

# Введение в медицинскую статистику

# Зачем это нужно?

- Собственное исследование
- Критическое чтение источников



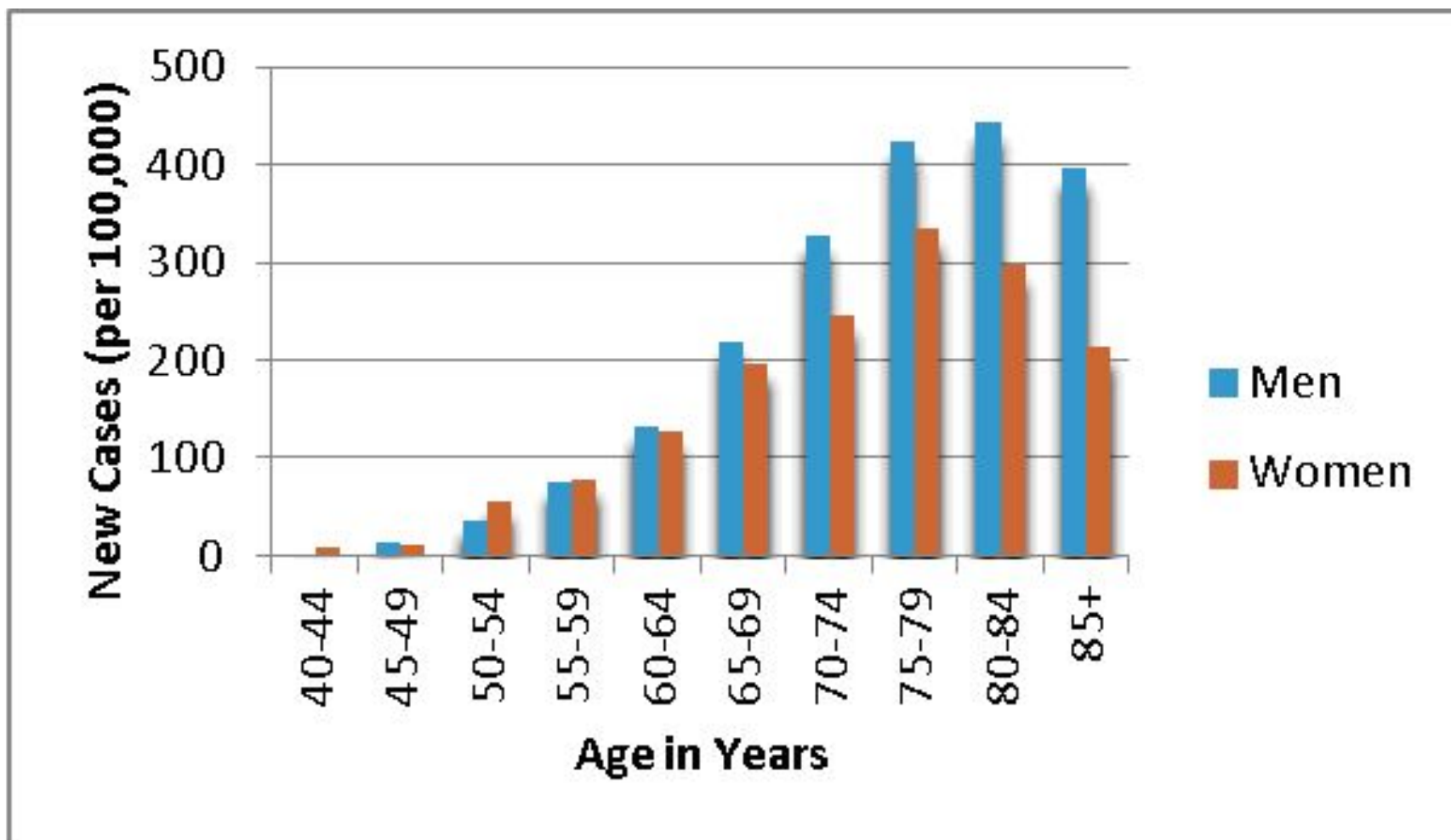
# Основные понятия

- Популяция = генеральная совокупность
- Выборка (главное свойство – репрезентативность)

