

Статистическая проверка гипотез



Квантиль и квартиль. Дизайн эксперимента и статистические гипотезы о данных. Ошибки 1-го и 2-го рода. Статистическая значимость p -value. A/B тестирование. Тесты на нормальность. Корреляционные тесты. Проверка гипотезы t -критерия Стьюдента. Одновыборочный t -критерий. Двухвыборочный t -критерий для независимых выборок. Множественный тест (ANOVA).

Юстина Иванова

Специалист по Анализу Данных



Инженер-программист МГТУ им. Баумана

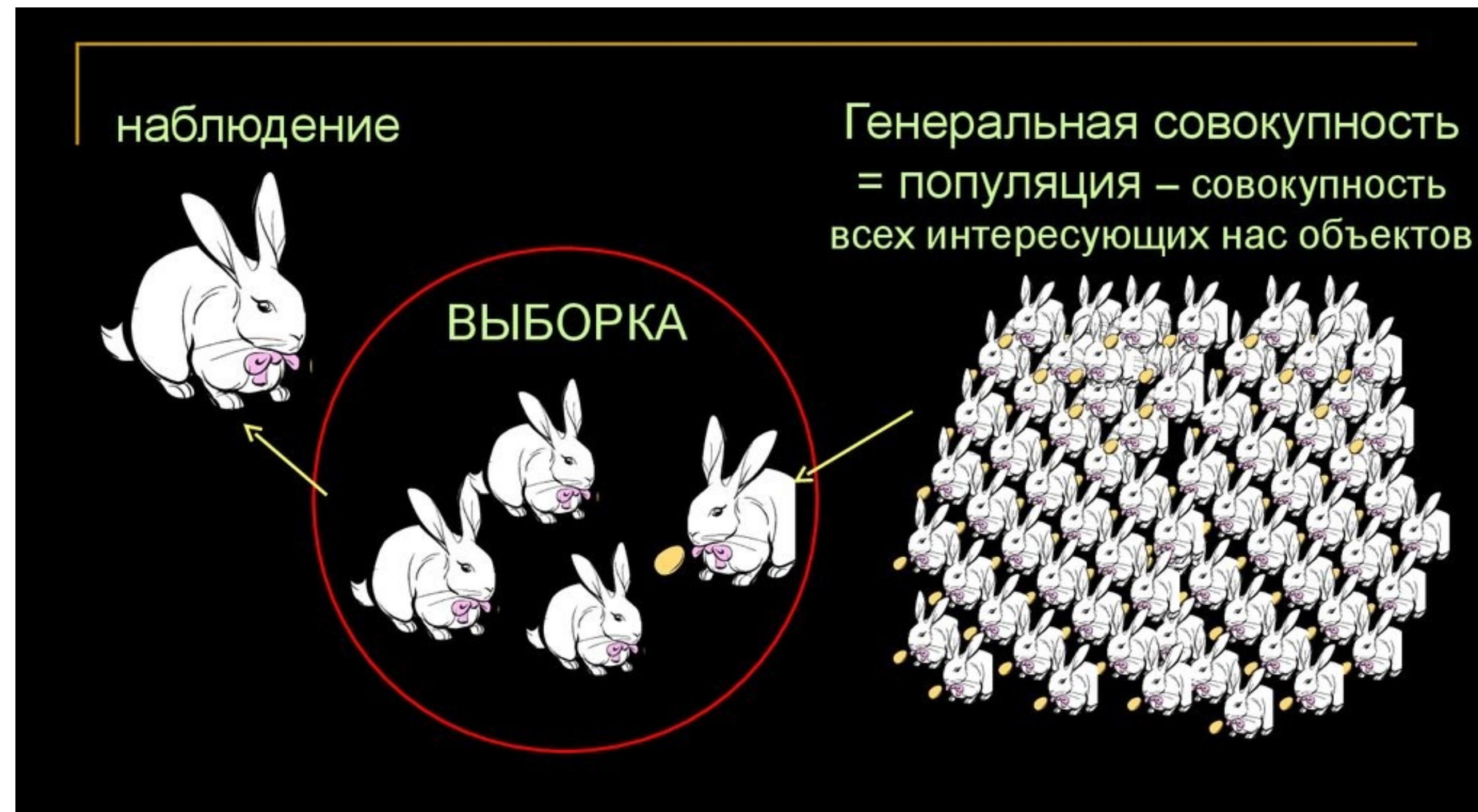
Master of Science in Artificial Intelligence
University of Southampton

Специалист по анализу данных
в компании ОЦРВ.

Юстина Иванова
студент-аспирант
University of Bolzano

Генеральная совокупность и выборка.

