Статистические показатели в форме средних величин

Т.Н. Тарасова – к.п.н., доцент кафедры административного и финансового права

План лекции

1 Степенные средние

1.1 Средняя арифметическая и её свойства
1.2 Средняя геометрическая
1.3 Средние более высоких порядков

2 Структурные средние

2.1 Определение моды вариационного ряда
2.2 Определение медианы вариационного ряда

3 Показатели вариации

1 Степенные средние

1.1 Средняя арифметическая и её свойства

Средняя величина

- это

обобщающий показатель,
который дает количественную
характеристику признака
в статистической совокупности
в условиях конкретного
места и времени.

Показатель в форме средней величины обладает следующими свойствами:

- представляет собой универсальную обобщающую характеристику статистической совокупности;
- отражает типичный уровень изучаемого признака;
- является центром распределения.

Условия правильного применения средней величины:

- исчисляется только для совокупностей, состоящих из однородных единиц;
- для совокупности, неоднородной в качественном отношении, необходимо разделение на однородные группы и вычисление для них групповых средних, характеризующих каждую из этих групп;
- кроме средней величины следует исчислять другие показатели, поскольку средняя величина сглаживает индивидуальные значения
- среднюю величину исчисляют не для отдельных единичных фактов, а для их совокупности.

Виды средних величин

Степенные

- Гармоническая
- Геометрическая
- Арифметическая
- Квадратическая
- Кубическая
- Биквадратическая

Структурные

- Мода
- Медиана
- Квартили
- Децили
- Квинтили
- Перцентили

Элементы степенной средней

Варианта (X)

Признак, для которого исчисляется средняя величина.

Число единиц (n)

Количество вариант в исследуемой совокупности

Веса, частоты (f)

Показатели повторяемости вариант в исследуемой совокупности

Средняя арифметическая простая:

$$\overline{x} = \frac{\sum x_i}{n},$$

где *x* - значение признака *i*-й единицы совокупности;

n – объём совокупности.

Пример: Доходы пяти банков по операциям с ценными бумагами за отчетный период составили: 0,6; 0,7; 0,9; 1,1; 1,3 млн. руб. Определить средний доход банка по данной операции.

$$\overline{\mathrm{X}} = \frac{0{,}6 + 0{,}7 + 0{,}9 + 1{,}1 + 1{,}3}{5} = \frac{4{,}6}{5} = 0{,}92\textit{ìëí .ðóá.}$$

Средняя арифметическая взвешенная

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i},$$

f – частота.

Пример.

Сделки по акциям элемента «Х» за торговую сессию

Сделка	Количество проданных акций, шт.	Курс продажи, руб.	Общая сумма сделок, xf
1	700	420	294000
2	200	440	88000
3	950	410	389500
Итого	1850		771500

$$\bar{X} = \frac{771500}{1850} = 417{,}03\, руб.$$

Свойства средней арифметической:

1.Произведение средней на сумму частот равно сумме произведений отдельных вариантов на соответствующие им частоты:

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

2. Сумма отклонений индивидуальных значений признака от средней арифметической равно нулю:

$$\sum (x - \bar{x}) f = 0$$

3. Если все значения признака уменьшить или увеличить на постоянное число А, то средняя арифметическая соответственно уменьшится или увеличится на ту же величину

$$\frac{\sum (x \pm A) f}{\sum f} = \bar{x} \pm A$$

4. Если все варианты значений признака уменьшить или увеличить в А раз, то средняя также соответственно увеличится или уменьшится в А раз

$$\frac{\sum \frac{x}{A} f}{\sum f} = \frac{\frac{1}{A}\sum xf}{\sum f} = \frac{1}{A}\bar{x}$$

.

