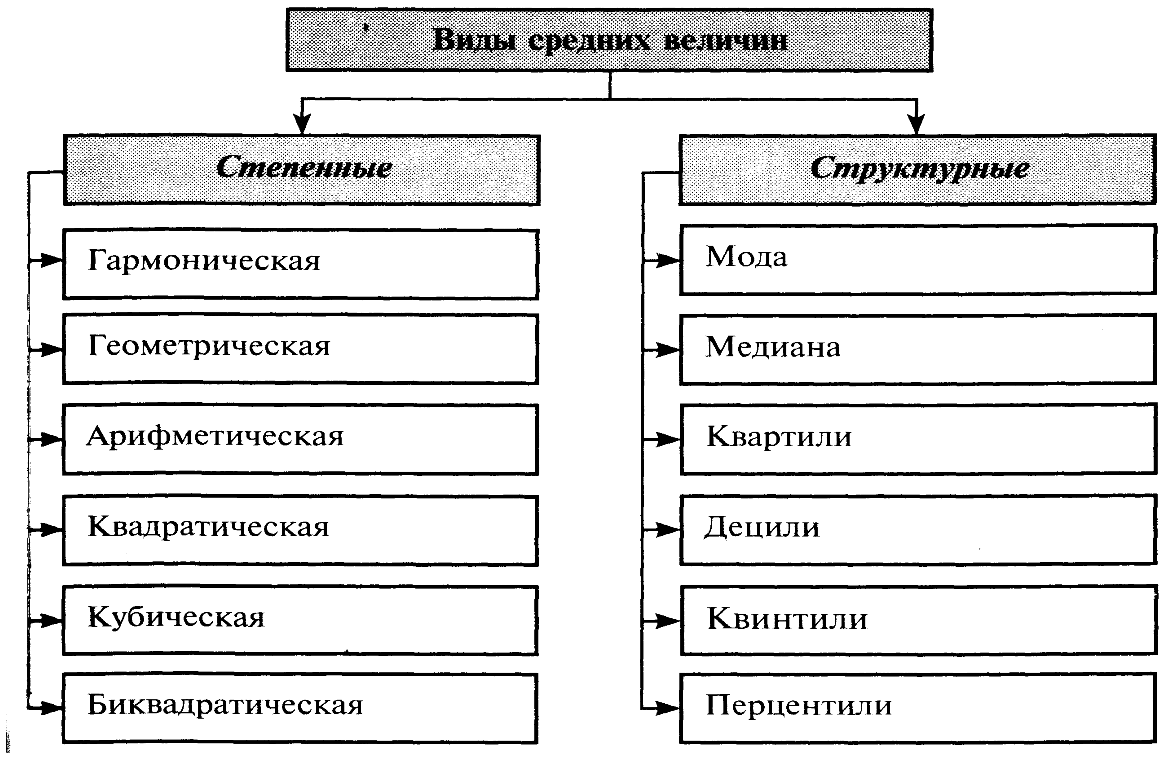
Лекции № 5

Абсолютные, относительные и средние статистические показатели

(Продолжение)

Показатели среднего значения ряда

Среди обобщающих величин, характеризующих статистическую совокупность, важнейшую роль играют показатели *среднего* или *центрального значения* ряда. Совокупность, изучаемая по количественному признаку, состоит из индивидуальных значений, на них оказывают влияние как общие причины, так и индивидуальные условия. При расчете среднего индивидуальные особенности погашаются, и его значение отражает то общее, что присуще всем наблюдаемым единицам.



Степенные средние

Общая форма степенной средней простой: $\overline{x}_{стер.\,пр.} = \sqrt[\alpha]{\dfrac{\sum_{i=1}^{n} x_i^{\alpha}}{n}}$.

Общая форма степенной средней взвешенной: $\overline{x}_{степ.\,взв.}=\sqrt[\alpha]{\dfrac{\sum\limits_{i=1}^{k}x_i^{\alpha}f_i}{\sum\limits_{i=1}^{k}f_i}}$.