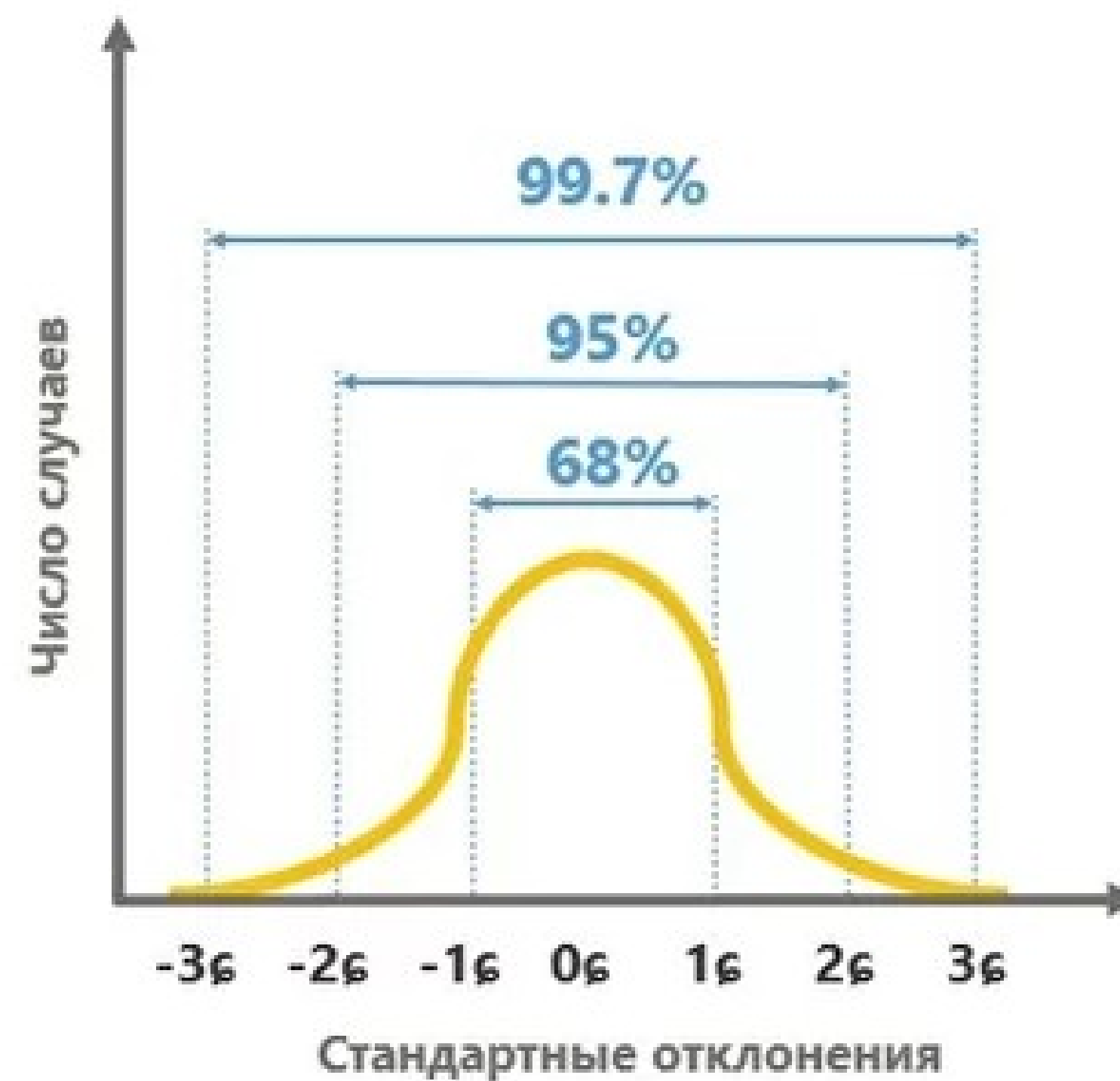
Генеральная совокупность — множество всех объектов, обладающих изучаемым признаком.



На основе свойств выборки делаем заключение о свойствах генеральной совокупности.