5.Если все веса уменьшить или увеличить в А раз, то средняя арифметическая от этого не изменится

$$\frac{\sum x\frac{f}{A}}{\sum\frac{f}{A}}=\frac{\frac{1}{A}\sum xf}{\frac{1}{A}\sum f}=\bar{x}$$

1.2 Средняя геометрическая

Средняя геометрическая невзвешенная (простая)

$$\overline{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot x_n} = \sqrt[n]{\prod_{i-1}^{n} x_i}$$

где

$\tilde{o}_i$ *− çíà÷åíèå ïðèçíàêà i − é åäèíèöû ñîâîêóïíîñòè ;*

n *− îáúåì ñîâîêóïíîñòè*

Средняя геометрическая взвешенная

$$\overline{x} = \sqrt[\sum f]{x_1^{f1} \cdot x_2^{f2} \cdot x_3^{f3} \cdot x_k^{fk}} = \sqrt[\sum f]{\prod_{i-1}^{f} x_i^{fi}},$$

где

x_i *− i − й вариант признака;*

f_i *− вес i − го варианта;*

k *− число вариантов осредняемого признака.*

1.3 Средние более высоких порядков

Средняя квадратическая

$$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}} - \textit{невзвешенная};$$

$$\bar{x} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}} - \textit{взвешенная};$$

Применяется при изучении вариации признака.

Средняя кубическая

$$\bar{x} = \sqrt[3]{\frac{\sum x_i^3}{n}} - \textit{невзвешенная};$$

$$\bar{x} = \sqrt[3]{\frac{\sum x_i^3 \cdot f_i}{\sum f_i}} - \textit{взвешенная};$$

Основная область применения степенных средних второго и более высоких порядков – расчет показателей вариации, взаимосвязи, структурных изменений, асимметрии и эксцесса.

Правило мажорантности средних

$$\overline{X}_{\text{ãåîìåòð}} \le \overline{X}_{\text{àðèôìåò}} \le \overline{X}_{\text{êâàäðàò}}$$

2 Структурные средние

2.1 Определение моды вариационного ряда

Мода (Мо) – вариант изучаемого признака, имеющий наибольшую частоту. Мода отражает то значение признака, которое является наиболее типичным, преобладающим, доминирующим.

При большом числе наблюдений совокупность может характеризоваться двумя и более модальными вариантами.

При определении моды по дискретному ряду распределения подсчитываются частоты, соответствующие каждому варианту изучаемого признака:

Вариант признака x_i	Частота f_i
x_1	f_1
x_2	f_2
x_3	f_3
.	..
.	
$x_{Mo}(Mo)$	.
.	f_{max}
.	
x_k	.
	.
	f_k
Итого	$\sum f_i$

Модальным вариантом, или *модой*, является вариант, которому соответствует максимальная частота

Пример

При обследовании 500 семей рабочих одной из отраслей промышленности установлены следующие показатели количества членов семей:

Количество членов семьи, человек	Число семей
2	50
3	80
4	260
5	40
6	30
7	20
8	10
9	10

Определить моду данного вариационного ряда распределения

При определении моды по *интервальному ряду распределения* необходимо первоначально определить модальный интервал - интервал, имеющий наибольшую частоту.

Значение моды внутри данного интервала определяется по формуле:

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

где x_0 – нижняя граница модального интервала;

i – величина интервала;

f_{Mo} – частота модального интервала;

$f_{\text{Mo-1}}$ – частота интервала, предшествующего модальному;

$f_{\text{Mo+1}}$ – частота интервала, следующего за модальным.

