



ЮЖНО-УРАЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ
МЕДИЦИНСКИЙ
УНИВЕРСИТЕТ

ОСНОВЫ СТАТИСТИКИ

Дисциплина «Стандарты диагностики и лечения. Медицина основанная на доказательствах»
Специальность 31.05.01 Лечебное дело
Лекция (2 часа)

А.И.Долгушина, д-р мед. наук,
Зав.кафедрой госпитальной терапии
Кафедра госпитальной терапии
Южно-Уральский государственный
медицинский университет, Челябинск, Россия

План лекции

1. Определение статистики
2. Работа с данными
 - Описательная статистика
 - Сравнение групп
 - Корреляция

Роль статистики в анализе исследований

- **Статистика** - это метод обработки данных, возникающих в контексте исследования
- **Статистика** - это лишь средство, а не результат исследования!
 - Компьютер может обработать данные (**Excel, Statistica for Windows, SPSS, STATTECH (Статтех)**). Но только исследователь знает, какими критериями статистического анализа ему следует воспользоваться.
 - В неумелых руках самое замечательное программное обеспечение прикладной статистики может стать «смертельным оружием»

Две концепции достоверности

- **Внутренняя достоверность:** в рамках исследования результаты точны, применяемые методы и расчёты адекватны, а интерпретации и выводы подкреплены данными, в том числе из соответствующей медицинской литературы.
- **Внешняя достоверность или возможность обобщения.** Достаточно ли продумана структура исследования, чтобы наблюдения и выводы можно было перенести на всю популяцию.

Работа с данными

I. Описательная статистика

- **Определить тип данных**
- **Представить данные наглядно (в виде таблицы, диаграммы, гистограммы, графика)**
- **Определить тип распределение данных**

Типы данные

Количественные

Непрерывные

ИМТ, средняя оценка...

Дискретные

Количество детей...

Качественные

Номинальные (классификационные)

Нельзя сравнивать «больше-меньше»

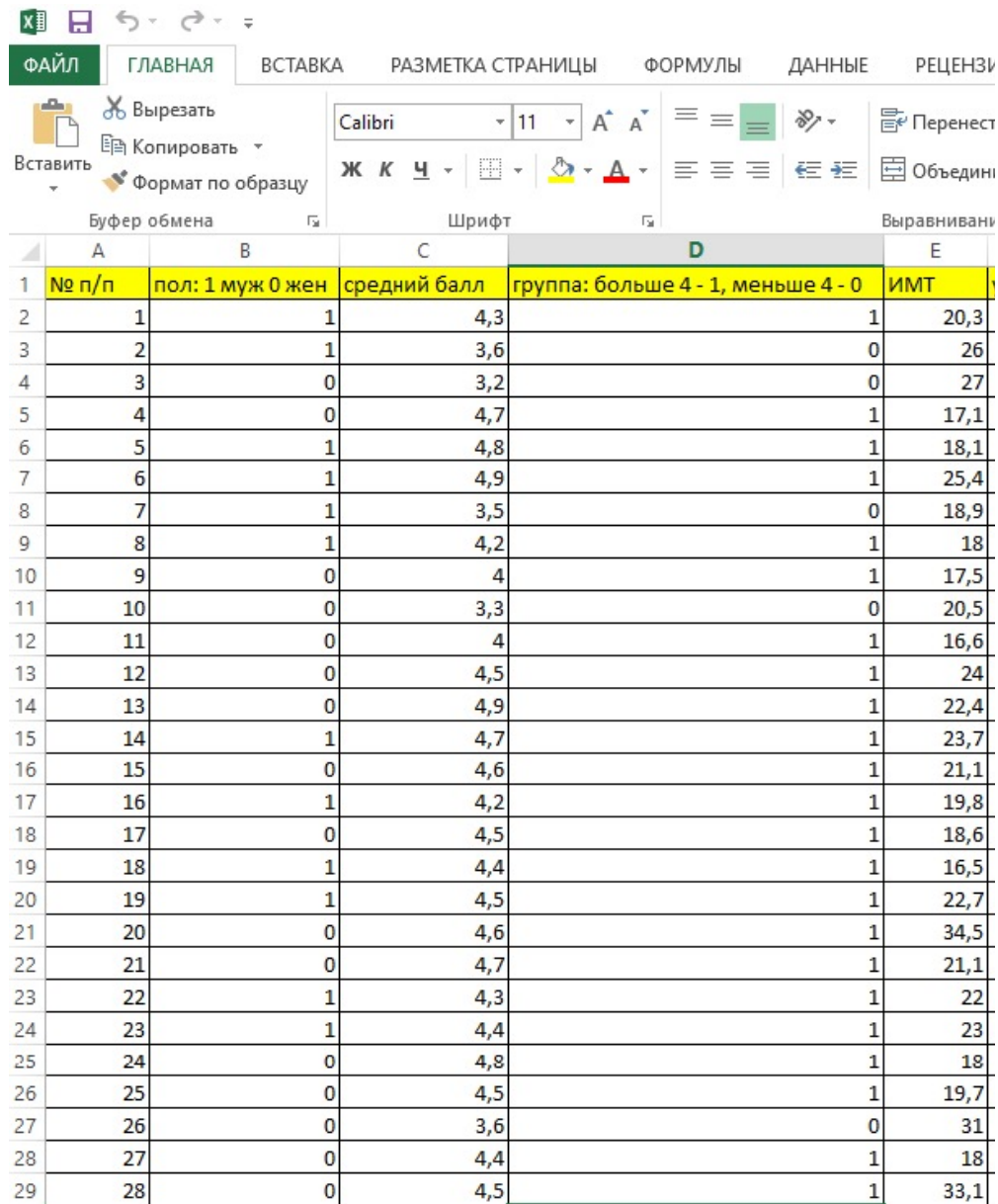
Пол...

Ординальные (порядковые)

Значения могут быть упорядочены

Оценка...

