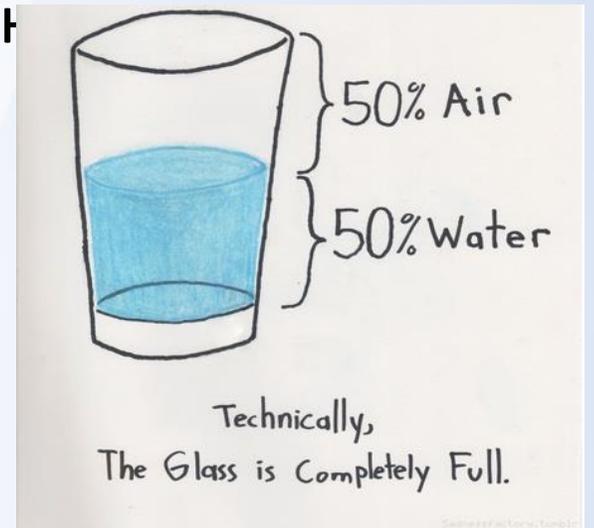


Статистика для чайников

и для адвокатов пациентских движений 😊

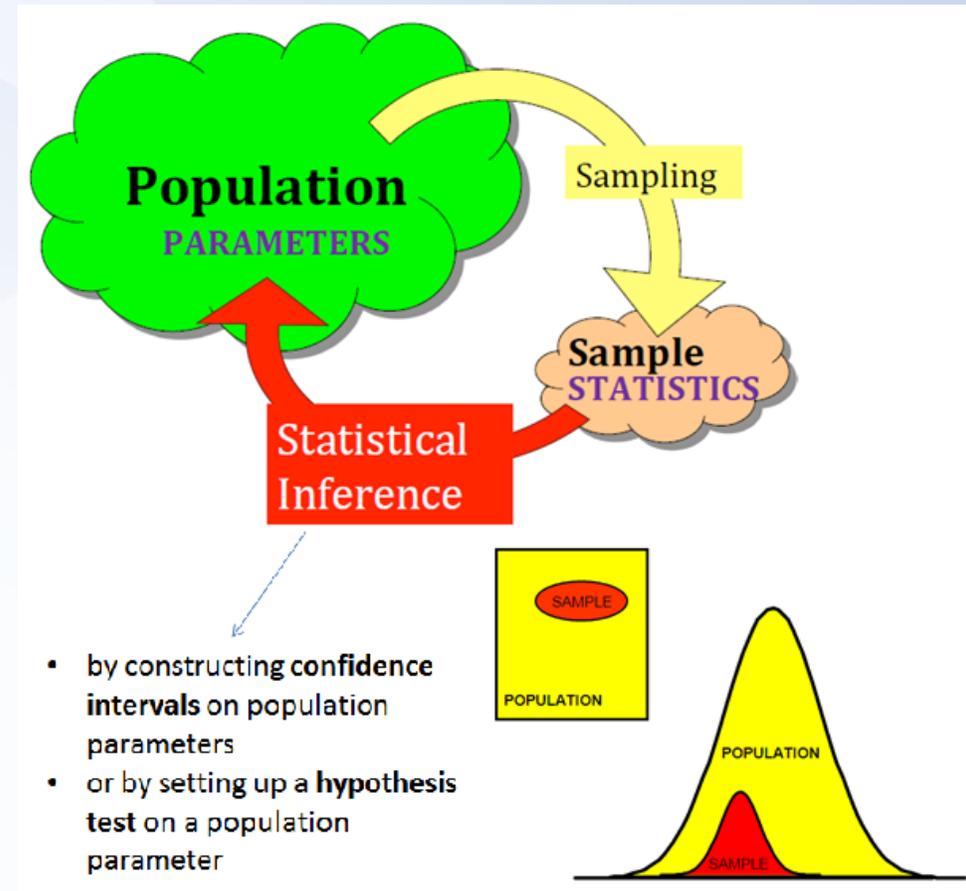
Что такое статистика и зачем она нужна?

- “Государственная наука»– была изначально создана для того, чтобы можно было оценить богатство и силу отдельных государств
- Научный метод и способ мышления: точность и объективность измерений (в рамках возможного)
- Наука о сборе, упорядочивании и трактовании данных
- **Статистика зависит от трактовки данных**



Виды статистики

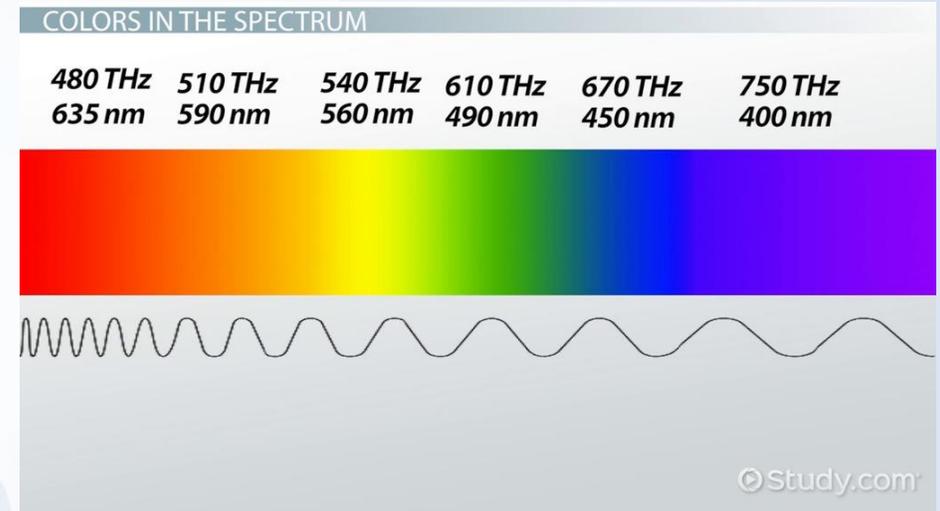
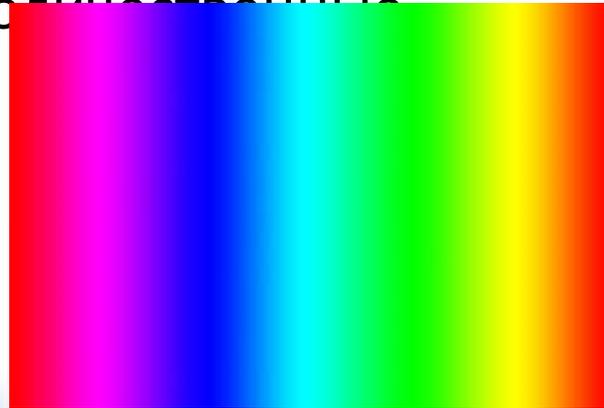
- **Описательная статистика**– методы упорядочивания, суммирования и презентации данных в информативной форме
- **Статистический вывод**– этот метод используется для получения выводов о совокупности на основе случайной выборки
- **Совокупность** – ряд лиц, предметов или часть расчетов из числа всех лиц, предметов или расчетов
- **Выборка** – часть сферы интересов



<https://mahritaharahap.wordpress.com/teaching-areas/inferential-statistics/>

Что такое данные?

