Средние величины и показатели вариации

Понятие средней величины

Средняя величина

Обобщающий показатель, который дает количественную характеристику признака в статистической совокупности в условиях конкретного время и места

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

$$\bar{X} = \sqrt[K]{\frac{\sum X^K}{n}}$$

где К – показатель степени
Применяется в случае, если каждая варианта X встречается в совокупности один или одинаковое число раз

Средняя степенная взвешенная

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

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.

Средняя гармоническая

К=-1;

$$\overline{X} = \frac{n}{\sum \frac{1}{x_i}}$$

или

$$\overline{X} = \frac{\sum \omega}{\sum \frac{\omega}{x_i}}$$

где $\omega = x_i * f_i$

Средняя гармоническая применяется в случае, если известны варьирующие обратные значения признака.

Средняя геометрическая

К=0;

$$\bar{X} = \sqrt[n]{\prod x_i}$$

или

$$\bar{X} = \sqrt[\sum f_i]{\prod x_i^{f_i}}$$

Наиболее широкое применение средняя геометрическая получила для определения средних темпов изменения в рядах динамики, а также в рядах распределения

Средняя арифметическая

К=1;

$$\overline{X} = \frac{\sum X_i}{n}$$

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.

Средняя квадратическая

К=2;

или $$\overline{X} = \frac{\sqrt[2]{\sum X_i^2}}{n}$$

$$\overline{X} = \sqrt[2]{\frac{\sum X_i^2 \cdot f_i}{\sum f_i}}$$

Средняя кубическая

К=3;

$$\overline{X} = \frac{\sqrt[3]{\sum X_i^3}}{n}$$

$$\overline{X} = \sqrt[3]{\frac{\sum X_i^3 \cdot f_i}{\sum f_i}}$$

или

Средняя биквадратическая

К=4;

или
$$\overline{X} = \frac{\sqrt[4]{\sum X_i^4}}{n}$$

$$\overline{X} = \sqrt[4]{\frac{\sum X_i^4 \cdot f_i}{\sum f_i}}$$

Правило мажорантности средних

Для одной и той же совокупности существуют строго определенные соотношения между разными видами средних. Эти соотношения называют *правилом мажорантности средних.*

$$\overline{X}_{ГАРМОНИЧЕСКАЯ} \leq \overline{X}_{ГЕОМЕТРИЧЕСКАЯ} \leq$$

$$\leq \overline{X}_{АРИФМЕТИЧЕСКАЯ} \leq \overline{X}_{КВАДРАТИЧЕСКАЯ}$$

Способ моментов

При исчислении средней величины в вариационном ряду с равными интервалами часто используется способ моментов

$$\overline{X} = i \cdot m_1 + A; \qquad m_1 = \frac{\sum \left(\frac{X_i - A}{i}\right) \cdot f_i}{\sum f_i}$$

где m_1 – величина момента первого порядка;

i – величина интервала;

A – центральная варианта ряда (условный 0)

Понятие моды

Мода

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

Понятие медианы

Медиана

- это

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

Мода

В интервальных рядах с равными интервалами мода вычисляется по формуле

$$M_o = X_o + i \cdot \frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m+1})}$$

где X_0 – минимальная граница модального интервала;

i – величина модального интервала;

f_m – частота модального интервала;

f_{m-1} – частота интервала, предшествующего модальному;

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

$$N_{ME} = \frac{n+1}{2}$$

где *n* – объем совокупности.

Полученное значение показывает, где точно находится номер медианной единицы (номер середины ряда). Медианное значение характеризуется тем, что его кумулятивная частота равна половине суммы всех частот или превышает ее.

Медиана

В интервальных рядах с равными интервалами медиана исчисляется 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; Σf – сумма частот ряда; S_{m-1} – сумма накопленных частот в интервалах, предшествующего медианному; f_m – частота медианного интервала.

Для определения медианного интервала необходимо рассчитать сумму накопленных частот. Медианный интервал характерен тем, что его кумулятивная частота равна полусумме всех частот ряда или превышает ее.

Квартили

Значения признака, делящее 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 (*Me*). Для расчета квартилей по интервальному вариационному ряду используют формулы

Квартили

$$Q_1 = X_{Q_1} + i\frac{\frac{1}{4}\sum f - S_{Q_1-1}}{f_{Q_1}} \qquad Q_3 = X_{Q_3} + i\frac{\frac{3}{4}\sum f - S_{Q_3-1}}{f_{Q_3}};$$

где X_{Q1} (X_{Q3}) – нижняя граница интервала, содержащего нижний (верхний) квартиль;
i – величина интервала;
S_{Q1-1} (S_{Q3-1}) – накопленная частота интервала, предшествующего интервалу, содержащий нижний (верхний) квартиль ;
f_{Q1}(f_{Q3}) – частота интервала, содержащего нижний (верхний) квартиль.

Децили

Варианты, делящие ранжированный ряд на 10 равных частей; они вычисляются по той же схеме, что и квартили:

$$d_1 = X_{d_1} + i\frac{\frac{1}{10}\sum f - S_{d_1 - 1}}{f_{d_1}}$$

$$d_2 = X_{d_2} + i\frac{\frac{2}{10}\sum f - S_{d_2 - 1}}{f_{d_2}}$$

Понятие квинтилей и перцентилей

Квинтили

- это

значения признака, делящие ряд на 5 равных частей.
Они вычисляются по той же схеме, что и квартили и децили.