Куда внести данные. Excel

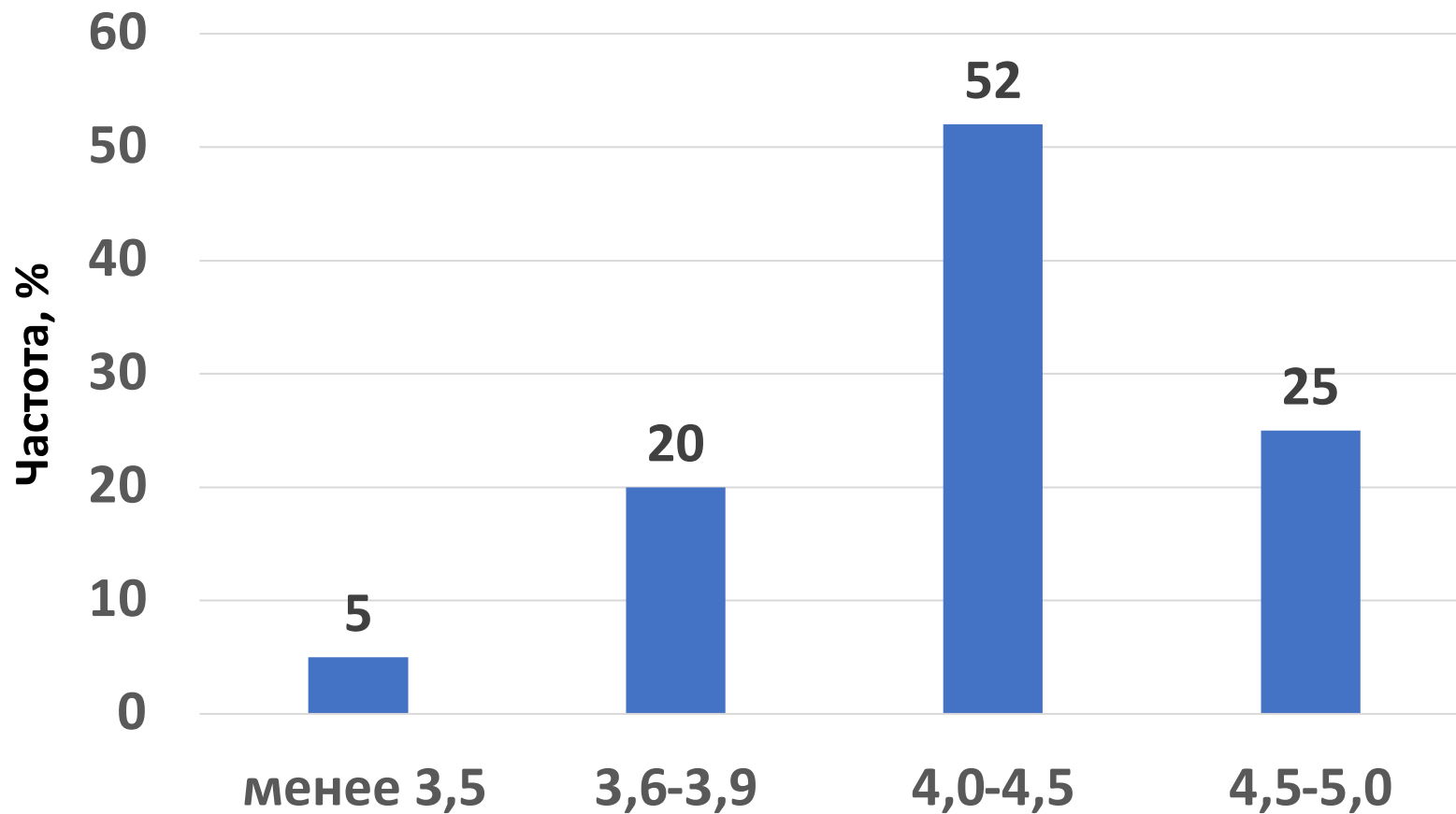


The image shows the Microsoft Excel interface with a data table. The ribbon includes 'ФАЙЛ', 'ГЛАВНАЯ', 'ВСТАВКА', 'РАЗМЕТКА СТРАНИЦЫ', 'ФОРМУЛЫ', 'ДАнные', and 'РЕЦЕНЗИ'. The 'ГЛАВНАЯ' ribbon is active, showing options for 'Вставить' (Paste) and 'Буфер обмена' (Clipboard). The font settings are 'Calibri', size '11'. The table has 5 columns: '№ п/п', 'пол: 1 муж 0 жен', 'средний балл', 'группа: больше 4 - 1, меньше 4 - 0', and 'ИМТ'. The data is as follows:

№ п/п	пол: 1 муж 0 жен	средний балл	группа: больше 4 - 1, меньше 4 - 0	ИМТ
1	1	4,3	1	20,3
2	1	3,6	0	26
3	0	3,2	0	27
4	0	4,7	1	17,1
5	1	4,8	1	18,1
6	1	4,9	1	25,4
7	1	3,5	0	18,9
8	1	4,2	1	18
9	0	4	1	17,5
10	0	3,3	0	20,5
11	0	4	1	16,6
12	0	4,5	1	24
13	0	4,9	1	22,4
14	1	4,7	1	23,7
15	0	4,6	1	21,1
16	1	4,2	1	19,8
17	0	4,5	1	18,6
18	1	4,4	1	16,5
19	1	4,5	1	22,7
20	0	4,6	1	34,5
21	0	4,7	1	21,1
22	1	4,3	1	22
23	1	4,4	1	23
24	0	4,8	1	18
25	0	4,5	1	19,7
26	0	3,6	0	31
27	0	4,4	1	18
28	0	4,5	1	33,1

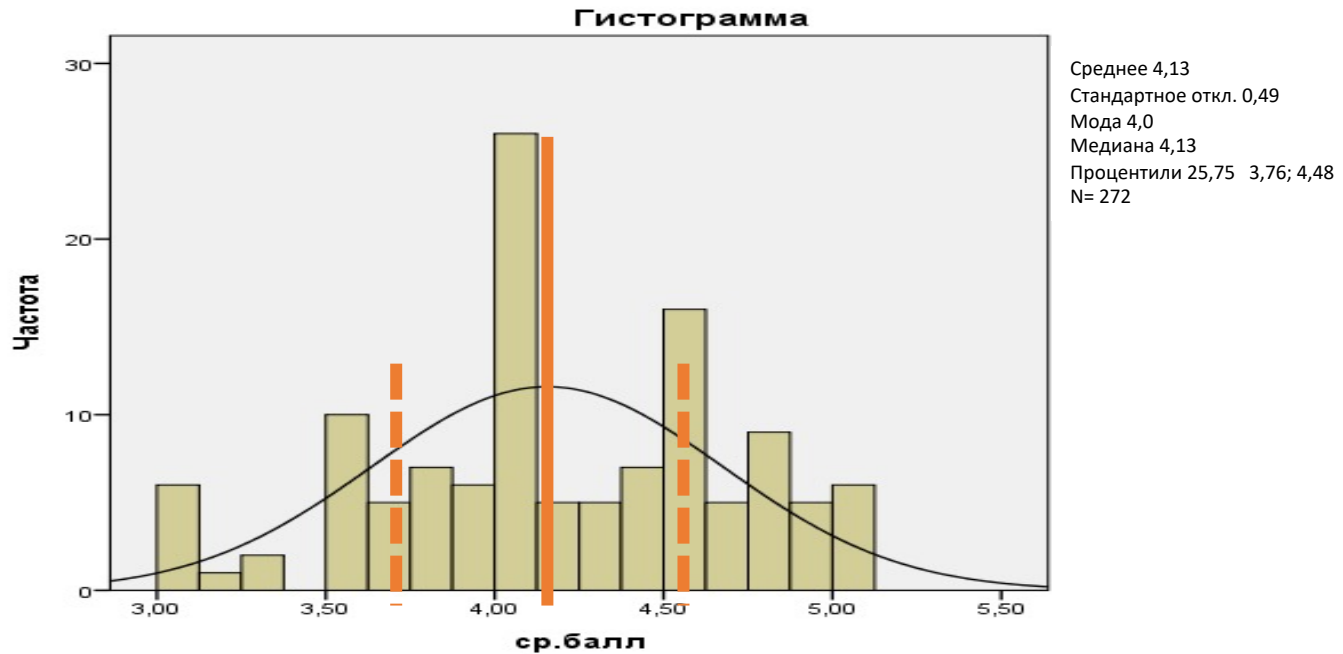
Как представить данные. **График**

Средний балл



Как представить данные. Гистограмма

Диаграмма, представляющая распределение переменной величины, в том случае если имеется информация об отдельных ее значениях