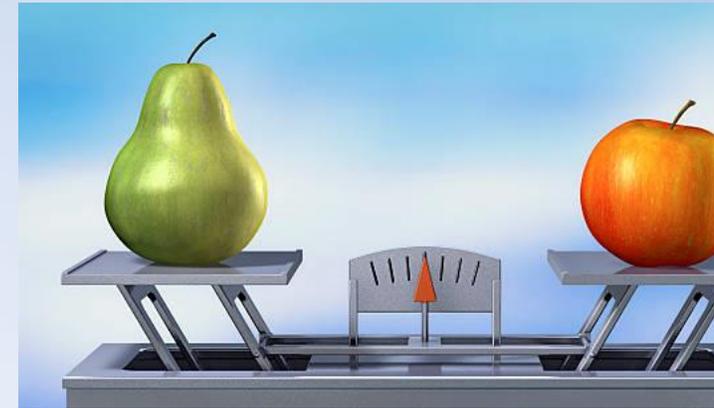
- Необработанная или неупорядоченная информация (такая как буквы, цифры или символы), которая относится к или представляет условия, идеи, предметы.
- Существуют разные виды данных
 - Непрерывные/дискретные
 - Аналоговые/цифровые
 - Качественные/количественные



<https://study.com/academy/lesson/the-nature-of-light-origin-spectrum-color-frequency.html>

Сбор данных

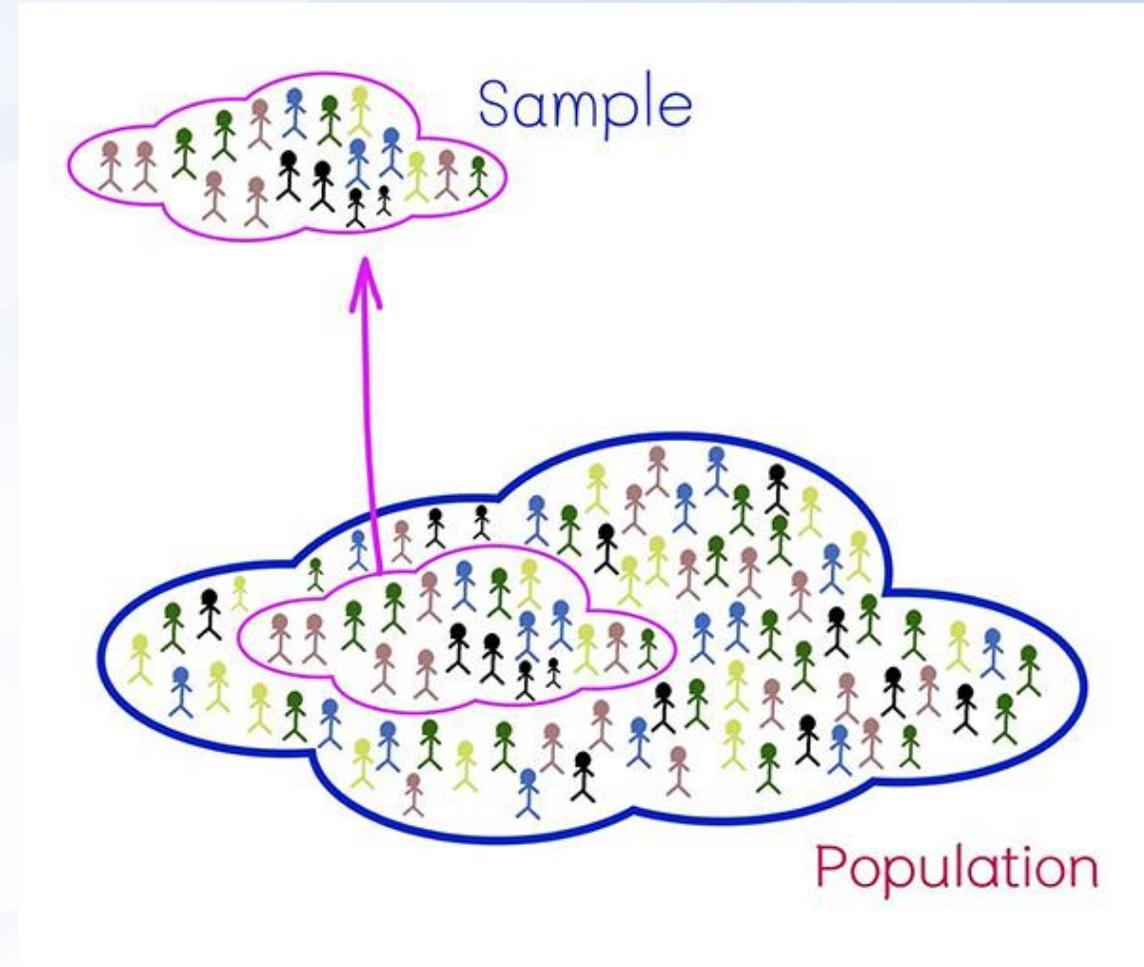
- **Систематический** сбор и упорядочивание данных
- Различные методы:
 - Исследования
 - Эксперимент
 - Наблюдение
- Научный метод сбора данных обуславливает логическую обоснованность данных и их систематическое упорядочивание для дальнейшего сравнения и анализа
- **Мы делаем это, чтобы ответить на ВОПРОС!** (а значит, вопрос должен быть задан)



Выборка

Можно использовать следующие методики выборки:

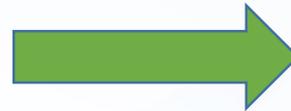
- **Случайная** (каждый представитель совокупности имеет равные шансы на участие)
- **Неслучайная**
- Построение выборки может привести к ошибкам выборки. К примеру, выборка может быть недостаточно большая, чтобы представить всю совокупность. Факторы, которые не относятся к процессу выборки, могут привести к ошибкам. Некачественные приборы (инструменты) для подсчётов могут привести к ошибке выборки.



Что такое вопрос?

- **Статистический вопрос:** вопрос, на который вы рассчитываете получить ряд ответов, и вы заинтересованы в тенденциях и обобщенности этих ответов.

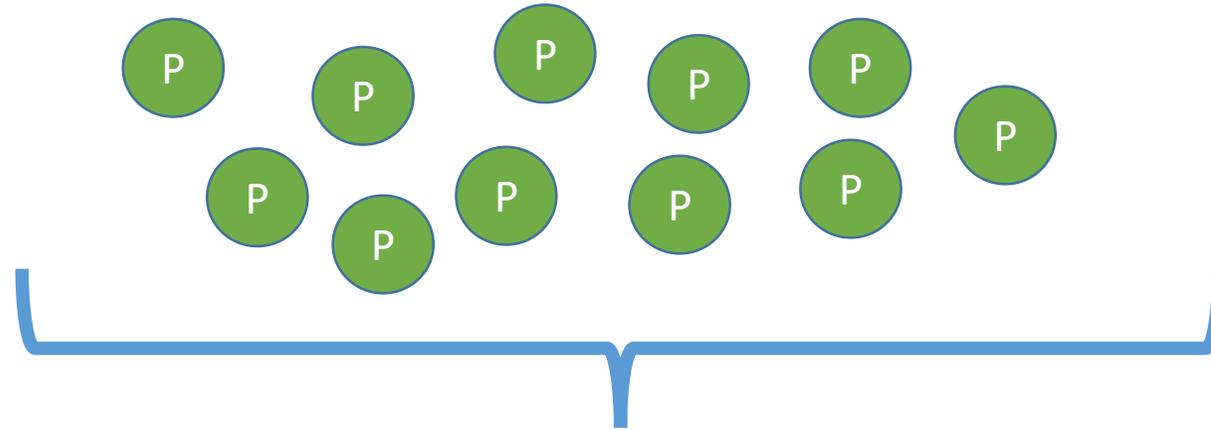
Как у вас сегодня дела?



