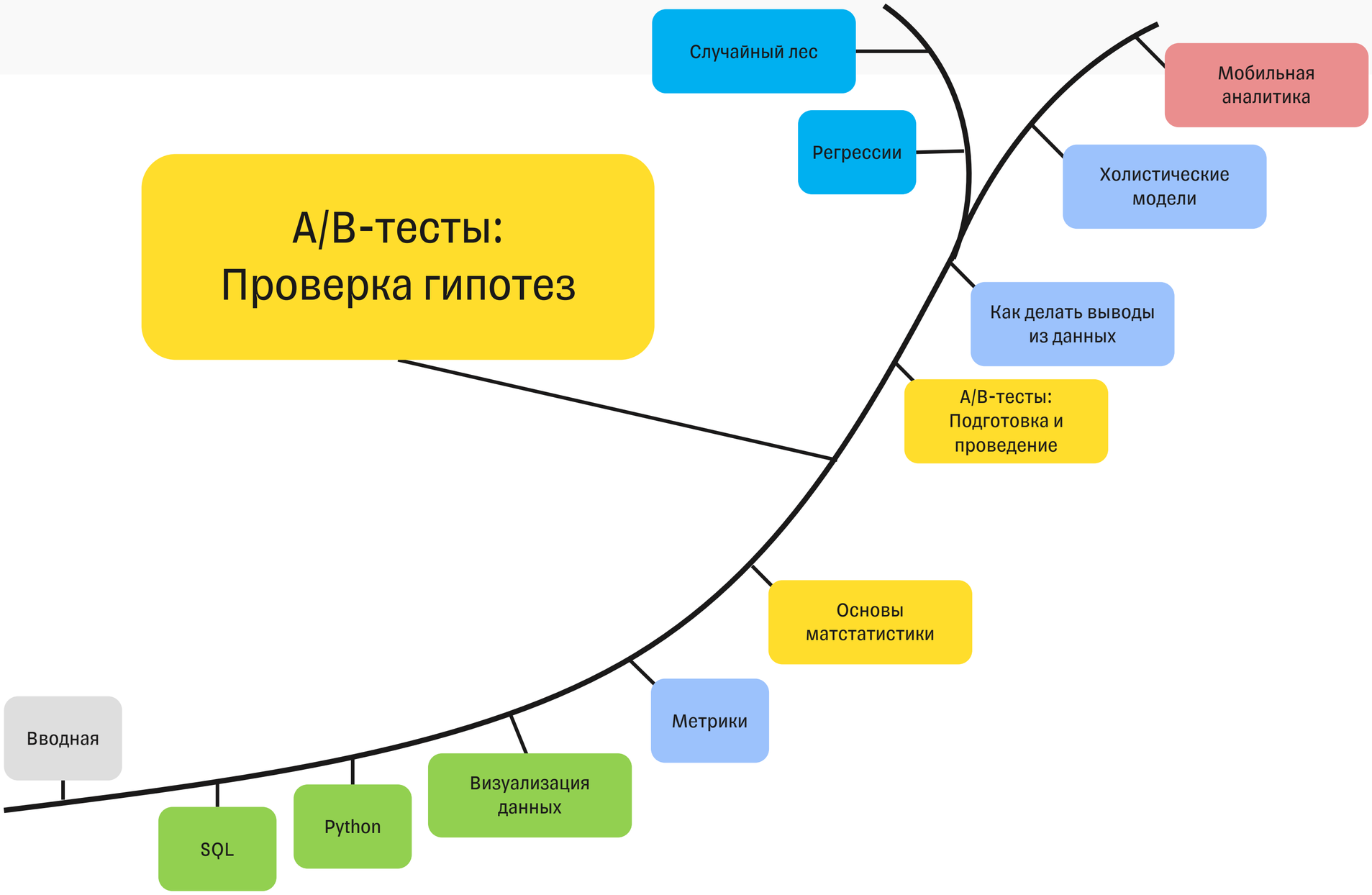

A/B-тесты:
Проверка гипотез
Случайный лес
Регрессии
Мобильная аналитика
Холистические модели
Как делать выводы из данных
A/B-тесты: Подготовка и проведение
Основы матстатистики
Метрики
Визуализация данных
Python
SQL
Вводная

ТИНЬКОФФ

Введение

Регламент

Участвуем в интерактивах

Задаём вопросы устно в конце блоков

Вы не только слушатели, но и участники

Что такое p-value?