Среднее: сумма значений ÷ число наблюдений **Средний балл 4,13**

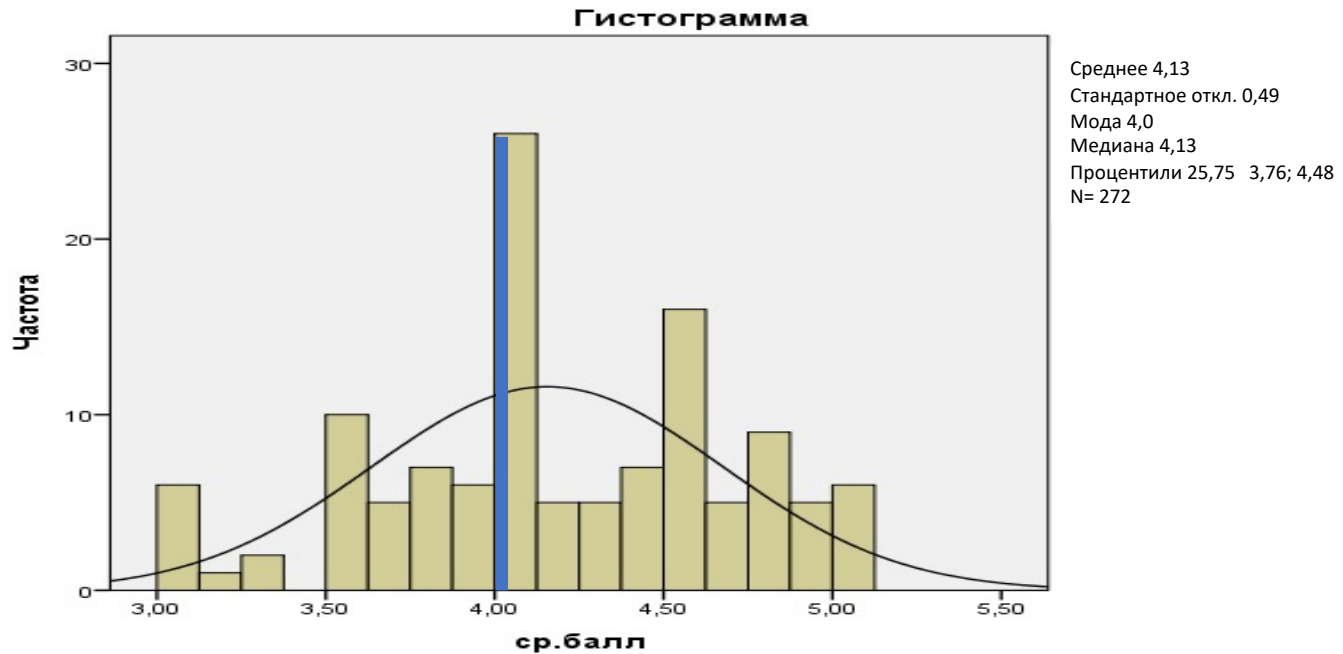
Стандартное отклонение: средний квадрат отклонений **0,49**

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

M±m Средний балл 4,13 ±0,49

Как представить данные. Гистограмма

Диаграмма, представляющая распределение переменной величины, в том случае если имеется информация об отдельных ее значениях



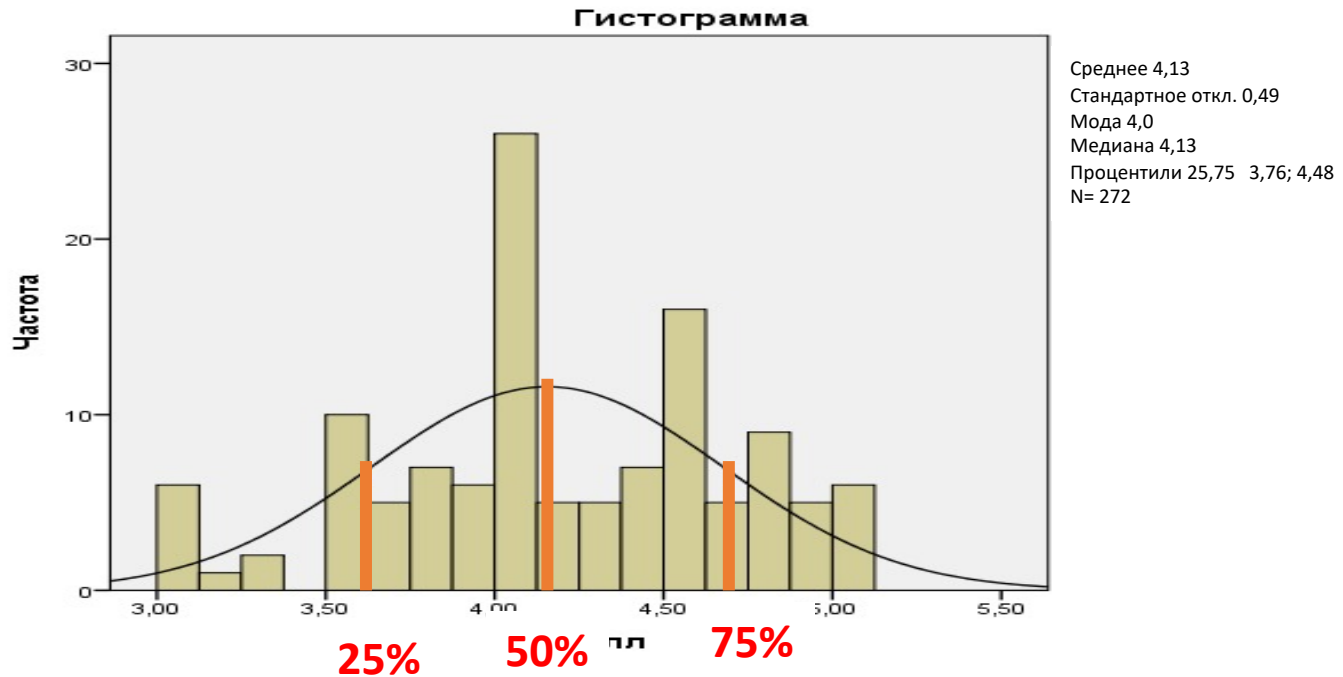
Мода: Самый частый показатель **Мода оценки 4**

Медиана: Точка, ниже и выше которой располагается ровно половина показателей, по 50%

Процентили - значение, ниже которого расположено n% (25 и 75%) наблюдений данной переменной

Как представить данные. Гистограмма

Диаграмма, представляющая распределение переменной величины, в том случае если имеется информация об отдельных ее значениях



