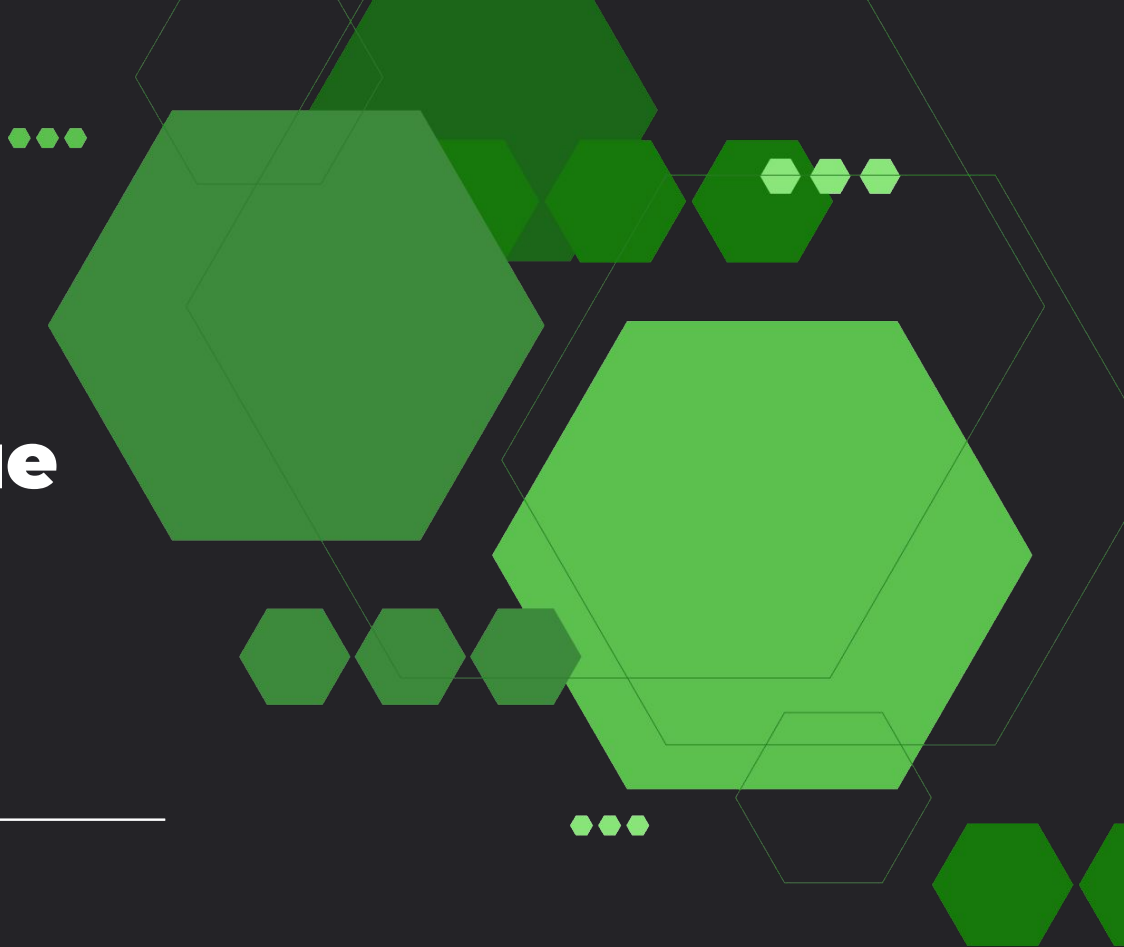


SKILLFACTORY

# Статистические тесты

---

Евгений Быков



## План

- Понятие статистической гипотезы, статистический критерий
- Уровень значимости, мощность теста, статистическая значимость
- Статистические тесты
  - Тест Стьюдента
  - Тест Манна-Уитни
  - Критерий Колмогорова
  - Критерий отношения правдоподобия
- Популярные статистические тесты в python
  - Тесты на нормальность
  - Параметрические тесты
  - Непараметрические тесты
  - Корреляционные тесты
- Советы по прохождению собеседований

