МЕТОДЫ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ

Тема 3

Меры центральной тенденции

- Среднее арифметическое
- Мода
- Медиана

$\overline{X}$

СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

Среднее арифметическое $\overline{X}$ определяется как сумма всех значений измеренного признака, деленная на количество суммированных значений

Некоторый признак X измерен в группе испытуемых численностью n, получена выборка значений этого признака: x_1, x_2, x_3, …x_n . Тогда среднее арифметическое значение признака X определяется по формуле:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

Пример: найти среднее арифметическое следующего ряда: 3, 5, 6, 3, 7, 8, 10.

$$\bar{x} = \frac{3+5+6+3+7+8+10}{7} = 6$$

СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

Если вариационный ряд признака X объемом n содержит k различных значений x_i с частотами f_i , то для вычисления среднего арифметического используется формула:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n}$$

СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

Варианты (x_i)	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Частоты (f_i)	**1**	**2**	**2**	**3**	**1**	**5**	**2**	**6**	**2**	**1**

$$\bar{x} = \frac{5 \cdot 1 + 6 \cdot 2 + 7 \cdot 2 + 8 \cdot 3 + 9 \cdot 1 + 10 \cdot 5 + 11 \cdot 2 + 12 \cdot 6 + 13 \cdot 2 + 14 \cdot 1}{25} =$$

$$= \frac{248}{25} = 9{,}92 \approx 10$$

МОДА

- **Мода** (Мо) — это такое числовое значение, которое встречается в выборке наиболее часто

Варианты (x_i)	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Частоты (f_i)	**1**	**2**	**2**	**3**	**1**	**5**	**2**	**6**	**2**	**1**

Мо = 12

МОДА

- В зависимости от того, сколько значений признака удовлетворяют определению моды, различают мономодальное, бимодальное, полимодальные распределения. Распределение, имеющее одну моду, называется мономодальным

МЕДИАНА

- **Медиана** (Ме) — это такое значение признака, которое делит упорядоченное множество данных пополам так, что одна половина всех значений оказывается меньше медианы, а другая — больше.
- Первым шагом при определении медианы является упорядочивание всех значений по возрастанию

МЕДИАНА

- если объем выборки является нечетным числом, то медиана есть центральное значение

- 9, 10, 12, 13, 17, 18, 20

- Me =13

МЕДИАНА

- если объем выборки является четным числом, то медиана есть точка, лежащая посередине между двумя центральными значениями
- 9, 10, 12, 13, 17, 18, 19, 21
- Me = (13+17)/2 = 15

Для номинативных данных единственной подходящей мерой центральной тенденции является мода.

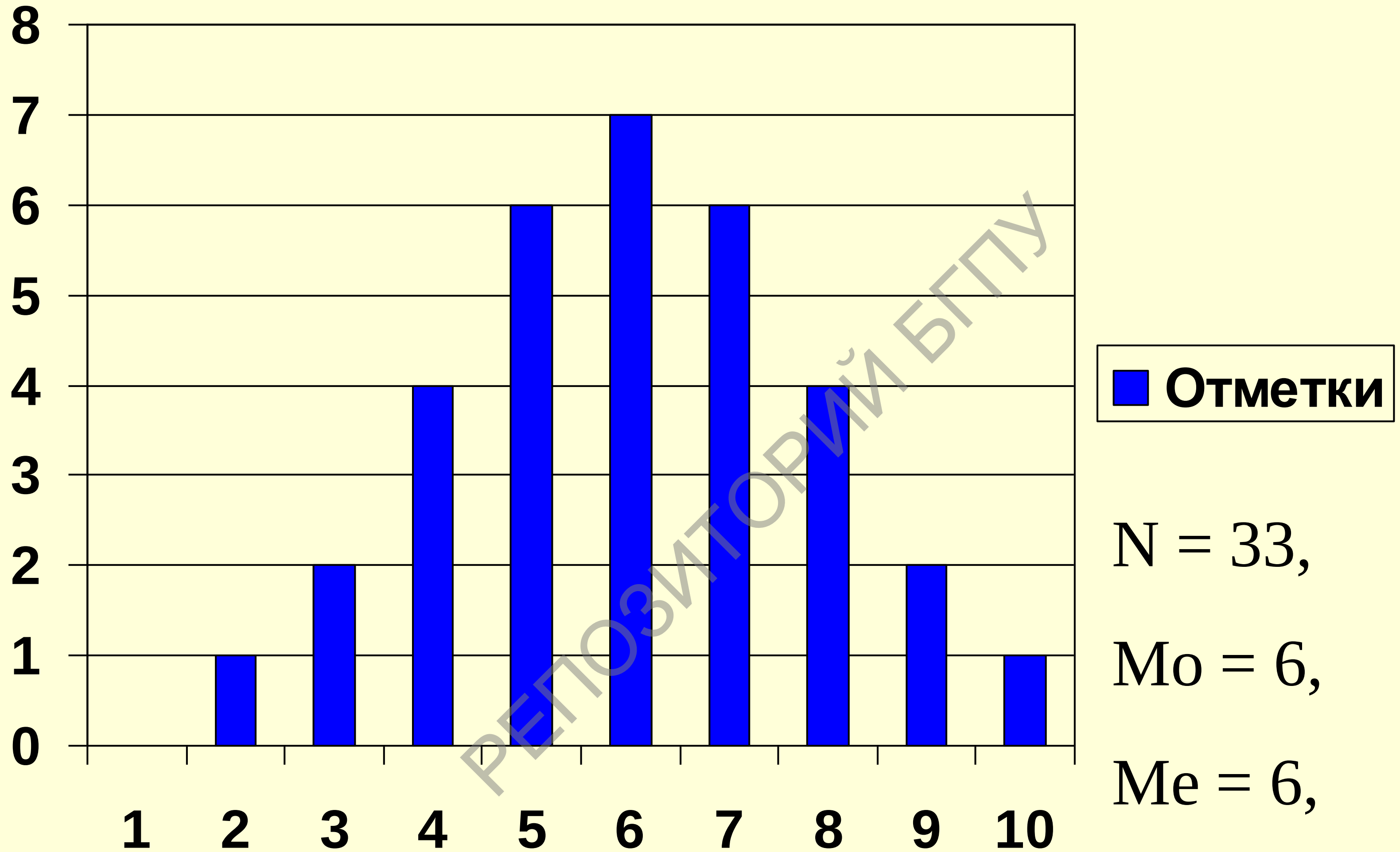
Для ранговых данных можно определять моду и медиану.

