



Статистическая обработка результатов

**медико-биологических
исследований**

Зачем нам статистика?...

Статистика в медицине является одним из инструментов анализа экспериментальных данных и клинических наблюдений, а также языком, с помощью которого сообщаются полученные математические результаты.



На каком принципе базируется статистика?

верное для случайной выборки верно и для генеральной совокупности (популяции), из которой эта выборка получена

Статистика предполагает, что...

выборка репрезентативна по отношению к изучаемой популяции, т.е. достаточно адекватно отражает все возможные аспекты изучаемого состояния или заболевания в популяции

Чтобы статистика не стала «ложью»...

необходимо чётко формулировать цели и строго соблюдать критерии включения и исключения как в исследование, так и в статистический анализ



ОПРЕДЕЛЕНИЕ СТАТИСТИКИ

Математический аппарат, позволяющий осуществить перевод количества в качество, т.е. на основании полученных данных найти у обследованных определенное свойство.

Задачи статистики

- I. Оценка статистической значимости выдвигаемой в процессе исследования гипотезы – сопряжено с планированием исследования.**
- II. Получение качественных и количественных характеристик исследуемого свойства**
- III. Оценка статистической значимости проведенных вычислений**

1 этап

Определение выборки

- **Выборка** – часть популяции, полученная путем отбора по одному или нескольким признакам.
- **Популяция** – совокупность всех больных, страдающих определенным заболеванием.

1 этап

Определение выборки

Популяционное значение параметра получить невозможно (исключения составляют случаи, когда исследование проводится на группе, которая включает всех членов популяции).

Популяционное значение параметра **можно оценить по выборке**. Точность такой оценки зависит от метода измерения (ошибки измерения), объема и репрезентативности выборки (ошибка выборки) и биологической вариации.

2 этап

Определение типа данных

- **Качественные (категории: целая разрывная последовательность) –**
 - бинарные:** пол (муж-1 жен-2), наличие-отсутствие признака (нет-0 да-1),
 - упорядоченные:** выраженность признака (нет-0 слабо-1 умеренно-2 сильно-3),
 - неупорядоченные:** семейное положение (женат-1 разведен-2 псих-одиночка-3)

2 этап

Определение типа данных

- **Количественные (числа: неразрывная последовательность) –**
 - дискретные:** количество доз препарата, количество нежелательных явлений (не могут быть выражены в нецелых единицах)
 - непрерывные:** возраст, давление, креатинин (могут быть выражены в нецелых единицах)

2 этап

Определение типа данных

- **Смешанный (двойной) тип – возможность выражения в виде качественных и количественных величин -**
Функциональный класс SN

3 этап

Определение типа распределения количественных данных

По кривой Гаусса:

- Нормальное распределение (параметрическое)
- Ненормальное распределение (непараметрическое, ранговое)

3 этап

Определение типа распределения количественных данных

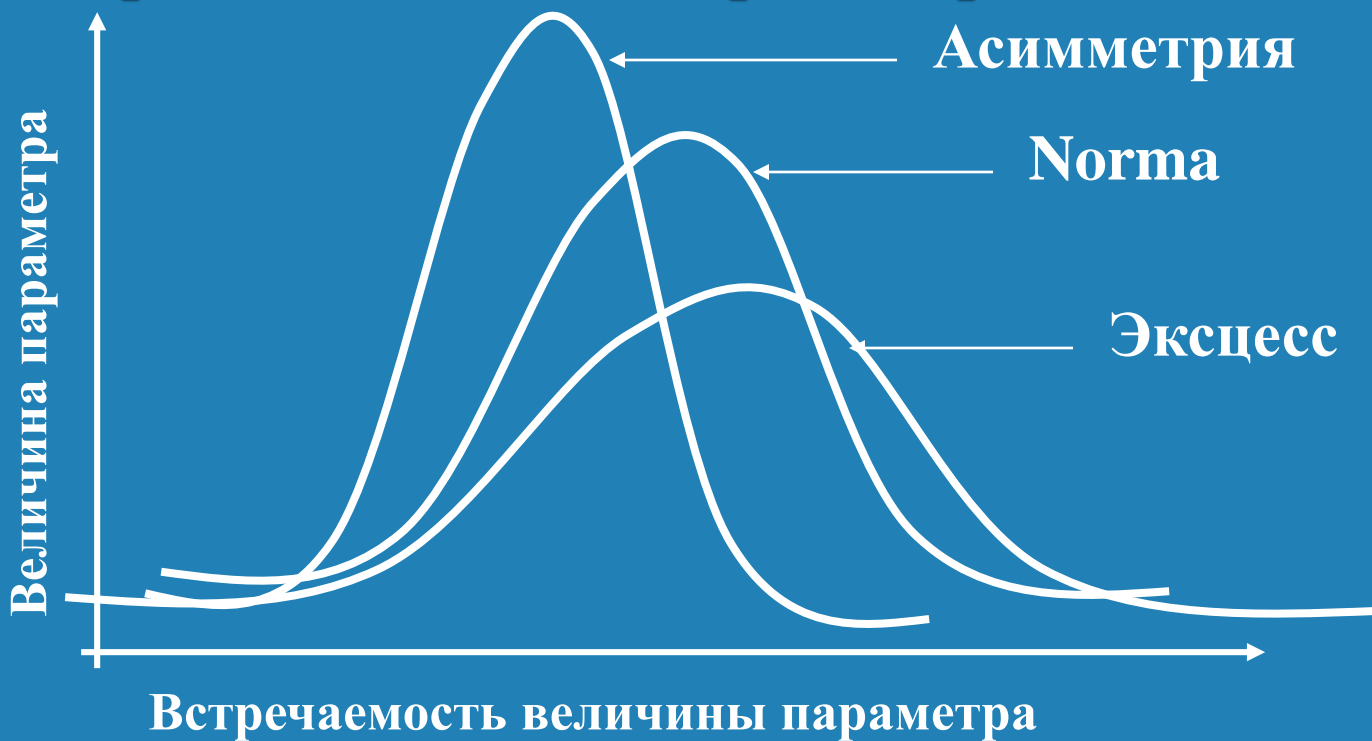
Критерии определения нормальности распределения

