1 = X_{d_1} + i \cdot \frac{\frac{1}{10}\sum f - S_{d_1-1}}{f_{d_1}};$$ $$d_2 = X_{d_2} + i \cdot \frac{\frac{2}{10}\sum f - S_{d_2-1}}{f_{d_2}}$$ и т.д.
Квинтили	Значения признака, делящие ряд на пять равных частей. Они вычисляются по той же схеме, что квартили и децили
Перцентили	Значения признака, делящие ряд на 100 равных частей

ТЕМА
«ПОКАЗАТЕЛИ ВАРИАЦИИ И АНАЛИЗ ЧАСТОТНЫХ РАСПРЕДЕЛЕНИЙ»

1. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ПОКАЗАТЕЛИ ВАРИАЦИИ.
2. ВИДЫ ДИСПЕРСИЙ. ПРАВИЛО СЛОЖЕНИЯ ДИСПЕРСИЙ.
3. МОМЕНТЫ РАСПРЕДЕЛЕНИЯ. ПОКАЗАТЕЛИ ФОРМЫ РАСПРЕДЕЛЕНИЯ.

Показатель	**Методика расчета и содержание**
Размах вариации	Характеристика границ вариации изучаемого признака. Определяется по формуле $$R = X_{\max} - X_{\min},$$ где X_{max} - максимальное значение варьирующего признака; X_{min} - минимальное значение варьирующего признака. Показывает, сколь велико различие между единицами совокупности, имеющими самое маленькое и самое большое значение признака, основан на крайних значениях варьирующего признака и не отражает отклонений всех вариант в ряду

Дисперсия	Средний квадрат отклонений индивидуальных значений признака от их средней величины. Вычисляется по следующим формулам. 1-й способ определения дисперсии $$\sigma^2 = \frac{\sum (X_i - \overline{X})^2}{n}, \text{ или } \sigma^2 = \frac{\sum (X_i - \overline{X})^2 \cdot f_i}{f_i},$$ где X_i - индивидуальные значения варьирующего признака (варианты); $\overline{X}$ - среднее значение варьирующего признака; n – количество разновидностей вариант; f_i – показатель повторяемости вариант (частоты, веса) 2-й способ определения дисперсии $$\sigma^2 = \overline{X^2} - (\overline{X})^2,$$ где $\overline{X^2}$ - средняя из квадратов индивидуальных значений; $(\overline{X})^2$ - квадрат средней величины признака 3-й способ определения дисперсии – метод моментов $$\sigma^2 = i \cdot (m_2 - m_1^2),$$ где i – величина интервала в интервальном ряду; m_1 - момент первого порядка; m_2 - момент второго порядка, который определяется по формуле $$m_2 = \frac{\sum \left(\frac{X_i - A}{i}\right)^2 \cdot f_i}{\sum f_i}$$

Среднее квадра-тическое отклонение	Обобщающая характеристика размеров вариации признака в совокупности определяется по формуле $$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}}, \text{ или } \sigma = \sqrt{\frac{\sum (X_i - \overline{X})^2 \cdot f_i}{f_i}}.$$

СРЕДНЕЕ ЛИНЕЙНОЕ ОТКЛОНЕНИЕ

$$L = \text{c}\frac{\sum|X_i - \bar{X}|}{n} \text{ или } L = \frac{\sum|X_i - \bar{X}| * f}{\sum f}$$

Коэффициент вариации	Характеристика меры вариации значений признака вокруг средней величины: $$V = \frac{\sigma}{\overline{X}} \cdot 100$$ Чем этот показатель меньше, тем однороднее совокупность, а средняя величина признака типична для данной совокупности. Чем коэффициент вариации больше, тем неоднороднее совокупность

ЛИНЕЙНЫЙ КОЭФФИЦИЕНТ ВАРИАЦИИ

$$\mathrm{v}_l = \frac{L}{\bar{X}} * 100 \text{ или } \mathrm{v}_l = \frac{L}{M_E} * 100$$