Мода: Самый частый показатель **Мода оценки 4**

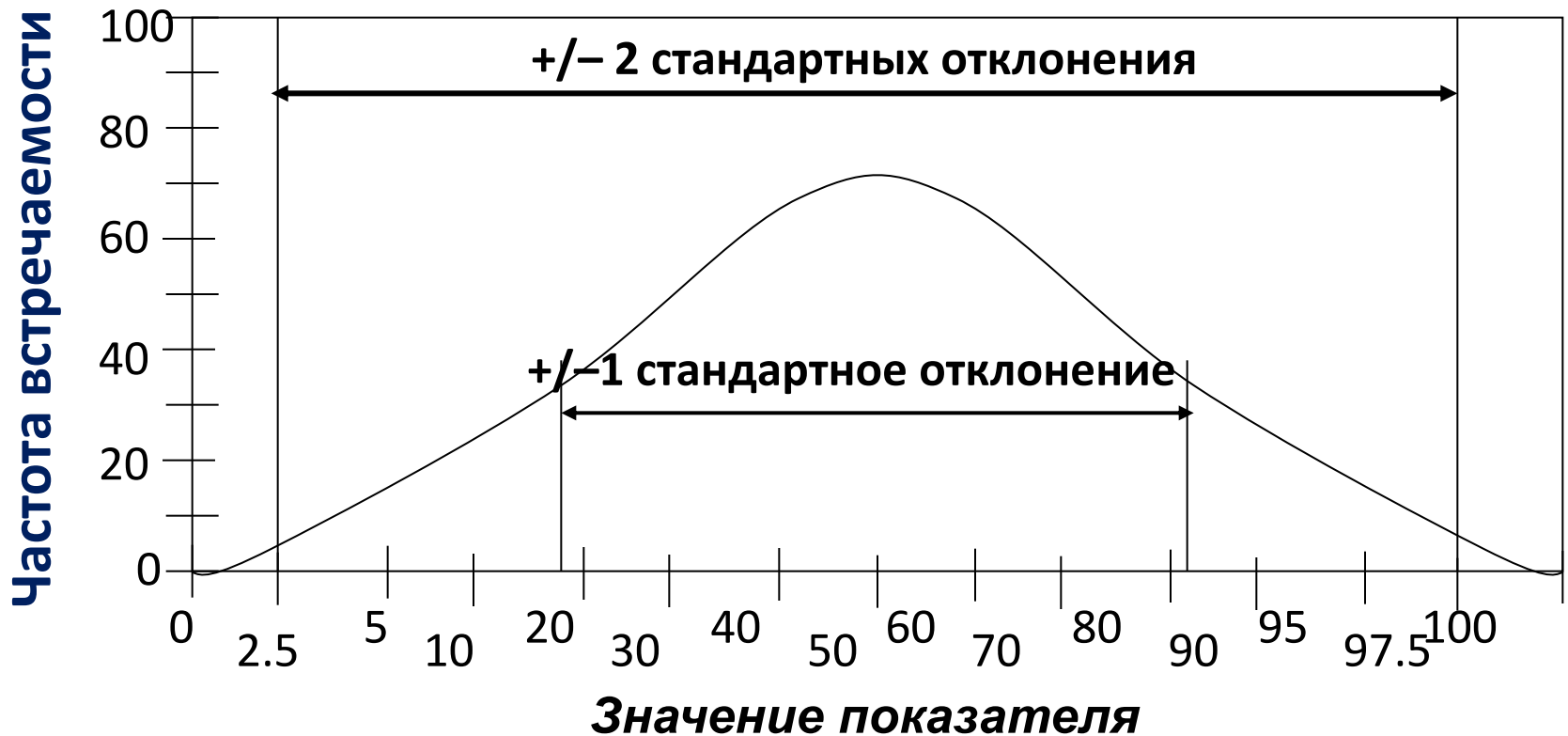
Медиана: Точка, ниже и выше которой располагается ровно половина показателей, по 50%

Процентили - значение, ниже которого расположено n% (25 и 75%) наблюдений данной переменной

Медиана балла 4,13 Процентили 25,75 3,76; 4,48

Выборочные распределения. Как определить тип?

КРИВАЯ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ



НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

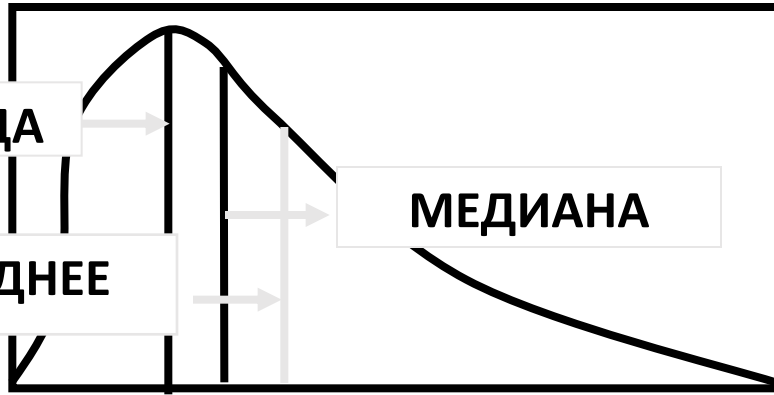
Нормальное распределение

- Графическое распределение абсолютно симметрично. (ось абсцисс – величина, ось ординат – количество пациентов) кривая Гаусса-Лапласа
- Среднее=мода=медиана
- Проверка – критерий Шапиро-Уилка, Колмогорова-Смирнова, Хи-квадрат и др.
- В дальнейшем для сравнения с другими выборками применяются параметрические методы