Пожалуйста, оцените
свое настроение от 1 до
10

- **Количественные ответы.** Может не всегда. В таком случае используется кодирование.

Клинические исследования – это эксперимент

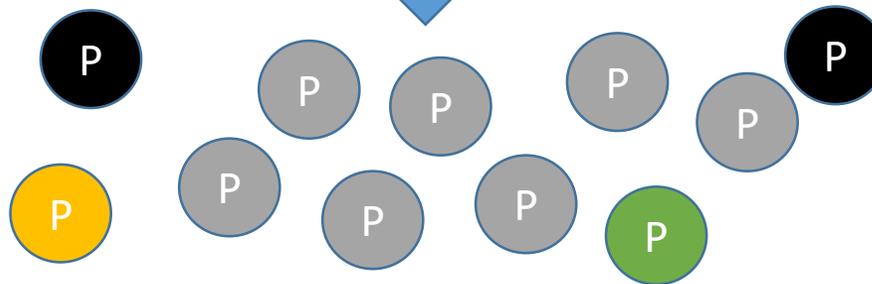


Что здесь происходит?

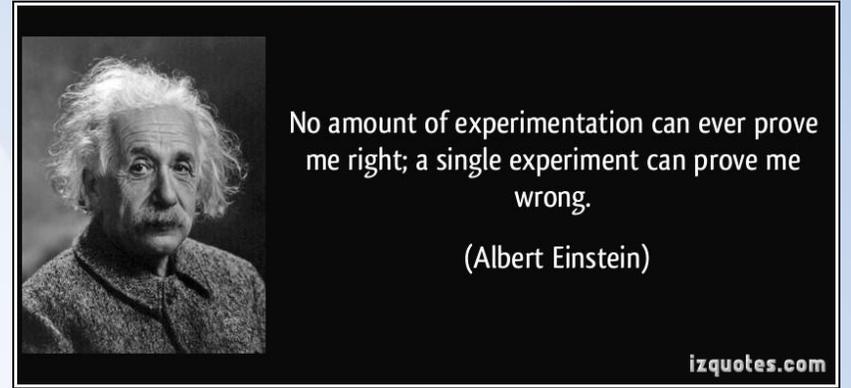


Клиническое исследование

Измерение
Статистика



Нулевая гипотеза



- Гипотеза = допущение, или ряд допущений, которые либо а) утверждают что-то на временной основе как часть научного следствия; либо б) подтверждают что-то крайне вероятное в свете установленных фактов.
- Вместо того, чтобы доказать гипотезу В, научный метод предполагает, что факты А – правдивы – что нет разницы между стандартами ухода и новым лечение. Это называется «нулевая» гипотеза. Ученые затем пытаются опровергнуть А.
- Нулевая гипотеза описывает текущее состояние или известный нынче факт, в который мы должны верить до тех пор, пока нет новых данных, полностью его опровергающих. Но если мы доказали так называемую «альтернативную гипотезу» (она так называется так в противоположность к «нулевой гипотезе»), тогда соответственно «обвиняемый – виновен».
- Процесс доказательства того, что нулевая гипотеза – ошибочна, более оправдан и более прост, нежели попытки доказать правоту альтернативной гипотезы.

Нулевая гипотеза

www.majordifferences.com



H_1 : Application of bio-fertilizer 'x' increase plant growth.

Alternative hypothesis

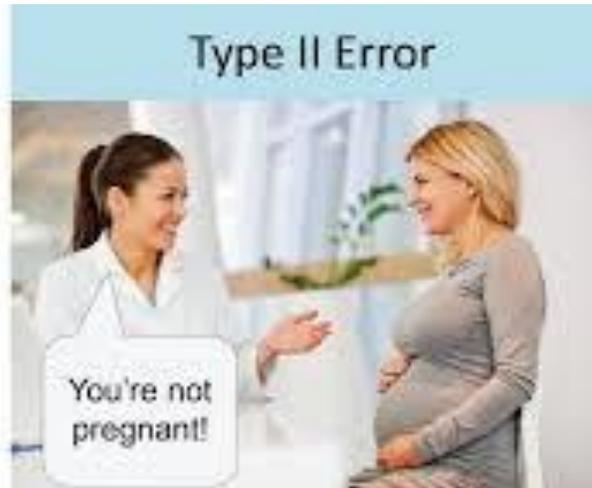
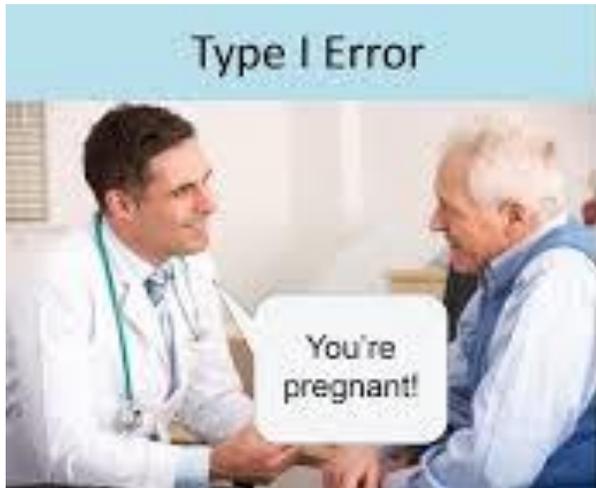
✓ The alternative hypothesis is a hypothesis which the researcher tries to prove.

H_0 : Application of bio-fertilizer 'x' do not increase plant growth.

Null hypothesis

✓ The null hypothesis is a hypothesis which the researcher tries to disprove, or nullify.

Виды ошибок: I и ii



		The Truth (Based on Entire Population)	
		Nothing Is There (H_0 Is True)	Something Is There (H_0 Is False)
Your Conclusion (Based on Your Sample)	I Don't See Anything (Nonsignificant)	Right!	Wrong (Type II Error)
	I See Something (Significant)	Wrong (Type I Error)	Right!

