



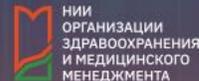
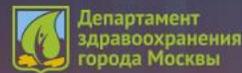
IV СЪЕЗД МЕДИЦИНСКИХ  
СТАТИСТИКОВ ГОРОДА МОСКВЫ

## ЦИФРОВАЯ СТАТИСТИКА

НОВЫЕ ЗАДАЧИ И ТРАЕКТОРИЯ  
ДВИЖЕНИЯ

# Описательная статистика медицинских данных

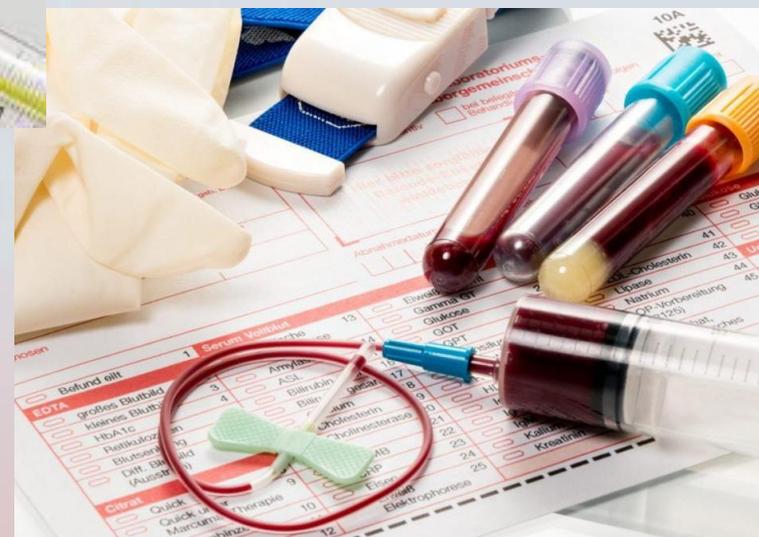
**Искандаров Ильдар**, доцент кафедры общественного здоровья и организации здравоохранения  
ФГБОУ ВО «Казанский государственный медицинский университет» МЗ РФ