Если $\alpha=1$ -средняя арифметическая, $\alpha=2$ -средняя квадратическая, $\alpha=-1$ -средняя гармоническая и т.д.

Квартили - значения признака, делящие ранжированную совокупность на четыре равные части.

Децили - значения признака, делящие ранжированную совокупность на десять равные части, квинтили – на пять равных частей, перцентили – на сто равных частей.

Среднее арифметическое

Средняя арифметическое является наиболее распространенным видом среднего.

Рассмотрим, как вычисляется этот показатель для несгруппированных данных, а также сгруппированных данных, представленных в виде таблиц дискретного и интервального частотных распределений. Для несгруппированных данных применяется *простое* среднее арифметическое:

$$m=\left(x_1+x_2+\cdots+x_N\right)\Big/N=\left(\sum_{i=1}^{N}x_i\right)\Big/N$$

где x_i - наблюдаемые значения признака, N - число единиц совокупности.

Пример 1. Возраст 11 студентов института составляет: 18; 24; 20; 35; 19; 23; 26; 23: 19; 21; 18 лет. Рассчитать среднее арифметическое.

$m=\left(18+24+20+35+19+23+26+23+19+21+18\right)/11=246/11\approx 22{,}4$ года

Для расчета среднего арифметического сгруппированных данных применяют формулу *взвешенного* среднего арифметического, в которой в качестве весов выступают частоты. Для дискретного частотного распределения формула имеет вид:

$$m = \left(f_1 x_1 + f_2 x_2 + \cdots + f_N x_N \right) / N = \left(\sum_{i=1}^{N} \left(f_i x_i \right) \right) \Big/ N$$

В случае интервального распределения усредняемой величиной в формуле будут средние значения интервалов x_{im}

$$m = \left(f_1 x_{m1} + f_2 x_{m2} + \cdots + f_N x_{mN} \right) / N = \left(\sum_{i=1}^{N} \left(f_i x_{mi} \right) \right) \Big/ N$$

Пример 2. Найти среднее арифметическое для времени выполнения теста двадцатью студентами (табл. 1.2).

$m = \left(2 \times 5 + 1 \times 6 + 5 \times 7 + 7 \times 8 + 3 \times 9 + 2 \times 10 \right) / 20 = 154/20 \approx 7.7$ минут

Пример 3. Найти среднее арифметическое для пробега 20 автомобильных шин (табл. 1.3).

$$m = \left(2 \times 8.5 + 1 \times 13.5 + 3 \times 18.5 + 5 \times 23.5 + 4 \times 28.5 + 3 \times 33.5 + 2 \times 38.5 \right) / 20 = 495/20 \approx 24.75$$

т = (2x8.5 +1x13.5+3x18.5+5x23.5+ 4x28.5 +3x33.5+ 2x38.5)/20 = 495/20« 24.75.

Медиана

Медианой Me называют значение признака, которое приходится на середину упорядоченного по возрастанию (ранжированного) ряда. Согласно этому определению, медиана делит совокупность на две равные по численности группы - у 50% единиц значение признака будет больше медианы, а у 50% - меньше (если медиана не совпадет ни с одним из значений). Медиана относится к так называемым структурным или порядковым показателям ряда.

Нахождение медианы для несгруппированных данных производится в два этапа -сначала все данные упорядочиваются по возрастанию, а затем находится само медианное значение $Me:$ для четного числа единиц совокупности $Me = \left(x'_{N/2} + x'_{N/2+1} \right) / 2$ для нечетного - $Me = x'_{(N+1)/2}$, где через x' обозначены значения упорядоченного ряда.

Пример 1. Возраст студентов. Упорядоченный по возрастанию ряд x'

имеет вид: 18; 18; 19; 19; 20; 21; 23; 23; 24; 26; 35. Так как число элементов нечетное $N = 11$. то $Me = x'_{(N+1)/2} = x'_{12/2} = x'_6 = 21$.