Новый продукт

Задача

- Продажа нового продукта
- 2 варианта продажи
- Нужно выбрать лучший

Исходные данные

1. Предложили 100 клиентам, продали 30.
2. Предложили 1000 клиентам, продали 250.

Решение

А какой лучше?

Первый, так как 30% > 25%?

Подходы к задаче

Подходы	Результаты	Проблема
Точечная оценка	$30\% > 25\%$, первая лучше	Оценки случайны
Интервальная оценка	$(21\%, 39\%)$ против $(22\%, 28\%)$	Интервалы пересекаются

Принятие решения

Выборки

- $X = (X_1, \dots, X_{100}) \in \{0,1\}^{100}$, $X_1 + \dots + X_{100} = 30$;
- $Y = (Y_1, \dots, Y_{1000}) \in \{0,1\}^{1000}$, $Y_1 + \dots + Y_{1000} = 250$.

Мы хотим придумать функцию $\varphi(X, Y)$, которая по выборкам X и Y позволит принять решение:

- $\varphi(X, Y) = 2$ – выбираем второй вариант продажи;
- $\varphi(X, Y) = 1$ – выбираем первый вариант продажи;
- $\varphi(X, Y) = 0$ – разница мала, выбрать не можем.

Как выбрать $\varphi(X, Y)$?

- $\varphi(X, Y) = 2$, если $\bar{X} - \bar{Y} < c_2$;
- $\varphi(X, Y) = 1$, если $\bar{X} - \bar{Y} > c_1$;
- $\varphi(X, Y) = 0$, иначе.

Нужен инструмент для принятия решений в случайном мире

Проверка гипотезы

Общая постановка задачи

Наблюдаем: выборку X из пространства $(\mathcal{X}, \mathcal{F}, \mathbb{P})$.

Даны: мн-ва непересекющихся вер. мер $\mathcal{P}_0, \dots, \mathcal{P}_k$.
Гипотеза $H_i: \mathbb{P} \in \mathcal{P}_i$.

Цель: определить такой *статистический критерий*

$$\varphi: \mathcal{X} \to \{0, \dots, k\},$$

что

$$\mathbb{P}(\varphi(X) \neq i) \approx 0, \qquad \mathbb{P} \in \mathcal{P}_i.$$

Самый естественный пример

Наблюдаем: величину X с распределением Бернулли p.

Даны: гипотеза $H_0: p = 0.3$ и гипотеза $H_1: p = 0.5$.

Как будет выглядеть хороший статистический критерий?

Имеем

$$\varphi(x) = \mathbb{I}\{x = 1\}.$$

Тогда

$$\mathbb{P}_{H_0}(\varphi(X) = 1) = \mathbb{P}_{H_0}(X = 1) = 0.3,$$
$$\mathbb{P}_{H_1}(\varphi(X) = 0) = \mathbb{P}_{H_1}(X = 0) = 0.5.$$

Случай двух гипотез

Наблюдаем: выборку X из пространства $(\mathcal{X}, \mathcal{F}, \mathbb{P})$.

Даны: гипотеза $H_0: \mathbb{P} \in \mathcal{P}_0$ и альтернатива $H_1: \mathbb{P} \in \mathcal{P}_1$.

Цель: определить такую *статистику критерия*

$$T: \mathcal{X} \to \mathbb{R}$$

и такие числа t_1, t_2, что

$$\mathbb{P}(T(X) \notin (t_1, t_2)) \approx 0, \qquad \mathbb{P} \in \mathcal{P}_0,$$
$$\mathbb{P}(T(X) \in (t_1, t_2)) \approx 0, \qquad \mathbb{P} \in \mathcal{P}_1.$$

Нормальный пример

Наблюдаем: величины $X_1, \dots, X_n \sim \mathcal{N}(\mu, 1)$.

Даны: гипотеза $H_0: \mu = 0$ и альтернатива $H_1: \mu = 1$.

Как будет выглядеть хороший статистический критерий?

Имеем

$$T(x) = \bar{x}, \qquad t_1 = -\infty.$$

Тогда $T(X) \sim \mathcal{N}(\mu, 1/n)$.

Дилемма:

- чем больше t_2, тем меньше $\mathbb{P}_{H_0}(T(X) \geq t_2)$;
- чем меньше t_2, тем меньше $\mathbb{P}_{H_1}(T(X) < t_2)$.