Выборочные распределения

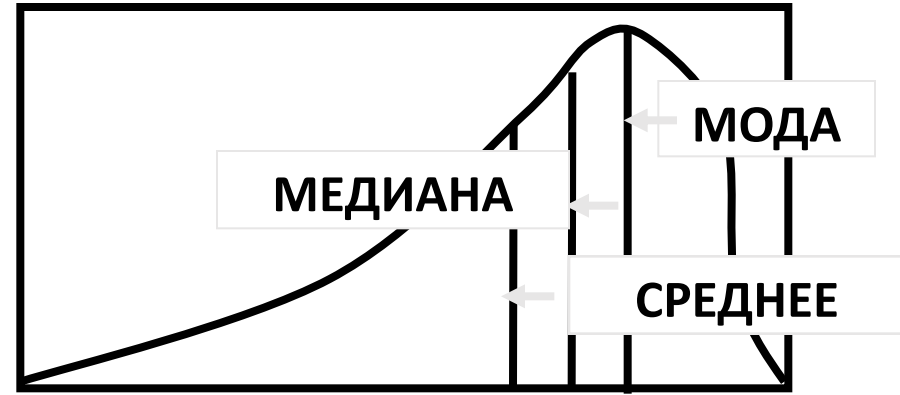
АСИММЕТРИЯ

Низкие показатели чаще высоких



Положительный сдвиг

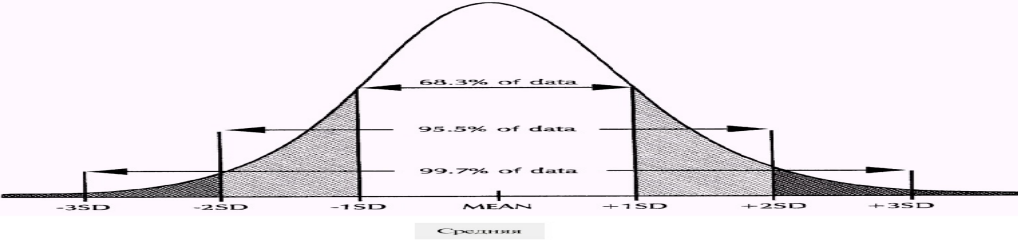
Высокие показатели чаще низких



Отрицательный сдвиг

**Распределение, отличное от нормального –
непараметрические методы статистики**

Описательная статистика. Вариабельность данных

Показатель	Тип данных	Определение
Среднее \pm стандартное отклонение (стандартная ошибка среднего) $M \pm m$	Нормальное распределение	 <p>A normal distribution curve with a horizontal axis labeled 'Среднее' (Mean) at the center. Vertical lines mark standard deviation intervals: -3SD, -2SD, -1SD, MEAN, +1SD, +2SD, and +3SD. Horizontal arrows indicate the percentage of data within these intervals: 68.3% between -1SD and +1SD, 95.5% between -2SD and +2SD, and 99.7% between -3SD and +3SD. The areas under the curve between these lines are shaded.</p>
Медиана и проценти (диапазон между 25 и 75 проценти	Любое распределение	 <p>A normal distribution curve with a horizontal axis marked with percentiles: 10%, 25%, 50%, 75%, and 90%. Vertical lines extend from these percentiles to the curve. The area under the curve between the 25% and 75% percentiles is shaded. Three orange circles are drawn around the 25%, 50%, and 75% percentile markers.</p>

Работа с данными

II. Сравнение данных

- **Формирование групп**
- **Формулировка гипотезы**
- **Сравнение групп при помощи р-значения (выбор статистического критерия)**
- **Оценка взаимосвязи (корреляция, регрессия)**

Все обследованные были разделены на группы...

Группа 1

Студенты, которые хотят заниматься наукой

N=140

Группа 2

Студенты, которые не хотят заниматься наукой

N=132

Таблица

Признак	Группа 1 Хотят заниматься наукой n= 140	Группа 1 Не хотят заниматься наукой n= 132	p
Мужчины	39 (28%)	40 (30%)	
Женщины	101 (72%)	92 (70%)	
ИМТ, кг/м² Me (ИИ)	23,88 (20,1- 28,1)	20,16 (18,6 - 26,7)	
Средний балл >> 4	113 (81%)	91 (70%)	

Формулируем гипотезу

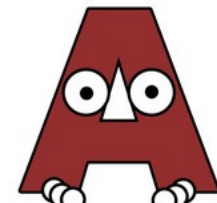
Гипотеза – вероятное знание (может быть опровергнута или подтверждена)

Нулевая гипотеза гласит, что **разницы** между сравниваемыми объектами не обнаруживается



Нулевая гипотеза

$p < 0,05$



Альтернативная гипотеза

Статистический анализ: нулевая гипотеза и статистическая значимость

- Несмотря на то, что порог статистической значимости может назначаться произвольно, по договорённости **статистической значимости** приписывают уровень 0,05. То есть **вероятность (p)** того, что происходит случайно, меньше 5%.
- Принято считать, что $P \leq 0,05$ демонстрирует “статистическую значимость”, но на самом деле это означает лишь то, что у некоего результата есть шанс проявиться 1 раз из 20 или меньше. Т.е. случайное событие произойдёт один раз из 19 (“не” значительная ассоциация).

Статистическая ошибка

- **Альфа (ошибка I-го рода) = позитивная ошибка**
 - Вероятность принятия альтернативной гипотезы (о существовании различий), когда она на самом деле неверна, т.е. различий нет
 - Объявлено о не существующих в реальности различиях в методах лечения.
 - Минимизировать с уровнем значимости $<0,05$
- **Бета (ошибка II-го рода) = негативная ошибка**
 - Вероятность принять нулевую гипотезу (об отсутствии различий), хотя она ложная, т.е. «на самом деле» различия есть
 - Не объявлено о различиях в методах лечения, в то время, как это надо было сделать.

Статистическая мощность

- **Вероятность того, что будет отвергнута гипотеза с заданным уровнем значимости.**
 - Мощность = $(1 - \text{ошибка-бета})$ - чем больше ошибка бета, тем ниже мощность.
 - Чем выше мощность, тем больше шансов обнаружить преимущество метода лечения, если оно есть
 - Обычно принимается 80%-я мощность (многие используют 90%).
 - Мощность увеличивается за счёт увеличения размера выборки и уменьшения разброса.

Формулировка гипотезы

- Студенты, которые хотят заниматься наукой и студенты, которые не хотят, не различаются по исследуемым параметрам, если $p \geq 0,05$

По:

- ✓ Полу
- ✓ ИМТ
- ✓ Среднему баллу



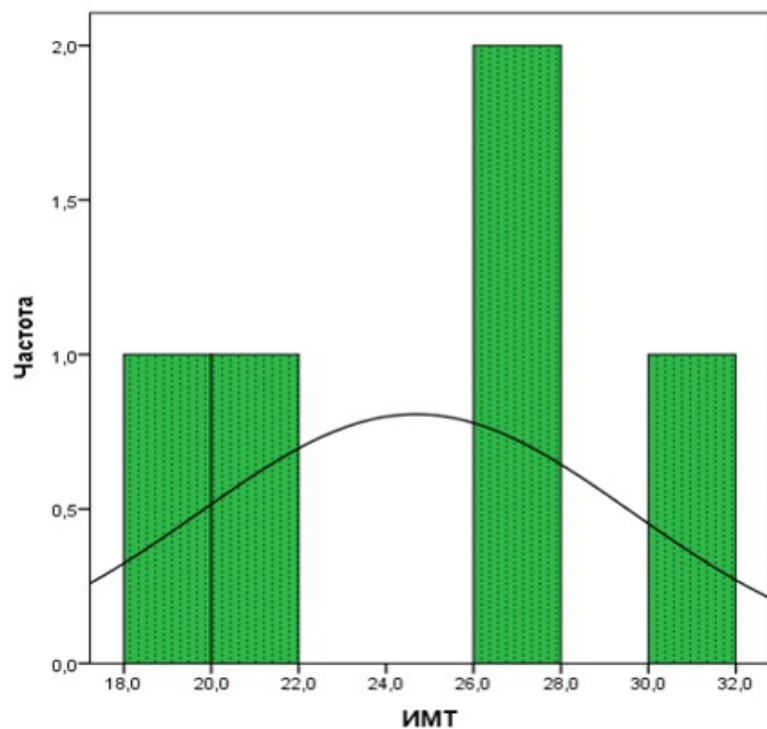
Выбор статистического критерия

Тип данных	Двумерный анализ	Многомерный анализ
Непрерывные данные, нормальное распределение	<ul style="list-style-type: none">• Непарные измерения: t-критерий Стьюдента• Парные измерения: парный t-критерий Стьюдента	<ul style="list-style-type: none">• Непарные измерения: ANOVA• Парные измерения: парный вариант ANOVA
Непрерывные данные, не имеющие нормального распределения	<ul style="list-style-type: none">• Непарные измерения: U-критерий Манна-Уитни• Парные измерения: критерий Вилкоксона	<ul style="list-style-type: none">• Непарные измерения: критерий Краскелла-Уоллиса• Парные измерения: критерий Фридмана
Номинальные данные	<ul style="list-style-type: none">• Таблицы сопряженности• Критерий Фишера (малые выборки)• Критерий хи-квадрат (χ^2-квадрат)	<ul style="list-style-type: none">• Таблицы сопряженности• Критерий Фишера (малые выборки)• Критерий хи-квадрат (χ^2-квадрат)