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

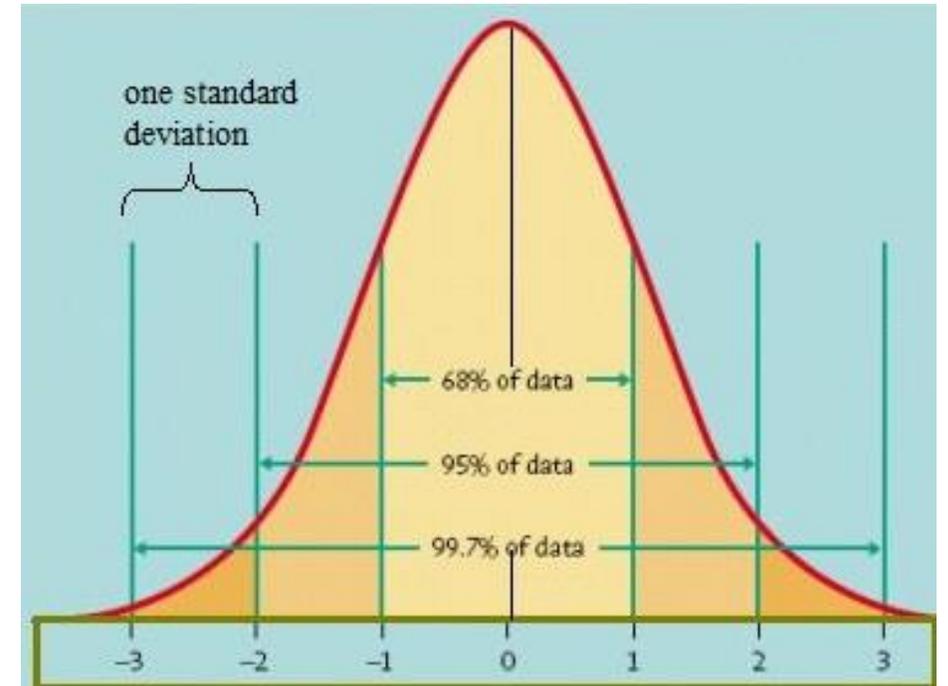
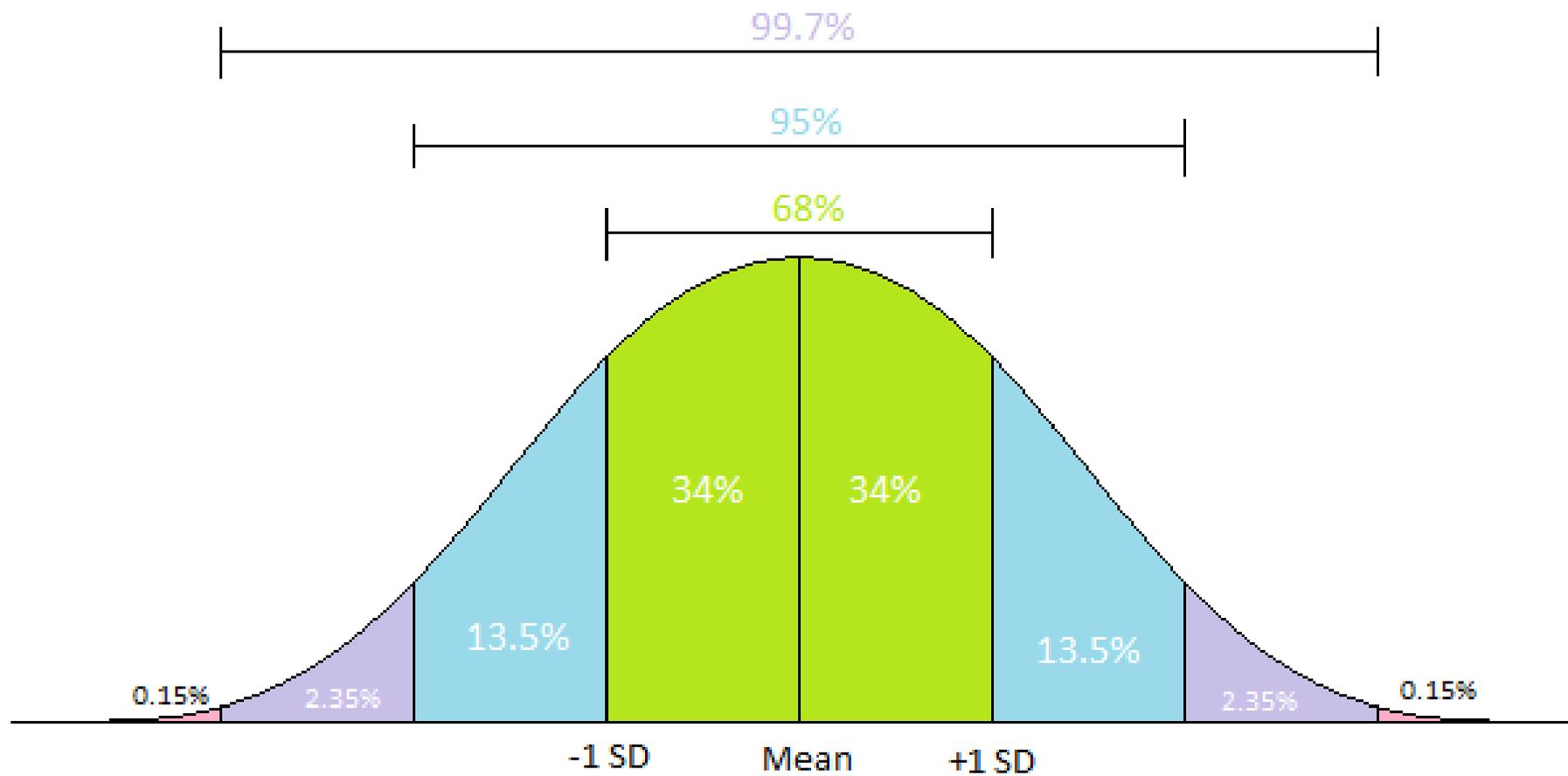


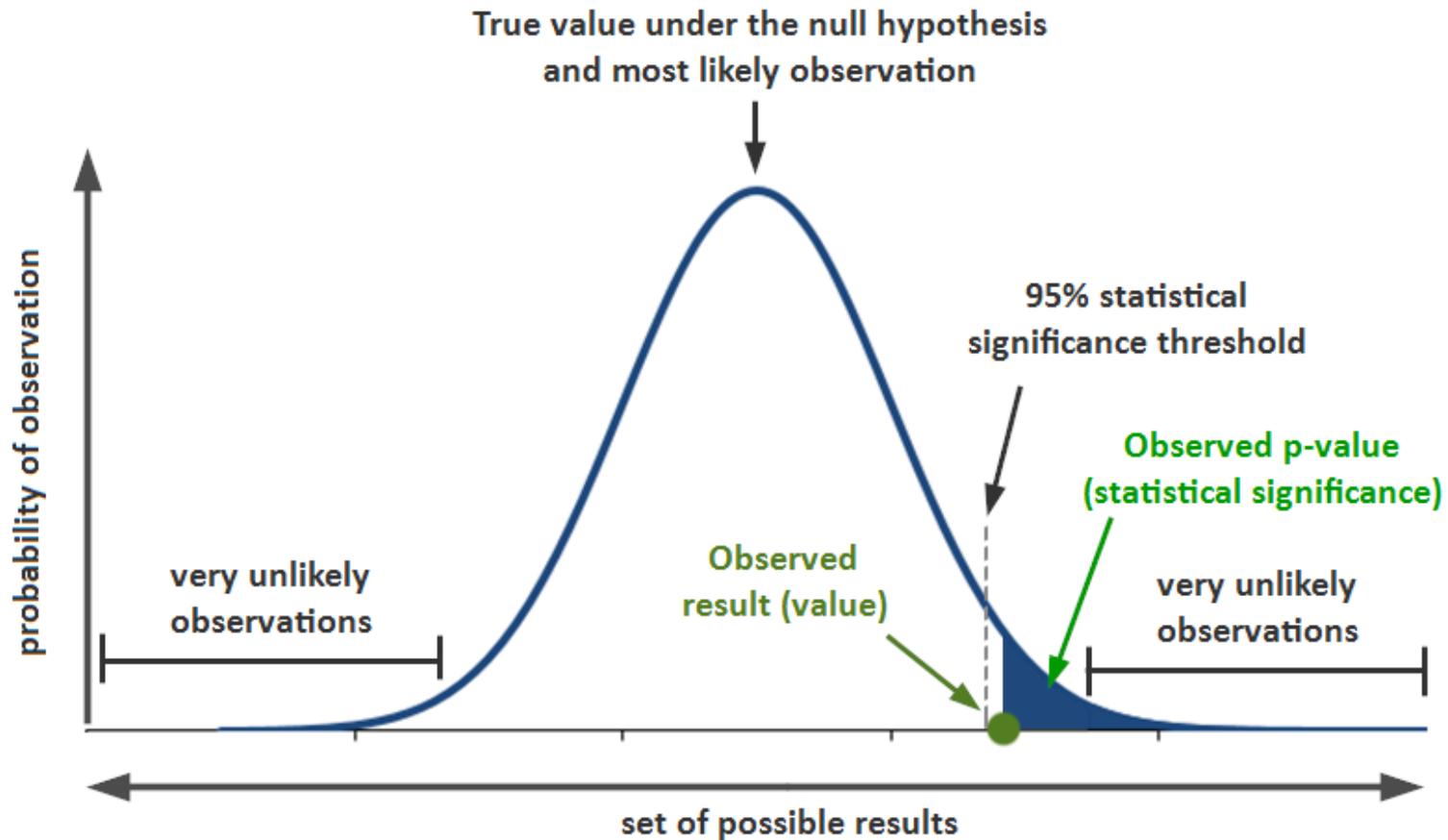
Photo courtesy of Judy Davidson, DNP, RN