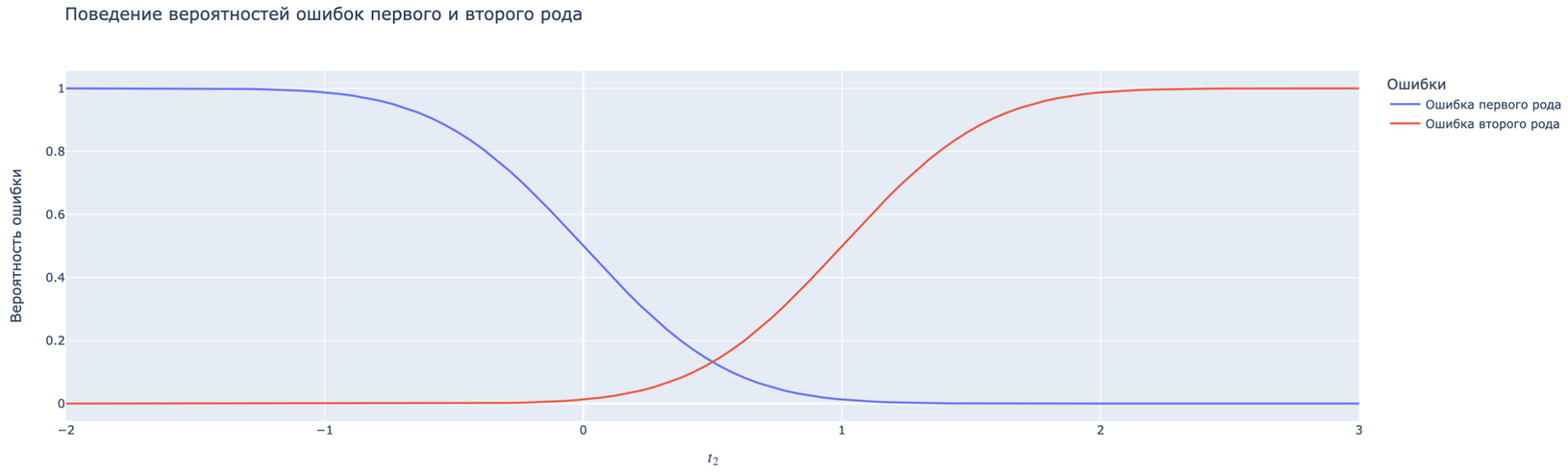
Ошибки первого и второго рода

Вероятность ошибки *первого рода:*

$$\alpha = \alpha(\mathbb{P}) = \mathbb{P}(\varphi(X) = 1), \qquad \mathbb{P} \in \mathcal{P}_0.$$

Вероятность ошибки *второго рода:*

$$\beta = \beta(\mathbb{P}) = \mathbb{P}(\varphi(X) = 0), \qquad \mathbb{P} \in \mathcal{P}_1.$$



Уровень значимости – верхняя граница для $\alpha(\mathbb{P})$.

Альтернативы

Пусть $\mathcal{P}_0, \mathcal{P}_1 \subset \{\mathbb{P}_\theta \mid \theta \in \mathbb{R}\}$, при любом $t \in \mathbb{R}$ функция

$$g_t(\theta) = \mathbb{P}_\theta(T(X) > t)$$

растёт по θ.

Тогда

- $\varphi(x) = \mathbb{I}\{T(x) \geq t_2\}$ – *правосторонняя* альтернатива ($H_0: \theta \leq \theta_0, H_1: \theta > \theta_0$);
- $\varphi(x) = \mathbb{I}\{T(x) \leq t_1\}$ – *левосторонняя* альтернатива ($H_0: \theta \geq \theta_0, H_1: \theta < \theta_0$);
- $\varphi(x) = \mathbb{I}\{T(x) \notin (t_1, t_2)\}$ – *двухсторонняя* альтернатива ($H_0: \theta = \theta_0, H_1: \theta \neq \theta_0$).

Конверсия

Пусть $X_1, \dots, X_n \sim Bern(p)$.
Отметим, что при любом $t \in \mathbb{R}$

$$g_t(p) = \mathbb{P}_p(T(X) > t) = \mathbb{P}_p(X_1 + \dots + X_n > nt)$$

растёт по p.

