

Средние величины

Средняя величина обобщает качественно однородные значения признака.

Виды средних:

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

средняя арифметическая

Взвешенная средняя
арифметическая

$$\bar{x} = \frac{\sum_{i=1}^n x'_i \cdot f_i}{\sum_{i=1}^n f_i}$$

Возраст	Число больных f_i	Середина инт. X_i'	$X_i' f_i$
17 - 20	48	18.5	888
20 - 30	120	25	3000
30 - 40	75	35	2625
40 - 50	62	45	2790
50 - 65	54	57.5	3105

359

12408

$X = 34.56$ года

средняя квадратическая

$$x_{KB} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$$

средняя степенная

$$x_{CT} = \sqrt[k]{\frac{\sum_{i=1}^n x_i^k}{n}}$$

средняя гармоническая

$$X_{\text{гарм}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

средняя геометрическая

$$X_{\text{геом}} = \sqrt[n]{X_1 \cdot X_2 \cdot \dots \cdot X_n}$$

Правило мажорантности средних величин:

$$\bar{X}_{\text{гарм}} \leq \bar{X}_{\text{геом}} \leq \bar{X}_{\text{ариф}} \leq \bar{X}_{\text{кв}} \leq \bar{X}_{\text{ст}}$$

Вариации массовых явлений

Вариацией значений какого - либо признака в совокупности называется различие его значений у разных единиц данной совокупности в один и тот же период или момент времени .

Вариационный ряд - упорядоченное распределение единиц совокупности по возрастающим / убывающим значениям признака и подсчет числа единиц с тем или иным признаком.

Вариационный ряд (ряд распределения)

```
graph TD; A[Вариационный ряд (ряд распределения)] --> B[Ранжированный]; A --> C[Дискретный]; A --> D[Интервальный];
```

Ранжированный

Дискретный

Интервальный



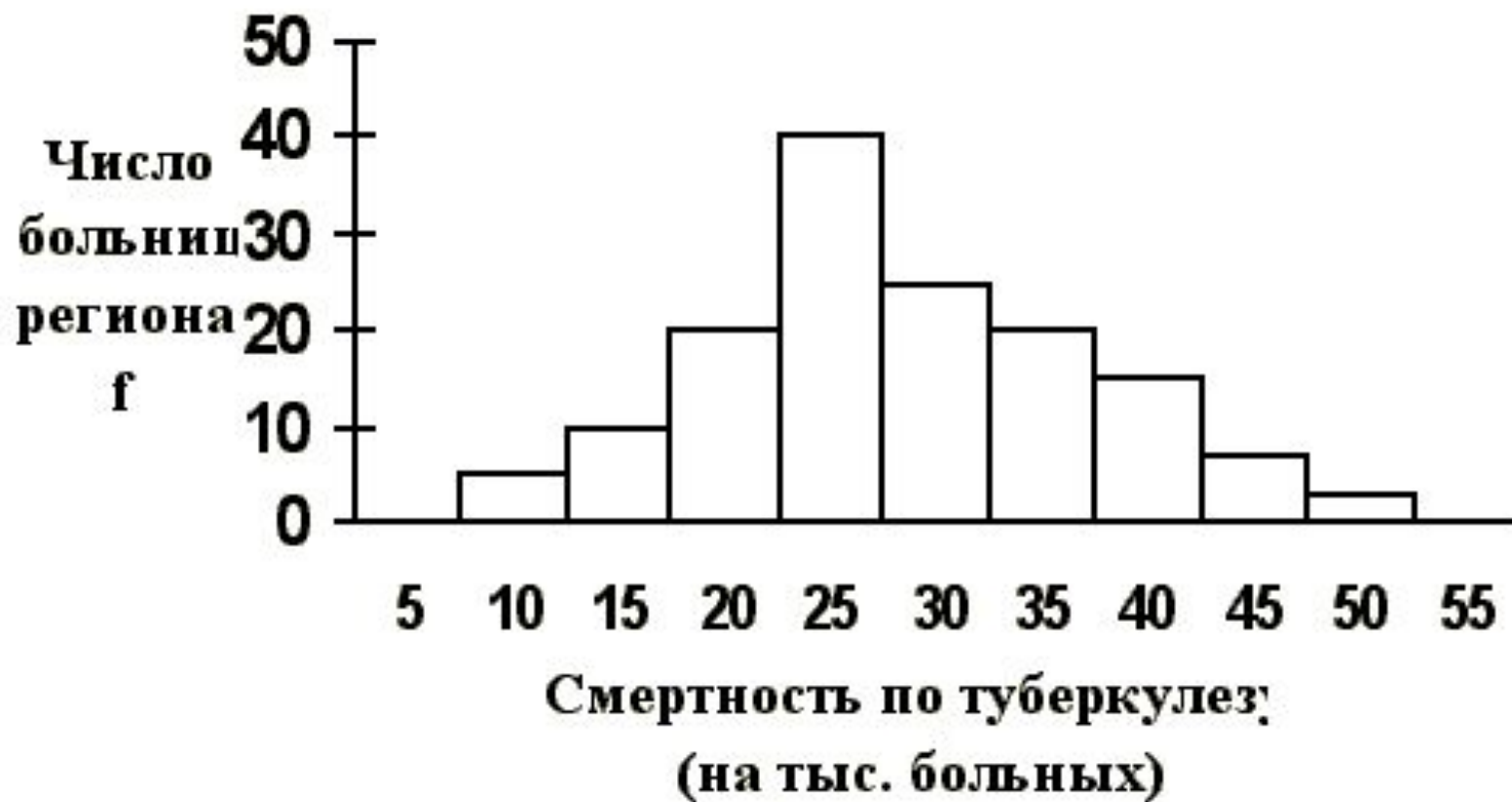
Наименование больницы	Число <u>койко-дней</u> в кардиологических отделениях больниц города
Больница N3	14.9
Железнодорожная клиника	16.0
Больница N1	18.6
Больница N2	25.2