# Статистика

- **Дескриптивная = описательная**
- Аналитическая

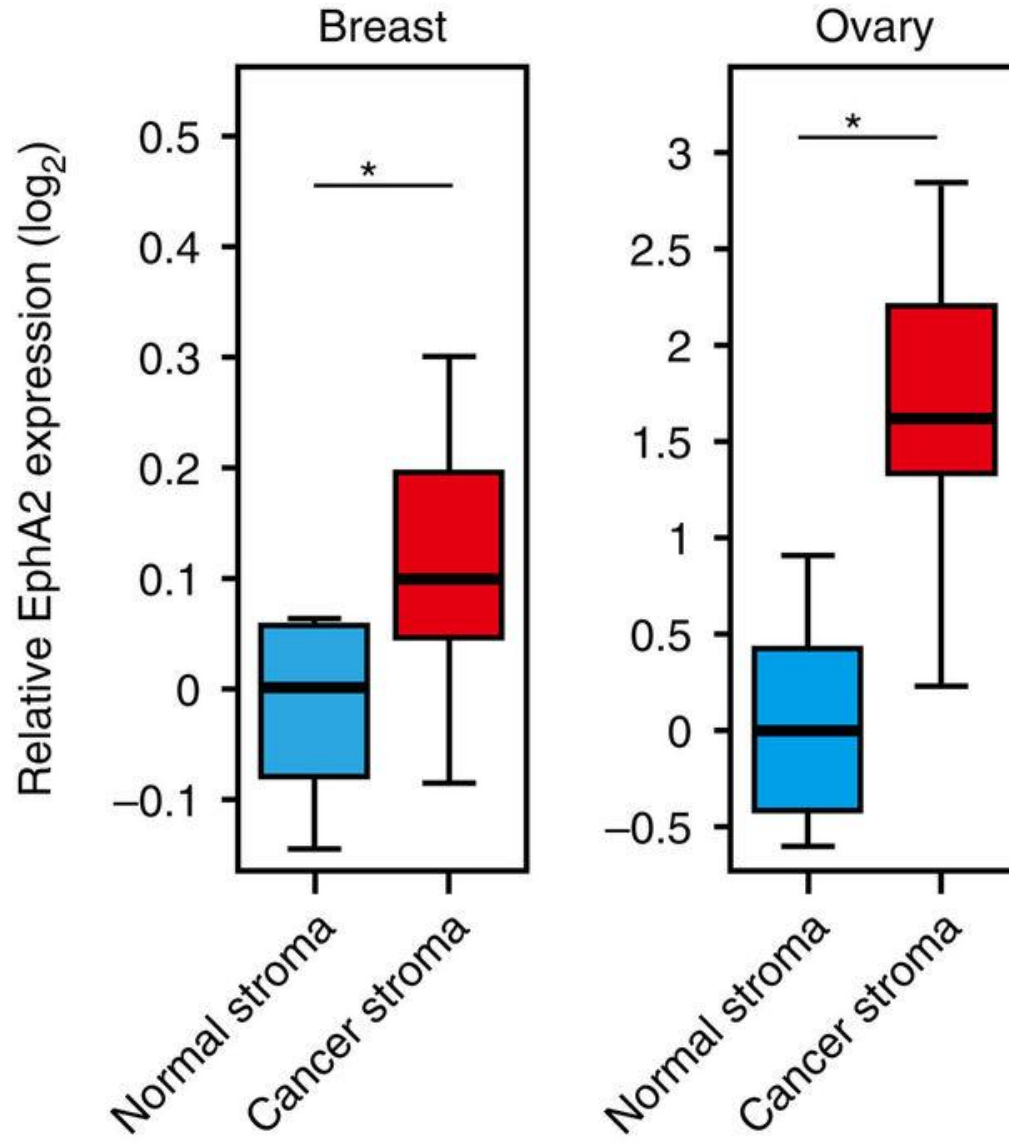
# Гистограмма частот



# Гистограмма частот



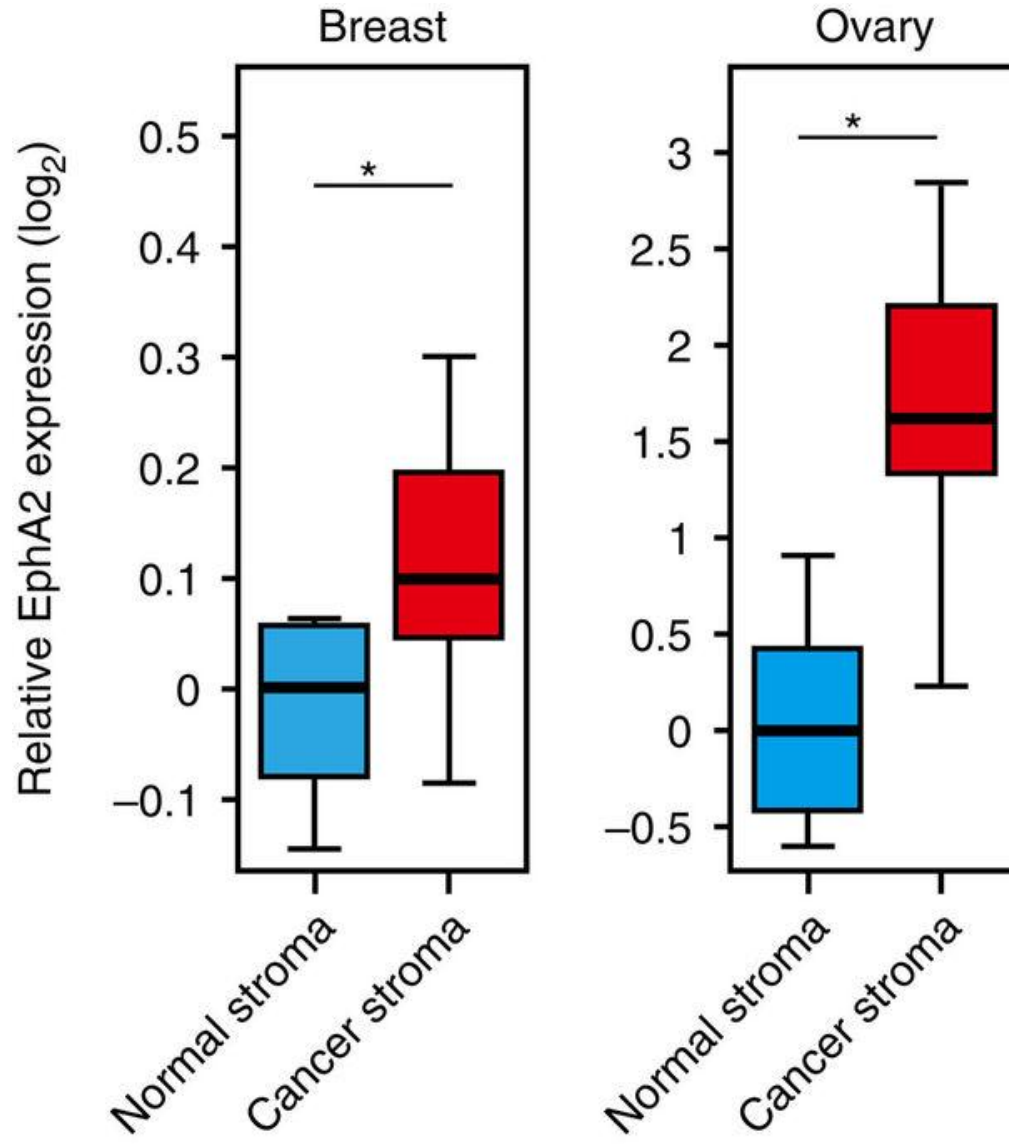
# Box plot



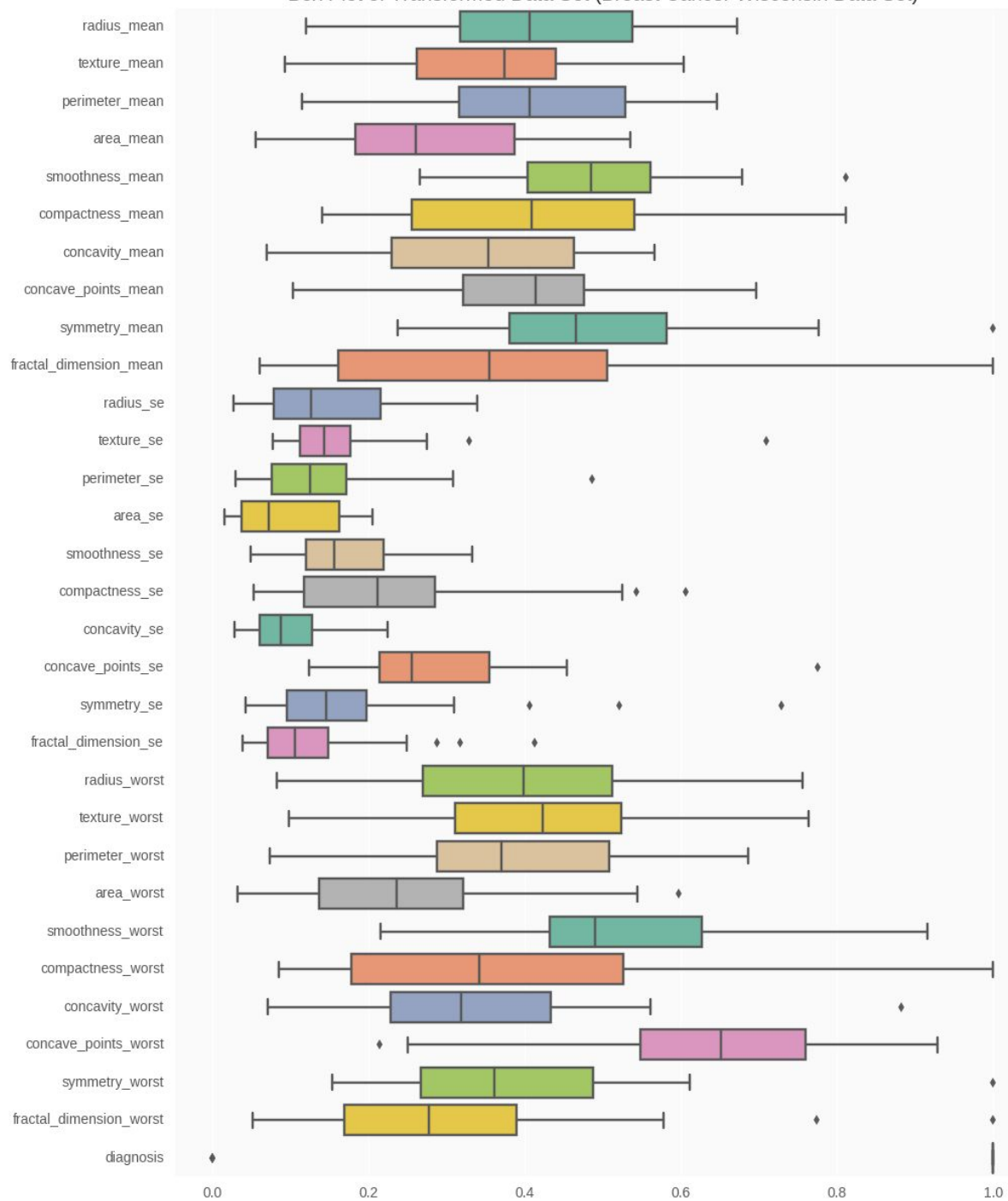


- Среднее арифметическое
- Мода – наиболее часто встречающаяся величина
- Медиана – делит выборку пополам по количеству случаев
- Разделение выборки на 4 равные части - квантили