# Виды данных

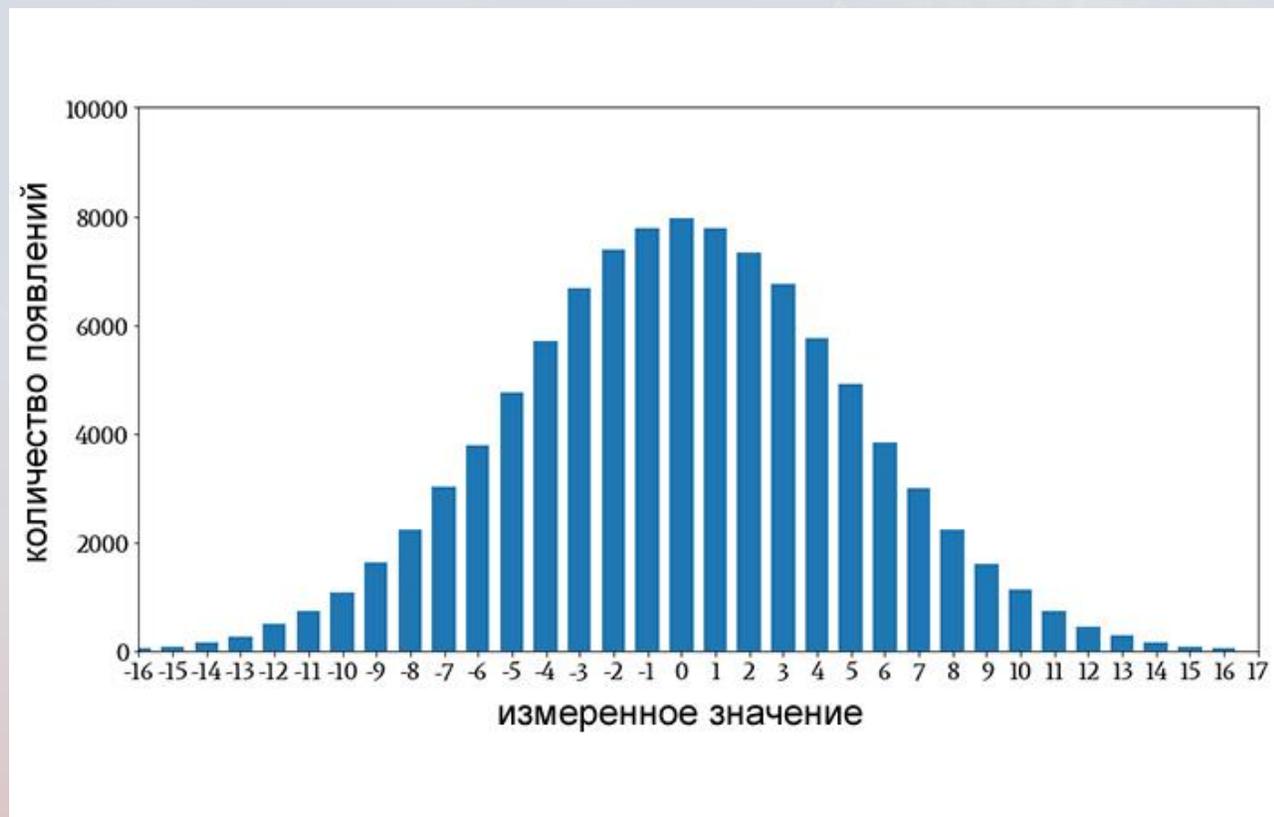


# Непрерывные количественные данные



# Дискретные количественные данные





121,122, 125, 125, 128, 128, 128, 130, 130, **130**, 130, 130, 132, 132,132, 135, 135, 138, 139

**M (mean) – средняя арифметическая, или средняя** Сумма всех значений, деленная на их количество

$$M = (121+122+125*2+128*3+130*5+132*3+135*2+138+139) / 19 = 130$$

**Me (median) – медиана**

Значение, находящееся в середине вариационного ряда.

