



# АВ тестирование

# Давайте представлюсь

**Шульчевский Дмитрий**

Старший аналитик-разработчик

## **5 лет в Яндексе**

- › Занимался разработкой метрик Поиска
- › Сейчас занимаюсь безопасностью поездок в Такси

## **3 года в Тинькофф**

- › Занимался кредитным скорингом

# Содержание

- 1 | АВ тесты
- 2 | Платформа АВ тестирования
- 3 | Разработка метрик

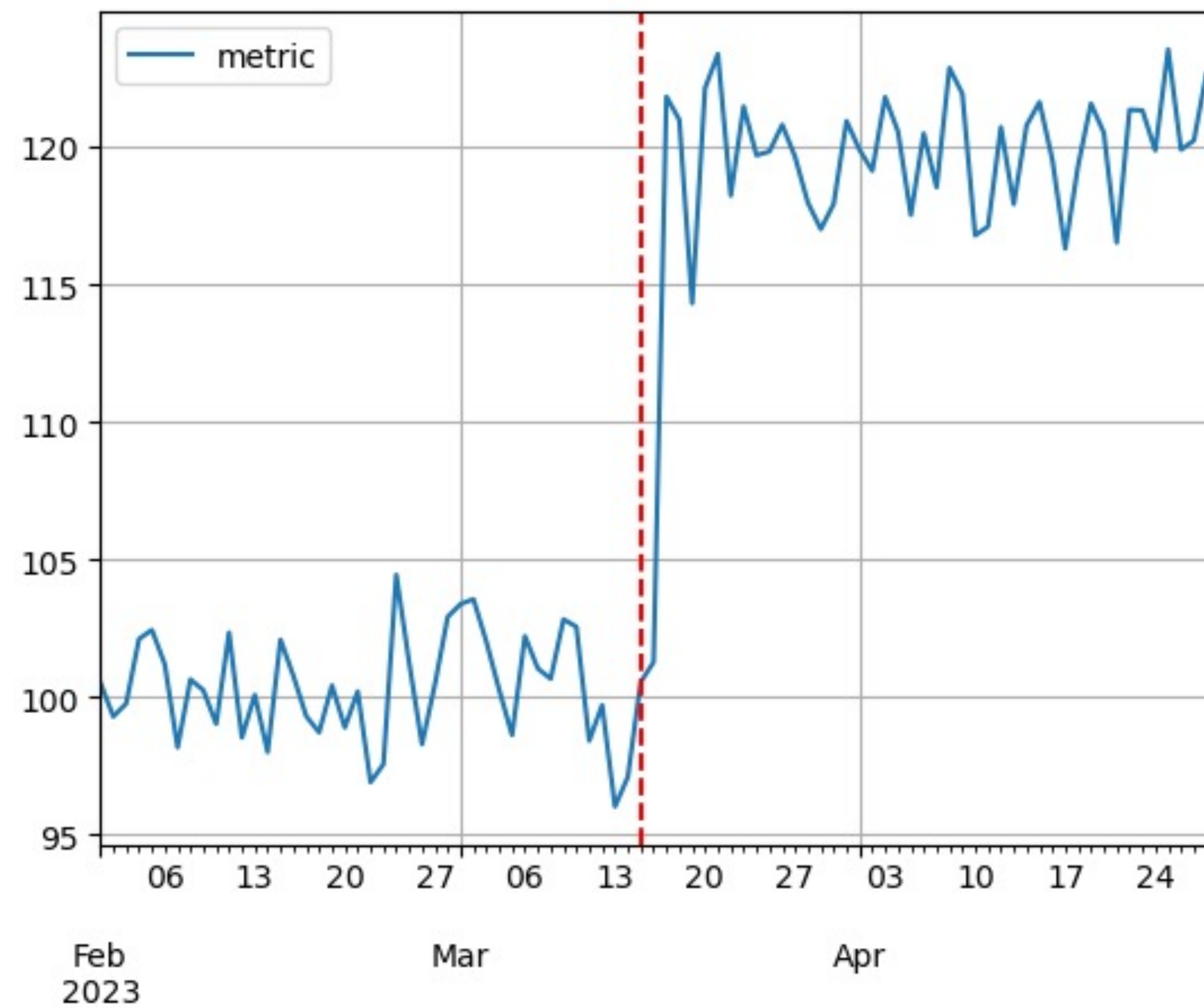
01



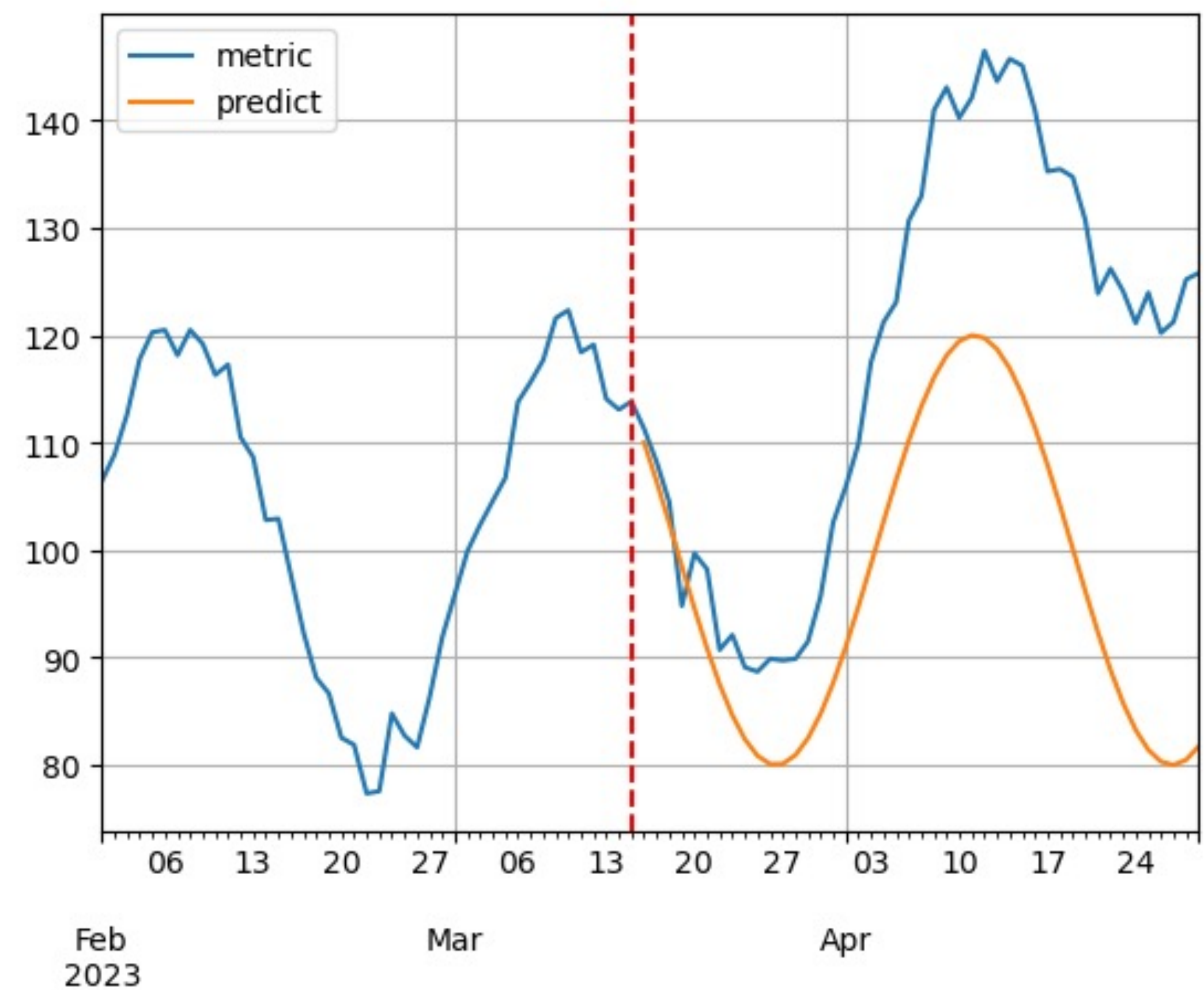