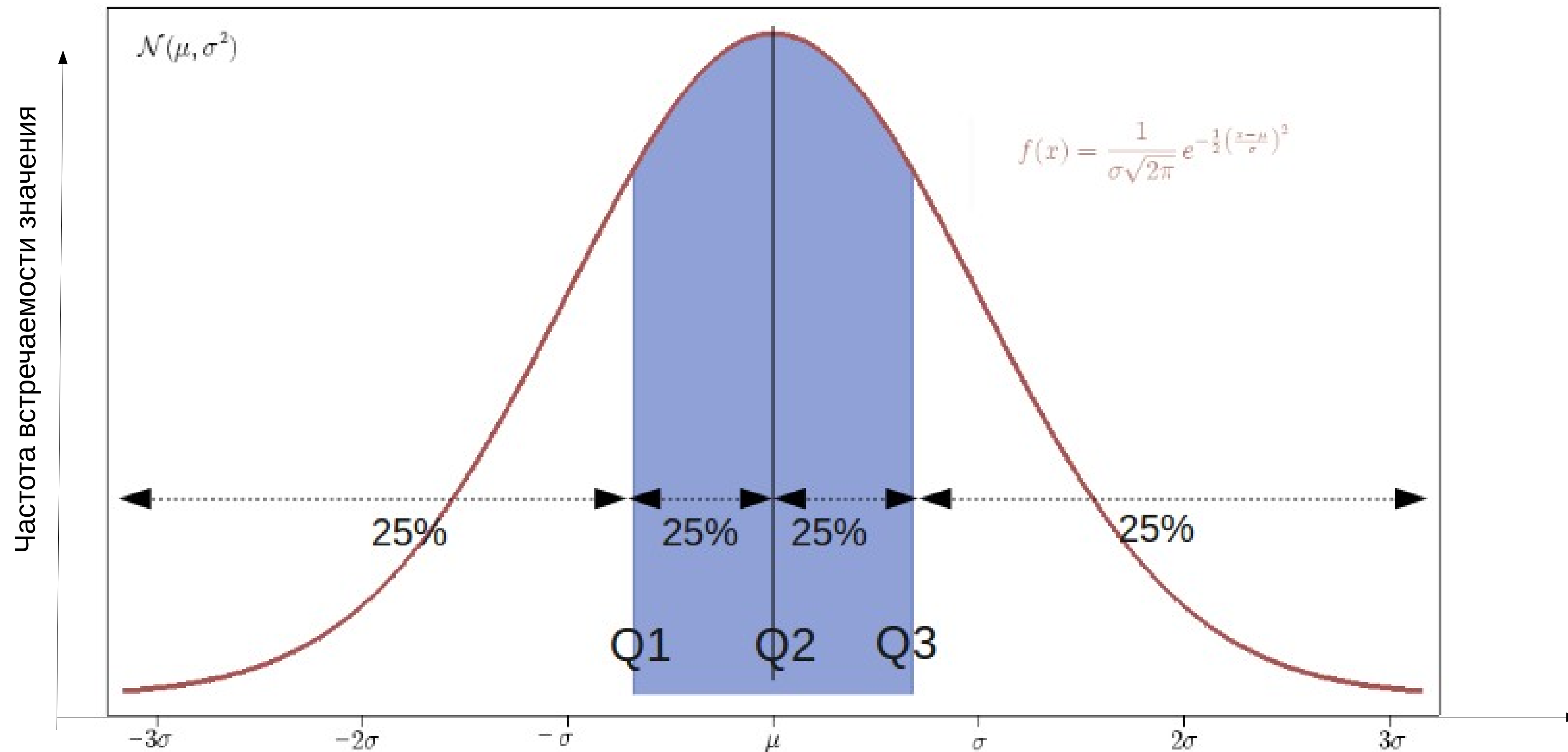
Доверительный интервал

Доверительный интервал — интервал, в котором лежит 95% данных.

