Средние величины. Ч.2

К.и.н., доцент кафедры Истории РБ,

археологии и этнологии

Р.Р.Газизов

Дисперсия .Среднее квадратическое отклонение
Средняя квадратическая
Средние показатели динамики

Дисперсия

- Простейшим способом изучения вариации признака в совокупности является размах вариации или ее амплитуда (R) Величина R определяется как разность между максимальным и минимальным значениями признака в изучаемой совокупности.

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

- Пример 3. Распределение рабочих N-ского предприятия по возрасту.

Возраст	До 20	20-30	30-40	40-50	Старше 50
Число рабочих	48	102	75	62	54

- **R=65-17=48 (лет)**

- На практике чаще всего прибегают к изучению среднего квадратического отклонения (стандартного отклонения) конкретных значений признака от его средней величины. Оно обозначается σ(сигма) или S и позволяет определить границы, в которых изменяются конкретные значения признака. Величина, насколько в среднем каждое значение признака отличается от его средней арифметической, находится по формуле:
- $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$, где
- x – конкретные значения признака;
- $\bar{x}$ - средняя арифметическая;
- n – число наблюдений.
- Стандартное квадратическое отклонение, возведенное в квадрат, называется дисперсией.

- В том случае, если мы имеем дело с группировкой, с интервальным рядом, формула видоизменяется:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 p}{\sum p}}, \quad \text{где}$$

- x – конкретные значения признака;
- $\bar{x}$ - средняя арифметическая;
- p – частота признака в группировке.

- Рассмотрим вычисление среднего квадратического отклонения для данных примера 3.

- $\sigma = \sqrt{\frac{(18,5+34,56)^2 * 48+(25-34,56)^2 * 120+(35-34,56)^{22} * 62+(57,5-34,56)^2 * 54}{48+120+75+62+54}} \approx 12,77$

- Полученное значение говорит о том, что в рассматриваемой совокупности рабочих N-ского предприятиях их возраст в среднем отклонялся на 12,77 лет от средней величины, равной 34,56 лет.

Достаточно просто вычисляется среднее квадратическое отклонение для определения размаха вариации качественных (альтернативных) признаков. Формула выглядит так:

$$\sigma = \sqrt{\frac{p_1 p_2}{n}},$$ где

p2 - частота второй варианты признака;
n – число наблюдений.

- ***Пример 4:***

- Даны сведения об успеваемости группы студентов в количестве 24 человек. После очередной экзаменационной сессии 6 человек имеют задолженности по тем или иным учебным дисциплинам, а 18 человек сдали успешно все экзамены.

- Дисперсии при интерпретации выражаются в тех же единицах, что и сами признаки. Это приводит, к тому, что будучи выражены в разных единицах измерения, средние квадратические отклонения несравнимы. То есть, нельзя сравнивать количество детей с земельной площадью. В случае необходимости пользуются коэффициентом вариации (V), определяемым как отношение стандартного отклонения к средней арифметической.

$$V = \sigma / \bar{x}$$

- Полученную величину можно выразить в процентах. Сопоставление коэффициентов вариации нескольких признаков расширяет возможности исследователя при анализе и интерпретации распределений признаков - их равномерности, нормальности, колеблемости.

Какое место занимает дисперсия в исторических исследованиях?

- Во-первых, она является необходимым и обязательным дополнительным показателем при сравнении средних и сопоставлении различных группировок.
- Во- вторых, с ее помощью проверяется и обосновывается правомерность применения математических методов. Дисперсия служит своеобразным индикатором однородности изучаемой совокупности и нормальности ее распределения.
- В-третьих, сравнение дисперсий различных признаков позволяет судить об их качественном значении в рассматриваемой системе. Дисперсии помогают не потерять сглаженное средними показателями своеобразие признаков изучаемого явления.

Средняя квадратическая

$$\overline{X}_{кв} = \sqrt{\frac{X_1^2 + X_2^2 + X_3^2 + \ldots + X_n^2}{n}}$$

- Пример 5:
- Предположим, что имеются три участка земли. Протяженность одного -100 м, второго - 200 м, третьего - 300 м. Надо определить среднюю протяженность земельного участка. Величина средней арифметической = 200 м [(100+200+300)/3]. Ее реальность можно проверить, подсчитав площадь земельных участков, предположив, что они квадратной формы. Реальная площадь - $100^7+200^2+ 3\ 00^2 = 140\ 000$ м^2, а площадь трех участков со стороной 200 м - $3(200)^2=120\ 000$ м^2. Получилось, что мы "потеряли" в виду усреднения 20 000 м$^{2.}$ Следовательно, средняя арифметическая нас не удовлетворяет.

Средние показатели динамики

- К средним показателям динамики относятся средний уровень ряда, средние абсолютные изменения и ускорения, средний темп роста. Все они выступают характеристиками тенденции.
- Средний уровень (y) интервального динамического ряда определяется как простая средняя арифметическая из уровней за равные промежутки времени или как средневзвешенная из уровней за неравные промежутки времени, длительность которых выступает в качестве "весов".

- Пример 6:

- **Добыча нефти в СССР в 1976 - 1980 гг.**

годы	1976	1977	1978	1979	1980
добыча нефти в млн.т	519,7	545,8	571,5	586,0	603,2

- Как показывают данные таблицы, промежутки времени в примере 6. равные: по одному году. Значит мы должны применить здесь формулу простой средней арифметической для определения среднегодового уровня добычи нефти за 5 лет.