Для дискретного частотного распределения наиболее просто находить медиану по значениям накопленных частостей. Просматривая колонку этих величин, нужно найти *медианную группу* I_{me} - группу, в которую попадает значение 50%. Значение признака медианной группы и будет медианой: $Me = x_{I_{me}}$.

Пример 2. Время выполнения теста (табл 1.2).

Значение 50% для данного распределения попадает в группу $I_{me} = 4$, которой соответствует диапазон накопленных частостей от 40% до *75%.* Следовательно, значение медианы $Me = x_4 = 8$.

Схема нахождения медианы интервального частотного распределения несколько более сложная. Сначала, так же как и в предыдущем случае, определяется медианная группа (интервал) I_{me}, а затем значение медианы находится с помощью соотношения:

$$Me = \frac{50 - cw_{I_{me}-1}}{w_{I_{me}}} \times H + x_{l,I_{me}}$$

где $cw_{I_{me}}$ обозначает накопленную частость группы, предшествующей медианой, $w_{I_{me}}$ - частость медианой группы, H - ширина интервалов, $x_{l,I_{me}}$ - нижняя граница медианной группе. Формула, довольно громоздкая на первый взгляд, представляет собой линейное интерполяционное соотношение, записанное для величины 50% в медианной группе.

Пример 3. Пробег автомобильных шин (табл. 1.3).

Медианной группой распределения является $I_{me} = 4$ - этой группе соответствует диапазон накопленных частостей от 30%) до 55%. Далее выписываем все величины, входящие в интерполяционную формулу: $cw_3 = 30\%$, $w_4 = 25\%$, $H = 5$, $x_{1,4} = 21$. Подставляя значения в формулу, получаем:

$Me = \frac{50-30}{25} \times 5 + 21 = 25$ Медиана распределения равняется 25.

Мода

Еще одним показателем центрального значения распределения является ***мода*** -наиболее часто встречающееся значение признака. По сравнению с другими показателями среднего, мода определяется наиболее легко и обладает большей устойчивостью к изменениям отдельных значений. Мода, однако, не всегда является единственной и существует не для всех распределений

Пример. Количество чашек кофе, которое заказали 18 случайно выбранных посетителей кафе, составило: 0; 0; 0; 1; 1; 1; 1; 1; 1; 2; 2; 2; 2; 3; 3; 4; 5; 5. Найти моду распределения.

Модой распределения Mo является значение 1, так как оно встречается наибольшее количество раз, 6.

Пример. В ходе испытаний для 10 автомобилей определена величина тормозного пути (в метрах) со скорости 60 км/час до полной остановки: 15; 18; 18; 18; 20; 24; 24; 24; 26; 26. Найти моду распределения.

В данном распределении два значения - 18 и 24 имеют максимальную частоту, равную 3, поэтому распределение имеет две моды $Mo_1 = 18$, $Mo_1 = 24$ и называется *бимодальным.*

Пример. В ходе эксперимента исследователь-биолог измерил время выживания (в минутах) 6 бактерий: $2;3;5;7;8;10$. Найти моду распределения.

Все значения в распределении встречаются по одному разу, поэтому выбрать значение с максимальной частотой невозможно - у распределения нет моды.

Для сгруппированных данных определение моды начинается с нахождения модальной группы I_{mo}, определяемой по максимальной частоте. Сама же мода равняется значению признака в модальной группе для дискретного распределения, $Mo = x_{I_{mo}}$ или среднему значению модальной группы для интервального распределения, $Mo = x_{m,I_{mo}}$.

Пример 2. Время выполнения теста (табл. 1.2).

Наибольшую частоту ($f = 7$) имеет четвертая группа, $I_{mo} = 4$, а значение моды $Mo = x_4 = 8$. Для данного распределение медиана и мода совпадают.