]

Число забитых мячей, X_i	Число игр с таким кол - вом X_i
0	21
1	41
2	41
3	37
4	19
5	10
6	6
7	2

□



Как определить число
требуемых интервалов в
интервальном вариационном
ряду?

Формула Стержеса :

$$k = \text{integer} [1 + 3.32 \lg(n)]$$

тогда, ширина интервала:

$$\tau = \frac{X_{\max} - X_{\min}}{k}$$

A spiral-bound notebook with a light beige, textured cover. The metal spiral binding is visible on the left side. The text is centered on the cover in a bold, black, serif font.

**Структурные
характеристики
вариационного ряда**

Медиана распределения

Медиана - это численное значение признака у той единицы изучаемой совокупности, которая находится в середине ранжированного ряда. Медиана делит совокупность на две равные части. Первая половина единиц статистической совокупности (**после ранжирования!**) имеет значение варьирующего признака меньше, чем медиана, элементы из второй половины совокупности - больше.

Пример: группа из 7 студентов в возрасте от 17 до 23 лет сидят в аудитории за семью столами.

Вариационный признак -
возраст студента.

Первичные данные

Номер стола	1	2	3	4	5	6	7
Возраст студента	21	17	19	22	18	20	23

Строим ранжированный (по возрасту) вариационный ряд

Номер стола	2	5	3	6	1	4	7
Возраст студента	17	18	19	20	21	22	23

Медиана равна 20 годам. Т.е. возраст четвертого (в ранжированном вариационном ряду) студента делит совокупность на две равные части.

Трое студентов моложе его, трое - старше.

Если число единиц наблюдения (число элементов статистической совокупности) четное, то медианой считается средняя арифметическая из значений признака у двух срединных членов совокупности.

Рассмотрим абсолютно аналогичный пример, но для случая, когда наблюдается группа из 8 студентов.

Первичные данные

Номер стола	1	2	3	4	5	6	7	8
Возраст студента	21	17	19	22	18	20	23	20.5

Ранжированный вариационный ряд (по возрасту)

Номер стола	2	5	3	6	8	1	4	7
Возраст студента	17	18	19	20	20.5	21	22	23

В этом случае медиана равна $\frac{[20 + 20.5]}{2} = 20.25$

Определение медианы по интервальному ряду

Предположим, что первичные данные обработаны, и по ним построен интервальный вариационный ряд.

Пример: статистическому наблюдению подвергаются больницы области. Число больниц - 143. Вариационный признак - число коек. Строится интервальный ряд:

Группы больниц по числу коек X_i	Число больниц f_i	Середина Интервала X_i'	$X_i' f_i$	Накопленная частота f_i'
10-15	6	12,5	75	6
15-20	9	17,5	157,5	15
20-25	20	22,5	450	35
◆ 25-30	41	27,5	1127,5	76
30-35	26	32,5	845,5	102
35-40	21	37,5	787,5	123
40-45	14	42,5	595	137
45-50	5	47,5	237	142
50-55	1	52,5	52,5	143
ИТОГО:	143		4327,5	

Предположим, что у нас нет в нашем распоряжении первичных данных. В этом случае мы не можем построить ранжированный вариационный ряд, как это было сделано в предыдущем примере.