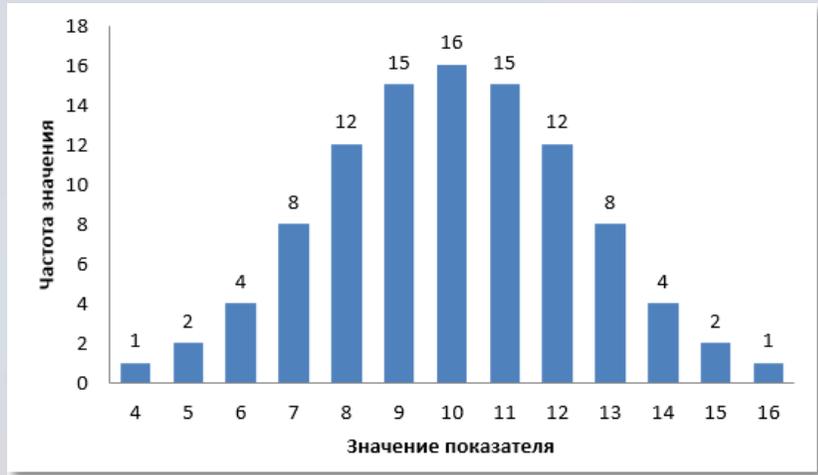
$$Me = V_{10} = 130$$

**Mo – мода**

Наиболее часто встречающееся значение

$$Mo = 130 \text{ (т.к. встречается 5 раз)}$$

# Меры центральной тенденции при разных распределениях



## При нормальном распределении:

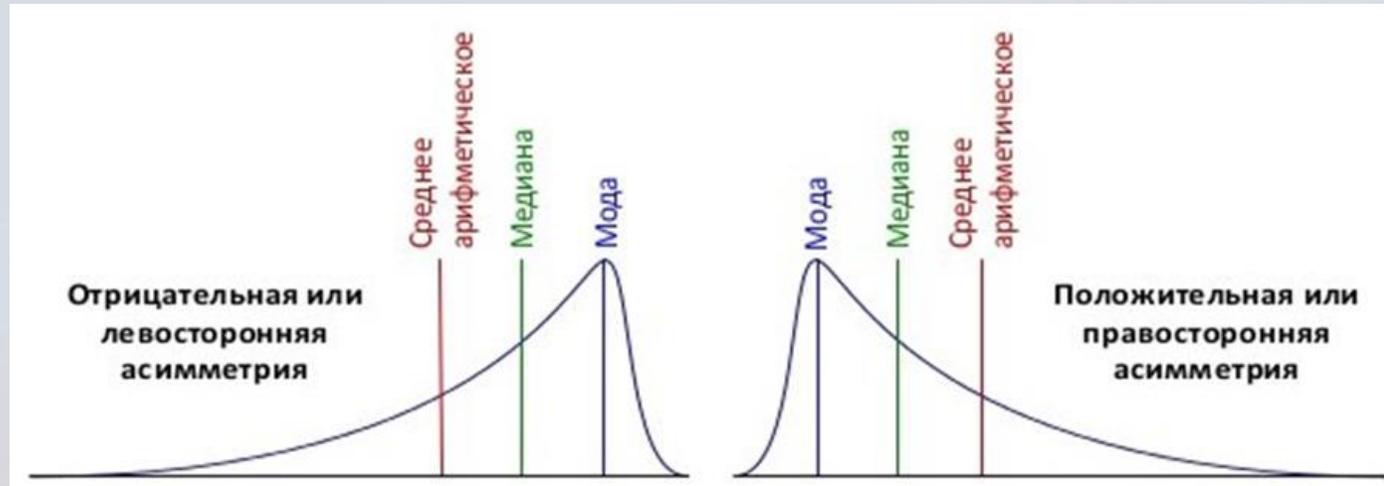
- Переменные описываются с помощью:
  - средней арифметической ( $M$ );
  - стандартного отклонения ( $\pm SD$ );
  - 95% доверительного интервала (95% CI)
  - столбиковых диаграмм (bar chart) с планками 95% ДИ



## При «ненормальном» распределении:

- **Среднее значение – малоинформативно**
- Переменные описываются с помощью:
  - медианы ( $Me$ );
  - интерквартильного размаха (IQR);
  - ящичных диаграмм (box-plot)

# Меры центральной тенденции при разной асимметрии



Больше случаев со значениями выше среднего

Больше случаев со значениями ниже среднего

v	d =  V-M
121	9
122	8
125	5
125	5
128	2
128	2
128	2
130	0
130	0
130	0
130	0
130	0
132	2
132	2
132	2
135	5
135	5
138	8
139	9

# Меры изменчивости

**SD ( $\sigma$ ) – стандартное отклонение (standard deviation)**

Квадратный корень из суммы квадратов отклонений каждого значения от среднего (d), деленной на n или n-1 (при n<30).

Единицы измерения – как у показателя.