- Средний уровень годовой добычи нефти за период 1976 - 1980 х.г. составил: 565,2 млн.т.

- Пример 7.

  - **Распределение занятости учебно-вспомогательного персонала в приемной комиссии.**

числа в июле	1-14	15-21	22-27	28-31
кол-во работн.	20	16	19	24

- В примере 7 временные промежутки разные - 14 дней, 7 дней, 6 дней, 4 дня. В данном случае надо использовать формулу средневзвешенной.

- В моментном ряду средняя величина характеризует обобщенное значение признака между начальным и конечным моментом наблюдения. Следовательно, начальный и конечный уровни лишь наполовину относятся к изучаемому отрезку времени, а на половину к прошлому и будущему периодам. Это обстоятельство определило ***формулу средней хронологической***.

- $\overline{y}_{\text{хр}} = \dfrac{\frac{y_1}{2}+y_2+y_3+\frac{y_n}{2}}{n-1}$

- **Пример 8.**
- На 1 января 1924 г. в Средне-Волжском районе было зарегестрировано 47 546 переселенцев-мужчин (это не значит, что все они прибыли сюда 1 января), на 1 января 1925 г.- 46 725 мужчин, на 1 января 1926 г.-64 368 мужчин. Определить среднегодовое количество прибывших в Средне-Волжский район мужчин в 1924-26 гг.

- В случае, если промежутки между датами моментного ряда не равны, хронологическая средняя вычисляется по формуле:

$$\bar{У}_{хр} = \frac{У_1T_1 + У_2(T_1 + T_2) + У_3(T_2 + T_3) + \ldots У_n T_{n-1}}{2(T_1 + T_2 + T_3 + \ldots + T_{n-1})},$$ где

- Т – промежутки между датами;
- Y12,Y3…Yn – уровни ряда;
- n – количество уровней.

- Часть математиков считают проблему вычисления среднего уровня моментного ряда при неравных временных промежутках спорной. Однако в исторических исследованиях использование этой формулы возможно при тщательном контроле исходных данных и результатов вычисления качественным анализом.

- Один из наиболее важных средних показателей динамического ряда - ***средний темп изменения*** (роста или сокращения).
- $\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}}$, где
- Y1 = начальный уровень динамического ряда;
- Yn = последний уровень динамического ряда;
- n – число временных промежутков.
- Вычислите среднюю геометрическую по данным предыдущего примера 8.

- С помощью средней геометрической величины мы получили среднюю скорость изменения признака. В нашем распоряжении есть еще один средний показатель - ***средний абсолютный прироста*** (абсолютное значение). Он показывает на какую величину (в единицах измерения уровней ряда) показатель одного временного периода больше или меньше любого предшествующего. При возрастании уровней абсолютное изменение принимает положительное, а при уменьшении - отрицательное значения.
- $\Delta\bar{y} = \frac{y_n - y_1}{n}$, где
- Y1 = начальный уровень динамического ряда;
- Yn = конечный уровень динамического ряда;
- n – число временных промежутков.

- Определение "начального" и "конечного" уровней динамического ряда в каждом вычислении зависит от задач исследования. По одной группировке можно определить несколько средних значений абсолютного прироста за разные временные промежутки.
- Подсчитате средний абсолютный прирост по данным примера 8.
- У1,=47546; Уn= 64368. Чему равно n?

- Как видим, в 1924-1926 гг. в Среднее Поволжье ежегодно прибывали в среднем по 8411 мужчин. Интерпретация $\Delta\overline{y}$ и $\overline{T}$ сопровождается обязательным указанием двух хронологических единиц - периода, который характеризуется (в нашем случае - 1924-1926 гг.) и периода, на который рассчитан средний показатель (в нашем случае - выяснялся ежегодный показатель).

- По среднему темпу роста можно без труда определить средний темп прироста, вычитая из значения Т единицу. В нашем примере Т = 1,16. Тогда средний темп прироста: 1,16-1 = 0,16. Полученное значение можно выразить в процентах, умножив его на 100% (у нас 0,16 100% = 16%).
- Разделив абсолютный прирост на темп прироста (за соответствующий период) получим среднее абсолютное значение прироста
- $\bar{\alpha} = \frac{\Delta\bar{y}}{\bar{\mathrm{T}}}$

- Приведенные показатели служат основными характеристиками, применяемыми для анализа динамических рядов. Они позволяют судить об абсолютном и относительном изменениях уровней ряда. В заключение необходимо сделать несколько замечаний.
- 1. Все перечисленные показатели обладают высокой точностью и достоверностью при небольших колебаниях в значениях признака.
- 2. Средние хронологические особенно полезно вычислять при сравнительном анализе двух и более динамических рядов.

Дополнительная литература

- 1. Джини К. Средние величины. - М., 1970.
- 2. Елисеева И.И., Юзбашев М.М. Общая теория статистики. - М., 1995. - С.66-103, 257-306.
- 3. Измайлова М.О., Рахманкулов И.Ш. Категория "средняя величина" и ее методологическое значение в научном исследовании. - Казань, 1982.
- 4. Славко Т.И. Математико-статистические методы в исторических исследованиях. - М., 1981. - С.47-57.
- 5. Пасхавер И.С. Средние величины в статистике. - М., 1979.
- 6. Общая теория статистики. - М., 1984. - С.54-78, 94-104, 195-201.

Спасибо за внимание!