## В этом модуле

1. посмотрим типовые задачи аналитика данных
2. научимся задавать правильные вопросы (т.е. формулировать статистические гипотезы)
3. научимся проверять гипотезы при помощи данных (использовать статистические критерии)
4. научимся понимать и объяснять результаты нашего анализа команде

## Примеры проблем

- 1) Есть чеки двух похожих магазинов X и Y за месяц. Есть ли среди них магазин-чемпион, или магазины справляются одинаково?
- 2) Есть два варианта сайта A и B. Какой из них дает бóльшую конверсию в покупку?
- 3) Монета выпала решкой 14 из 20 раз. Честная ли монета?
- 4) Привычная конверсия в покупку составляет  $p = 0.14$ , после проведения рекламной кампании продукт купили 104 из 590 посетителей. Была ли кампания успешна?
- 5) Проверка остатков линейной регрессии на нормальность?

## Статистическая гипотеза

- гипотеза о виде распределения и свойствах случайной величины, которое можно **подтвердить** или **опровергнуть** применением статистических методов к данным выборки.

Например:

1.  $H_0 = \{F = F_0\}, H_1 = \{F = F_1\};$

2.  $H_0 = \{\mathbb{E}X = \mathbb{E}Y\}, H_1 = \{\mathbb{E}X \neq \mathbb{E}Y\}$

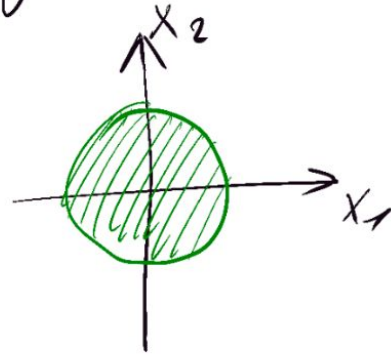
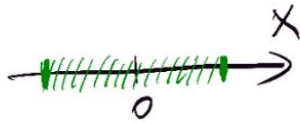
3.  $H_0 = \{\mathbb{P}(X) = \mathbb{P}(X|Y)\}$

## Придумываем статистический критерий

$$X_0 \sim N(\mu, 1)$$

$$X_1, X_2 \sim N(\mu, 1)$$

$$\begin{cases} H_0: \mu = 0 \\ H_1: \mu \neq 0 \end{cases}$$



## Статистический критерий

математическое правило, позволяющее по реализациям выборок отвергнуть или принять нулевую гипотезу с заданным уровнем значимости

Дана выборка  $X_1, \dots, X_n \sim F$

Хотим проверить гипотезу  $H_0$  (с альтернативной гипотезой  $H_1$ ).

Пусть можно задать функцию  $T(X^n)$ , такую что:

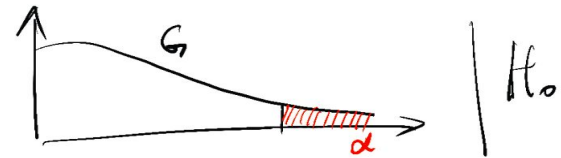
1. Если гипотеза  $H_0$  верна, то  $T(X^n) \Rightarrow G$ , где  $G$  - непрерывное распределение
2. Если гипотеза  $H_0$  неверна  $|T(X^n)| \rightarrow \infty$  при  $n \rightarrow \infty$

Определим переменную  $C$  из равенства  $\alpha = P(|y| \geq C)$ , где  $y \sim G$

Тогда статистический критерий будет выглядеть следующим образом:

$$\delta(X^n) = \begin{cases} H_0, & \text{если } |T(X^n)| < C \\ H_1, & \text{если } |T(X^n)| \geq C \end{cases}$$

$$T(x_1, \dots, x_n) \rightarrow \mathbb{R}^1$$



## Ошибка первого и второго рода

**Ошибка первого рода** ( $\alpha$ -ошибка, ложноположительное заключение) — ситуация, когда отвергнута верная нулевая гипотеза.