$$SD = \sqrt{(81+64+25*2+4*3+4*3+25*2+64+81)/18} = \pm 4,8 \text{ мм рт.ст.}$$

**Чем шире размах значений и чаще встречаются значения, отличные от средней, тем выше SD**

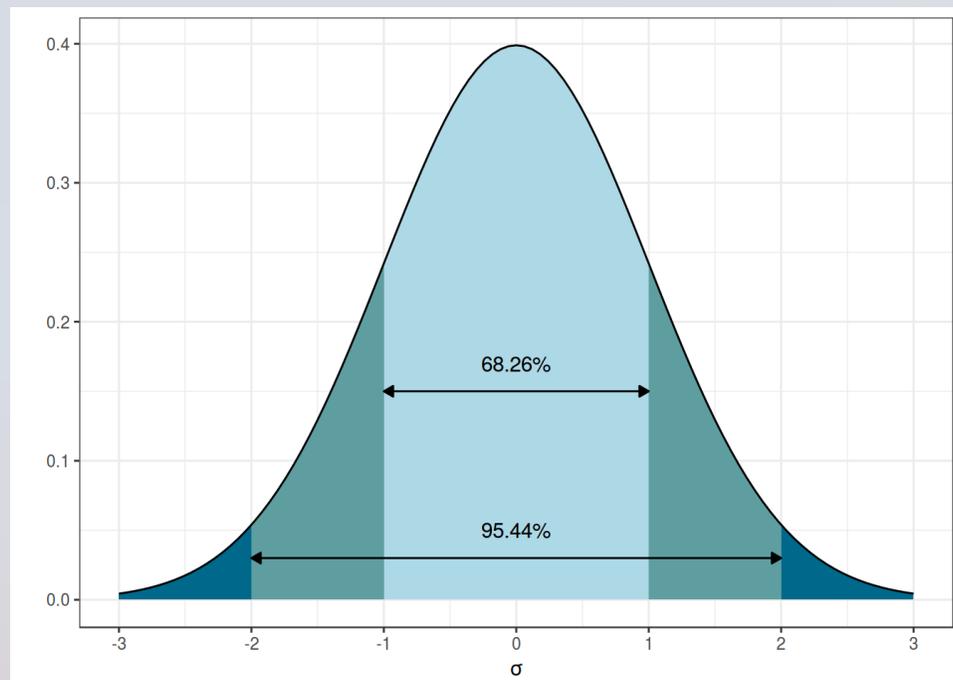
**Ряд 1:** 1, 2, 3, 4, 5, 6, 7, 8, 9

**Ряд 2:** 3, 4, 4, 5, 5, 5, 6, 6, 7

$$M_1 = M_2 = 5$$

$$SD_1 \gg SD_2$$

# От вариабельности – к структуре



## Правило «трёх сигм», или правило Гаусса

В диапазонах значений:

$M \pm 1 \cdot SD$  – 68,3% исследуемых,

$M \pm 2 \cdot SD$  – 95,4% исследуемых,

$M \pm 3 \cdot SD$  – 99,7% исследуемых

САД(ориг.) = **130±5** мм рт.ст. ( $M \pm SD$ )

САД(джен.) = **130±15** мм рт.ст. ( $M \pm SD$ )

**В каком диапазоне находится САД у 95% пациентов в каждой группе?**

Оригинальный препарат – 120-140, Дженерик – 100-160

**Table 1. Demographic and Clinical Characteristics at Baseline in the Induction Trial (Randomly Assigned Patients).<sup>⊖</sup>**

Characteristic	Placebo (N=319)	Ustekinumab	
		130 mg (N=320)	6 mg/kg <sup>†</sup> (N=322)
Male sex — no. (%)	197 (61.8)	190 (59.4)	195 (60.6)
Age — yr	41.2±13.5	42.2±13.9	41.7±13.7
Weight — kg	72.9±16.8	73.7±16.8	73.0±19.3
Duration of disease — yr	8.0±7.2	8.1±7.2	8.2±7.8
Total Mayo score <sup>‡</sup>	8.9±1.6	8.9±1.6	8.9±1.5
Score of 6–10, indicating moderate disease — no./total no. (%)	263/319 (82.4)	271/320 (84.7)	276/321 (86.0)
Disease limited to left side of colon — no./total no. (%)	167/316 (52.8)	183/318 (57.5)	168/320 (52.5)
C-reactive protein — mg/liter <sup>§</sup>			
Median	4.7	4.5	4.8
IQR	1.4–10.0	1.6–9.9	1.8–13.7
Fecal calprotectin — mg/kg <sup>¶</sup>			
Median	1224.0	1382.0	1506.5
IQR	496.0–2224.0	564.5–2681.0	621.5–3192.5
Medications for ulcerative colitis taken at baseline			
≥1 Medication — no. (%)	283 (88.7)	290 (90.6)	294 (91.3)
Aminosalicylates — no. (%)	207 (64.9)	215 (67.2)	238 (73.9)
Corticosteroids — no. (%) <sup>  </sup>	157 (49.2)	173 (54.1)	168 (52.2)
Median dose (IQR) — mg/day	20.0 (10.0–20.0)	20.0 (10.0–20.0)	20.0 (10.0–20.0)
Immunomodulator — no. (%) <sup>**</sup>	89 (27.9)	93 (29.1)	89 (27.6)
No history of disease refractory to treatment with biologic agents — no. (%)	158 (49.5)	156 (48.8)	156 (48.4)
Had not received biologics	151 (47.3)	145 (45.3)	147 (45.7)
Had received biologics but did not have documented treatment failure	7 (2.2)	11 (3.4)	9 (2.8)
History of treatment failure with biologics — no. (%) <sup>††</sup>	161 (50.5)	164 (51.2)	166 (51.6)
Only TNF antagonist	112 (35.1)	107 (33.4)	106 (32.9)
Vedolizumab	49 (15.4)	57 (17.8)	60 (18.6)
≥1 TNF antagonist, regardless of vedolizumab	159 (49.8)	162 (50.6)	164 (50.9)
Any TNF antagonist and vedolizumab	47 (14.7)	55 (17.2)	58 (18.0)