КОЭФФИЦИЕНТ ОСЦИЛЛЯЦИИ

$$\mathrm{v}_r = \frac{R}{\bar{\bar{X}}} * 100$$

Математические свойства дисперсии

Свойства	Методика расчета
Если из всех значений вариант отнять какое-то постоянное число А, то средний квадрат отклонений от этого не изменится	$\sigma^2_{(X-A)} = \sigma^2_x$
Если все значения вариант разделить на какое-то постоянное число А, то средний квадрат отклонений уменьшится от этого в $А^2$ раз, а среднее квадратическое отклонение – в А раз	$\sigma^2_{\left(\frac{X}{A}\right)} = \sigma^2_x : A^2;$ $\sigma_{\left(\frac{X}{A}\right)} = \sigma_x : A$
Если исчислить средний квадрат отклонений от любой величины (А), в той или иной степени отличающейся от средней арифметической ($\overline{X}$), то он всегда будет больше среднего квадрата отклонений, исчисленного от средней арифметической	$\sigma^2_A > \sigma^2_x$

Альтернативный признак

Среди множества признаков, изучаемых статистикой, выделяют такие, которыми обладают одни единицы совокупности и не обладают другие, называемые *альтернативными.*

Понятие альтернативного признака

Альтерна-тивный признак	- это	признак, которым обладают одни единицы и не обладают другие единицы совокупности

Вариация альтернативного признака количественно проявляется в значении нуля у единиц, которые этим признаком не обладают, или единицы у тех, которые данный признак имеют.

Средняя и дисперсия для альтернативного признака

Средняя и дисперсия альтернативного признака

Среднее значение альтернативного

$$\overline{X} = \frac{1\cdot p + 0\cdot q}{p+q} = p,$$

так как

$0\cdot q = 0$, а $p+q=1$

Дисперсия альтернативного

$$\sigma_p^2 = \frac{(1-p)^2\cdot p + (0-p)^2\cdot q}{p+q} =$$

$$q^2\cdot p + p^2\cdot q = p\cdot q\cdot(p+q) =$$

$$= p\cdot q$$

Три вида дисперсий

Для сгруппированной статистической совокупности возможно вычисление трех видов дисперсии: общей, внутригрупповой и межгрупповой

ЗАКОН СЛОЖЕНИЯ ДИСПЕРСИЙ

Общая дисперсия $\sigma^2 = \dfrac{\sum\left[x:(-\bar{x})\right]^2 \cdot f_i}{\sum f_i}$ **или** $\sigma^2 = \delta^2 + \bar{\sigma}^2$**,**

Характеризует вариацию признака во всей совокупности под влиянием всех факторов,

Межгрупповая дисперсия (δ^2 ***)***

Характеризует вариацию изучаемого признака под влиянием признака-фактора, положенного в основание группировки:

$$\delta^2 = \frac{\sum(\overline{X}_i - \overline{X})^2 \cdot f_i}{\sum f_i},$$

где $\overline{X}$ - общая средняя;

$\overline{X}_i$ - средняя i-группы;

Средняя из внутригрупповых дисперсий ($\overline{\sigma}_i^2$ ***)***

Отражает случайную вариацию, обусловленную неучтенными факторами и не зависящую от признака-фактора, положенного в основание группировки

$$\overline{\sigma}_i^2 = \frac{\sum \sigma_i^2 \cdot f_i}{\sum f_i},$$

где $\overline{\sigma}_i^2 = \dfrac{\sum(\overline{X}_i - \overline{X}_i)^2 \cdot f_i}{\sum f_i}$ -

Эмпирическое корреляционное отношение

Разные виды дисперсий широко используют для исчисления показателей тесноты связи между признаками.