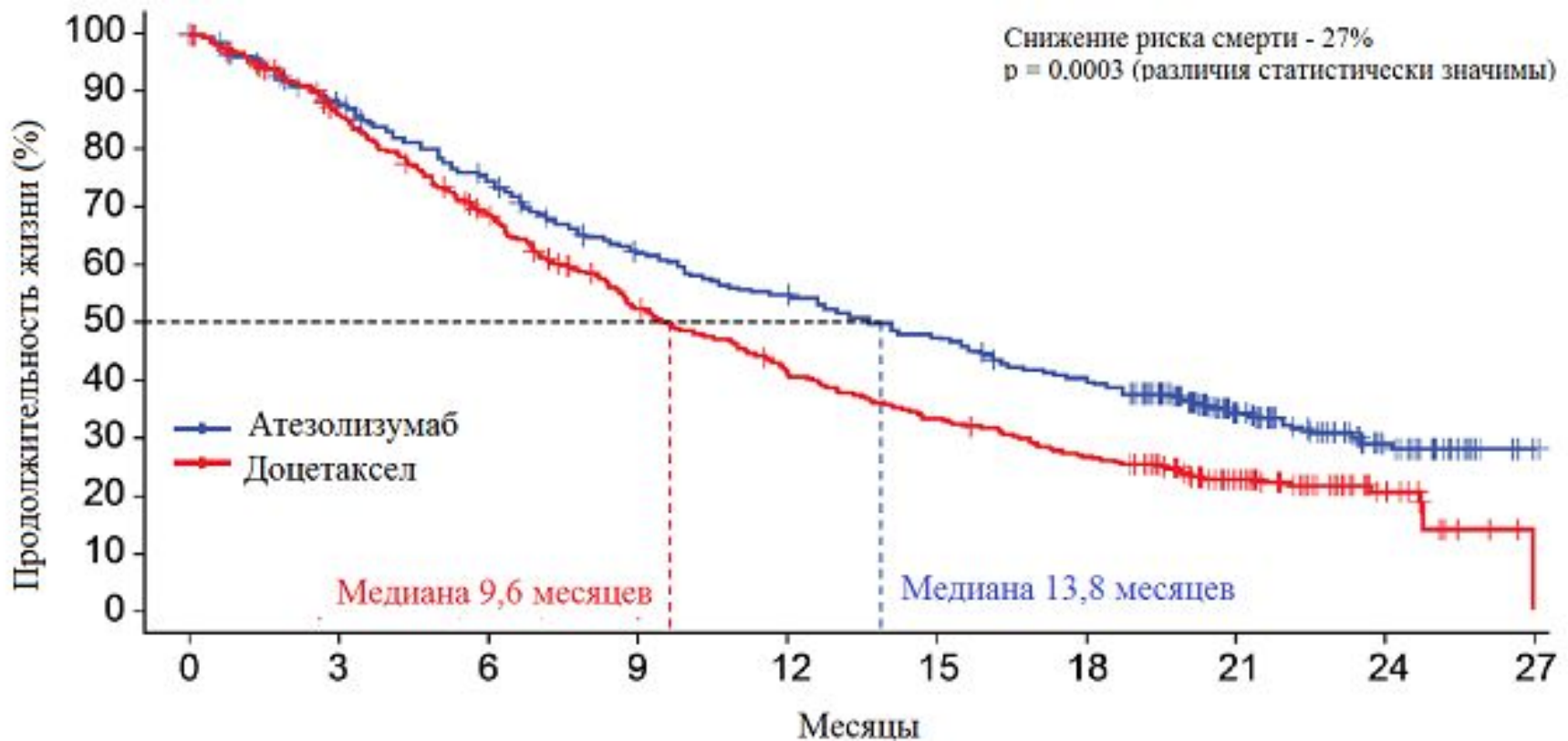
# Box plot



Box Plot of Transformed Data Set (Breast Cancer Wisconsin Data Set)