Тогда

- $H_0: p \leq p_0,\ H_1: p > p_0$ - *правосторонняя* альтернатива;
- $H_0: p \geq p_0,\ H_1: p < p_0$ - *левосторонняя* альтернатива;
- $H_0: p = p_0,\ H_1: p \neq p_0$ - двухсторонняя альтернатива.

Ухудшающий тест

Задача

- Поддерживаем костыльный процесс продажи
- Хотим понять, нужно ли поддерживать
- Проведём тест выключения канала

Исходные данные

1. Исторически конверсия 28%.
2. Без канала – 957 клиентов, в среднем конверсия 24%.

Как проверить?

Мы тестируем отклонение вниз, поэтому используем левостороннюю альтернативу.

Алгоритм составления критерия

01

Определите как будут приниматься решения

- В чём цель вашего эксперимента?
- Что вы хотите проверить или оценить?
- Что должно произойти, чтобы вы выбрали то или иное дело?

02

Составьте статистическую модель

- Какая модель учтёт всё то, что нужно для принятия решений?
- Какие возможные значения у величин?
- Какие распределения могут иметь величины?

03

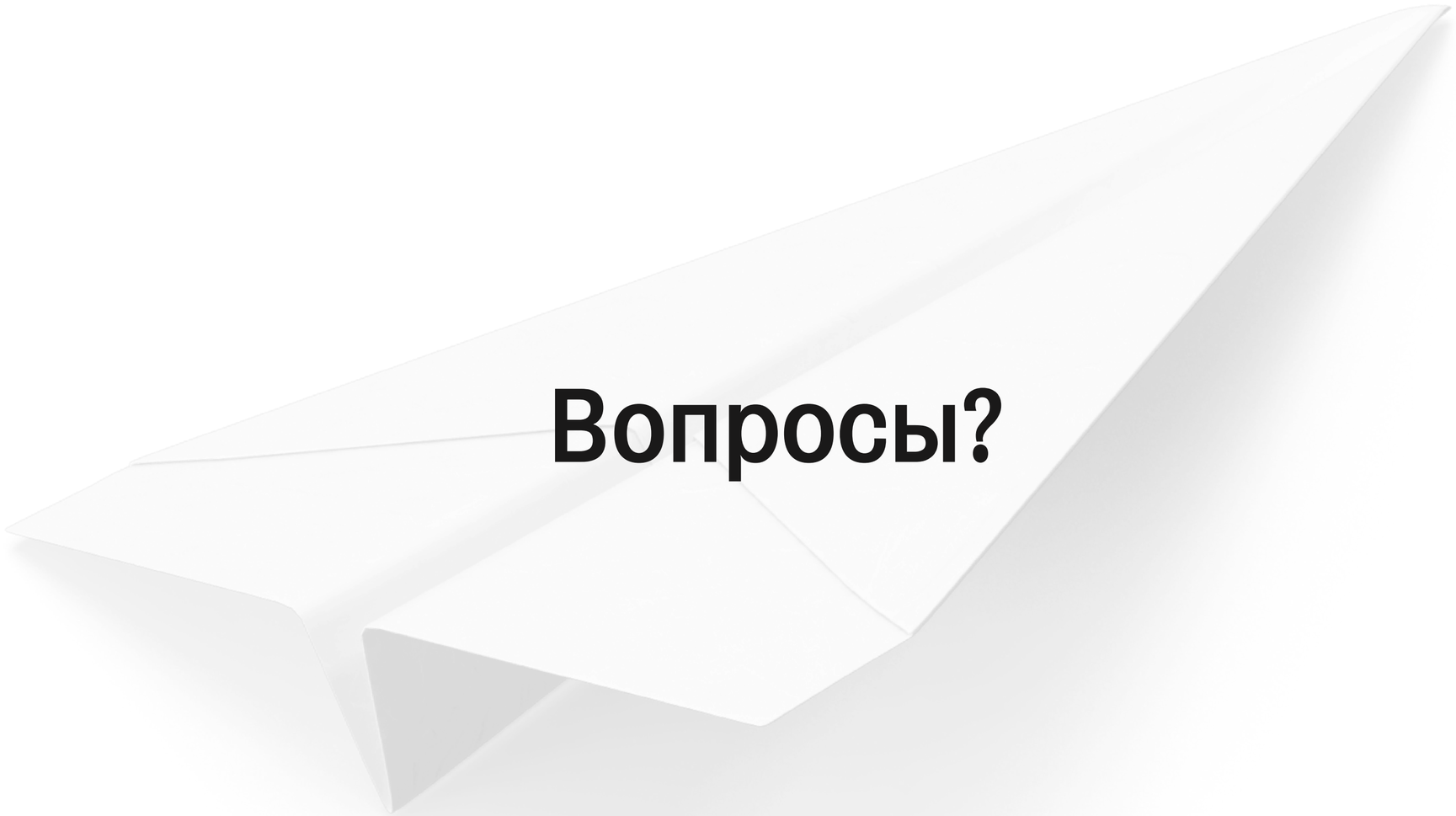
Установите гипотезы и альтернативы

- Являются ли альтернативы односторонними или двухсторонними?
- Правда ли при выборе одной из гипотез будет принято решение?

04

Задание критерия

- Какой уровень значимости?
- Есть ли известный критерий, подходящий для проверки альтернатив?
- Если нет, как адаптировать известные критерии под гипотезы и альтернативы?



Вопросы?

P-value

Желаемый инструмент

Фиксируем $\alpha \in (0, 1)$

Критерий $\varphi(x) = \mathbb{I}\{p(x) \leq \alpha\}$

$$\mathbb{P}(\mathrm{I}\ type\ error) \leq \alpha, \quad \mathbb{P} \in \mathcal{P}_0.$$