СТАНДАРТНОЕ ОТКЛОНЕНИЕ(σ)

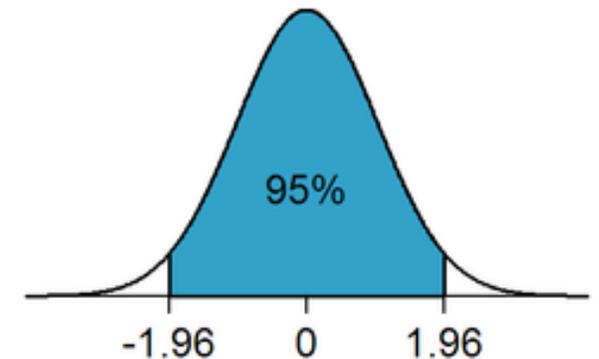


Статистическая значимость

Probability & Statistical Significance Explained



У нас есть результат. Но значимый ли он
И насколько можем мы в нем быть уверенными



Об этом говорит величина p .
Условная, но полезная.
Она говорит о том, случайны или нет

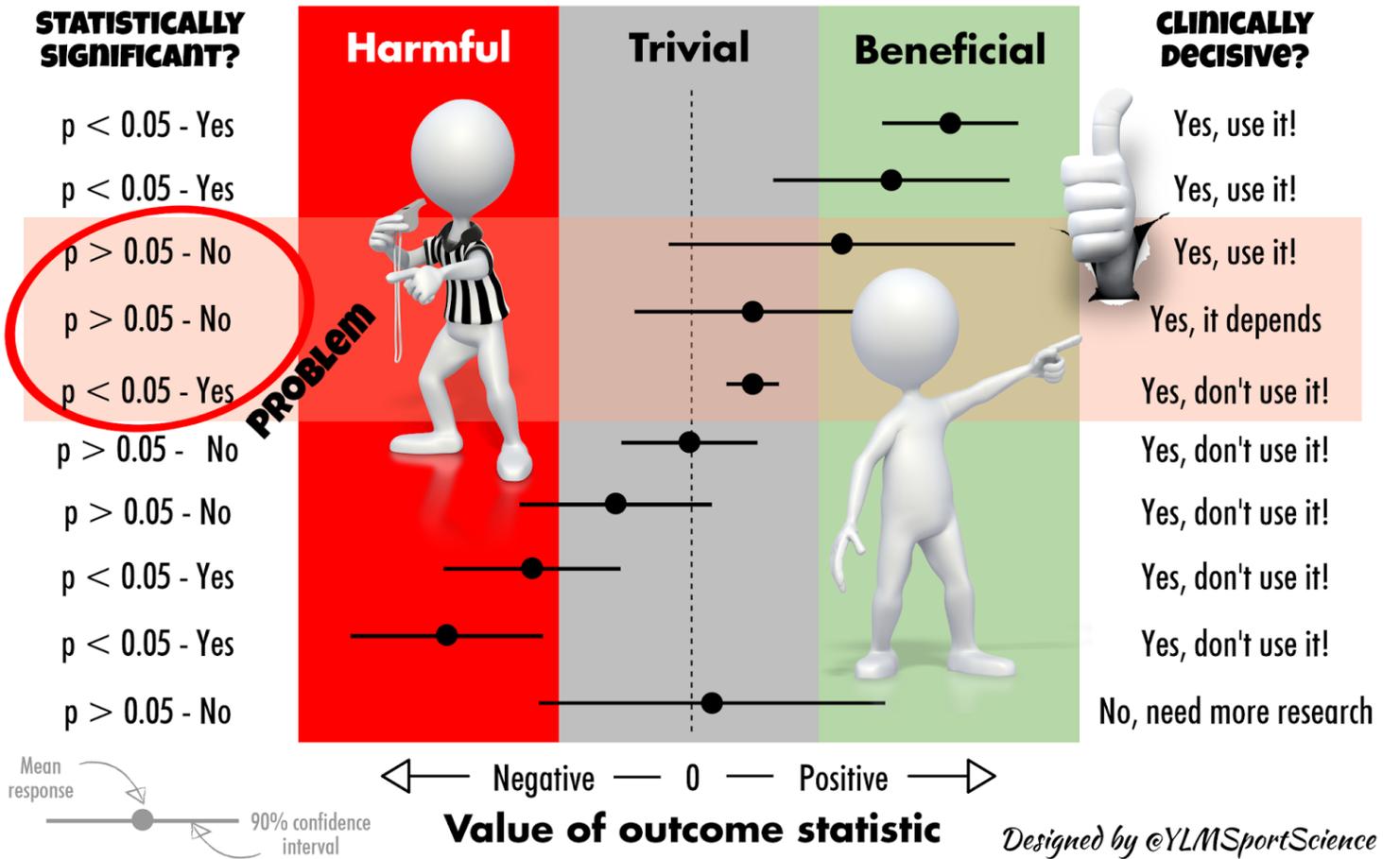
Клиническая значимость

STATISTICS

MAKING INFERENCES: CLINICAL VS STATISTICAL SIGNIFICANCE

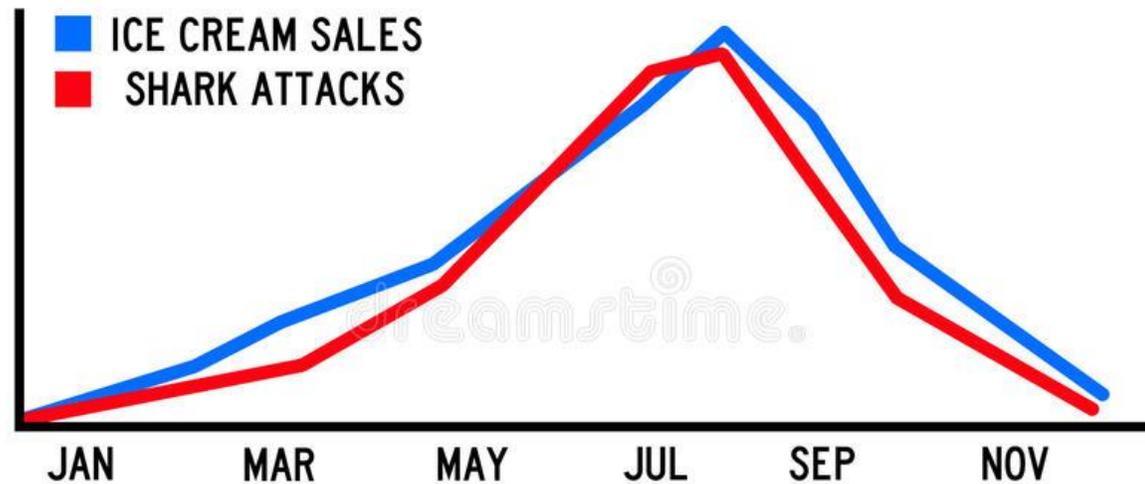
Reference: by Batterham & Hopkins, IJSP 2006

У нас есть результат. Он клинически значим? Мы можем его для чего-то использовать?