Распределение ИМТ

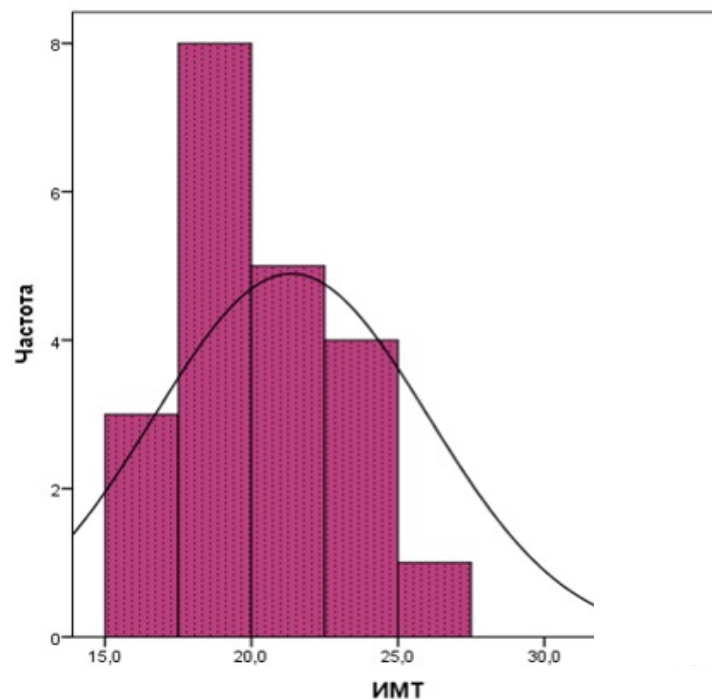
Как сравнить два распределения?

Группа 1
Наука +



Me (ИИ) 23,88 (20,1- 28,1)

Группа 2
Наука -

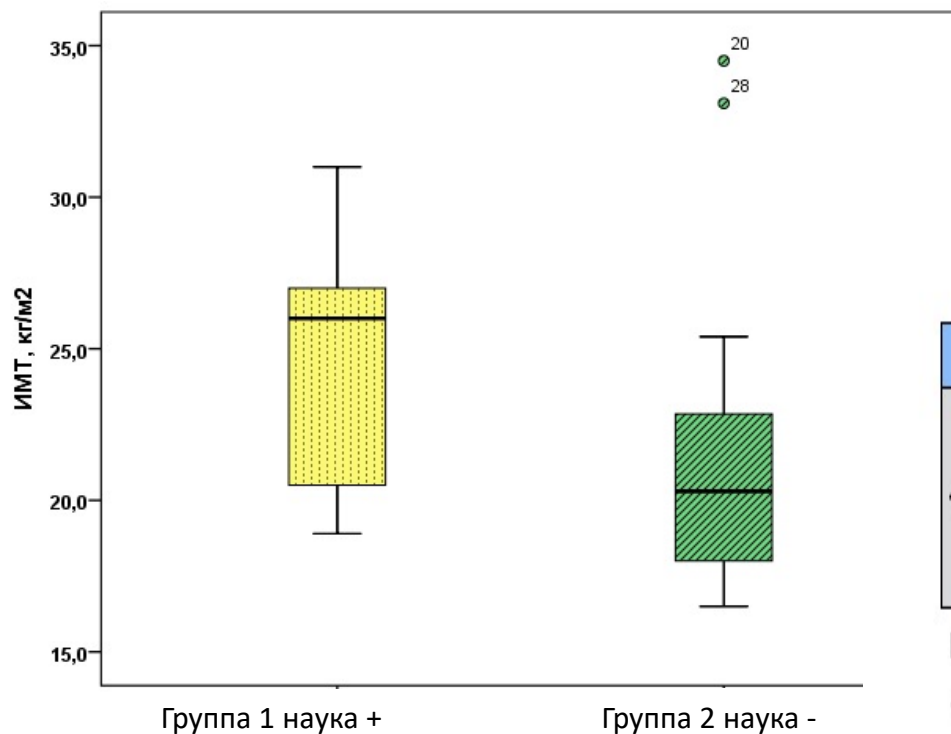


Me (ИИ) 20,16 (18,6 - 26,7)

Выбор статистического критерия

Тип данных	Двумерный анализ	Многомерный анализ
Непрерывные данные, нормальное распределение	<ul style="list-style-type: none">• Непарные измерения: t-критерий Стьюдента• Парные измерения: парный t-критерий Стьюдента	<ul style="list-style-type: none">• Непарные измерения: ANOVA• Парные измерения: парный вариант ANOVA
Непрерывные данные, не имеющие нормального распределения	<ul style="list-style-type: none">• Непарные измерения: U-критерий Манна-Уитни• Парные измерения: критерий Вилкоксона	<ul style="list-style-type: none">• Непарные измерения: критерий Краскелла-Уоллиса• Парные измерения: критерий Фридмана
Номинальные данные	<ul style="list-style-type: none">• Таблицы сопряженности• Критерий Фишера (малые выборки)• Критерий хи-квадрат (χ^2-квадрат)	<ul style="list-style-type: none">• Таблицы сопряженности• Критерий Фишера (малые выборки)• Критерий хи-квадрат (χ^2-квадрат)

Проверка гипотезы с помощью U-критерия Манна-Уитни



Итоги по проверке гипотезы

	Нулевая гипотеза	Критерий	Знач.	Решение
1	Распределение индекс массы тела является одинаковым для категорий группа: больше 4 - 1, меньше 4 - 0.	Критерий U Манна-Уитни для независимых выборок	0,041	Нулевая гипотеза отклоняется.

Выводятся асимптотические значимости. Уровень значимости равен ,05.

¹Приводится точная значимость критерия.