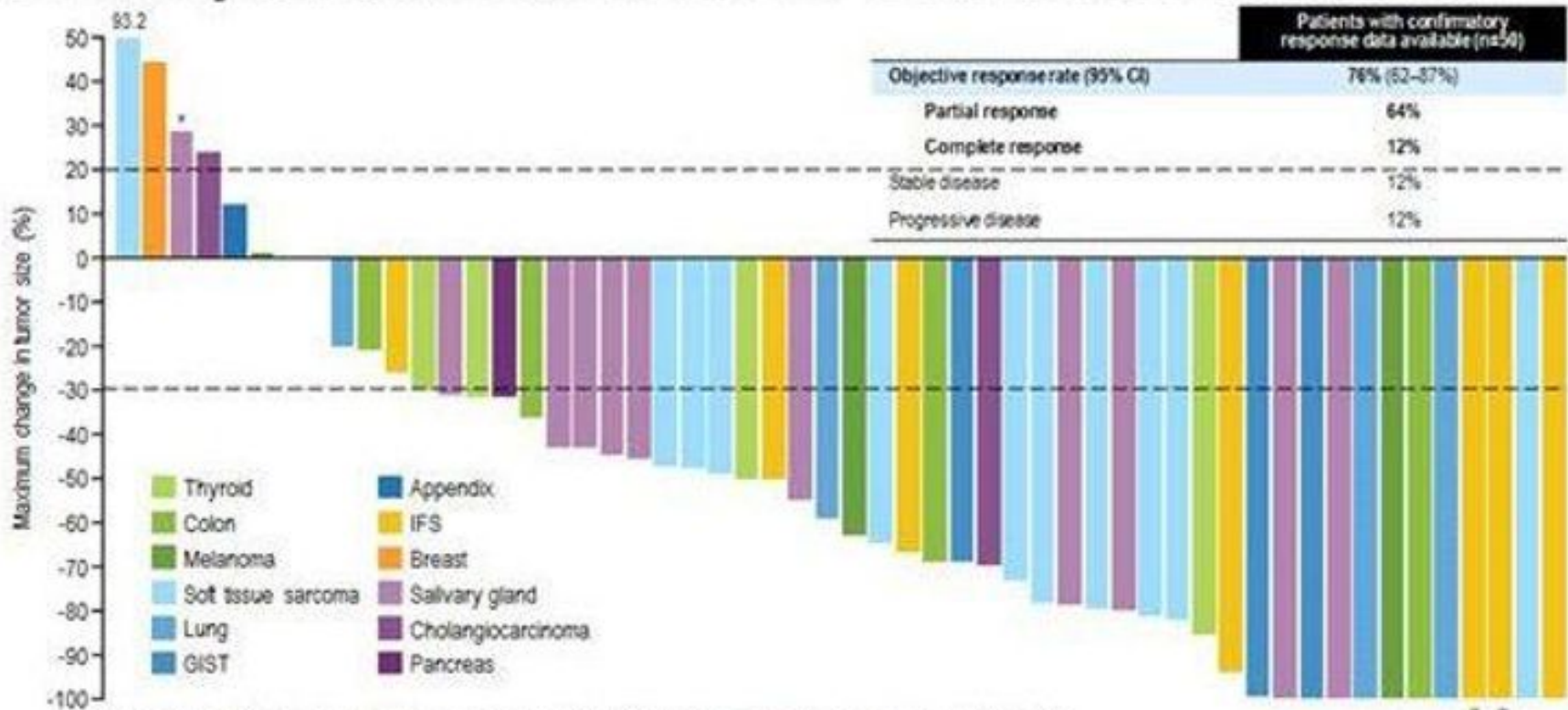
# Кривая Каплана-Майера



- Анализ не только выживаемости
- Возможность работы с цензурированными данными

# Waterfall гистограмма

## Efficacy of larotrectinib in TRK fusion cancers



\*Patient had TRK solvent front resistance mutation (NTRK3 G623R) at baseline due to prior therapy, #Pathologic CR

Note: One patient not shown here. Patient experienced clinical progression and no post-baseline tumor measurements were recorded.

# Статистика

- Дескриптивная = описательная
- **Аналитическая**

# Аналитическая статистика

- Позволяет делать выводы
- Мера уверенности в «истинности» результата – р-значение (должно быть  $\leq 0.05$ )

# Виды данных

- Номинативные (качественные, категориальные)
  - \*бинарные
- Количественные (метрические)
- Ранговые



# Виды количественных данных

- Непрерывные
- Дискретные (прерывные)

# Программы

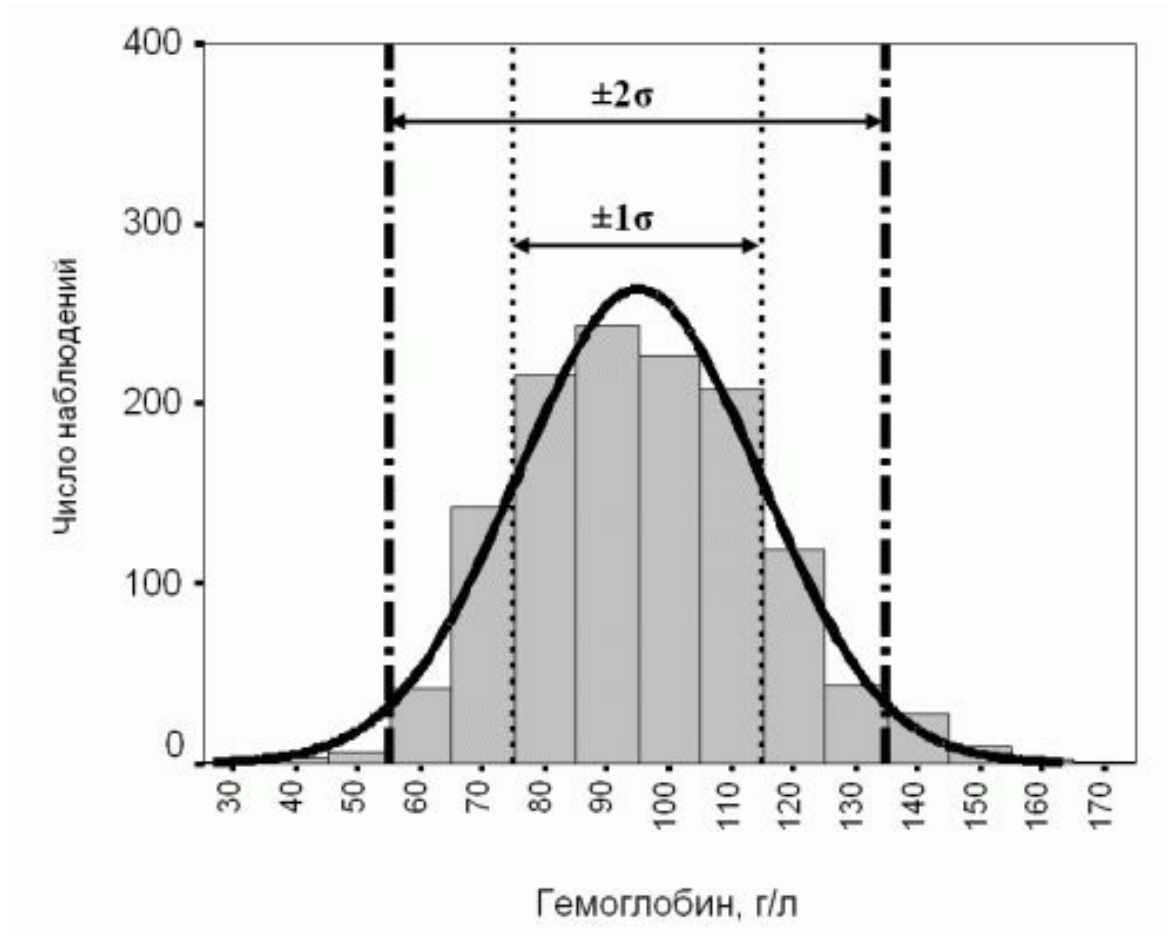
- Statistica
- SPSS
- R
- SAS

# Методы

- Параметрические – для нормально распределенных величин
- Непараметрические – для ненормально распределенных величин
- Взаимодействие величин (корреляционный анализ, анализ выживаемости, регрессионный анализ...)

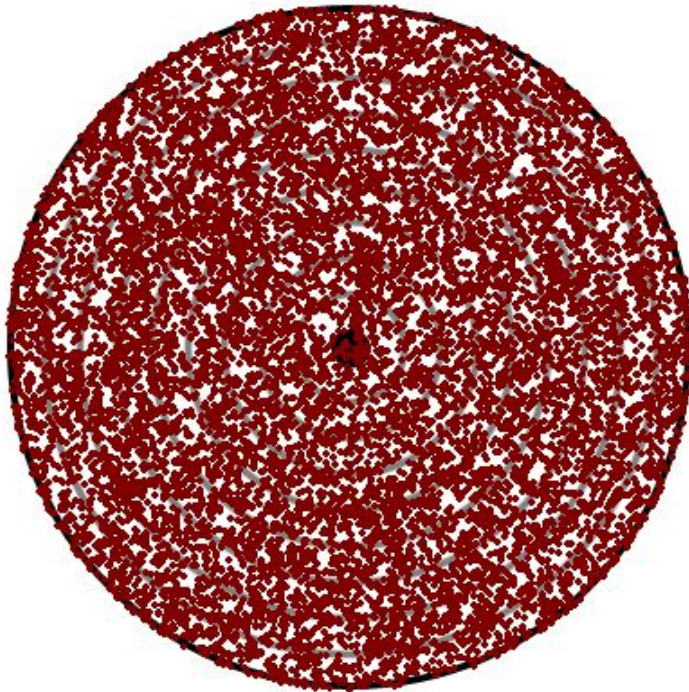
На выходе получаем p-value

# Нормальное распределение



Только для количественных  
данных!