\* Plus-minus values are means ±SD. IQR denotes interquartile range, and TNF tumor necrosis factor.

<sup>†</sup> Weight-range–based doses of ustekinumab approximate 6 mg per kilogram of body weight (with 260 mg prescribed for patients weighing ≤55 kg, 390 mg for patients weighing >55 kg and ≤85 kg, and 520 mg for patients weighing >85 kg).

<sup>‡</sup> Total scores on the Mayo scale range from 0 to 12, with higher scores indicating more severe disease.

<sup>§</sup> Data for C-reactive protein concentrations were available for 951 patients: 316 receiving placebo, 315 receiving 130 mg of ustekinumab, and 320 receiving 6 mg of ustekinumab per kilogram.

<sup>¶</sup> Data for fecal calprotectin concentrations were available for 855 patients: 289 receiving placebo, 296 receiving 130 mg of ustekinumab, and 300 receiving 6 mg of ustekinumab per kilogram.

Количественные данные представлены с помощью **M±SD**

Количественные данные представлены с помощью **Me [IQR]**

Категориальные данные представлены с помощью **абс. (%)**

121,122, 125, 125, 128, 128, 128, 130, 130, 130, 130, 130, 132, 132,132, 135, 135, 138, 139

### **ИКР – интерквартильный размах (range, IQR)**

- Разность верхнего и нижнего квартилей (или 75% и 25% перцентиля)
- Измеряются в тех же единицах, что и сам показатель
- Для целочисленных показателей могут иметь дробную часть только:  
...,0 ...,25 (...),3) ...,5 ...,75 (...),8)

Обычно представляются данные не о разности между квартилями, а указываются сами значения квартилей:  $Q_1$ - $Q_3$

Me = 130 ( $Q_1$ - $Q_3$ : 128-132)

Ряд 1: 1, 2, 3, 4, 5, 6, 7, 8, 9

Ряд 2: 3, 4, 4, 5, 5, 5, 6, 6, 7

121,122, 125, 125, 128, 128, 128, 130, 130, **130**, 130, 130, 132, 132,132, 135, 135, 138, 139