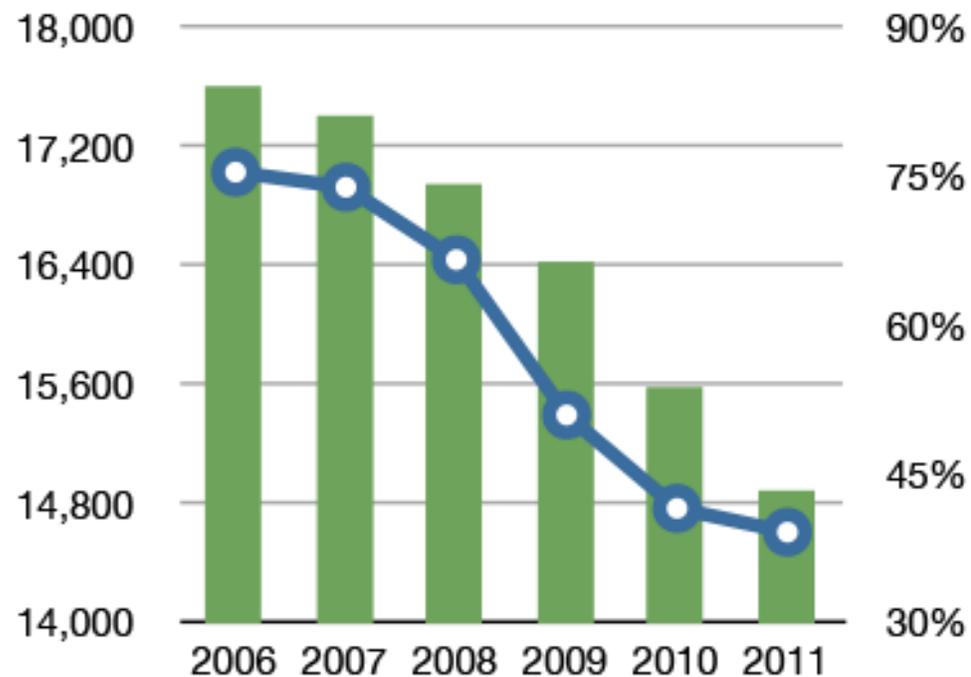
Взаимосвязь или причина?

CORRELATION IS NOT CAUSATION!



Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

Internet Explorer vs Murder Rate



○ Murders in US ■ Internet Explorer Market Share

Более продвинутые концепты

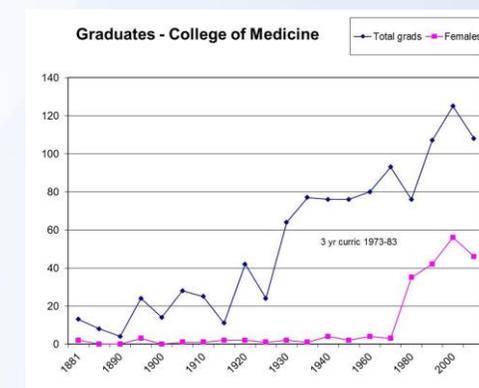
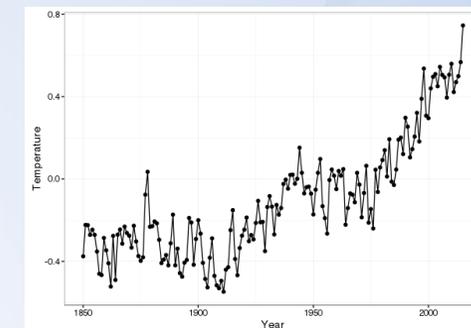
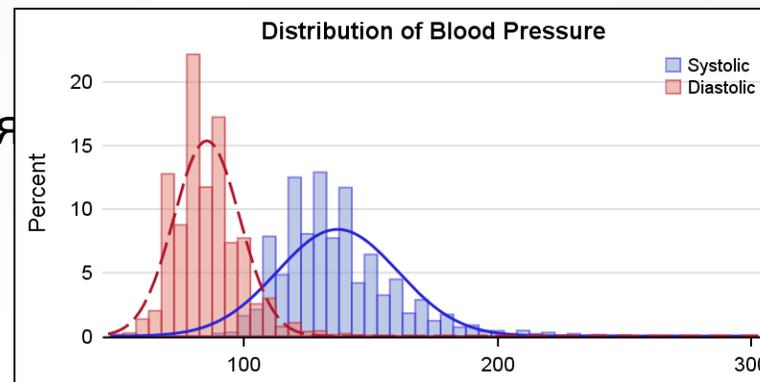
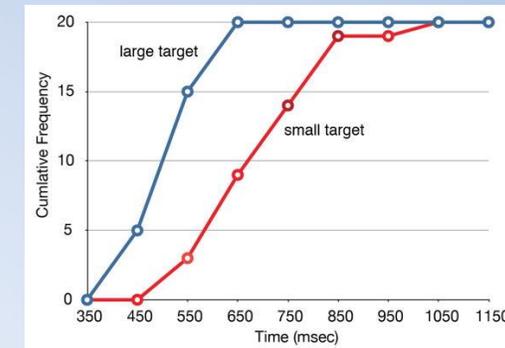
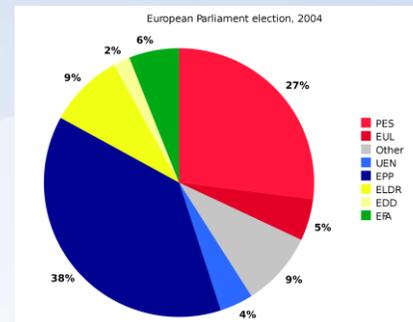
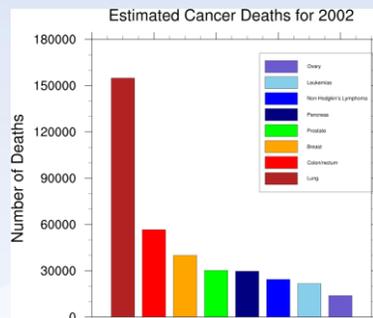
- Заняться таблицами?
- Прочитать графики
- Всплеск значений (наблюдения за пределами значений)
- Кривые Каплана-Мейера
- Соотношение рисков