Пример

Определить моду продолжительности стажа работы работников предприятия

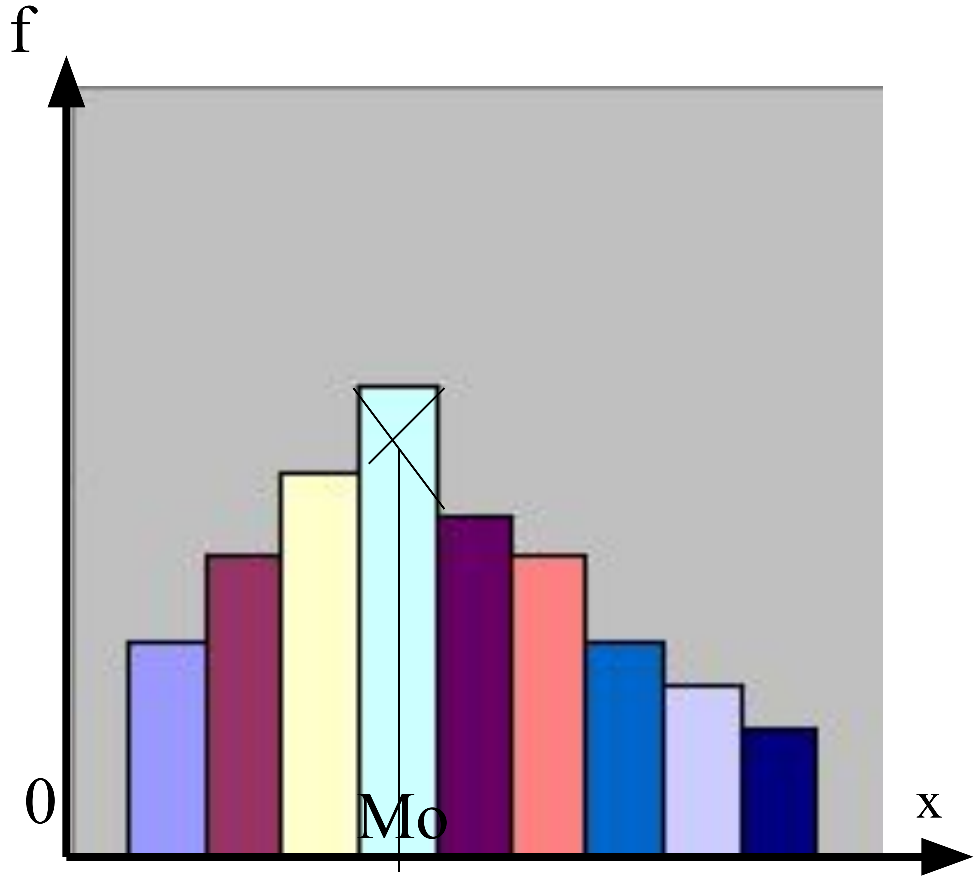
Распределение работников предприятия по продолжительности стажа работы

Группы работников по продолжительности стажа работы, лет	Число работников, человек
До 2	4
2-4	23
4-6	20
6-8	35
8-10	11
Свыше 10	7

Мода продолжительности стажа работы работников торгового предприятия составит:

$$M_0 = 6 + 2\frac{35-20}{(35-20)+(35-11)} = 6{,}77\,года$$

Для графического определения моды необходимо простроить *гистограмму* и соединить верхние углы модального столбика и соприкасающихся с ним столбиков таким образом, как это представлено на рисунке.



2.2 Определение медианы по ряду распределения

Медиана (Ме) – вариант изучаемого признака, находящийся в середине ранжированного (упорядоченного) ряда всех его значений.

Основное свойство медианы:

Сумма модулей отклонений всех значений признака от медианы всегда меньше, чем сумма таких отклонений от любой другой произвольной постоянной:

$$\sum \left| x_i - \grave{I}\mathring{a} \right| = \min$$

При определении медианы по дискретному ряду распределения рассчитываются накопленные частоты.

Вариант признака, x_i	Частота, f_i	Накопленная частота, f_i^H
x_1	f_1	$f_i^H = f_1$
x_2	f_2	$f_2^H = f_1^H + f_2$
x_3	f_3	$f_3^H = f_2^H + f_3$
$\vdots$	$\vdots$	$\vdots$
x_{Me-1}	f_{Me-1}	$f_{Me-1}^H \langle \frac{1}{2}\sum f_i$
x_{Me}(Me)	f_{Me}	$f_{Me-1}^H \rangle \frac{1}{2}\sum f_i$
$\vdots$	$\vdots$	$\vdots$
x_k	f_k	f_k^H
Итого	$\sum f_i$	X

Медианным вариантом, или *медианой*, будет первый вариант, накопленная частота которого превышает половину суммы всех частот.