Для метрических данных можно определять все меры центральной тенденции.

Для порядковых и метрических переменных (измеренных в шкале интервалов или равных отношений), распределение которых унимодальное и симметричное, мода, медиана и среднее совпадают.

Чем больше отклонение от симметричности, тем больше расхождение между значениями этих мер центральной тенденции.

По этому расхождению можно судить о том, насколько симметрично или асимметрично распределение



$N = 33$,

$Mo = 6$,

$Me = 6$,

$\overline{X} = 6$

Свойства мер центральной тенденции:

- На величину среднего арифметического влияет каждое отдельное значение выборки. То есть среднее арифметическое значение весьма чувствительно к «выбросам» — экстремально малым или большим значениям переменной.
- На величину моды и медианы не влияет величина каждого отдельного значения.

Пример:

9 человек имеют месячный доход от 500.000 до 600.000 рублей, в среднем 560.000 рублей,

а доход десятого составляет 1.500.000 рублей.

Средний доход для этих 10 человек составляет 654.000 рублей.

Средние арифметические значения нескольких выборок можно сравнивать, если:

- группы достаточно большие, чтобы судить о форме распределения
- распределения симметричны
- отсутствуют «выбросы»

Если хотя бы одно из перечисленных условий не выполняется, то следует ограничиться модой и медианой

Меры изменчивости:

- Дисперсия
- Среднее квадратичное (стандартное) отклонение
- Коэффициент вариации

Меры изменчивости применяются для численного выражения величины вариации признака между его отдельными значениями

Дисперсия — мера изменчивости для метрических данных, пропорциональная сумме квадратов отклонений измеренных значений от их среднего арифметического.

Обозначают дисперсию по-разному: D или S^2.

$$D = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

Пример вычисления дисперсии признака X для выборки:
3, 5, 6, 3, 7, 8, 10

Вычисляем среднее арифметическое: $\overline{x}$= (3+5+6+3+7+8+10)/7 = 6.

№	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1	3-6=-3	9
2	5-6=-1	1
3	6-6=0	0
4	3-6=-3	9
5	7-6=1	1
6	8-6=2	4
7	10-6=4	16
Σ		**40**

Вычисляем дисперсию D=40/(7-1) = 6,667.

Стандартное отклонение (среднеквадратическое отклонение) — положительное значение квадратного корня из дисперсии:

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

$$\sigma = \sqrt{6{,}667} = 2{,}58$$

На практике чаще всего используется именно стандартное отклонение, а не дисперсия

Коэффициент вариации *CV* позволяет судить об однородности выборки:

$$CV = \frac{\sigma}{\bar{x}} \cdot 100\% .$$

$$CV = \frac{2{,}58}{6} \cdot 100\% = 43{,}0\%$$

CV ≤ 10%, то выборка считается однородной относительно изучаемого свойства,

10 < CV ≤ 20% — варьирование результатов среднее,

CV ≥ 20% — варьирование результатов большое, т.е. выборка является неоднородной