Таблицы

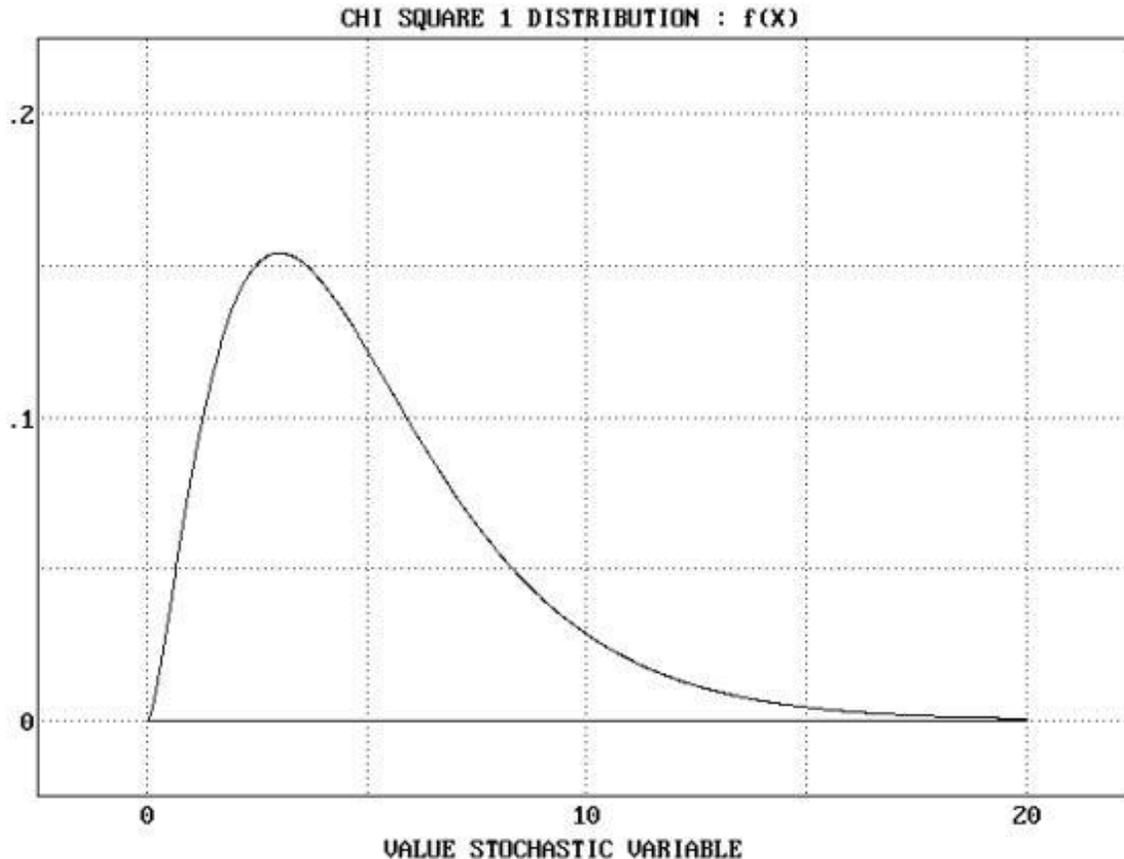
- Для них есть приложение! – Используйте такие программы, как Excel или SPSS или другие приложения для статистических данных
- Обратите внимание на анализ и резюме
- Держите таблицы под рукой и проверяйте с их помощью данные

Диаграммы и графики

- Частотное распределение помогает кратко представить изменяющиеся данные
- Картинки обычно более эффективны для отображения больших массивов данных
 - Гексаграммы
 - Частотные графики
 - Огива (интегральная кривая)
 - Круглые диаграммы
 - Столбиковые диаграммы
 - Временные ряды
 - Кривая Каплана-Мейера
 - И т.д.

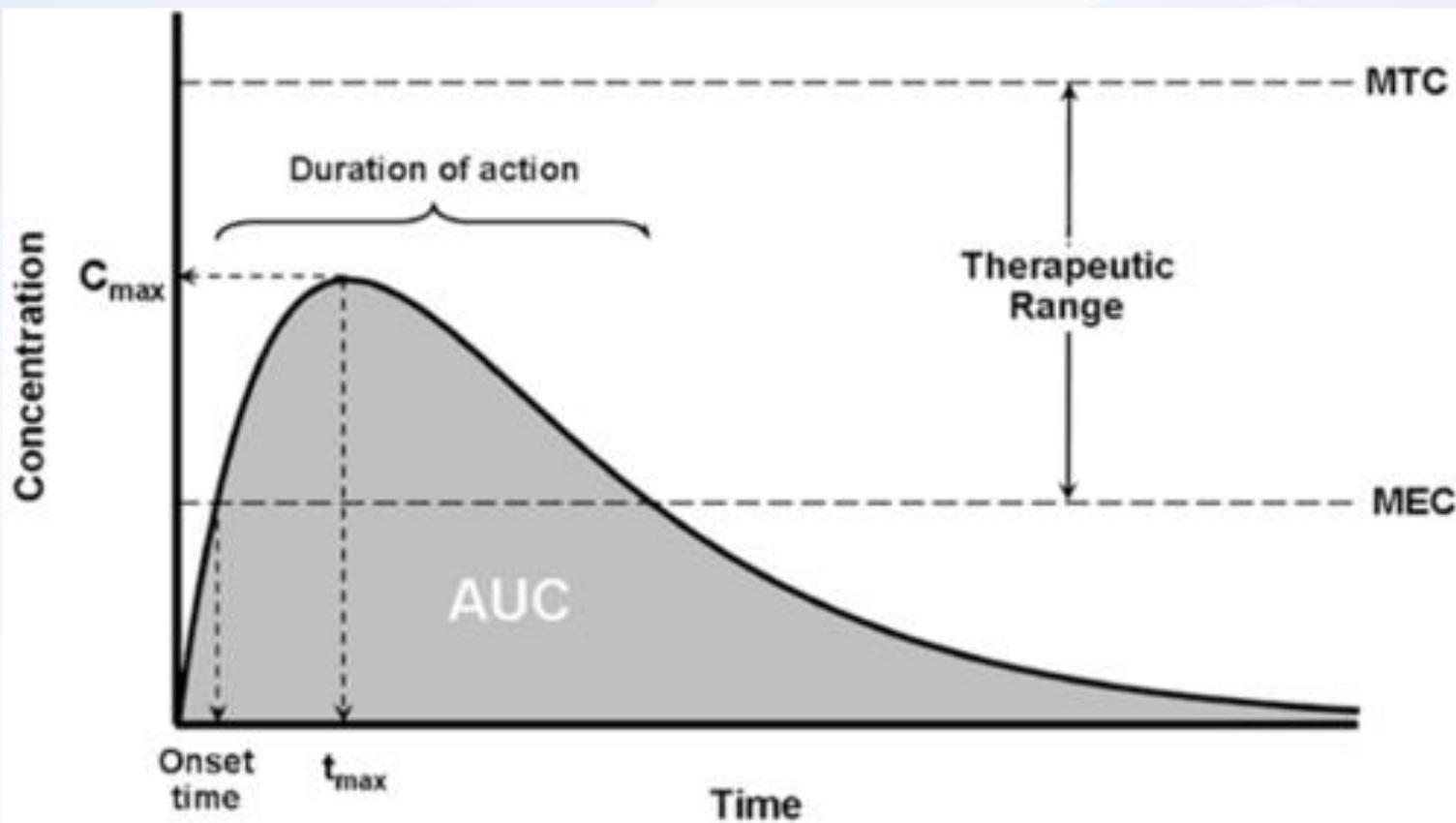
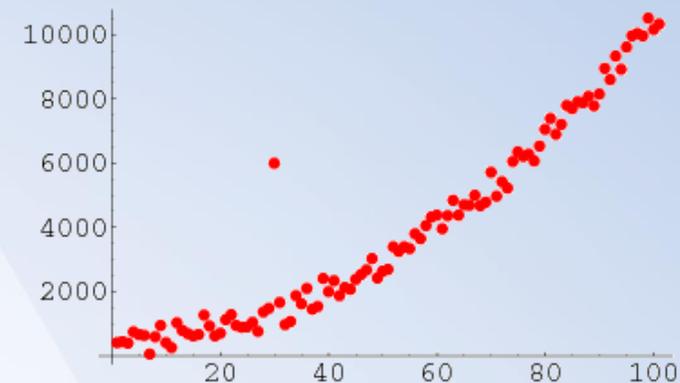


