

$$M_e = x_0 + i \times \frac{\frac{1}{2} \sum f_i - S_{M_e-1}}{f_{M_e}}$$

$$M_0 = x_0 + i \times \frac{(f_{M_0} - f_{M_0-1})}{(f_{M_0} - f_{M_0-1}) + (f_{M_0} - f_{M_0+1})}$$

Решение задач

дополненная
частота
и др.

$$M_e = x_0 + h * \frac{\frac{\sum f}{2} - S_{m-1}}{f_m}$$

- ▣ Предположим, есть список чисел, представляющий собой количество книг, которые каждый студент прочитал за последний месяц.
- ▣ После сортировки у нас получился следующий набор чисел: 3, 3, 5, 6, 6, 6, 8.

3, 3, 5, 6, 6, 6, 8

3, 3, 5, 6, 6, 6, 8

- вверху левой колонки пишем “Количество книг”, а вверху правой колонки — “Частота”.
- Во второй строке пишем первое количество прочитанных книг, то есть число 3, во второй строке колонки “Частота” пишем цифру 2.

| No. of Books | Frequency |
|-----------------|-----------|
| 3 | 2 |
| 5 | 1 |
| 6 | 3 |
| 8 | 1 |

- ▣ **Накопленная частота отвечает на вопрос "сколько раз встречается в списке данное значение *или* меньшая величина?".**
- ▣ **Всегда начинайте с наименьшего значения в наборе данных.**
- ▣ **Поскольку в нашем примере нет меньших значений, для данной величины**
- ▣ **накопленная частота
равна
абсолютной.**

| No. of Books | Frequency (F) | Cumulative Frequency (CF) |
|--------------|---------------|---------------------------|
| 3 | 2 | 2 |
| 5 | 1 | |
| 6 | 3 | |
| 8 | 1 | |

- ▣ Чтобы определить **накопленную частоту** для второго значения списка, необходимо прибавить его абсолютную частоту к накопительной частоте предыдущего значения.
- ▣ Иными словами, следует взять последнюю накопленную частоту и прибавить к ней абсо

| No. of Books | Frequency (F) | Cumulative Frequency (CF) |
|--------------|---------------|---------------------------|
| 3 | 2 | 2 |
| 5 | 1 | 3 |
| 6 | 3 | |
| 8 | 1 | |

Diagram illustrating the calculation of Cumulative Frequency (CF) for the second value (5 books):

The CF for 5 books is calculated as the CF for 3 books (2) plus the frequency of 5 books (1), resulting in 3. This is shown by the equation $2 + 1 = 3$ with arrows pointing from the '2' in the CF column for 3 books and the '1' in the Frequency column for 5 books to the equation.

| No. of Books | Frequency (F) | Cumulative Frequency (CF) |
|-----------------|------------------|---------------------------------|
| 3 | 2 | 2 |
| 5 | 1 | $2 + 1 = 3$ |
| 6 | 3 | $3 + 3 = 6$ |
| 8 | 1 | $6 + 1 = 7$ |

Проверьте полученные результаты:

- В итоге мы сложили абсолютные частоты всех значений списка.
- Конечная накопленная частота должна соответствовать числу значений в списке.
- Есть **два способа** проверить, так ли это:
- Сложим абсолютные частоты всех значений:
 $2 + 1 + 3 + 1 = 7$, в результате получится накопленная частота.
- Посчитаем число значений в наборе данных.
- В нашем примере список имел следующий вид: **3, 3, 5, 6, 6, 6, 8**.
- В этом списке семь величин, и итоговая накопленная частота также равна 7.

Frequencies:

$$2 + 1 + 3 + 1 = 7$$

3, 3, 5, 6, 6, 6, 8

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|

Дискретные данные

- ▣ Дискретные данные можно посчитать, они не дробятся на более мелкие составляющие.
- ▣ Количество собак является дискретным множеством.
- ▣ Нет такого понятия, как половина собаки.