Пример 3. Пробег автомобильных шин (табл. 1.3)

Наибольшую частоту ($f = 5$) имеет четвертая группа, $I_{mo} = 4$, $Mo = x_{m,4} = 23.5$.

Особенности и использования различных средних

Среднее арифметическое:

- рассчитывается по значениям признака всех единиц совокупности;
- является однозначной характеристикой;
- не может быть рассчитано для распределений с открытыми интервалами;
- сильно подвержено влиянию единичных особо больших и малых значений.

Медиана:

- слабо изменяется под влиянием единичных очень больших и малых значений;
- позволяет определить, попадает ли заданное значение в верхнюю или нижнюю половину распределения;
- используется для расчета среднего значения для частотных распределений с открытыми интервалами.

Мода:

- используется, когда нужно найти наиболее типичное значение;
- является наиболее просто рассчитываемым показателем среднего значения;
- может использоваться для распределений по атрибутивному признаку;
- не всегда является однозначной;
- существует не для всех распределений.

Форма распределения и соотношение различных средних

Частотные распределения могут иметь различную форму. Три основных типа распределения: симметричное, положительно-асимметричное и отрицательно-асимметричное, характеризуются различными соотношениями между показателями среднего значения.

Для *положительно-асимметричного* распределения большинство единиц наблюдения находится слева от среднего арифметического в области малых значений, рис. 1.7. В этом случае среднее арифметическое находится справа от медианы, а мода - слева от медианы: $Mo < Me < m$.

Для *симметричного* распределения значения расположены равномерно по обе стороны от среднего арифметического, рис. 1.8. Кроме того, для унимодального распределения медиана совпадает и со средним арифметическим, и с модой, и все они располагаются в центре распределения, $Mo \approx Me \approx m$.

Для *отрицательно-асимметричного* распределения большая часть единиц наблюдения лежит справа от среднего арифметического и расположена в области больших значений, рис. 1.9. Среднее арифметическое находится слева от медианы, а мода -справа от нее, $m < Me < Mo$.