**Ошибка второго рода** ( $\beta$ -ошибка, ложноотрицательное заключение) — ситуация, когда принята неверная нулевая гипотеза.

		Верная гипотеза	
		$H_0$	$H_1$
Результат применения критерия	$H_0$	$H_0$ верно принята	$H_0$ неверно принята (Ошибка второго рода)
	$H_1$	$H_0$ неверно отвергнута (Ошибка первого рода)	$H_0$ верно отвергнута



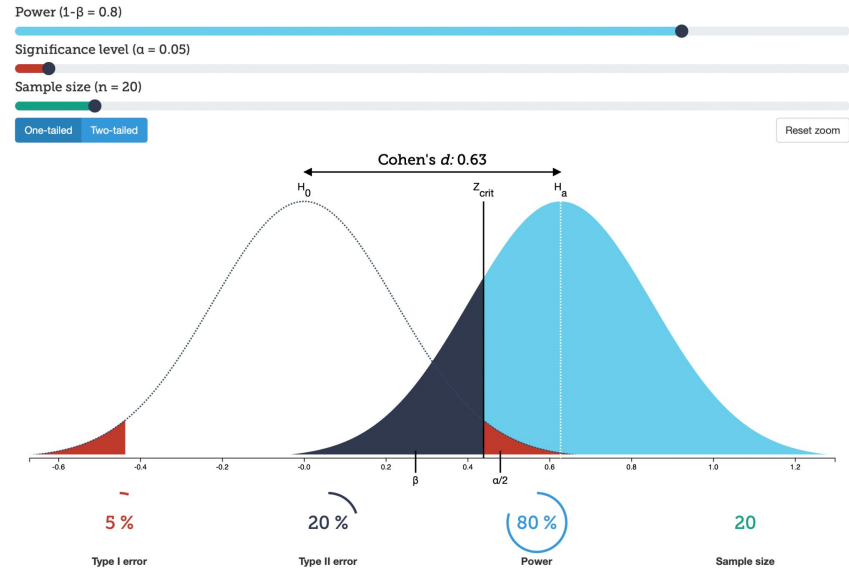
## Уровень значимости и мощность теста

**Уровень значимости** - вероятность отклонить нулевую гипотезу при условии ее истинности (вероятность совершения ошибки первого рода)

**Статистическая мощность** - вероятность отклонения нулевой гипотезы в случае, когда альтернативная гипотеза верна (чем выше мощность, тем меньше вероятность совершить ошибку второго рода)

## Выбор уровня значимости

Выбор  $\alpha$ -уровня неизбежно требует компромисса между значимостью и мощностью, и следовательно между вероятностями ошибок первого и второго рода.



Источник: <https://rpsychologist.com/d3/nhst/>

## Этапы проверки статистических гипотез

- 1) Выдвижение основной и альтернативной гипотез
- 2) Выбор уровня значимости  $\alpha$  (вероятность допустить ошибку первого рода)
- 3) Расчет статистики критерия (зависит от выборки, по ее величине можно сделать вывод об истинности нулевой гипотезы)
- 4) Построение критической области
- 5) Делаем вывод об истинности гипотезы на выбранном уровне значимости (по попаданию или не попаданию значения статистики в критическую область)

## Тест Стьюдента

Две выборки:  $X_1, \dots, X_{n_1} \sim N(\mu_1, \sigma_1)$  и  $Y_1, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2)$ .

Гипотезы:  $H_0 : \mathbb{E}X = \mathbb{E}Y$  и  $H_1 : \mathbb{E}X \neq \mathbb{E}Y$

Средние:  $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i, \quad \bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$

Оценки дисперсий:  $s_X^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2, \quad s_Y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$

Статистика теста: 
$$t(X^n, Y^n) = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}}$$

## Тест Манна-Уитни

Две выборки:  $X_1, \dots, X_{n_1} \sim F_X$  и  $Y_1, \dots, Y_{n_2} \sim F_Y$

Гипотезы:  $H_0 : F_X(t) = F_Y(t)$  и  $H_1 : F_X(t) = F_Y(t + \Delta), \Delta \neq 0$ .