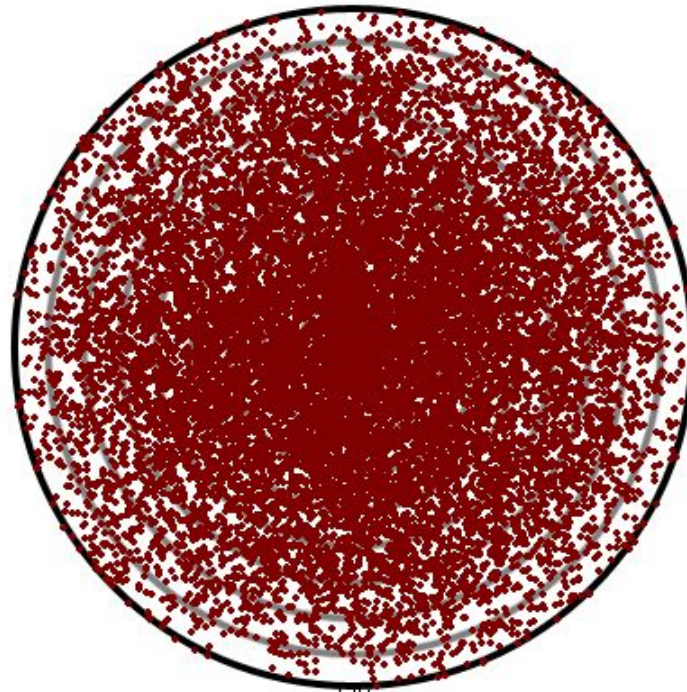
# Нормальное распределение $\neq$ равномерное



Равномерное распределение, ось X



Равномерное распределение, ось Y

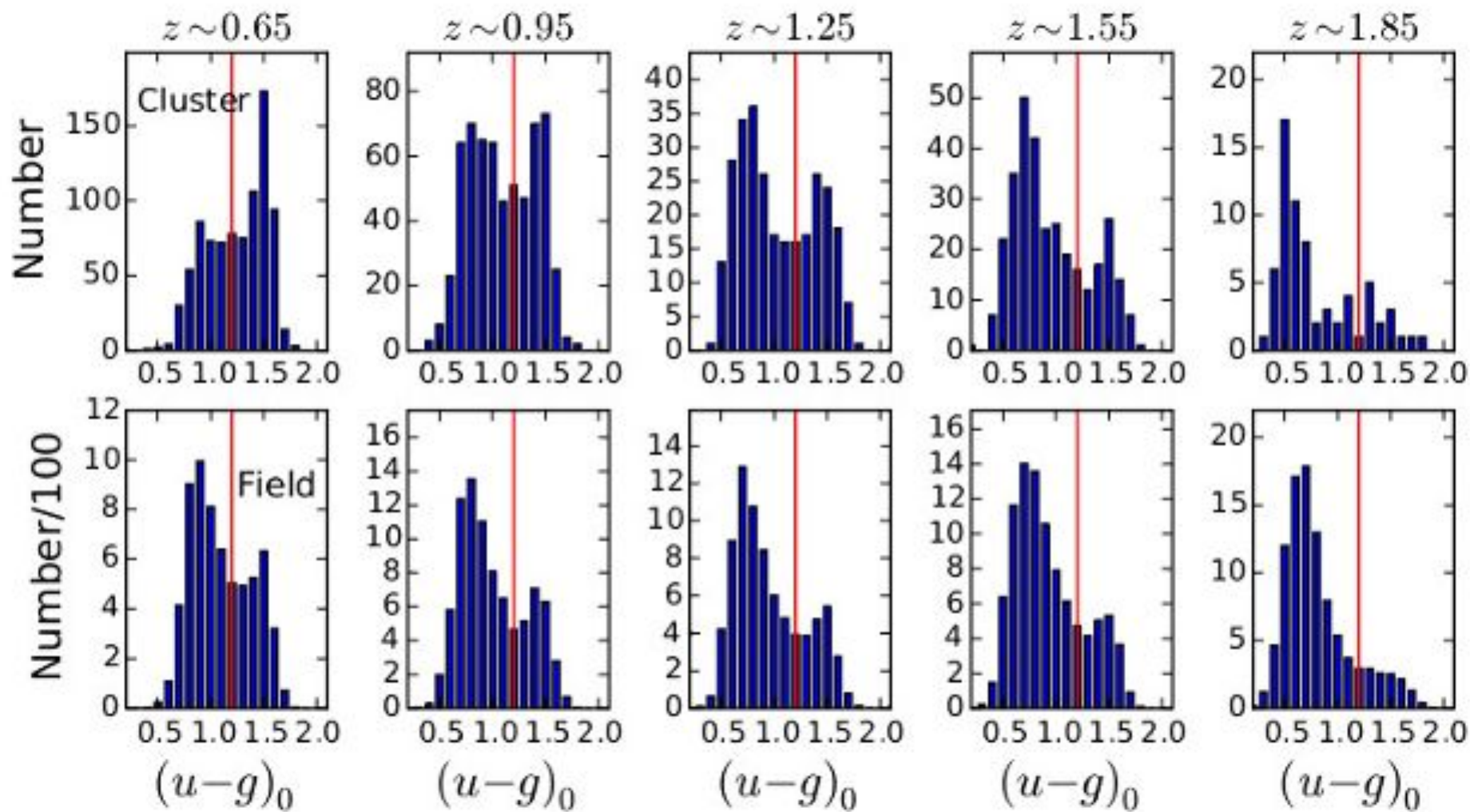


Гауссово распределение, ось X



Гауссово распределение, ось Y

# Ненормальное распределение



# Оценка нормальности распределения

- Визуальная
- Критерий Колмогорова-Смирнова

# Виды выборок

- Связанные
- Несвязанные



# Связанные выборки

	Уровень глюкозы до приема пищи	Уровень глюкозы после приема пищи
Пациент №1	3.3	8.4
Пациент №2	3.7	9.3
Пациент №3	4.3	7.7

# Несвязанные

Уровень глюкозы до приема пищи, пациенты без диабета	Уровень глюкозы до приема пищи, пациенты с диабетом
3.3	8.4
3.7	9.3
4.3	7.7