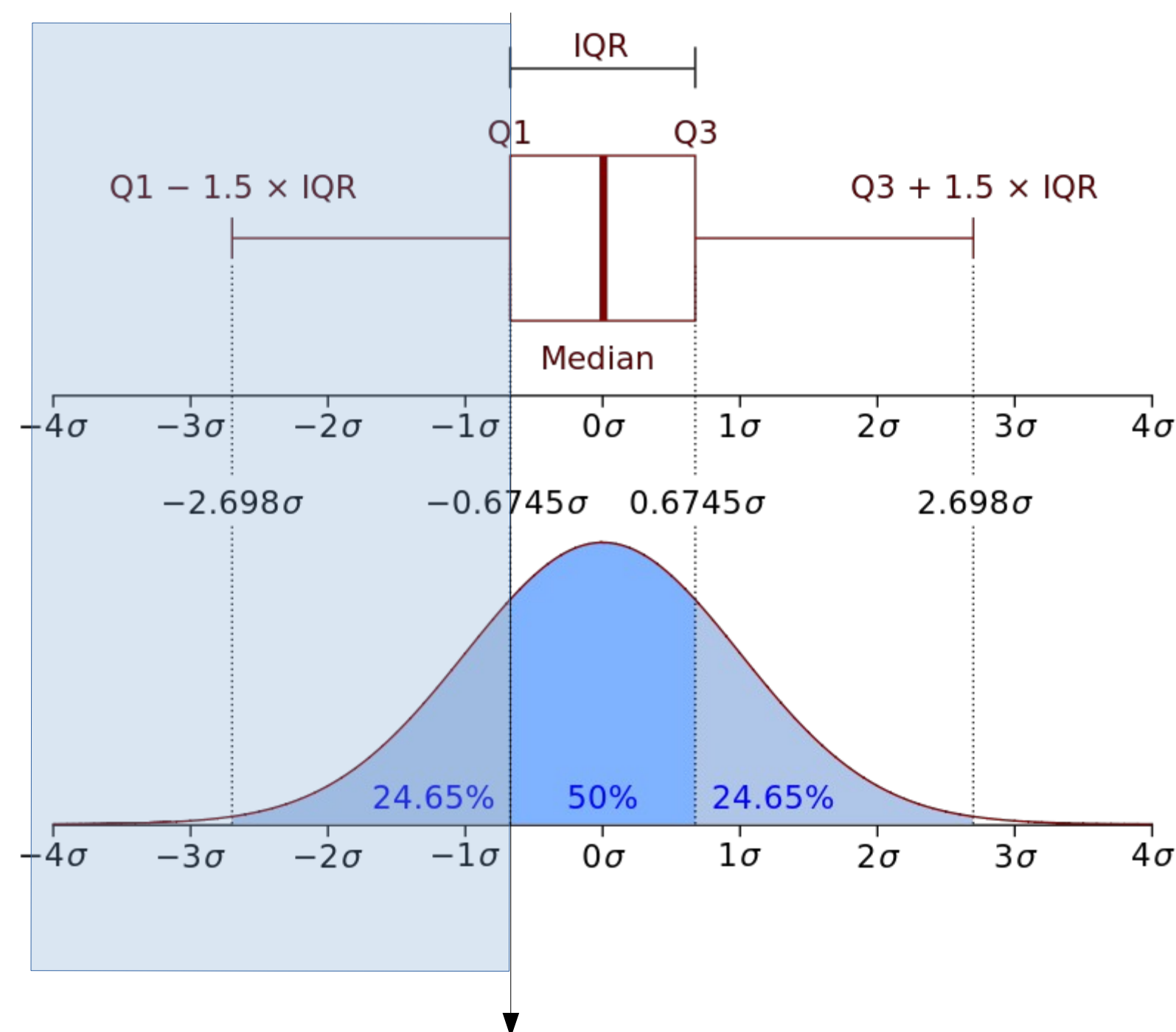


Квантиль

значение, которое заданная случайная величина не превышает с фиксированной вероятностью



Значение квантиля

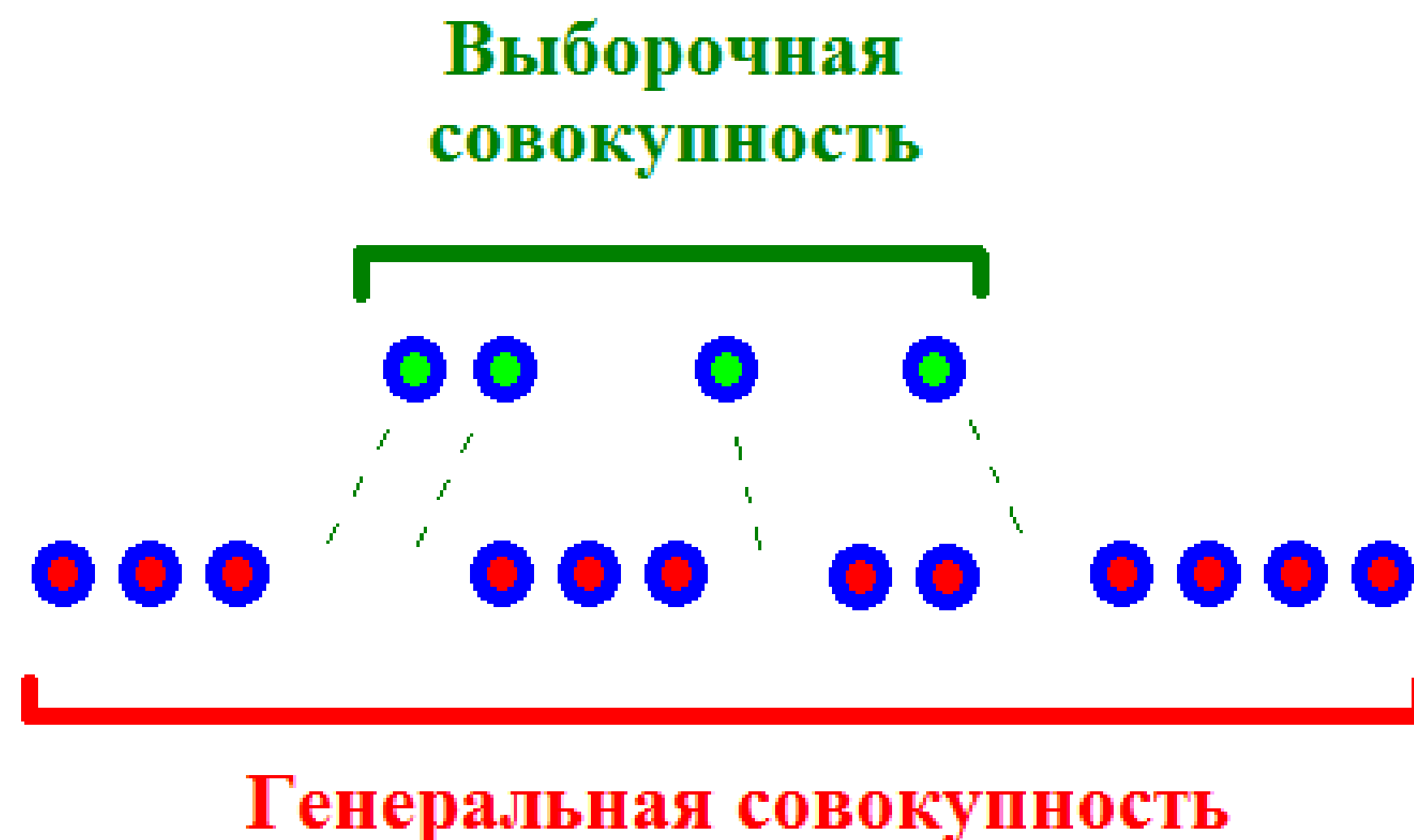


1ый квантиль

Предоставляют важную информацию о структуре **вариационного** (колонку таблицы) ряда признака. Вместе с медианой они делят вариационный ряд на 4 равные части. Квантилей две, их обозначают символами Q, верхняя и нижняя квантиль. 25% значений меньше, чем нижняя квантиль, 75% значений меньше, чем верхняя квантиль.