Непрерывные данные

- ▣ Непрерывные данные часто не поддаются конечному счету, между двумя произвольными величинами обязательно найдутся другие возможные значения.
- ▣ Высота снега представляет собой непрерывное множество.
- ▣ Высота снега возрастает постепенно и непрерывно, а не на дискретные величины.
- ▣ Если вы измерите высоту снега в сантиметрах, то точное значение может оказаться, например, 20,6 сантиметра.

Частоты – это...

- ▣ Существуют относительные показатели, характеризующие долю частоты отдельных вариантов в общей сумме частот.
- ▣ Эти относительные показатели именуют частотами и условно обозначают через ,
- ▣ т.е. .
- ▣ Сумма всех частостей равна единице.
- ▣ Частоты могут быть выражены и в процентах, и тогда их сумма будет равна 100%.

Мода и медиана

- **Мода** – это величина признака (варианта), которая **чаще** всего встречается в данной совокупности, т.е. имеет **наибольшую частоту**.
- Мода имеет большое практическое применение и в ряде случаев только мода может дать характеристику общественных явлений.
- **Медиана** – это варианта, которая находится в середине упорядоченного (ранжированного) вариационного ряда.
- Рассмотрим расчет моды и медианы в **дискретном вариационном ряду**:

Определить моду и медиану:

| Стаж, лет, X | Число рабочих, чел, f | Накопленные частоты |
|--------------|-----------------------|---------------------|
| 1 | 2 | 2 |
| 3 | 4 | 6 |
| 4 | 5 | (11) |
| 8 | 4 | 15 |
| 10 | 1 | 16 |
| ИТОГО: | 16 | |

- ✓ Мода $M_o = 4$ года, так как этому значению соответствует наибольшая частота $f = 5$.
- ✓ Т.е. наибольшее число рабочих имеют стаж 4 года.

Медиана

- Для того, чтобы вычислить медиану, найдем предварительно половину суммы частот. Если сумма частот является числом нечетным, то мы сначала прибавляем к этой сумме единицу, а затем делим пополам.
- $Me = 16 / 2 = 8$
- Медианой будет восьмая по счету варианта.
- Для того, чтобы найти, какая варианта будет восьмой по номеру, будем накапливать частоты до тех пор, пока не получим сумму частот, равную или превышающую половину суммы всех частот. Соответствующая варианта и будет медианой.

$$Me = \frac{\Sigma f + 1}{2}$$

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

Медиана

- ▣ $Me = 4$ года.
- ▣ Т.е. половина рабочих имеет стаж меньше четырех лет, половина больше.
- ▣ Если сумма накопленных частот против одной варианты равна половине сумме частот, то медиана определяется как средняя арифметическая этой варианты и последующей.

Свойство медианы

- Главное свойство Me в том, что сумма абсолютных отклонений значений признака от медианы меньше, чем от любой другой величины: .

$$\sum_{i=1}^n |x_i - Me| = \min$$

- Это свойство Me может быть использовано, например, при определении места строительства общественных зданий, т. к. Me определяет точку, дающую наименьшее расстояние, допустим, детских садов от местожительства родителей, жителей населенного пункта от кинотеатра, при проектировке трамвайных, троллейбусных остановок и т.д.

Медиана. Пример 2:

- Найти медиану дискретного ряда
- 16,13,15,10,19,22,25,12,18,14,19,14,16,10.
- **Решение.** Ранжируем ряд:
10,10,12,13,14,14,15,16,16,18,19,19,22,25, выборка содержит четное число элементов $n=14$, следовательно медиана лежит между двумя средними элементами выборки - между 7-элементом и 8-элементом:
 - 10,10,12,13,14,14,15,16,16,18,19,19,22,25
 - и равна среднему арифметическому этих элементов:
 - $Me=(15+16)/2=15,5$