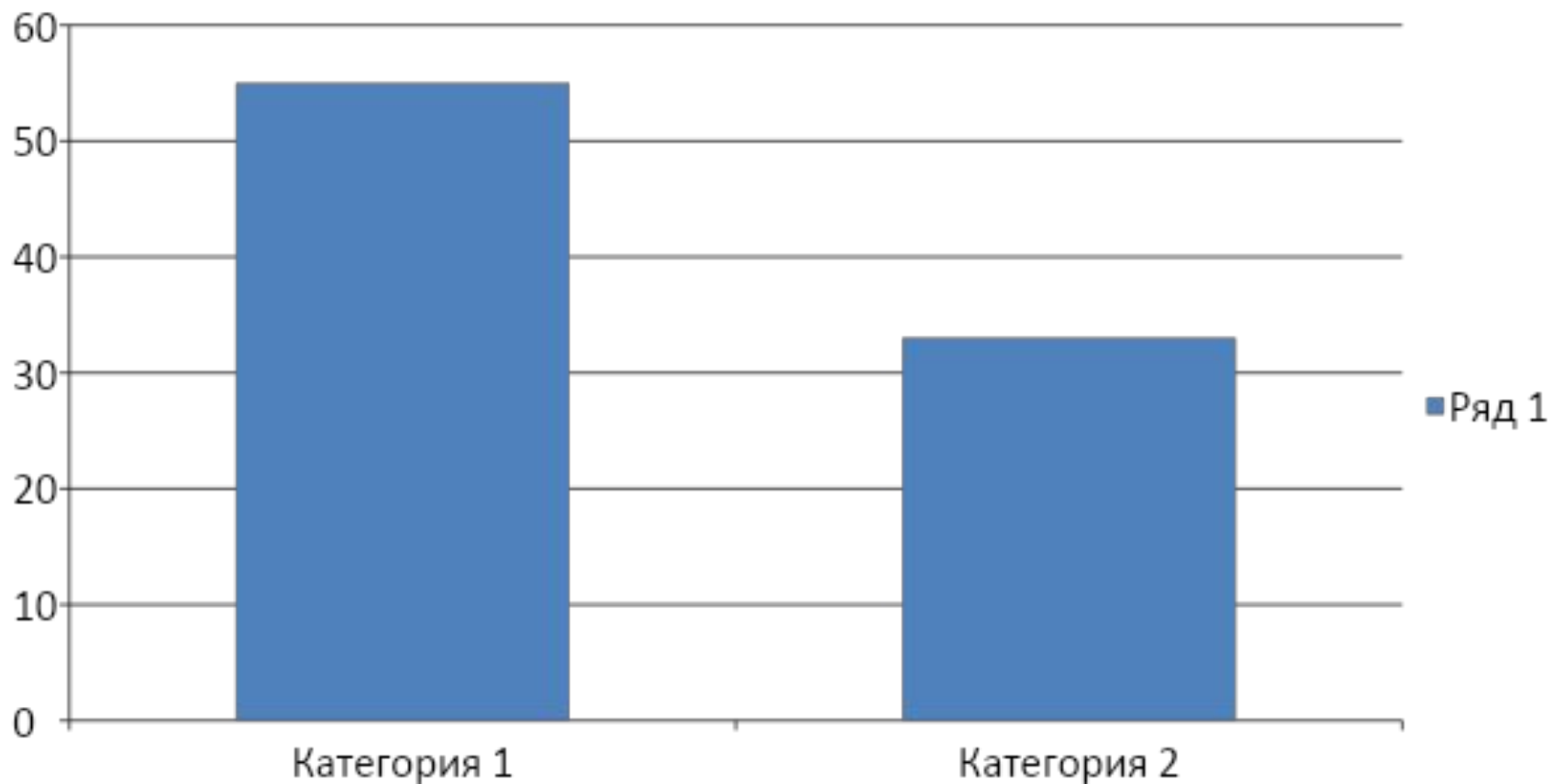
# Методы параметрической статистики

- Две группы
  - Несвязанные выборки: критерий Стьюдента
  - Связанные выборки: парный критерий Стьюдента
- Три группы и более
  - Несвязанные выборки: дисперсионный анализ, критерий Стьюдента с поправкой Бонферони/Ньюмена-Кейлса
  - Связанные выборки: дисперсионный анализ повторных измерений...

# Методы непараметрической статистики

- Две группы
  - Несвязанные выборки: критерий Манна-Уитни
  - Связанные выборки: критерий Уилкоксона
- Три группы и более
  - Несвязанные выборки: критерий Крускала-Уоллиса
  - Связанные выборки: критерий Фридмана

# Графическое представление



# Анализ качественных данных

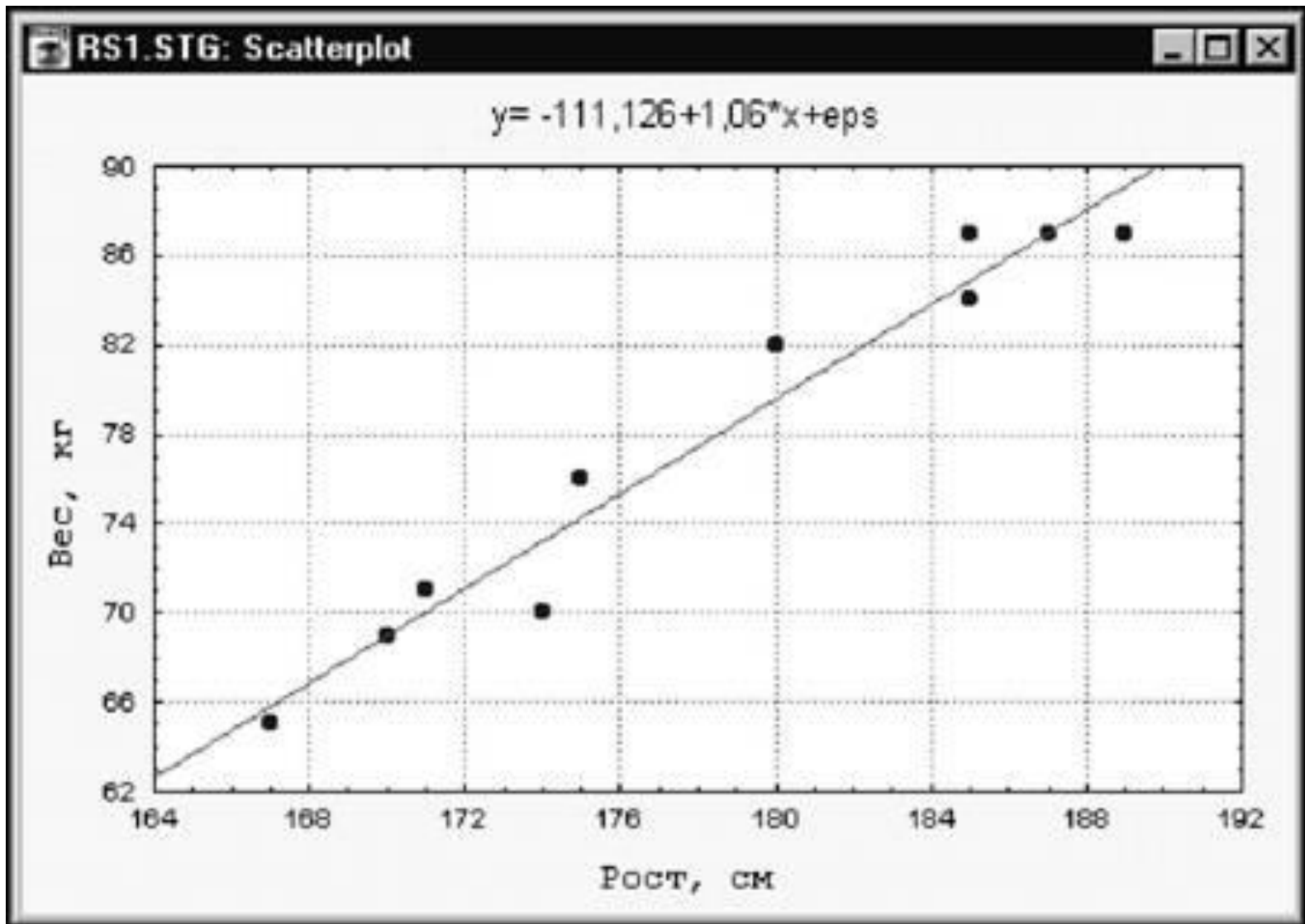
- Две группы
  - Несвязанные выборки: критерий хи-квадрат, критерий Z, точный критерий Фишера
  - Связанные выборки: критерий Мак-Нимара
- Три группы и более
  - Несвязанные выборки: хи-квадрат
  - Связанные выборки: критерий Кокрена

# Методы

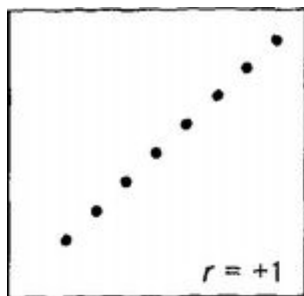
- Параметрические – для нормально распределенных величин
- Непараметрические – для ненормально распределенных величин
- **Взаимодействие величин**  
(корреляционный анализ, анализ выживаемости, регрессионный анализ...)