**АВ тесты**

**Как измерить эффект от изменения продукта?**

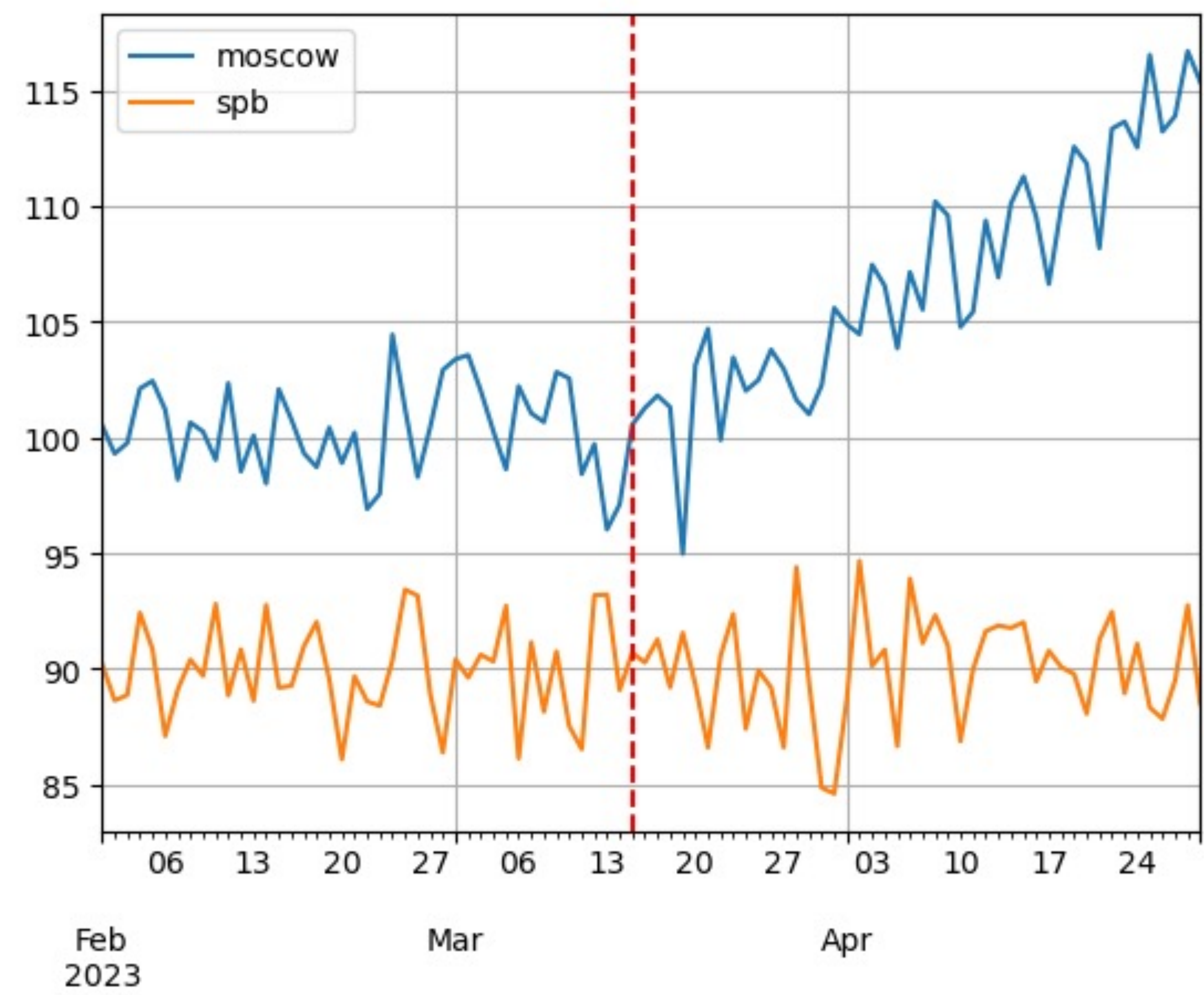
# “Ступеньки” на метриках



# Сравнение с прогнозом