- По критерию Колмогорова-Смирнова
- По величине асимметрии и эксцесса

3 этап

Определение типа распределения количественных данных

Критерии определения нормальности распределения



3 этап

Определение типа распределения количественных данных

Критерии определения нормальности распределения

- **Критерий Колмогорова-Смирнова** характеризует достоверность отличия исследуемого распределения от теоретически нормального распределения, при нормальном распределении больше $>0,05$
- **Асимметрия** характеризует степень несимметричности распределения относительно его среднего. При нормальном распределении не должен превышать среднеквадратичное отклонение в 2 и более раза.
- **Эксцесс** характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением. При нормальном распределении не должен превышать среднеквадратичное отклонение в 2 и более раза.

4 этап

Представление данных

Описательная статистика (дескриптивная статистика):

**предоставление сжатой и
концентрированной характеристики
изучаемого явления в числовом и
графическом виде**

4 этап

Представление данных

Параметры качественных данных:

- **Количество значений (для пола-2)**
- **Частота (количество одинаковых значений, count) – к примеру мужчин-4, женщин-6**
- **Доля в % от общего числа значений – к примеру мужчин-40%, женщин-60% (при $n < 100$ без десятичных знаков)**

4 этап

Представление данных

Меры центральной тенденции количественных данных:

- Среднее арифметическое (Mean)
- Медиана (Median)
- Мода (Moda)

А также указываются -

- Количество значений (N, count)
- Минимальное значение (Min)
- Максимальное значение (Max)

4 этап

Представление данных

Среднее арифметическое (M)

это показатель центральной тенденции, полученный делением суммы всех значений данных на число этих данных.

Среднее арифметическое используется для представления количественных переменных с нормальным распределением.

4 этап

Представление данных

Ширина распределения средней

« $M \pm S$ », или в зарубежной традиции $M (S)$, где M – среднее, а S – стандартное отклонение (Standard Deviation, равное корню квадратному из дисперсии).

« $M \pm m$ », где m - стандартная ошибка среднего (Standard Error of Mean), определяемая следующим образом: $m = s / \sqrt{n}$.

4 этап

Представление данных

Особенности представления медицинских лабораторных данных:

Важное значение имеет не мера центральной тенденции, а диапазон, в который укладывается большинство значений исследуемого признака, т.е. ширина распределения.

M (95% ДИ: 2,5%-97,5%)

M (90% ДИ: 5%-95%)

4 этап

Представление данных

Медиана (Me, Md)

это показатель центральной тенденции, который соответствует центральному значению в последовательном ряду всех полученных значений или среднему значению наиболее часто встречающихся значений выборки.

4 этап

Представление данных

Ширина распределения медианы – КВАНТИЛИ:

частота попадания значений переменной в определённые интервалы

Используется разделение на 10 (по 10%) или на 4 интервала (25%, 50%, 75%).

4 квантиля – это квартили, наиболее распространённый метод представления:
Me (Q 25%-75%)

4 этап

Представление данных

Мода (Mo)

это наиболее частое значение в выборке, или среднее значение класса с наибольшей частотой.

как центральная тенденция используется чаще всего для того, чтобы дать общее представление о распределении. В некоторых случаях у распределения могут быть две моды - бимодальное распределение, что указывает на наличие двух относительно самостоятельных групп и требует дополнительного анализа характеристик выборки.

5 этап

Определение метода анализа и сравнения данных (достоверности)

- **Непараметрический анализ:**
 - ✓ качественные данные,
 - ✓ количественные дискретные данные,
 - ✓ количественные непрерывные данные с ненормальным типом распределения
- **Параметрический анализ:**
 - ✓ количественные непрерывные данные с нормальным типом распределения

5 этап

Определение метода анализа и сравнения данных (достоверности)

- **Непараметрический анализ:**

- ✓ Критерий хи (кси)-квадрат, таблица сопряженности,
- ✓ Знаковый критерий Вилкоксона,
- ✓ Ранговый критерий Вилкоксона или Мана-Уитни (зависимые/независимые выборки)
- ✓ Доверительные интервалы

- **Параметрический анализ:**

- ✓ Критерий Стьюдента, Т-тест (независимые/зависимые выборки)
- ✓ Доверительные интервалы