$$\mathbb{P}(\mathrm{II}\ type\ error) < 1 - \alpha, \quad \mathbb{P} \in \mathcal{P}_1.$$

P-value

Пусть T – статистика критерия с правосторонней альтернативой.

Тогда p-value – значение статистики

$$p(x) = \sup\{\, \mathbb{P}_\theta\big(T(Y) > T(x)\big) \mid \theta \le \theta_0 \,\} = \mathbb{P}_{\theta_0}\big(T(Y) > T(x)\big)$$

на выборке X.

Свойства:

1. p-value $p(X)$ - случайная величина;
2. верно $p(X) \in [0, 1]$;
3. если $\theta = \theta_0$ и T - непрерывна, то $p(X) \sim R[0, 1]$.

Нормальный пример

Наблюдаем: величины $X_1, \dots, X_n \sim \mathcal{N}(\mu, 1)$.

Даны: гипотеза $H_0: \mu \leq 0$ и альтернатива $H_1: \mu > 0$.

Как будет выглядеть p-value критерия?

Имеем $T(x) = \bar{x}$ и правостороннюю альтернативу.
При $\mu = 0$ имеем $T(X) \sim \mathcal{N}(0, 1/n)$.

Отсюда

$$p(x) = \mathbb{P}_0\big(T(Y) > T(x)\big) = 1 - F_{0,1/n}(\bar{x}),$$

где F_{μ,σ^2} - функция распределения $\mathcal{N}(\mu, \sigma^2)$.

Какое распределение имеет p-value при альтернативе?

Эмпирическая функция распределения

Инструмент: функция распределения (ФР) содержит всю информацию о распределении.

Проблема: непонятно откуда её взять.

Инструмент: эмпирическая функция распределения (ЭФР)

$$\hat{F}_n(x) = (\mathbb{I}\{X_1 \leq x\} + \cdots + \mathbb{I}\{X_n \leq x\})/n.$$

Свойства:

- $\hat{F}_n(x)$ - функция распределения $R\{X_1, \dots, X_n\}$;
- $n\,\hat{F}_n(x)$ имеет биномиальное $(n, F(x))$ распределение;
- $\mathbb{E}\,\hat{F}_n(x) = F(x)$;
- в силу ЗБЧ $\hat{F}_n(x) \to F(x)$.