Статистические гипотезы о данных

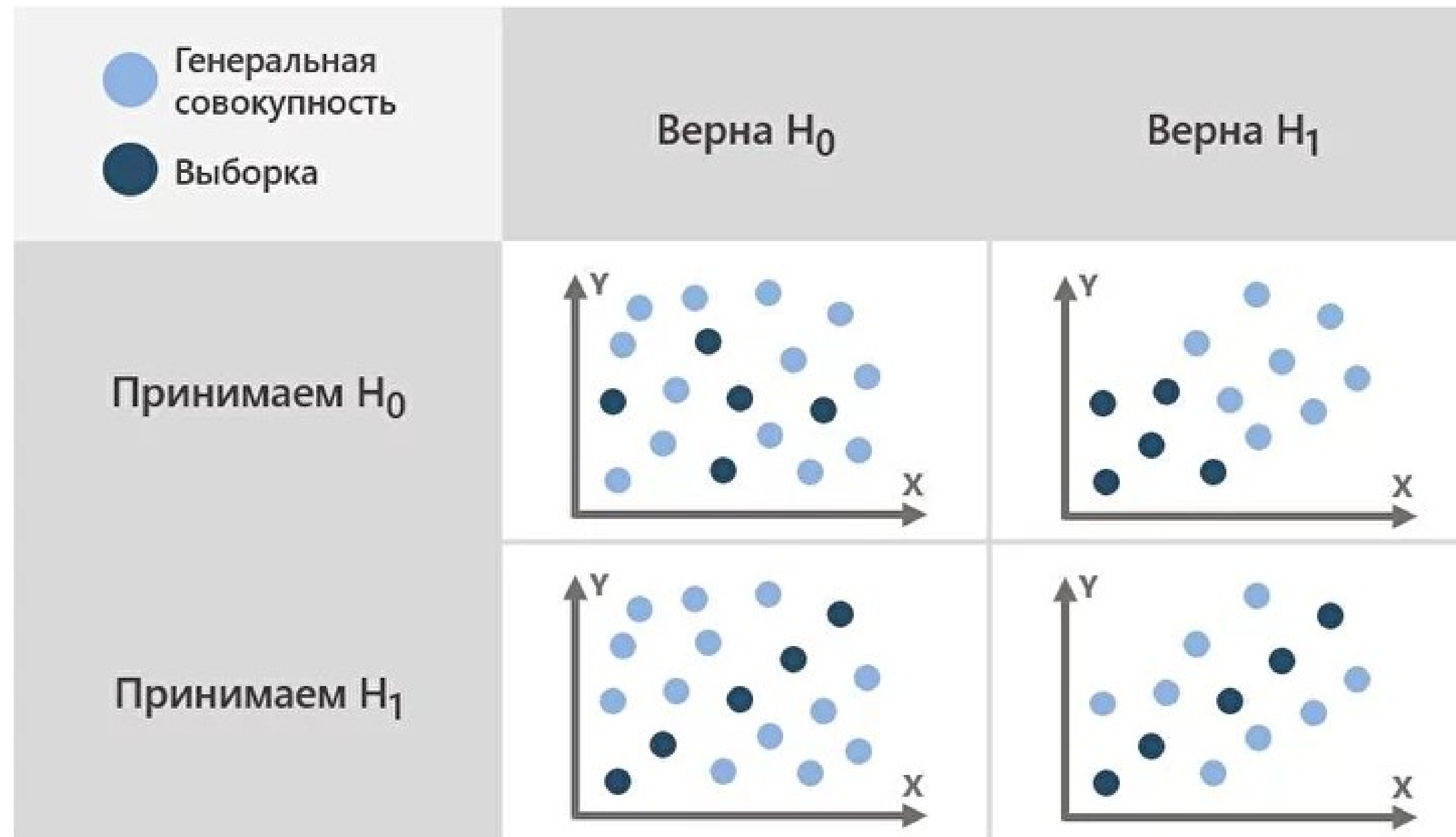
Выборочная совокупность — множество всех объектов, отобранных случайно из генеральной совокупности для изучения.



Нулевая гипотеза (H_0) — гипотеза о сходстве

Альтернативная гипотеза, конкурирующая, (H_1) — гипотеза о различиях

Нулевая и альтернативная гипотезы



Примеры основной и альтернативной гипотезы

Основная гипотеза:

$$H_0 : a = 368$$

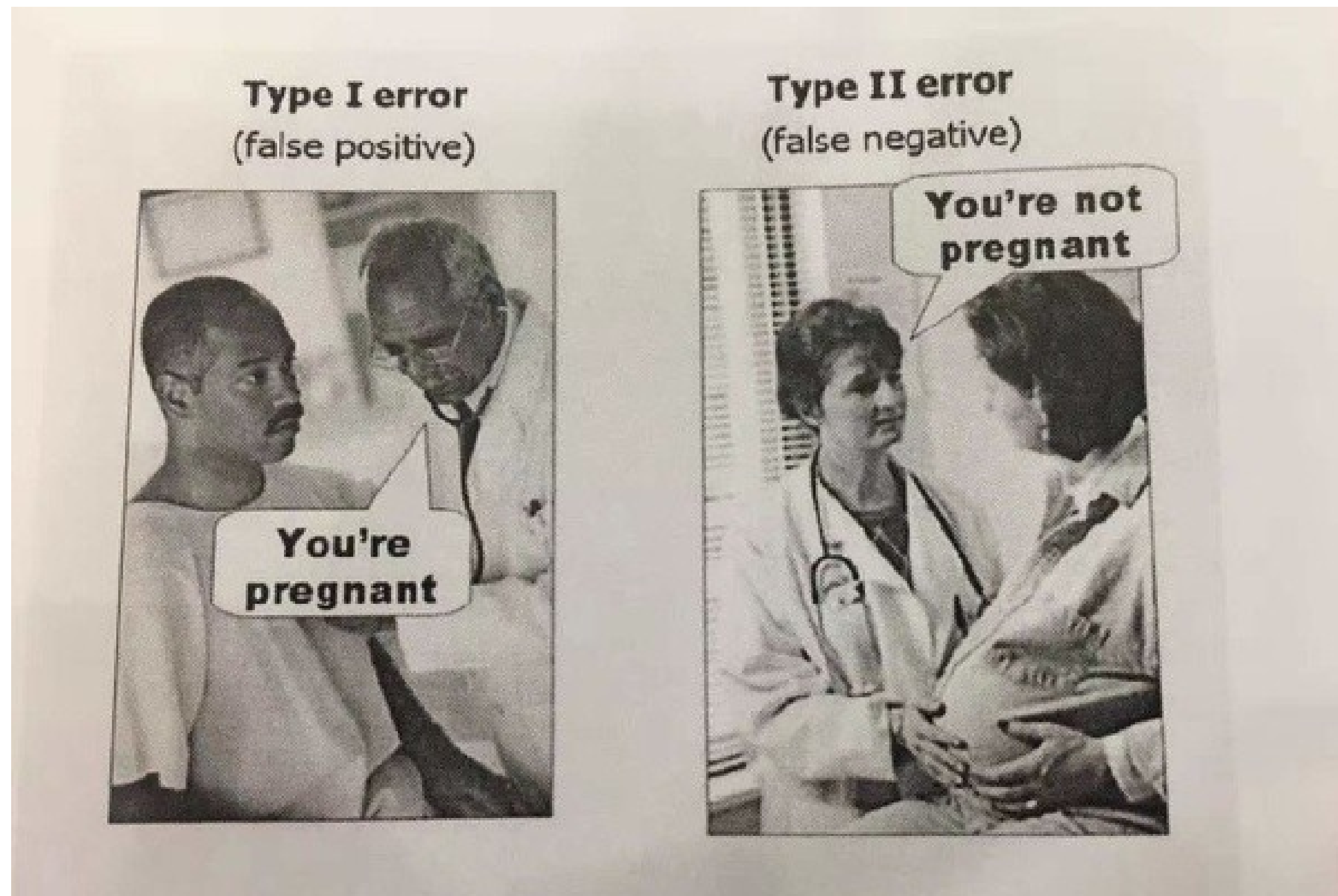
Средний вес выпускаемых коробок равен 368 г,
конвейер работает нормально

Альтернативная гипотеза:

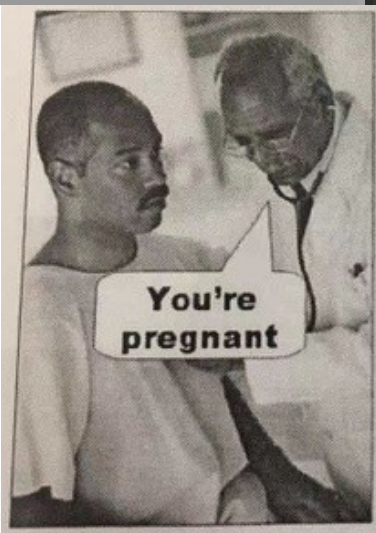

$$H_1 : a \neq 368$$

Средний вес выпускаемых коробок отличен от 368 г,
конвейер требует наладки

Пример: тест беременность.



Статистические гипотезы о данных

	Disease present	Disease absent
Positive	a True positive	b False positive 
Negative	 c False negative	d True negative

Ошибка 1 рода:
Вероятность отвергнуть гипотезу, но в действительности она верна

Критически значимый уровень
alpha = 0.05

Ошибка 2 рода:
Вероятность принять гипотезу, но в действительности она неверна
beta — вероятность ошибки.
Мощность исследования = 1-beta.

https://www.youtube.com/watch?v=4eyEp_NTxAU

<https://www.nejm.org/doi/full/10.1056/NEJM199908193410823>

Статистическая значимость

СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ (ЗНАЧЕНИЕ P)
– РАСЧЕТНАЯ ВЕРОЯТНОСТЬ ОШИБКИ
ПЕРВОГО РОДА, КОТОРАЯ РАССЧИТЫВАЕТСЯ С
ПОМОЩЬЮ РАЗЛИЧНЫХ СТАТИСТИЧЕСКИХ
КРИТЕРИЕВ



$$P < 0,05$$

Подсчитывается с помощью разных критериев.