Эмпирическое корреляционное отношение $\eta = \sqrt{\dfrac{\delta^2}{\sigma^2}}$	изменяется в пределах [0,1]	и характеризует влияние признака, положенного в основание группировки. Если $\eta = 0$, то группировочный признак не оказывает влияния на результативный. Если $\eta = 1$, то результативный признак изменяется только в зависимости от признака, положенного в основание группировки, а влияние прочих факторных признаков равно нулю

Шкала Чэддока

Шкала значений эмпирического корреляционного отношения

Эмпирическое корреляционное отношение может быть только положительным. Качественная интерпретация показателя осуществляется посредством *шкалы Чэддока.*

Категория	**Границы значений эмпирического корреляционного отношения**
Связь очень слабая	0,1-0,3
Умеренная	0,3-0,5
Заметная	0,5-0,7
Тесная	0,7-0,9
Весьма тесная	0,9-0,99

ЭМПИРИЧЕСКИЙ КОЭФФИЦИЕНТ ДЕТЕРМИНАЦИИ

Эмпирический коэффициент детерминации

На основе эмпирического корреляционного отношения рассчитывают *коэффициент детерминации.*

Эмпирический коэффициент детерминации

$$\eta^2 = \frac{\delta^2}{\sigma^2} \cdot 100$$

→ Показывает долю (удельный вес) общей вариации изучаемого признака, обусловленную вариацией группировочного признака

Оценка вариационного ряда на асимметрию и эксцесс

Моментный коэффициент асимметрии

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

Где M_3 - центральный момент третьего порядка.

Если $As < 0$, то это левосторонняя асимметрия.

Если $As > 0$, то это правосторонняя асимметрия

Структурный коэффициент асимметрии

- Английский статистик К. Пирсон предложил формулу:

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

$\bar{x}$- средняя величина

M_o- мода

σ- среднее квадратическое отклонение.

Оценка вариационного ряда на асимметрию и эксцесс

Левосторонняя

As<0

Правосторонняя

As>0

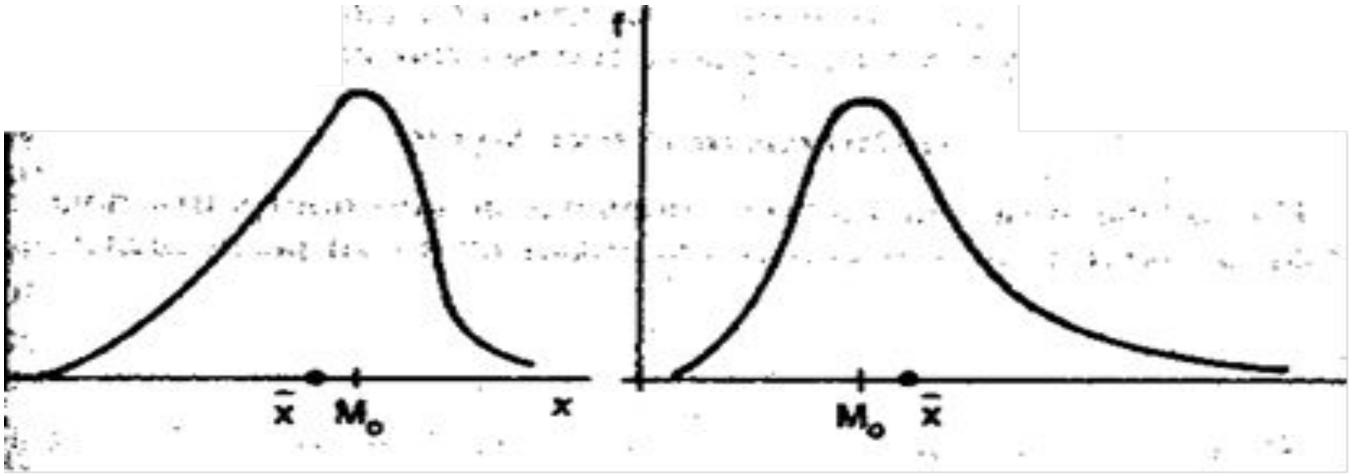

Рис. Асимметрия распределения

СООТНОШЕНИЕ МЕЖДУ СРЕДНЕЙ ВЕЛИЧИНОЙ, МОДОЙ И МЕДИАНОЙ

Мода < Медиана< Средняя величина – правосторонняя асимметрия.