### **A – амплитуда, размах**

Разность максимального и минимального значений

$$A = 139 - 121 = 18$$

### **Min - минимум**

наименьшее значение переменной

$$\text{Min} = 121$$

### **Max – максимум**

наибольшее значение переменной

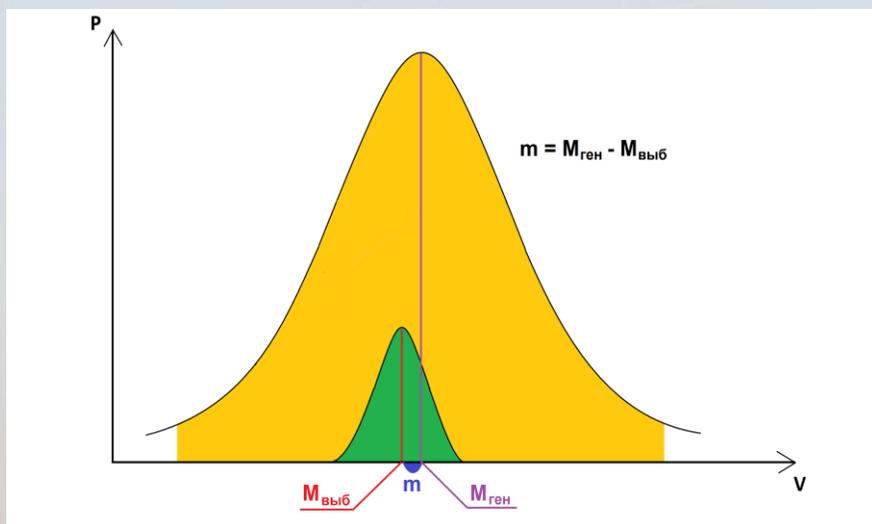
$$\text{Max} = 139$$

# Стандартная ошибка

## Стандартная (средняя) ошибка (SE, $m$ , standard error)

мера соответствия выборочных данных генеральной совокупности, мера репрезентативности

$SE$  – «Показатель-полуфабрикат». «Готовый продукт» - ДИ



$$SE = SD/\sqrt{n} = 4,8/4,24 = 1,13$$

Соответственно,  
 $SE$  увеличивается с ростом  $SD$   
и уменьшается с ростом числа  
исследуемых  $n$

# 95% доверительный интервал

## Доверительный интервал (Confidence Interval)

Интервал значений, в котором с определенной долей вероятности (обычно 95%) можно ожидать нахождения **среднего показателя для генеральной совокупности** (при множестве повторов эксперимента). Рассчитывается исходя из SE.

95% ДИ:  $M - t * SE; M + t * SE$

$t$  – коэффициент,  $\approx 2$

ОШИБКА!

«95% ДИ – это интервал, в котором находятся 95% данных»

«95% ДИ – это интервал  $M \pm 2\sigma$  из правила Гаусса»

Статистическая обработка данных

https://stattech.ru/databases/627273e02d23ae00018803ee

STATTECH

BASE

Отменить действие (ЖЗ) | Вернуть действие (ЖЗ) | Сбросить все зависимости

Описательная статистика | Настройки анализа | Обновить | Закрывать анализ

Всего в базе: 27 столбцов и 40 строк

группы	Без группы	Исходное ИМТ	ИМТ че
ле	Курение	Количественная	Количество
ная	0 (Некурящие)	21.564544913741823	18.55
1	1 (Курящие)	26.638917793964616	26.63
2	1 (Курящие)	27.513384889946465	26.02
3	1 (Курящие)	28.194284249647573	28.19
4	1 (Курящие)	17.555619394171533	17.8
5	1 (Курящие)	25.95155709342561	26.2
6	1 (Курящие)	21.63114885435026	21.6
7	0 (Некурящие)	27.74474831549742	26.4
8	0 (Некурящие)	26.296566837107374	24.85
9	1 (Курящие)		

✓ Анализ выполнен успешно

Показатель	Категории	Исходное ИМТ (кг/м2)			p
		M ± SD	95% ДИ	n	
Курение	Некурящие	26 ± 5	24 – 28	21	0,842
	Курящие	26 ± 4	24 – 28	19	

При оценке исходного ИМТ в зависимости от наличия курения, нам удалось выявить значимых различий ( $p = 0,842$ ) (используемый метод — критерий Стьюдента).

Курение: Некурящие (желтый), Курящие (синий)

+ Добавить строку

Описательная статистика

✓ Анализ выполнен успешно

Показатель	Категори
------------	----------

### Описательная статистика

Выбрать все

- Номер
- Пол
- Трудовой статус
- Возраст
- Возраст\_группы
- Ожирение
- Курение
- Исходное ИМТ
- ИМТ через 6 мес.
- ИМТ через 12 мес.
- Степень избытка массы тела
- Глюкоза до лечения
- Глюкоза после лечения
- Гипергликемия до лечения
- Гипергликемия после лечения
- Наличие и степень АГ
- Препарат

Отмена OK