Моделирование распределения p-value

Резюме

Случайная величина

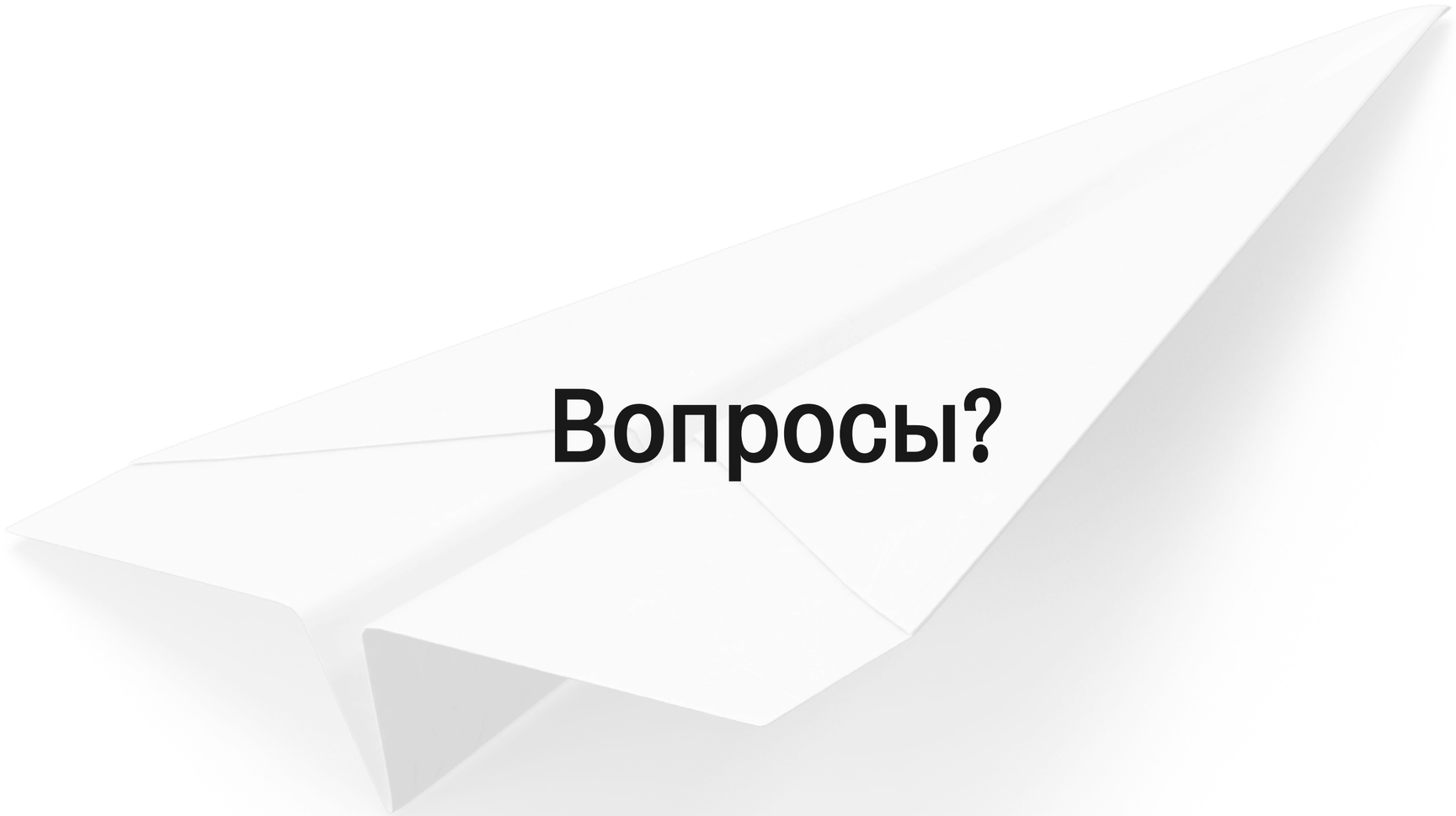
- P-value – случайная величина на $[0, 1]$
- При гипотезе распределение p-value должно быть ближе к 1
- При альтернативе распределение p-value должно быть ближе к 0

Простота использования

- В «худшем случае» при гипотезе p-value имеет равномерное распределение
- При уровне значимости α критерий $\varphi(x) = \mathbb{I}\{p(x) < \alpha\}$ имеет ошибку первого рода не больше α

Сравнение критериев

- С помощью ЭФР p-value можно проверить корректность критерия
- С помощью ЭФР p-value можно сравнивать эффективность критериев



Вопросы?

Как p-value связано с ROC?