Мода>Медиана>Средняя величина-левосторонняя асимметрия.

ЭКСЦЕСС РАСПРЕДЕЛЕНИЯ

- Под эксцессом понимают островершинность или плосковершинность распределения по сравнению с нормальным распределением при той же вариации.

$$E_x = \frac{м_4}{\sigma_4} - 3,$$ где

$м_4$. − центральный момент четвертого порядка.

Ex > 0, распределение является островершинным.

Ex < 0, распредление является плосковершинным.

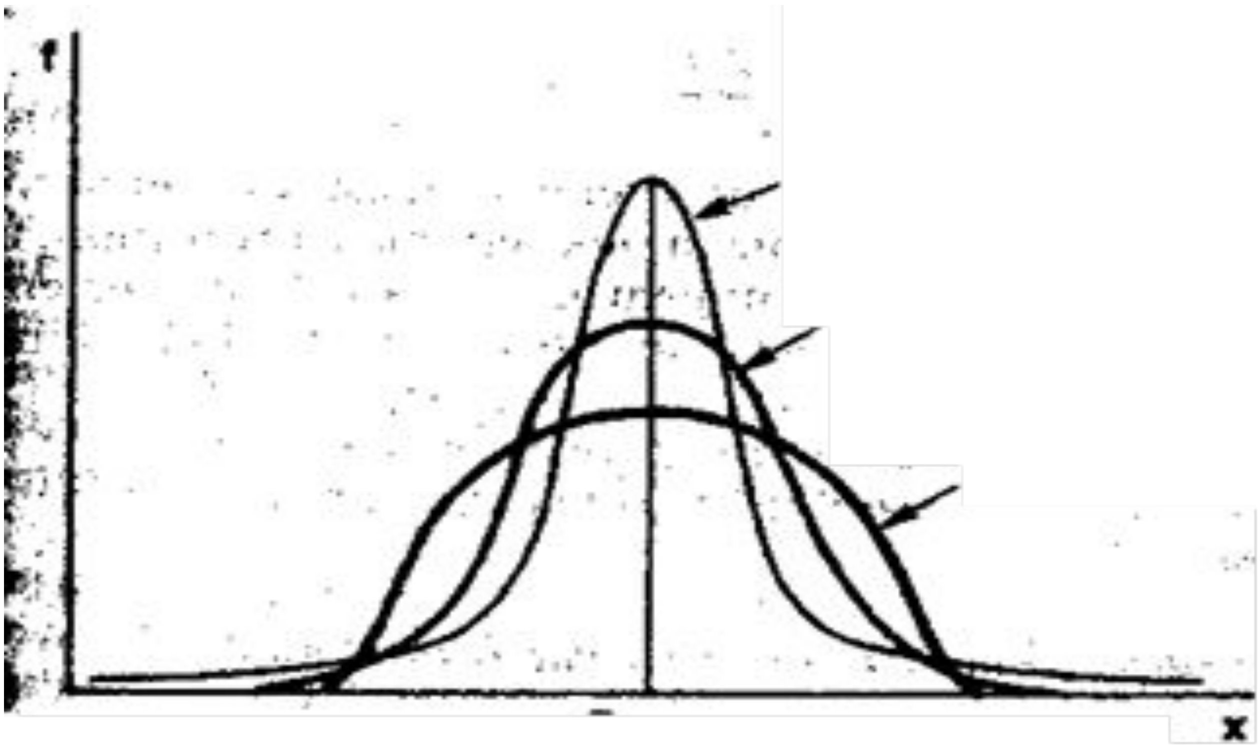

Рис. Эксцесс распределения

СПАСИБО ЗА ВНИМАНИЕ!