Сравнение номинальных данных

Таблицы сопряженности

Группа 1
наука +

Группа 2
наука-

ср.балл $\gg 4$

113

91

ср.балл $< 4,0$

27

41

113	91
27	41

Выбор статистического критерия

Тип данных	Двумерный анализ	Многомерный анализ
Непрерывные данные, нормальное распределение	<ul style="list-style-type: none">• Непарные измерения: t-критерий Стьюдента• Парные измерения: парный t-критерий Стьюдента	<ul style="list-style-type: none">• Непарные измерения: ANOVA• Парные измерения: парный вариант ANOVA
Непрерывные данные, не имеющие нормального распределения	<ul style="list-style-type: none">• Непарные измерения: U-критерий Манна-Уитни• Парные измерения: критерий Вилкоксона	<ul style="list-style-type: none">• Непарные измерения: критерий Краскелла-Уоллиса• Парные измерения: критерий Фридмана
Номинальные данные	<ul style="list-style-type: none">• Таблицы сопряженности• Критерий Фишера (малые выборки)• Критерий кси-квадрат (χ^2-квадрат)	<ul style="list-style-type: none">• Таблицы сопряженности• Критерий Фишера (малые выборки)• Критерий кси-квадрат (χ^2-квадрат)

Критерии хи-квадрат

	Значение	ст.св.	Асимптотическая значимость (2-сторонняя)	Точная знч. (2-сторонняя)	Точная знч. (1-сторонняя)
Хи-квадрат Пирсона	,101 ^a	1	,750		
Поправка на непрерывность ^b	,000	1	1,000		
Отношения правдоподобия	,102	1	,750		
Точный критерий Фишера				1,000	,572
Линейно-линейная связь	,098	1	,755		
Количество допустимых наблюдений	28				

а. Для числа ячеек 2 (50,0%) предполагается значение, меньше 5. Минимальное предполагаемое число равно 2,32.

б. Вычисляется только для таблицы 2x2

Таблица

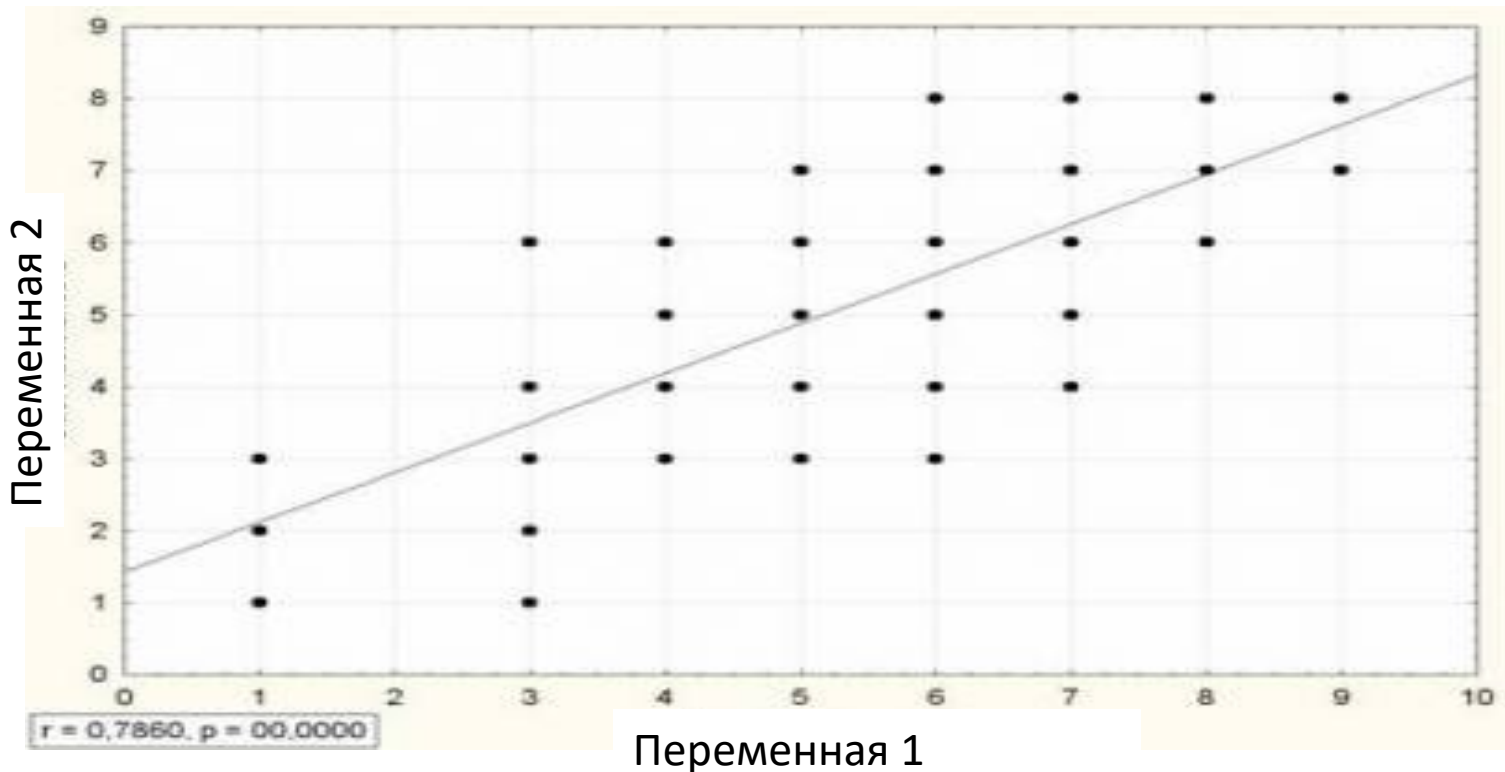
Признак	Группа 1 Хотят заниматься наукой n= 140	Группа 1 Не хотят заниматься наукой n= 132	p
Мужчины	39 (28%)	40 (30%)	0,890
Женщины	101 (72%)	92 (70%)	
ИМТ, кг/м ² Me (ИИ)	23,88 (20,1- 28,1)	20,16 (18,6 - 26,7)	0,041
Средний балл >> 4	113 (81%)	91 (70%)	0,750