Для **дискретного ранжированного ряда с нечетным числом членов** медианой является варианта, расположенная в центре ряда.

Пример

Процент выполнения плана товарооборота за месяц 13 торговых предприятий составил (%): 95; 98; 101; 104; 109; 115; 119; 126; 135; 144; 176; 202; 223. Определить медиану.

Решение

Медианой будет седьмая варианта, которая делит упорядоченный ряд пополам и соответствует 119% выполнения плана товарооборота.

Для **дискретного ранжированного ряда с чётным числом членов**, медианой будет варианта, рассчитанная как средняя арифметическая двух смежных центральных вариант

Пример

Сведения о стаже работы шести работников предприятия:

1, 3, 4, 5, 7, 9 лет.

Определить медиану стажа работы работников.

Решение

$$M_e = (4 + 5) \div 2 = 4{,}5 \textit{года}$$

При определении медианы по *интервальному ряду распределения* необходимо рассчитывать медианный интервал, т. е. первый интервал, накопленная частота которого превышает половину общей суммы частот.

Значение медианы внутри данного интервала определяется по формуле:

$$Me = x_0 + i\frac{\frac{1}{2}\sum f_i - f_{Me-1}^{H}}{f_{Me}},$$

где x_0 – *нижняя граница медианного интервала;*

i – *величина интервала;*

f_{Me} – *частота медианного интервала;*

f_{Me-1}^{H} – *накопленная частота интервала, предшествующего медианному.*

Пример

Определить медиану продолжительности стажа работы работников торгового предприятия

Распределение работников торгового предприятия по продолжительности стажа работы

Группы работников по продолжительности стажа работы, лет	Число работников, человек	Накопленная частота
До 2	4	4
2-4	23	27
4-6	20	47
6-8	35	82
8-10	11	
Свыше 10	7	
Итого	100	

$$M_e = 6 + 2\frac{\frac{100}{2} - 47}{35} = 6{,}17 \text{ года}$$

3 Показатели вариации

- Если отдельные значения изучаемого признака существенно отличаются от средней величины, то наряду с самой средней величиной выявляют и величину отклонения (вариации) отдельных признаков.
- Поэтому средние характеристики дополняют *показателями вариации признака.*

1) *размах вариации (R)*

равен разности между наибольшим и наименьшим значениям признака:

$$R = X\ max - X\ min$$

$$\overline{d} = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right|}{n}$$

2) *среднее линейное отклонение (d):*

для сгруппированных данных:

$$\overline{d} = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right| f_i}{\sum_{i=1}^{n} f_i}$$

для несгруппированных данных:

$$\overline{d} = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right|}{n}$$

3) *дисперсия (D):*

для сгруппированных данных:

$$D = \frac{\sum_{i-1}^{k} (x_i - \overline{x})^2 \cdot f_i}{\sum_{i=1}^{k} f_i}$$

для несгруппированных данных:

$$D = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}{n}$$

4) *среднее квадратическое отклонение (σ):*

- для сгруппированных и несгруппированных данных:

$$\sigma = \sqrt{D}$$

Свойства дисперсии и среднего квадратического отклонения:

1) дисперсия и среднее квадратическое отклонение постоянной величины равны нулю;

2) дисперсия и среднее квадратическое отклонение не меняются, если все варианты увеличить или уменьшить на какое-то постоянное число;

3) если все варианты умножить на какое-то постоянное число $A \neq 0$, то дисперсия увеличится в A квадрат раз, а среднее квадратическое отклонение – в A раз.

5) коэффициент вариации:

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

- Если V достаточно большой (более 40 %) – это означает, что типичность средней невысока и, наоборот, если его значение мало, то средняя величина признается типической и надежной характеристикой изучаемой совокупности.