Хи-квадрат (χ^2)



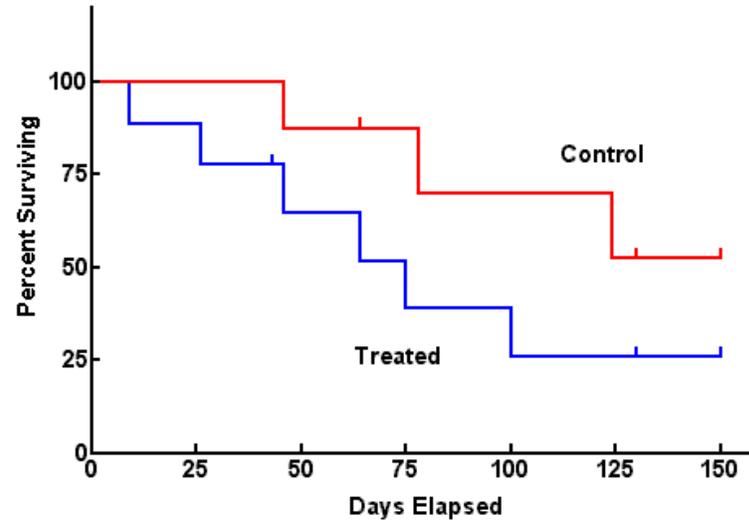
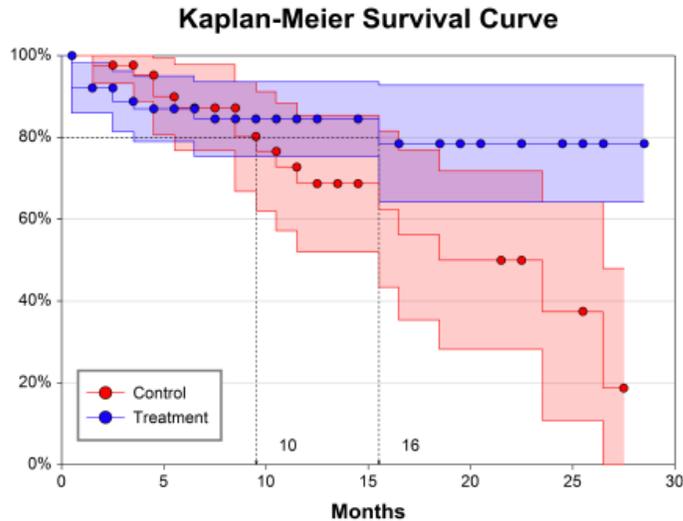
- Любое число в квадрате – положительное число
- Поэтому площадь под кривой начинается с нуля и стремится к бесконечности
- Чтобы быть статистически значимой, необходимо превышать уровень в 5% ($\alpha = .05$)
- Сравнивает наявную частоту с тем, что ожидали

Всплеск значений и площадь под кривой



Посмотрите на
площадь под кривой
Но также
посмотрите на
всплеск значений!

Кривая Каплана-Мейера



- Отрезок времени от ввода данных до окончания болезни, применим для лечения и контрольной группы
- Исходя из этой кривой, мы видим:
 - Медиану времени (время, за которое 50% вопросов решены)
 - Среднее время (среднее время для решения вопроса)
- Позволяет сравнивать информацию о пациентах на протяжении всего исследования и предоставляет информацию о пациентах, которые больше не участвуют в исследовании и о них ничего не известно
- Может применяться и в других сферах, таких как инженерия (изначально Каплан работал над вакуумными трубками)

- Независимо друг от друга Эдвард Каплан и Пол Мейер в 1958 году сделали открытие, над которым потом совместно работали
- Оценивает уровень выживаемости пациентов на основе данных двух групп – пациентов, которые принимают участие в исследовании и тех, которые прекратили свое участие (цензурированные)
- Учитывает тот факт, что некоторые пациенты могли умереть во время исследований, в то время как другие пережили исследования
- Использует информацию о тех, кто умер и тех, кто выжил, чтобы высчитать пропорцию пациентов, которые живы в любое отдельно взятое время исследования
- Когда выборка достаточно велика, приближается к настоящей функции выживания для совокупности

Отношение рисков

$$\text{Hazard ratio} = \frac{\text{hazard in exposed group}}{\text{hazard in unexposed group}}$$

- Отношение рисков это условие при котором происходит определенное событие
- Отношение рисков = уровень риска лечения/уровень риска плацебо, то есть отношение конкретного события, которое происходит в группе, которая получает лечение, к событию, которое происходит в группе, получающей плацебо
- Необходимость интерпретации отношения рисков через призму величины времени
- Отображает отношения времени «выживания» к событию
- Не указывает то, как быстро что-то происходит
- Широко используется для презентации данных клинических исследований (не одно и тоже с коэффициентом относительного риска)
- Показывает в числовом выражении «вероятность выигрыша», а не условия победы (смотри серединное значение)
- Отношения рисков 1 = равный уровень события между двумя группами
- Отношения рисков 2 = в два раза больше пациентов из активной группы будут участвовать в событии в сравнении с контролем в следующем отрезке времени
- Отношение рисков 0.5 = в два раза меньше пациентов из активной группы участвуют в событии в сравнении с контролем в следующем отрезке времени

Дополнительные источники информации

- Вебинар Bettina Ryll's о том, как правильно понимать науку:
https://youtu.be/H_V0tUBjBBM
- Курс по статистике на YouTube:
https://youtu.be/zouPoc49xbk?list=PL8dPuuaLjXtNM_Y-bUAhbISAdWRnmBUcr
- Книга по основам статистики:
<https://www.mv.helsinki.fi/home/jmisotal/BoS.pdf>