ТИНЬКОФФ

Гипотезы и критерии

Группы гипотез и критерии

Группа	Гипотеза	Критерии
Критерии согласия	$H_0: F \in \mathcal{P} = \{\, \mathbb{P}_\theta \mid \theta \in \Theta \}, \qquad H_1: F \notin \mathcal{P}$	• QQ-plot • Проверка нормальности • Проверка экспоненциальности • Проверка совпадения с фиксированным распределением
Критерии независимости	$H_0: \mathbb{P}_{(X,Y)} = \mathbb{P}_X \times \mathbb{P}_Y, \qquad H_1: \mathbb{P}_{(X,Y)} \neq \mathbb{P}_X \times \mathbb{P}_Y$	• Корреляции • Ядерные критерии • Энергетический критерий
Критерий однородности	$H_0: F = G, \qquad H_1: F \neq G$	• Проверка совпадения распределений • Проверка совпадения средних • Проверка совпадения медиан

Подход к критериям однородности

Определяем *характеристику* критерия $d(F, G)$

Статистика критерия $\mathrm{d}' = d(\hat{F}_n, \hat{G}_m)$,
где $\hat{F}_n$ и $\hat{G}_m$ - ЭФР выборок

Доказываем, что d' при больших n, m и $F = G$
имеет распределение T

Придумываем оптимальный метод расчёта d'

В итоге критерий $\varphi(x) = \mathbb{I}\{d' > t_{1-\alpha}\}$, где t_β -
β-квантиль распределения T

Критерии однородности

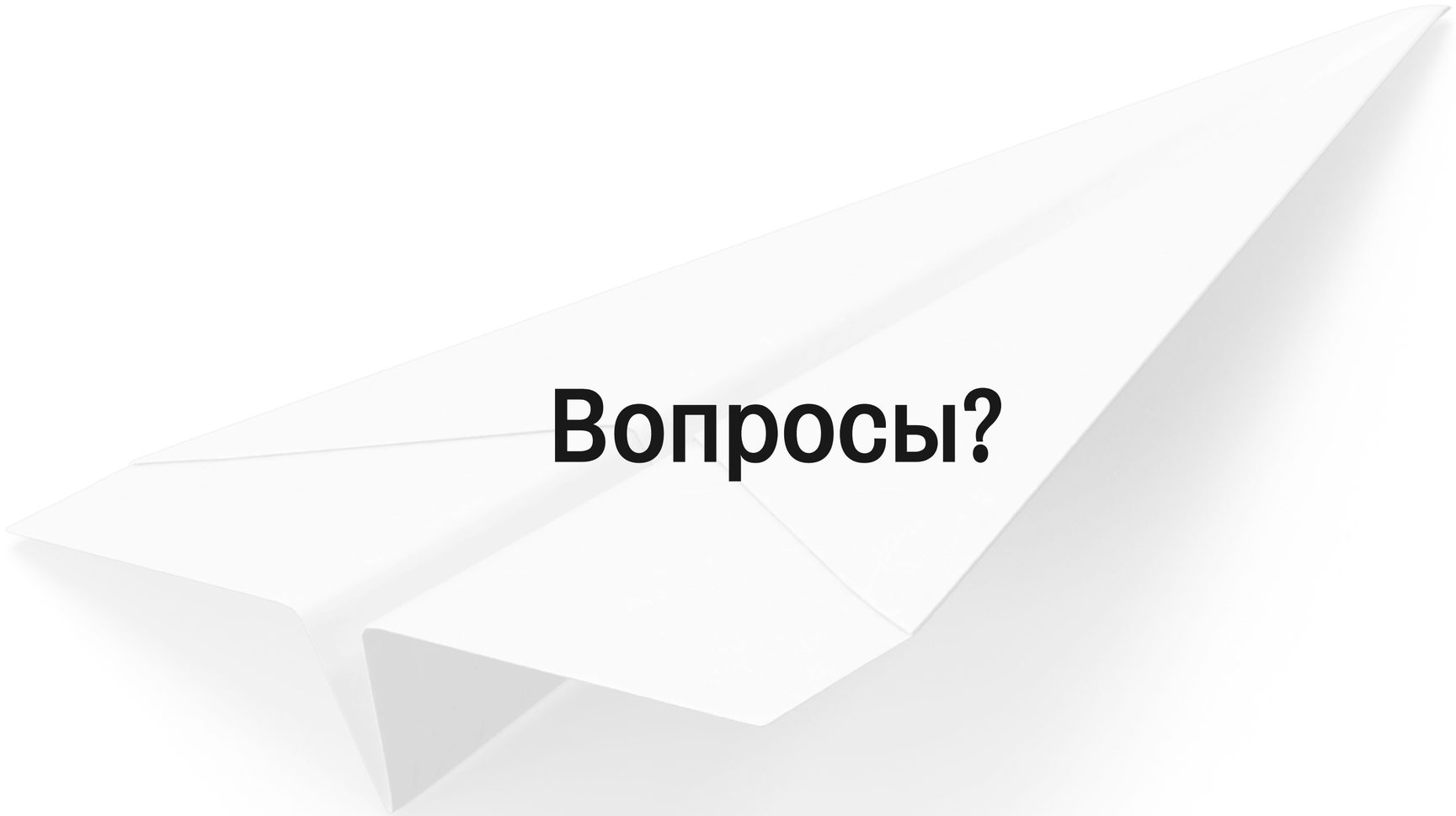
Тест	Характеристика	Альтернатива	Реализация в Python
Колмогоров-Смирнов	$\sup\{\lvert F(x) - G(x)\rvert : x \in \mathbb{R}\}$	$F \neq G$	ks_2samp
Леман-Розенблатт	$\int_{-\infty}^{+\infty} \big(F(x) - G(x)\big)^2 dH(x)$	$F \neq G$	cramervonmises_2samp
Андерсон-Дарлинг	$\int_{-\infty}^{+\infty} \frac{\big(F(x) - G(x)\big)^2}{H(x)\big(1 - H(x)\big)} dH(x)$	$F \neq G$	anderson_ksamp
Манн-Уитни-Уилкоксон	$\int_{-\infty}^{+\infty} G(x) dF(x) - \frac{1}{2}$	$\mathbb{P}(X < Y) \neq \frac{1}{2}$	mannwhitneyu
Z-тест, T-тест	$\frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\mathbb{D}X + \mathbb{D}Y}}$	$\mathbb{E}X \neq \mathbb{E}Y$	ztest, ttest_ind
Краскел-Уоллис	$\mathrm{MED}(Y) - MED(X)$	$MED(X) \neq MED(Y)$	kruskal
Хи-квадрат	$\sum_{i=1}^{n} \big(\mathbb{P}(X \in A_i) - \mathbb{P}(Y \in A_i)\big)^2$	$\mathbb{P}(X \in A_i) \neq \mathbb{P}(Y \in A_i)$	crosstab