Расчет медианы в интервальном ряду

$$M_e = x_e + h_e * \frac{\frac{\sum f_i}{2} - S_{M_{e-1}}}{f_{M_e}}$$

x_e - начало медианного интервала

h_e - длина модального интервала

$S_{M_{e-1}}$ - кумулятивная частота интервала, предшествующего медианному

f_{M_e} - частота медианного интервала (не накопленная)

Расчет моды в интервальном ряду

$$M_o = x_o + h_o * \frac{f_{M_o} - f_{M_{o-1}}}{(f_{M_o} - f_{M_{o-1}}) + (f_{M_o} - f_{M_{o+1}})}$$

x_o -начало модального интервала

h_o -длина модального интервала

f_{M_o} -частота модального интервала

$f_{M_{o-1}}$ -частота интервала, предшествующего модальному

$f_{M_{o+1}}$ -частота интервала, следующего за модальным

- Сначала выберем модальный интервал, максимальная частота в нашем случае равна 10. Таким образом, получаем:

| Группы предприятий по стоимости ОПФ, у.е. | Число предприятий, f | Середина интервалов, x | Накопленная частота, S |
|---|------------------------|--------------------------|--------------------------|
| 14-16 | 2 | 15 | 2 |
| 16-18 | 6 | 17 | 8 |
| 18-20 | 10 | 19 | 18 |
| 20-22 | 4 | 21 | 22 |
| 22-24 | 3 | 22 | 25 |
| Итого: | 25 | | |

Мода. Расчеты

- 1) По максимальной частоте найдем модальный интервал: $F_{\max}=10 \rightarrow I = 18-20$
- 2) По соответствующей формуле
- Мода** $= 18 + 2 * (10 - 6) / ((10 - 6)(10 - 4)) = 18,33$ млн. руб. – наиболее часто встречающаяся стоимость ОПФ среди 25 предприятий.

$$M_o = x_o + i \times \frac{(f_{M_o} - f_{M_o-1})}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}$$

Медиана. Расчеты:

- Нам даны интервалы.
- 1) Найдем медианный интервал по накопленной частоте. Нужная накопленная частота определяется путем суммирования частот f до тех пор, пока очередная накопленная частота впервые не превысит половину совокупности $n + 1/2$ или $n/2$.
- Для нечетного ряда $(25+1)/2 = 13 \rightarrow S = 18 \rightarrow 18-20$ - медианный интервал, так как в пределах этого интервала расположена варианта, которая делит совокупность на две равные части.

Медиана. Расчеты:

- 2) По соответствующей формуле
- Медиана $M_e = 18 + 2[(25+1)/2 - 8/10] = 18,9$ млн.руб.
- Из 25 малых предприятий региона 12 пр. имеют стоимость ОПФ менее 18 млн.руб., а 12 пр. более.

$$M_e = x_0 + i \times \frac{\frac{1}{2} \sum f_i - S_{M_{e-1}}}{f_{M_e}}$$

Медиана. Пример 3:

| Возрастные группы | Число студентов | Сумма накопленных частот ΣS |
|-------------------|-----------------|-------------------------------------|
| До 20 лет | 346 | 346 |
| 20 — 25 | 872 | 1218 |
| 25 — 30 | 1054 | 2272 |
| 30 — 35 | 781 | 3053 |
| 35 — 40 | 212 | 3265 |
| 40 — 45 | 121 | 3386 |
| 45 лет и более | 76 | 3462 |
| Итого | 3462 | |

Решение:

- ▣ Медианный интервал находится в возрастной группе **25-30 лет**, так как в пределах этого интервала расположена варианта, которая делит совокупность на две равные части

$$(\sum f_i / 2 = 3462 / 2 = 1731).$$

Решение:

- Далее подставляем в формулу необходимые числовые данные и получаем значение медианы:

$$M_e = x_0 + h \frac{\frac{\sum f_i}{2} - S_{m-1}}{f_m} = 25 + 5 \frac{\frac{3462}{2} - 1218}{1054} = 27,4 \text{ года.}$$

Это значит что одна половина студентов имеет возраст до 27,4 года, а другая свыше 27,4 года.

Спасибо за внимание!