Перцентили

- это

Значение признака, делящий ряд на 100 равных частей.

Понятие вариации

Вариац
ия

-
эт
о

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

Показатели вариации

- **Абсолютные**
  - размах вариации
  - среднее линейное отклонение
  - дисперсия
  - среднее квадратическое отклонение
- **Относительные**
  - коэффициент вариации
  - коэффициент осцилляции
  - линейный коэффициент вариации

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 средних величин. Вычисляется по следующим формулам:

1-й способ:

$$\sigma^2 = \frac{\sum (X_1 - \overline{X})^2}{n}$$

или

$$\sigma^2 = \frac{\sum (X_1 - \overline{X})^2 \cdot f_i}{\sum f_i}$$

где X_i – индивидуальное значение варьирующего признака (варианты);

$\overline{X}$ - среднее значение варьирующего признака;

n – количество разновидностей вариант;

f_i - показатель повторяемости вариант (частоты, веса).

2-ой способ определения дисперсии

$$\sigma^2 = \overline{X^2} - \left(\overline{X}\right)^2$$

где $\overline{X^2}$ – средняя из квадратов индивидуальных значений;

$(\overline{X})^2$ – квадрат средней величины признака.

3-й способ определения дисперсии - метод моментов

$$\sigma^2 = i \cdot (m_2 - m_1)$$

где m_1 – величина момента первого порядка;

i – величина интервала в интервальном ряду;

m_2 – величина момента второго порядка:

$$m_2 = \frac{\sum \left(\frac{X_i - A}{i}\right)^2 \cdot f_i}{\sum f_i}$$

Среднее квадратическое отклонение

Обобщающая характеристика размеров вариации признака в совокупности определяется по формуле

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum\left(X_i - \overline{\overline{X}}\right)}{n}}$$

Показывает, на какую величину в среднем значение признака отличается от стандартного значения, и выражается в тех же единицах, что и признак.

Среднее линейное отклонение

Показывает на какую величину отклоняется признак в изучаемой совокупности от средней величины признака:

$$L = \frac{\sum \left| X_i - \overline{X} \right|}{n}$$

Показатель рассчитывается по модулю.

Коэффициент вариации

Характеристика меры вариации значений признака вокруг средней величины:

$$V = \frac{\sigma}{\overline{\overline{X}}} \cdot 100$$

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 и коэффициент осцилляции

Линейный коэффициент вариации:

$$V_L = \frac{L}{\bar{X}} \cdot 100$$

Коэффициент осцилляции:

$$V_n = \frac{R}{\bar{X}} \cdot 100$$

Математические свойства дисперсии

Свойство	Методика расчета
Если из всех значений вариант отнять какое-то постоянное число А, то средний квадрат отклонений от этого не изменится	$\sigma^2{(X-A)} = \sigma_x^2$
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	$\sigma^2\left(\frac{X}{A}\right) = \frac{\sigma_x^2}{A^2}$
Если исчислить средний квадрат отклонений от любой величины, в той или иной степени отличающейся от средней арифметической, то он всегда будет больше среднего квадрата отклонений, исчисленного от средней арифметической	$\sigma_A^2 \rangle \sigma_x^2$

Свойство минимальности дисперсии

Свойство минималь-ности дисперсии

- это

дисперсия средней всегда меньше дисперсий, исчисленных от любых других величин.

Понятие альтернативного признака

Альтернативный признак

признак, которым обладают одни единицы и не обладают другие единицы совокупности

Средняя и дисперсия альтернативного признака

Среднее значение

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

так как

$$0 \cdot q = 0 \qquad p + q = 1$$

Дисперсия

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

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

$$= p \cdot q$$

p – доля единиц, обладающих признаком, в численности всей совокупности;

q – доля единиц совокупности, не обладающих этим признаком.

Закон сложения (разложения) дисперсий

Общая дисперсия

$$\sigma^2 = \frac{\sum\left[x/(-\bar{x})\right]^2 \cdot f_i}{\sum f_i}$$

или

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

характеризует вариацию признака во всей совокупности под влиянием всех факторов, обусловивших эту вариацию

Межгрупповая дисперсия δ^2

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

Межгрупповая дисперсия

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

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

где $\overline{X}$ - общая средняя;

$\overline{X}_i$ - средняя *i* - группы;

f_i – частота *i* - ой группы.

Средняя из внутригрупповых дисперсий

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

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

где $\overline{\sigma}_i^2 = \frac{\sum \left(X_i - \overline{X}_i\right)^2 \cdot f_i}{\sum f_i}$ - внутригрупповая дисперсия *i* – ой группы

Эмпирическое корреляционное отношение

Эмпирическое корреляционное отношение $\eta = \sqrt{\dfrac{\delta^2}{\sigma^2}}$

изменяется в пределах [0,1]

и характеризует влияние признака, положенного в основание группировки. Если η=0, то группировочный признак не оказывает влияния на результативный. Если η=1, то результативный признак изменяется только в зависимости от признака, положенного в основание группировки, а влияние прочих факторных признаков равно нулю.

Шкала значений эмпирического корреляционного отношения

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

Категория	Границы значений
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

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

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