На выходе получаем p-value

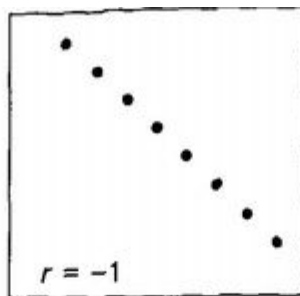
# Корреляция



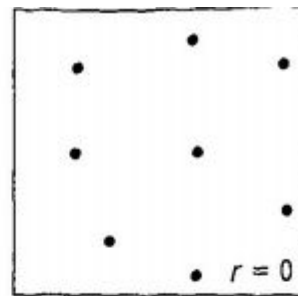
# Корреляция



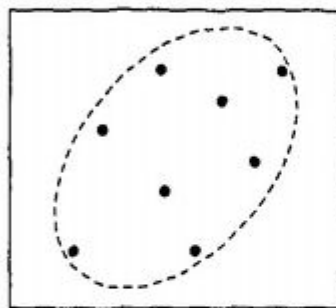
а



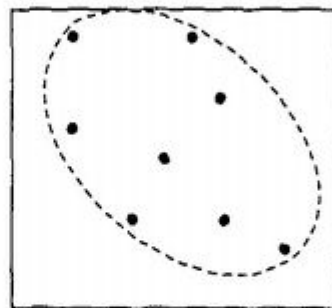
б



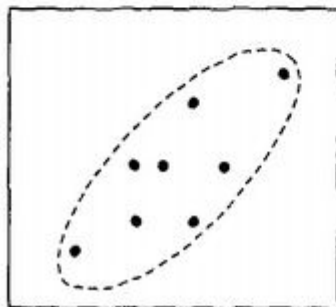
в



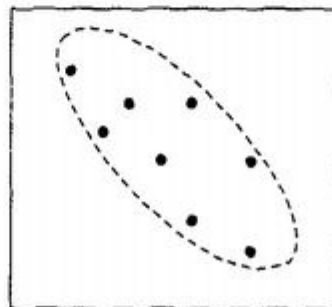
г



д



е



ж



# Корреляция

- Количественный признак, нормальное распределение: коэффициент Пирсона
- Количественный/ранговый признак, ненормальное распределение: коэффициент Спирмена

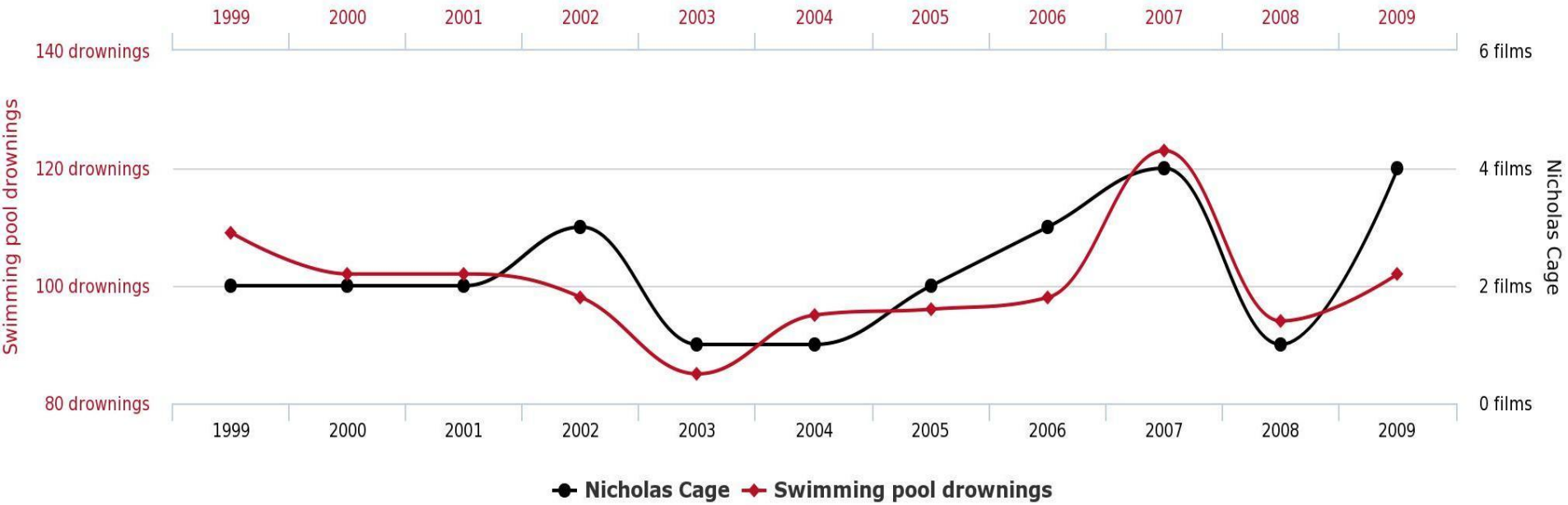
# Сила связи

Сильная	более 0,70
Средняя	от 0,50 до 0,69
Умеренная	от 0,30 до 0,49
Слабая	от 0,20 до 0,29
Очень слабая	меньше 0,19

# Number of people who drowned by falling into a pool

correlates with

## Films Nicolas Cage appeared in



tylervigen.com



# Анализ выживаемости

- Графическое представление – кривая Каплана-Майера
- Нельзя сравнивать кривые только на вид!
- Сравнение – лог-ранк тест