Механика: 1) составляем единый ряд из выборок, 2) отдельно суммируем номера (ранги) для выборок -  $R_1$  и  $R_2$  3) Вычисляем  $U_1$  и  $U_2$

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}, \quad U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

Статистика:  $U = \min\{U_1, U_2\}$

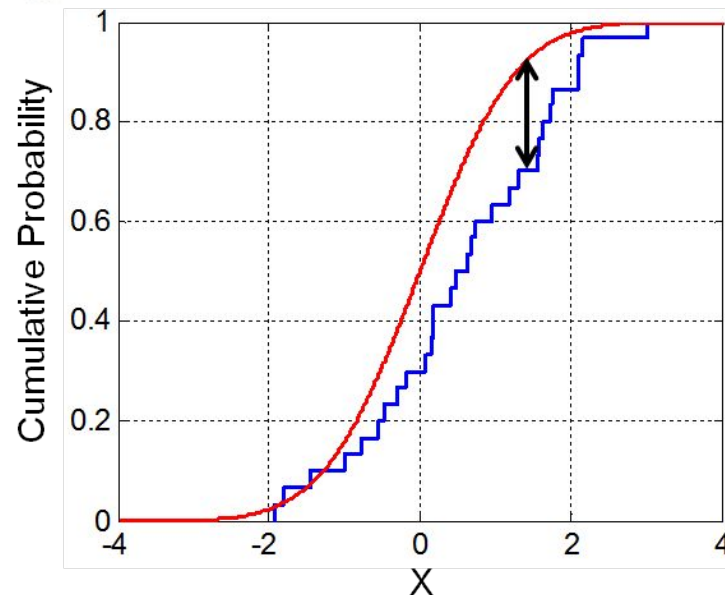
$$U(X^n, Y^n) \sim N\left(\frac{n_1 n_2}{2}, \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}\right)$$

## Критерий Колмогорова

Дана выборка  $X_1, \dots, X_n \sim F$ .

Гипотезы:  $H_0 : F = F_0$  и  $H_1 : F \neq F_0$ .

$$t(X^n) = \sqrt{n} \sup |\hat{F}_n(x) - F_0(x)|.$$



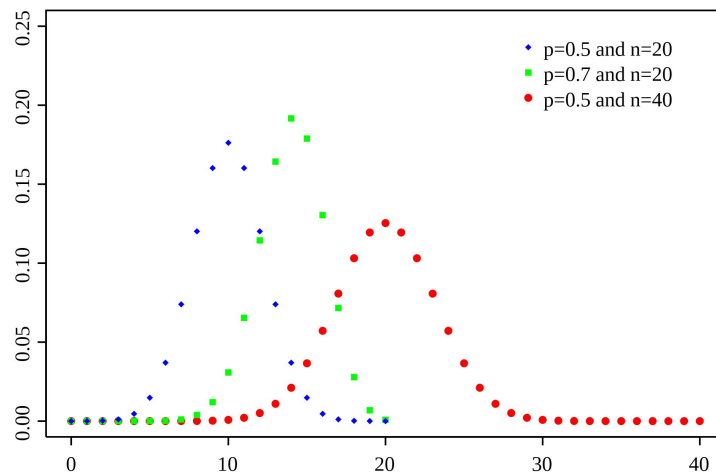
## Что такое p-value

- 😊 Вероятность получить для данной “вероятностной модели распределения значений случайной величины” такое же или более экстремальное значение статистики, по сравнению с ранее наблюдаемым, при условии, что нулевая гипотеза верна
- Если нулевая гипотеза верна, сколько могло случиться менее вероятных событий чтобы удивить нас и заставить усомниться в нулевой гипотезе.
- Мера величины доказательств, направленных против правдоподобия нулевой гипотезы
- НЕ(!) вероятность правоты нулевой гипотезы