Как выглядят односторонние альтернативы?

Выбор критерия однородности

Ситуация	Альтернатива	Реализация в Python
Частота VS Значение	$p \neq (<,>)\ p_0$	proportions_ztest(s, n, p0, alternative=)
Среднее VS Значение	$\mu \neq (<,>)\ \mu_0$	ztest(x, value=mu0, alternative=)
Частота VS Частота	$p_1 \neq (<,>)\ p_2$	proportions_ztest([s1, s2], [n1, n2], alternative=)
Среднее VS Среднее	$\mu_1 \neq (<,>)\ \mu_2$	есть дисперсия - ttest_ind(equal_var=False) есть только матожидание - permutation_test() нет матожидания, сдвиг монотонный - mannwhitneyu()
Распределение VS Распределение	$F_1 \neq F_2$	ks_2samp(x, y) anderson_ksamp([x, y]) cramervonmises_2samp(x, y) Energy(compute_distance=).test(x, y) MMD(compute_kernel=, **kwargs).test(x, y)

Моделирование критериев гипотез однородности



Вопросы?

p-value и ROC никак не связаны

Резюме

Статистика нужна для принятия решений

Формулируйте гипотезы и альтернативы исходя из принимаемых решений.

Выбирайте нужный вам критерий

Большинство критериев реализовано, их нужно грамотно использовать, ориентируясь на соответствующие им гипотезы и альтернативы.

Не стесняйтесь моделировать

В реальных задачах нужно проверять инструменты и выбирать нужный. Для этого моделируйте эффекты и используйте ЭФР p-value.

Литература

Лагутин М.Б.

Наглядная математическая статистика

Антон Коробейников

Проверка статистических гипотез

Мальцев А.С., Соколиков Е.А.

Проверка применимости критерия с помощью моделирования

Ryan Ashton

How to Select a Hypothesis Test – with Python Examples

Jason Brownlee

17 Statistical Hypothesis Tests in Python (Cheat Sheet)

ТИНЬКОФФ

Он такой один