# Выживаемость

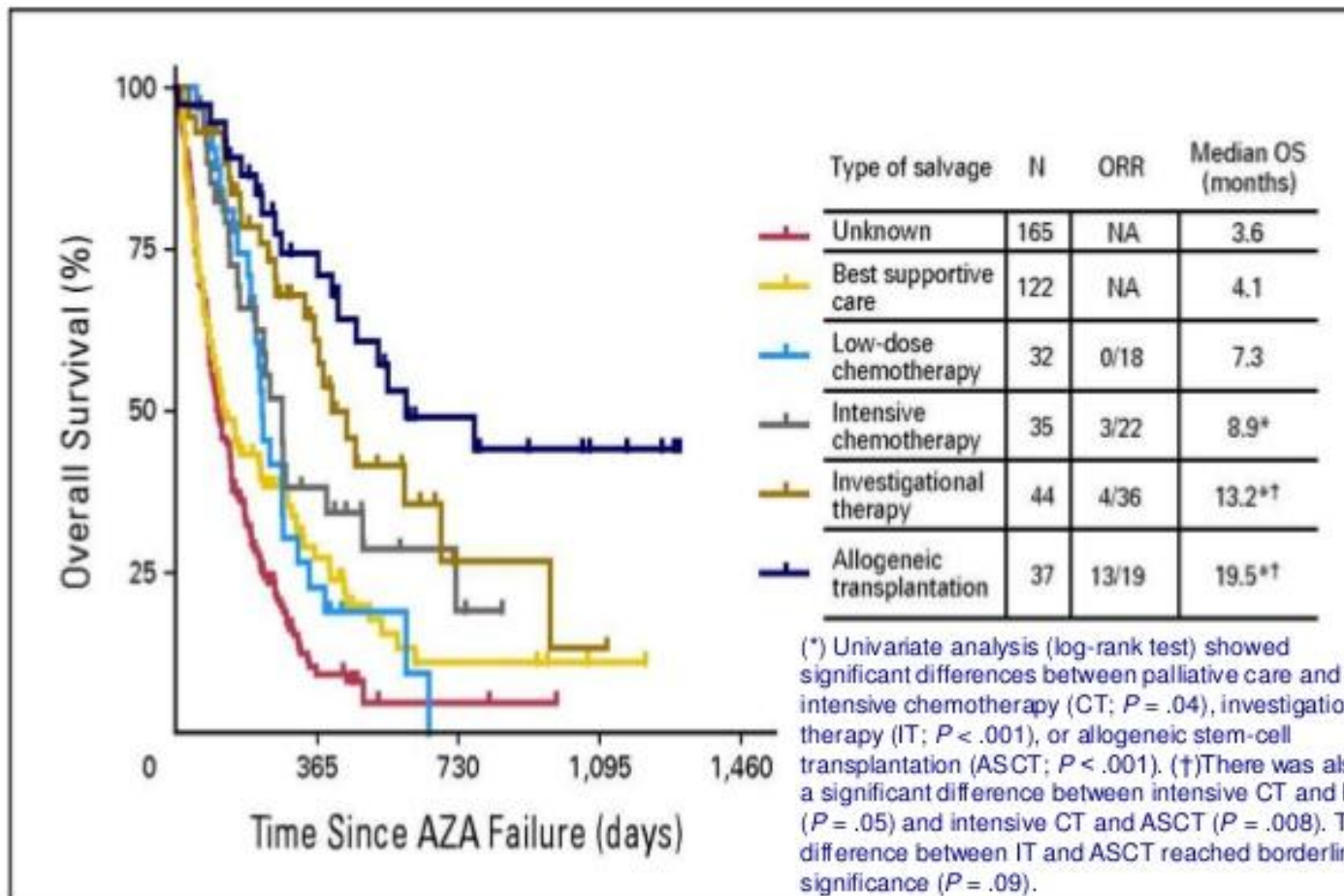
- Общая
- Кумулятивная

# Анализ выживаемости

	Безрецидивная выживаемость, мес	Статус рецидива нет 0/рецидив есть 1
Пациент №1	7	0
Пациент №2	11	1
Пациент №3	8	1
Пациент №4	33	0

# Анализ выживаемости

Survival analysis according to the salvage treatment regimens.



# Типы переменных

- Зависимая (то, что оцениваем)
- Независимая (фактор, который может повлиять на результат)



# Регрессионный анализ

- Линейная регрессия – зависимая переменная количественная
- Логистическая регрессия – зависимая переменная качественная
  - Бинарная
  - Мультиномиальная

Нет возможности работы с цензурированными данными

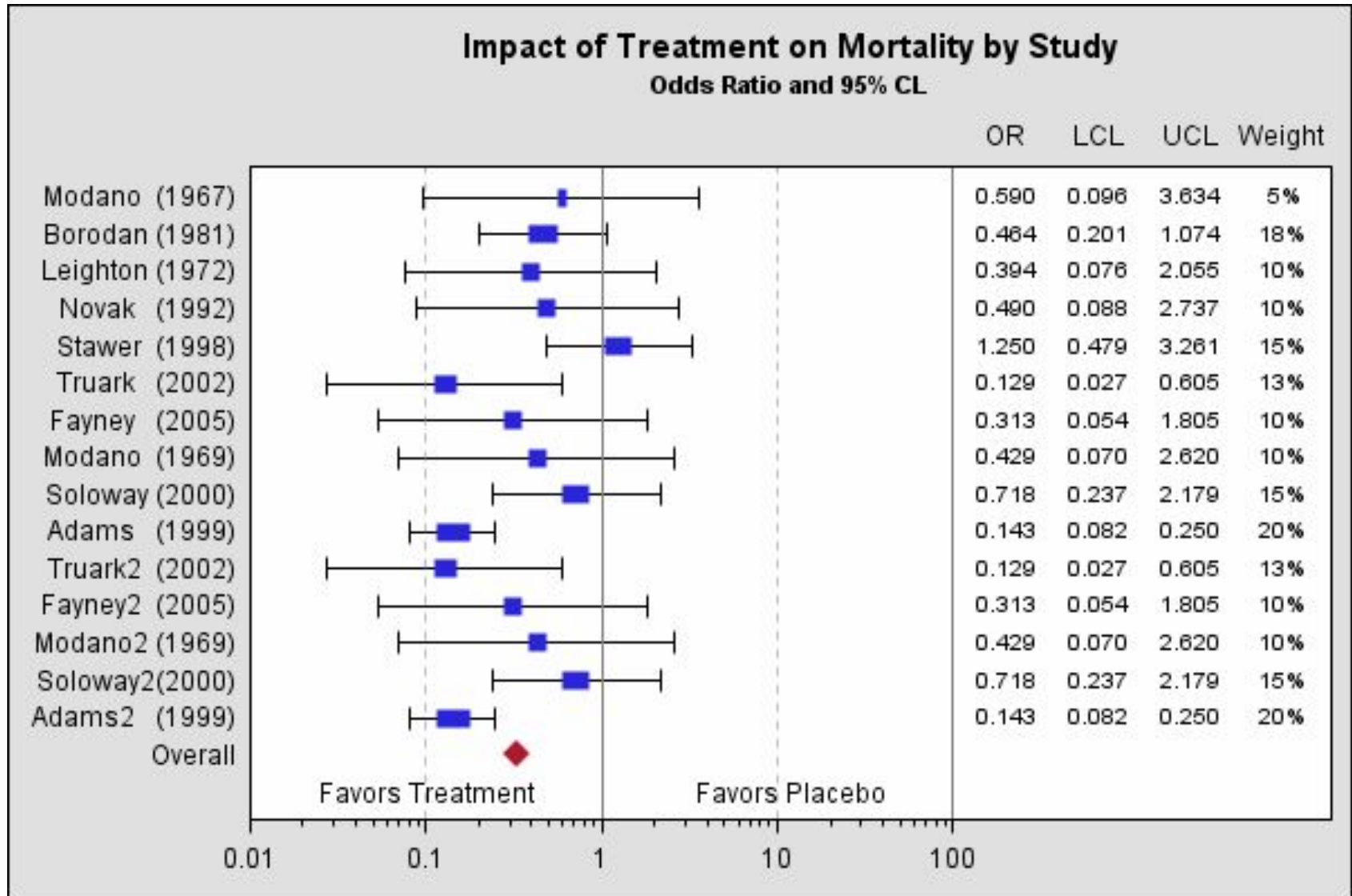
# Регрессия Кокса

- Регрессия Кокса – риск наступления события. Возможность работы с цензурированными данными

# Регрессия Кокса

- Коэффициент риска (hazard ratio) – отношение вероятности наступления некоторого события для первой группы объектов к вероятности наступления этого же события для второй группы объектов.
- Доверительный интервал (confidence interval) - диапазон вокруг значения величины, в котором находится истинное значение этой величины (с определенным уровнем доверия).  
ДИ 95 % (CI 95%)

# Forest-plot



# Полезные ссылки

- <http://welcome.stepik.org/ru>
- <https://ru.coursera.org/>
- Методические основы проведения клинических исследований – kurskmed
- С. Гланц. Медико-биологическая статистика
- Youtube