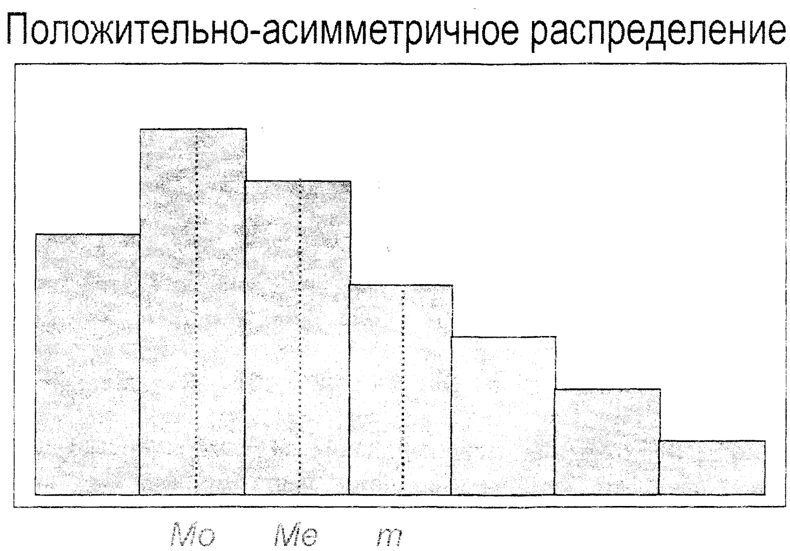


Рис. 1.7. *Положительно-асимметричное распределение*

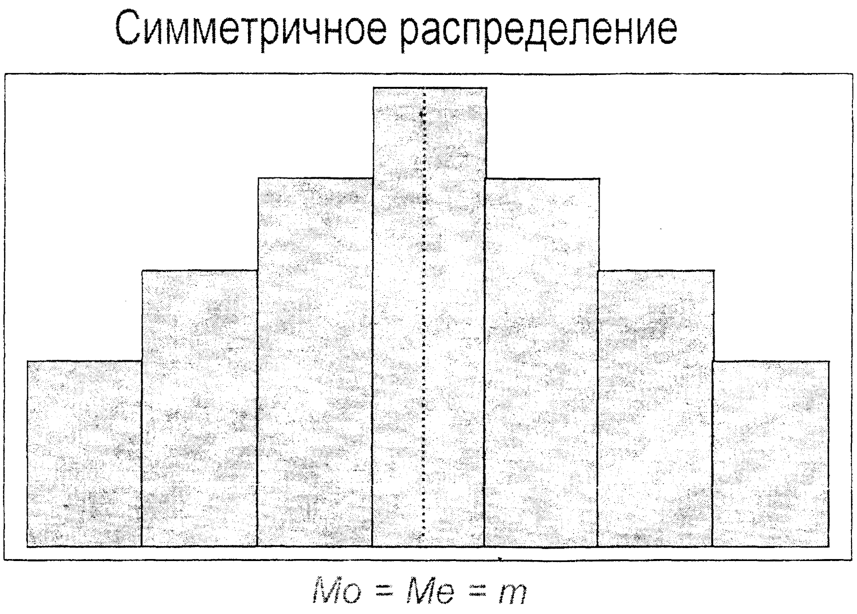


Рис. 1.8. *Симметричное распределение*

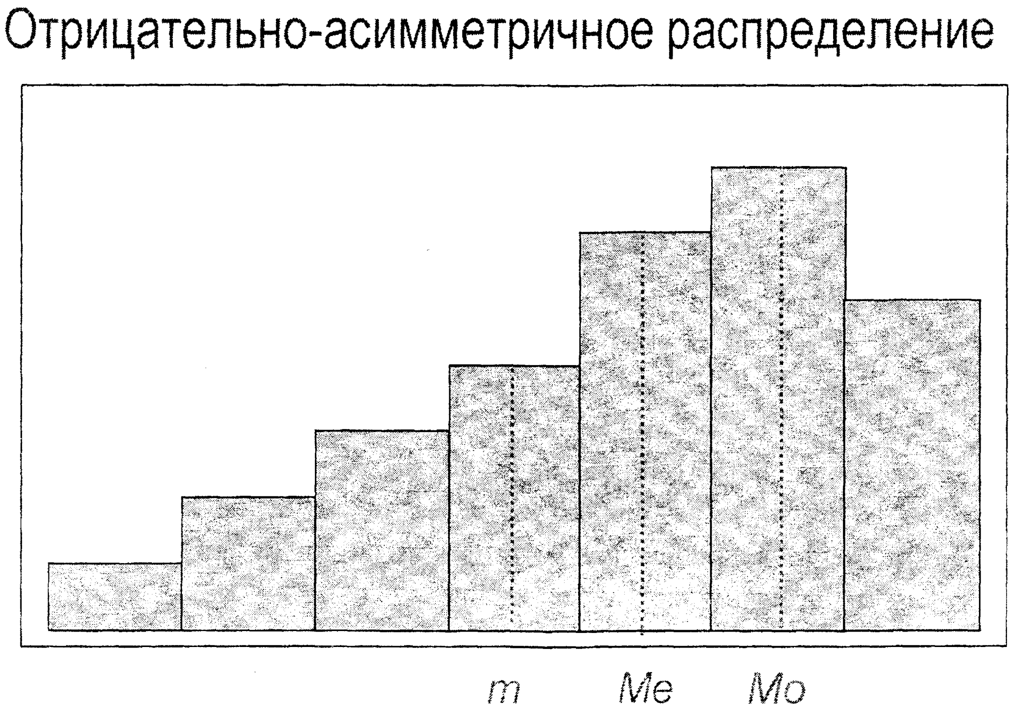


Рис. 1.9. *Отрицательно-асимметричное распределение*