Статистическая связь случайных величин

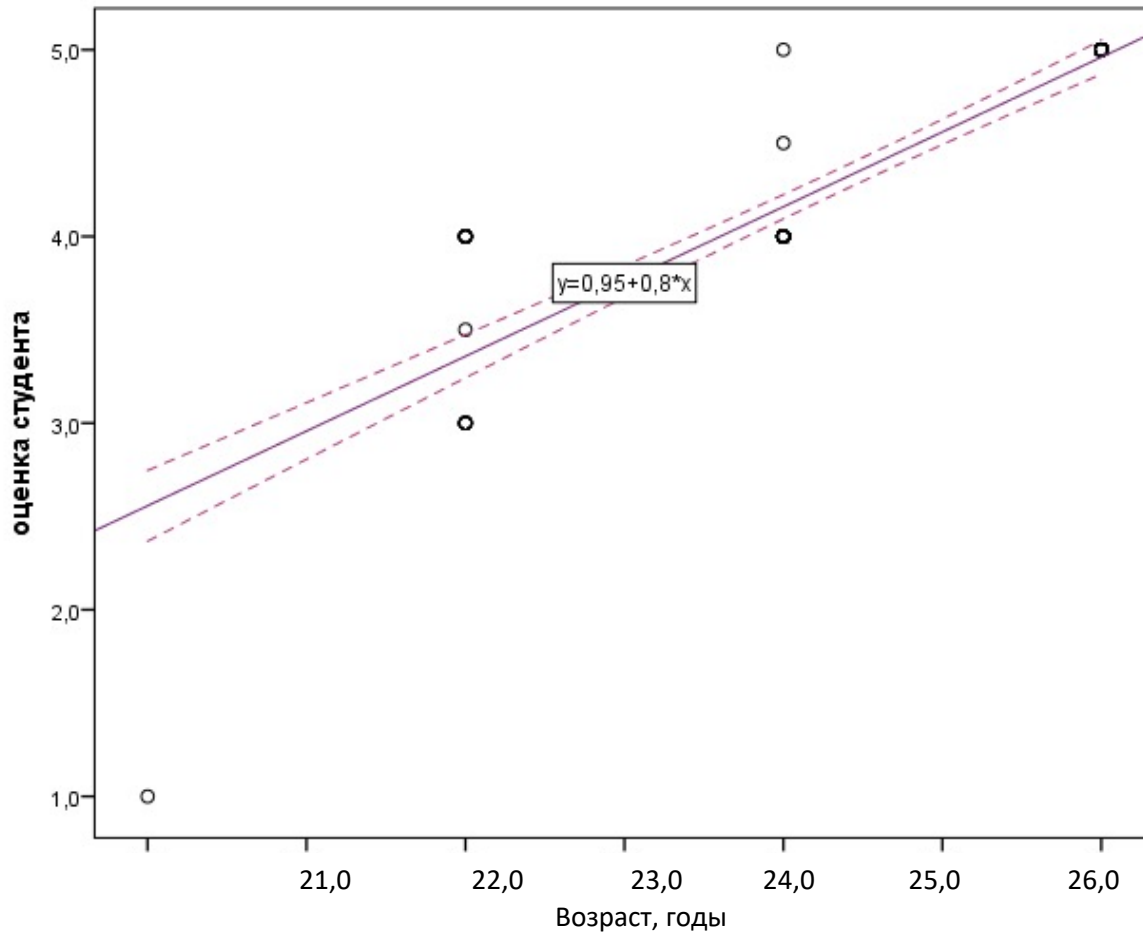
- **Корреляция** – метод обработки статистических данных, с помощью которого измеряется теснота связи между двумя и более переменными
- **Регрессия** – статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную

Корреляция

- Коэффициент корреляции Пирсона (нормальное распределение)
- Коэффициент корреляции Спирмана (непараметрические данные)

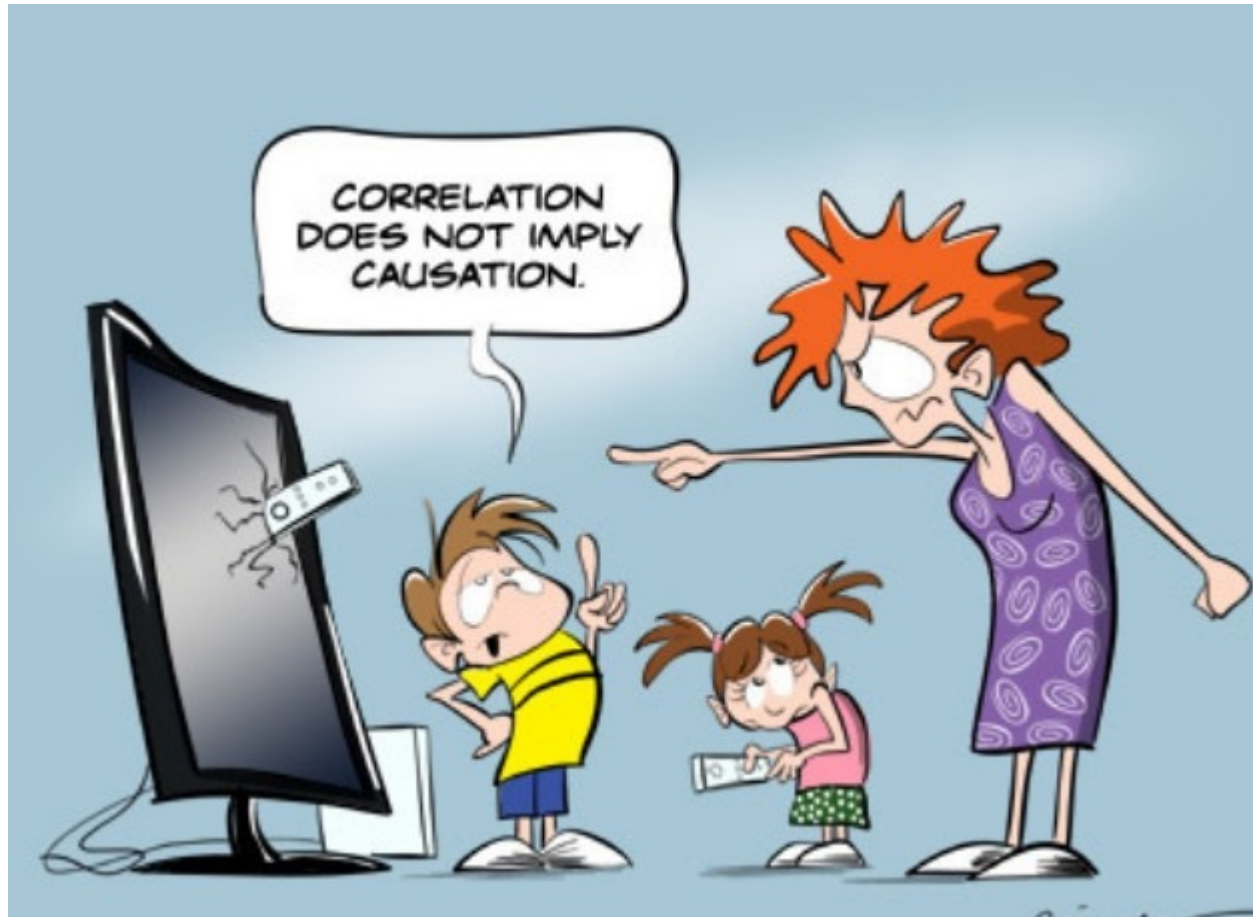


Корреляция между оценкой студента и возрастом (количественные переменные, коэффициент Спирмена)



Данные 2018 г.

**Корреляция не делает вывод о
причинно-следственных связях!!!!**



Эффективность клинических испытаний

- Сравнение результатов лечения (благоприятных и неблагоприятных) с помощью статистических методов
- Показатель NNT – число больных, которых необходимо лечить, чтобы достичь определенного благоприятного исхода или предотвратить один неблагоприятный исход у одного больного, в сравнении с контрольной группой

Число больных, которых необходимо лечить, чтобы достичь определенного эффекта у одного больного (NNT)

- **1/снижение абсолютного риска**

- Низкое значение (приближается к 1) – большая эффективность «Совершенный препарат»

Пример:

- В результате лечения препаратом X абсолютный риск смерти от острого инфаркта миокарда снизился с 0,5 до 0,3 (т.е. на 0,2)

$$\text{NNT} = 1/0,2 = 5.0$$

- 5 человек надо лечить препаратом X, что бы «спасти» 1 больного