В нашем распоряжении есть только обработанные до нас данные, которые уже сведены к интервальному ряду. Например, интервальный ряд (в виде гистограммы) был взят нами из периодической литературы.

Сами исходные данные не публиковались.

Медиана распределения вычисляется с использованием интервального ряда по формуле:

$$M_e = X_0 + \frac{\sum_{i=1}^k f'_i - f'_{M_e-1}}{2 f_{M_e}} \cdot \tau$$

X_0 - низшая граница интервала, в котором находится медиана;

$f'_{(M_e-1)}$ - накопленная частота в интервале, предшествующем медианному; f_{M_e} - частота в медианном интервале; τ - величина интервала;

k - число групп

Вычислим медиану по приведенной выше формуле:

- 1) $143/2 \sim 72$ \rightarrow медиана находится в **четвертом** интервале (т.к. полученное число **72** ближе всего к **76** в столбце накопленных частот). Этот интервал отмечен значком **◆** в таблице.
- 2) в четвертом интервале: низшая граница $X_0 = 25$, частота $f_{Me} = 41$
- 3) в предшествующем (третьем) интервале накопленная частота равна $f'_{(Me-1)} = 35$
- 4) ширина каждого интервала = **5** (см. первый столбец интервалов)
- 5) Окончательно, $Me = 25 + [(72 - 35) / 41] * 5 = 29.5$ коп.

Квартили распределения

Вычисляются абсолютно
аналогично медиане по формулам:

$$Q_1 = x_0 + \frac{\left[\sum_{j=1}^k f'_j / 4 \right] - f'_{(Q_1-1)}}{f_{Q_1}} \cdot \tau =$$

$$= 25 + \frac{(35.75 - 35)}{41} \cdot 5 = 25.09$$

$$Q_2 \equiv M_e$$

$$Q_3 = x_0 + \frac{\left[3 \sum_{j=1}^k f'_j / 4 \right] - f'_{(Q_3-1)}}{f_{Q_3}} \cdot \tau =$$

$$= 35 + \frac{(107.25 - 102)}{21} \cdot 5 = 36.25$$

Общее название для вышеприведенных структурных характеристик вариационного ряда - квантили. Если ряд делится на 4 части то в этом случае квантили называются квантилями (см. формулы выше), на 5 частей - квинтили ; на 10 - децили ; на 100 - перцентили .

Мода распределения .

Модальный интервал - интервал с наибольшей частотой .

Мода:

$$M_0 = X_0 + \frac{f_{M_0} - f_{(M_0-1)}}{[f_{M_0} - f_{(M_0-1)}] + [f_{M_0} - f_{(M_0+1)}]} \cdot \tau$$

по-прежнему, x_0 -нижняя граница
модального интервала,

f_{M_0} – частота в модальном интервале.

Рассмотрим пример с обследованием 143 больниц.

Группы больниц по числу коек X_i	Число больниц f_i	Середина Интервала X_i'	$X_i' f_i$	Накопленная частота f_i'
10-15	6	12,5	75	6
15-20	9	17,5	157,5	15
20-25	20	22,5	450	35
◆ 25-30	41	27,5	1127,5	76
30-35	26	32,5	845,5	102
35-40	21	37,5	787,5	123
40-45	14	42,5	595	137
45-50	5	47,5	237	142
50-55	1	52,5	52,5	143
ИТОГО:	143		4327,5	

Модальный интервал - четвертый. Наибольшая частота (41) относится к этому интервалу. Т.е., в рассматриваемом примере модальный и медианный интервалы совпали. Это часто встречается, но так бывает не всегда!

Частоты в интервалах в предшествующем (число 20) и следующем (это число 26) за модальным интервалом отмечены в таблице бирюзовой заливкой.

Вычислим моду:

$$M_0 = 25 + \frac{(41 - 20)}{(41 - 20) + (41 - 26)} \cdot 5 = 27.9$$

$$M_0 < M_e < \bar{x} \quad \text{для нормального распределения. 7.}$$