Виды статистических критериев

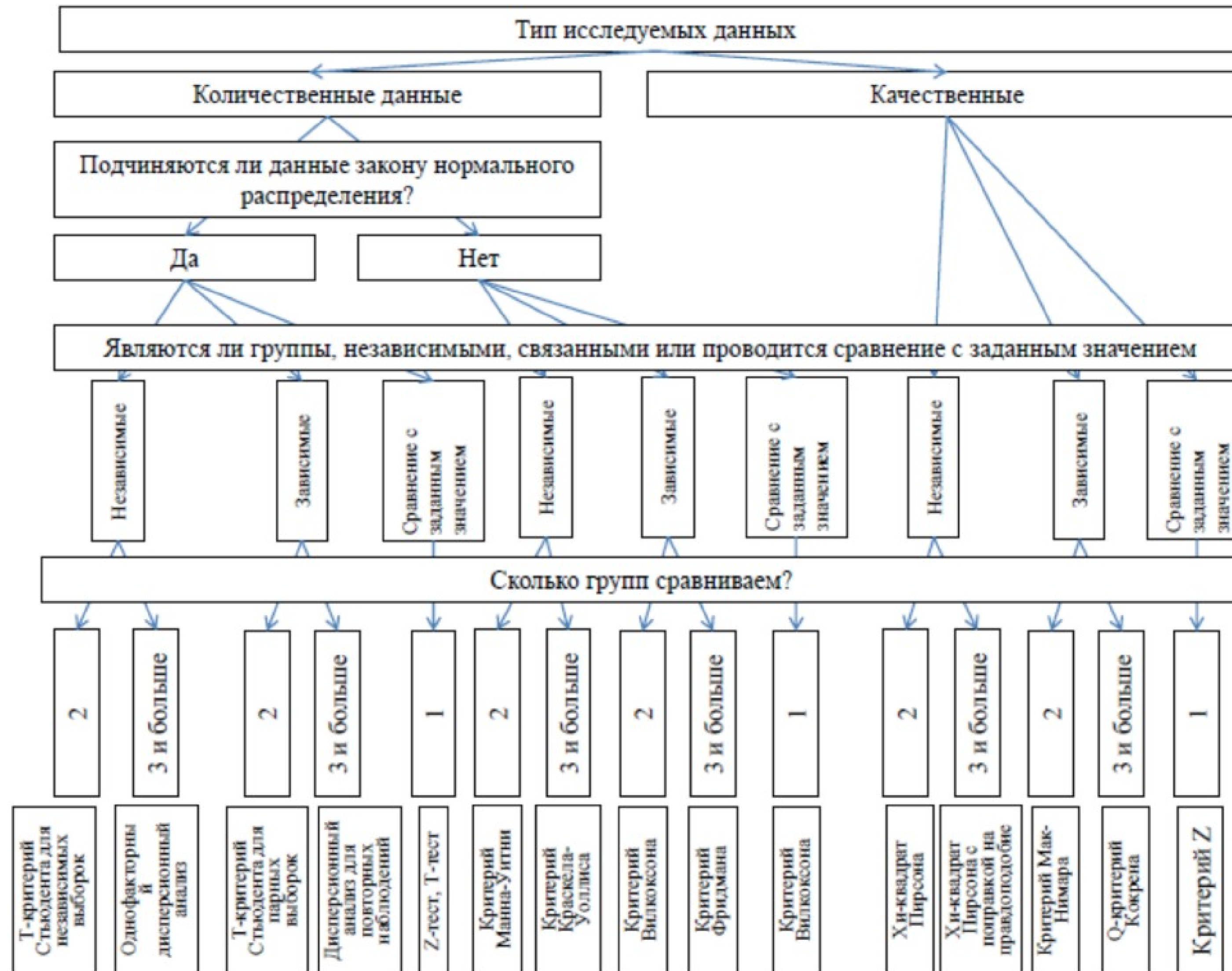
Критерии согласия -
исследуемая случайная величина подчиняется предполагаемому закону.

```
1 If Data Is Gaussian:  
2     Use Parametric Statistical Methods  
3 Else:  
4     Use Nonparametric Statistical Methods
```

Параметрические критерии -
включают в расчет параметры вероятностного распределения признака (средние и дисперсии).

Непараметрические критерии -
которые не включают в расчёт параметры вероятностного распределения и основаны на оперировании частотами или рангами.

Схема применения критериев



Параметрические критерии

Параметрические критерии - группа статистических критериев, которые включают в расчет параметры вероятностного распределения признака (средние и дисперсии).

t-критерий Стьюдента

Критерий Фишера

Критерий отношения правдоподобия

Критерий Романовского

t-критерий Стьюдента

Случайная величина t имеет распределение Стьюдента с $n-1$ степенями свободы, где n — размер выборки.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Критические значения t для заданного уровня доверия можно взять из таблицы

Т-критерий Стьюдента

Если t -критерий НЕ превышает пороговое t , при $p\text{-level} = 0.05$ — это значит, что у нас **нет оснований для отклонения нулевой гипотезы**.