## Считаем p-value для монетки

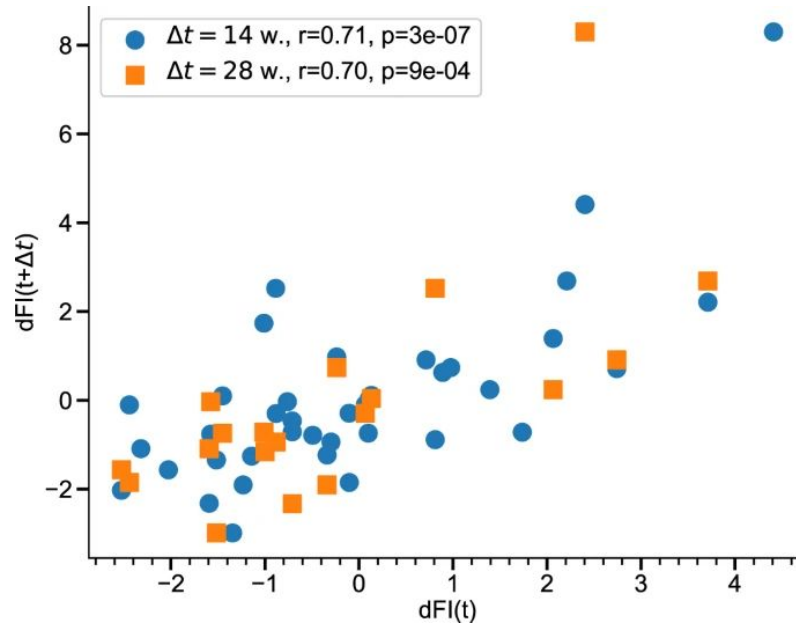
Монетка выпала орлом 14 из 20 раз. Честная ли монетка?

$$\frac{1}{2^{20}} \left[ \binom{20}{14} + \binom{20}{15} + \dots + \binom{20}{20} \right] = \frac{60,460}{1,048,576} \approx 0.058$$





## Проверка наличия эффекта с помощью p-value



source: <https://www.nature.com/articles/s41467-022-34051-9>

# Популярные статистические тесты

1. Тесты нормальности
  1. **Тест Шапиро-Вилка**
  2. Тест д'Агостино на  $K^2$
  3. Тест Андерсона-Дарлинга
2. Корреляционные тесты
  1. **Коэффициент корреляции Пирсона**
  2. Ранговая корреляция Спирмена
  3. Ранговая корреляция Кендалла
  4. Тест хи-квадрат
3. Параметрические тесты
  1. **Т-тест (Тест Стьюдента)**
  2. Парный т-тест
  3. Анализ дисперсионного теста (ANOVA)
  4. Повторные измерения ANOVA Test
4. Непараметрические тесты
  1. **U-тест Манна-Уитни**
  2. Тест Уилкоксона со знаком
  3. Kruskal-Wallis H Test
  4. Тест Фридмана

## Популярные параметрические тесты

	Независимая переменная	Зависимая переменная	Пример
Парный t-тест	Категориальная 1 переменная	Количественная Группы происходят из одной совокупности	Как влияют <b>две разные программы подготовки</b> к экзаменам на <b>средние баллы</b> учащихся одного класса?
Независимый t-тест	Категориальная 1 переменная	Количественная Группы происходят из разных совокупностей	Какова разница в <b>средних баллах</b> за экзамены у учеников <b>двух разных школ</b> ?
ANOVA	Категориальная 1 или несколько переменных	Количественная 1 переменная	Какова разница в <b>среднем уровне заработка</b> у выпускников <b>трех разных программ</b> ?
MANOVA	Категориальная 1 или несколько переменных	Количественная 2 или более переменных	Как влияет <b>вид цветка</b> на <b>длину и ширину лепестков и длину стебля</b> ?