Если фактическое **t** превышает критическое табличное значение при **$p=0.05$** , мы **отклоняем нулевую гипотезу**, — это означает, что мы обнаружили значимую закономерность!

Непараметрические критерии

Непараметрические критерии

Группа статистических критериев, которые не включают в расчёт параметры вероятностного распределения и основаны на оперировании частотами или рангами.

Q-критерий Розенбаума

U-критерий Манна — Уитни

Критерий Уилкоксона

Критерий Пирсона

Критерий Колмогорова — Смирнова

А/В тесты

А/В тестирование — это мощный маркетинговый инструмент для повышения эффективности работы вашего интернет-ресурса.

Ниже на картинках приведены примеры распределения значений показателя в сегментах.



Пример А/В теста: Wallmonkeys

Компания WallMonkeys решила оптимизировать веб-сайт на клики и конверсию.



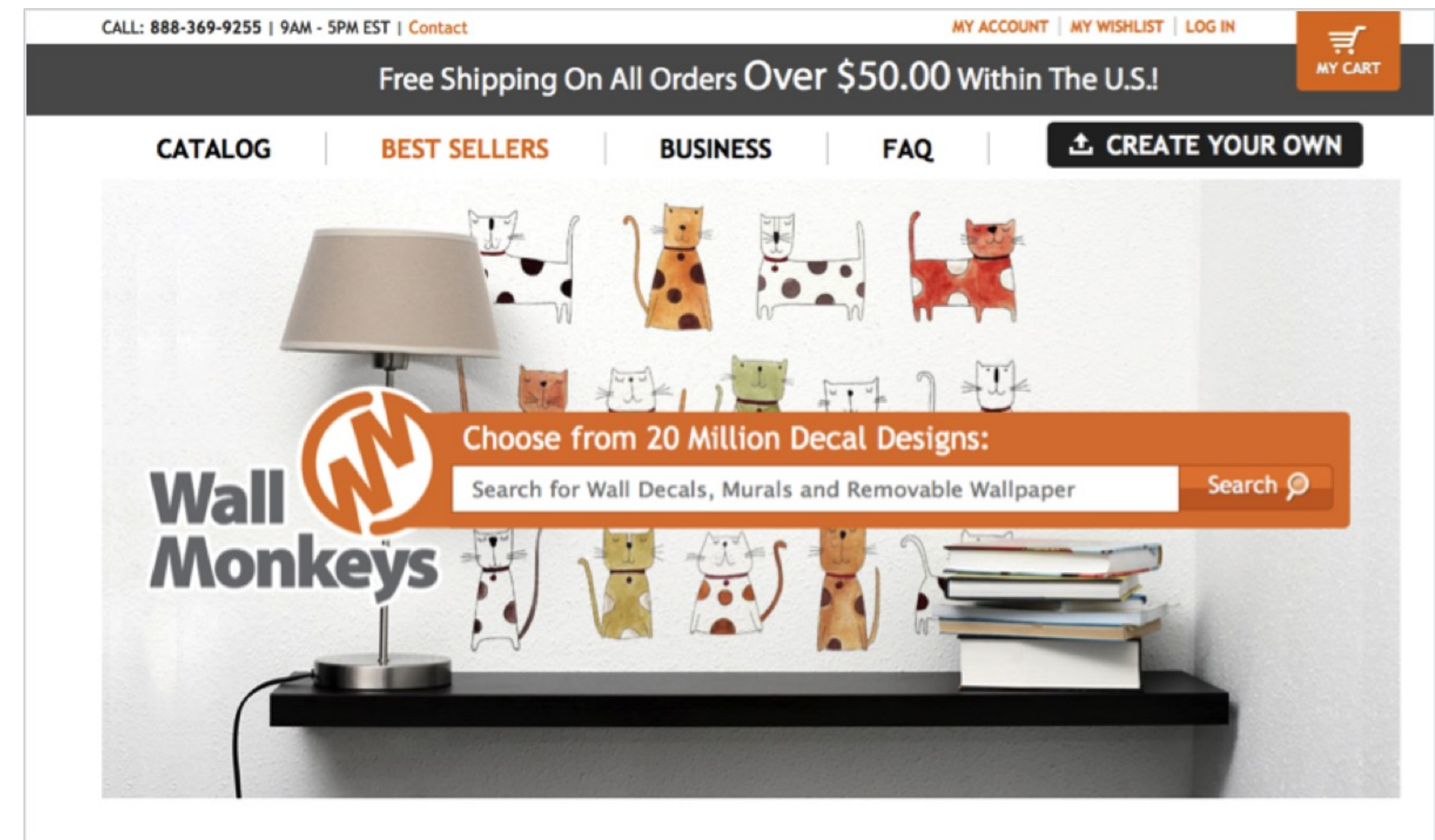
<https://www.crazyegg.com/blog/ab-testing-examples/>

Пример А/В теста: WallMonkeys

1 тест: 27% кликов.



2 тест: 550% кликов



Спасибо за внимание!