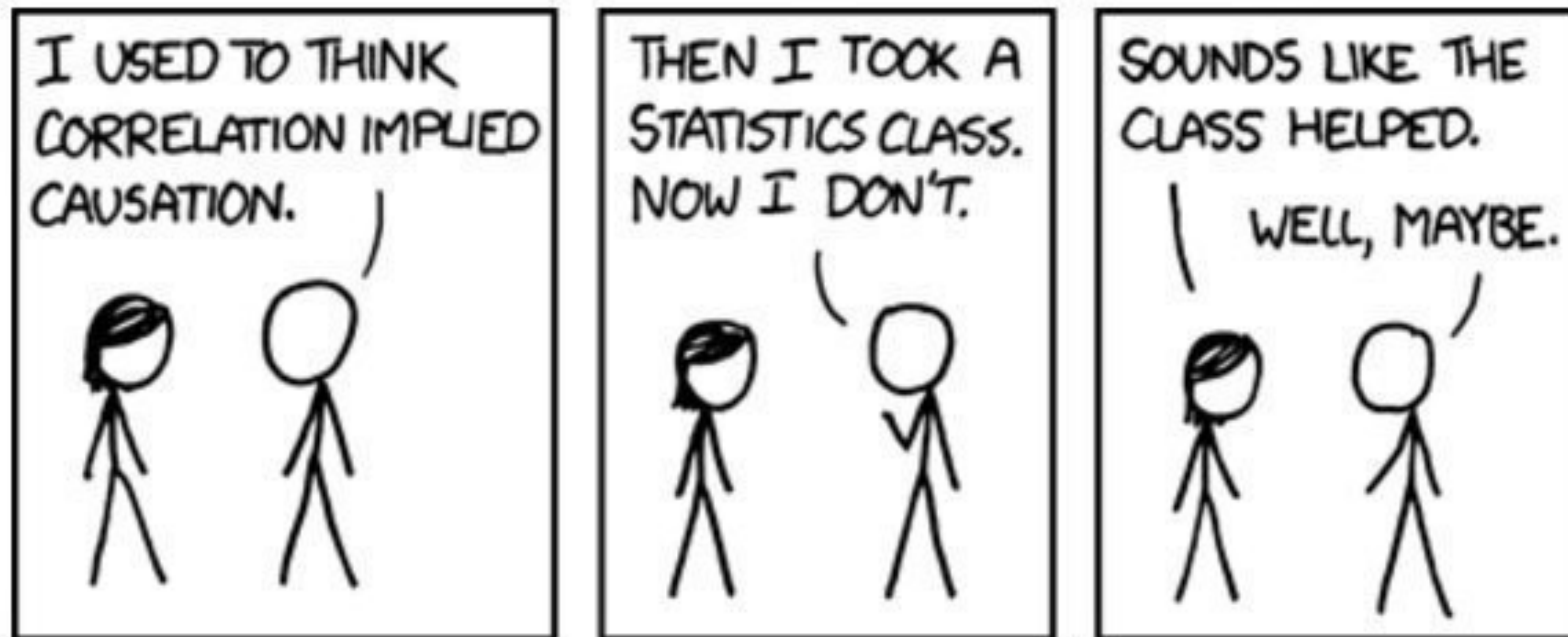


# Разбиение по группам

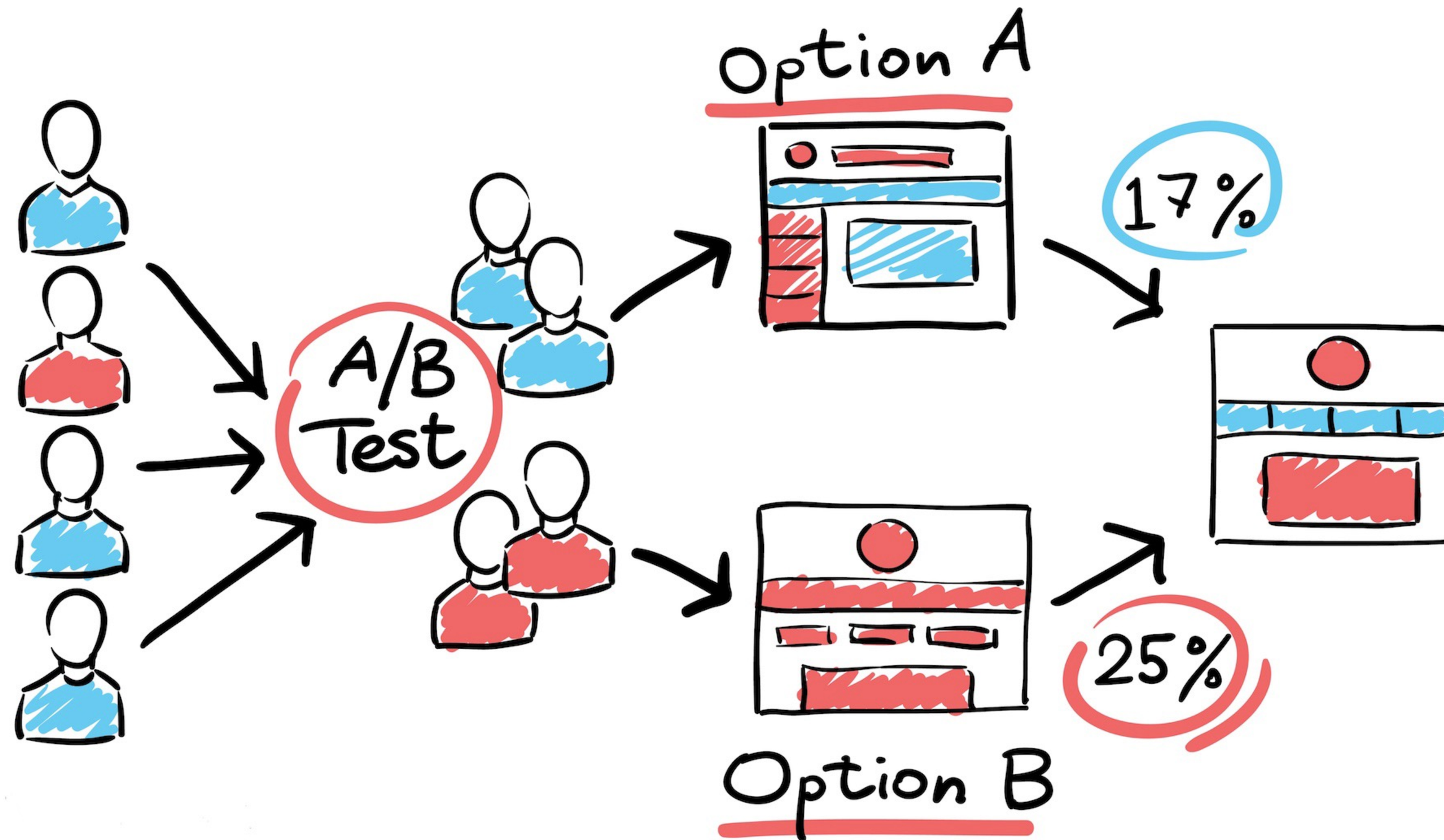




# Корреляция и причинно следственная связь

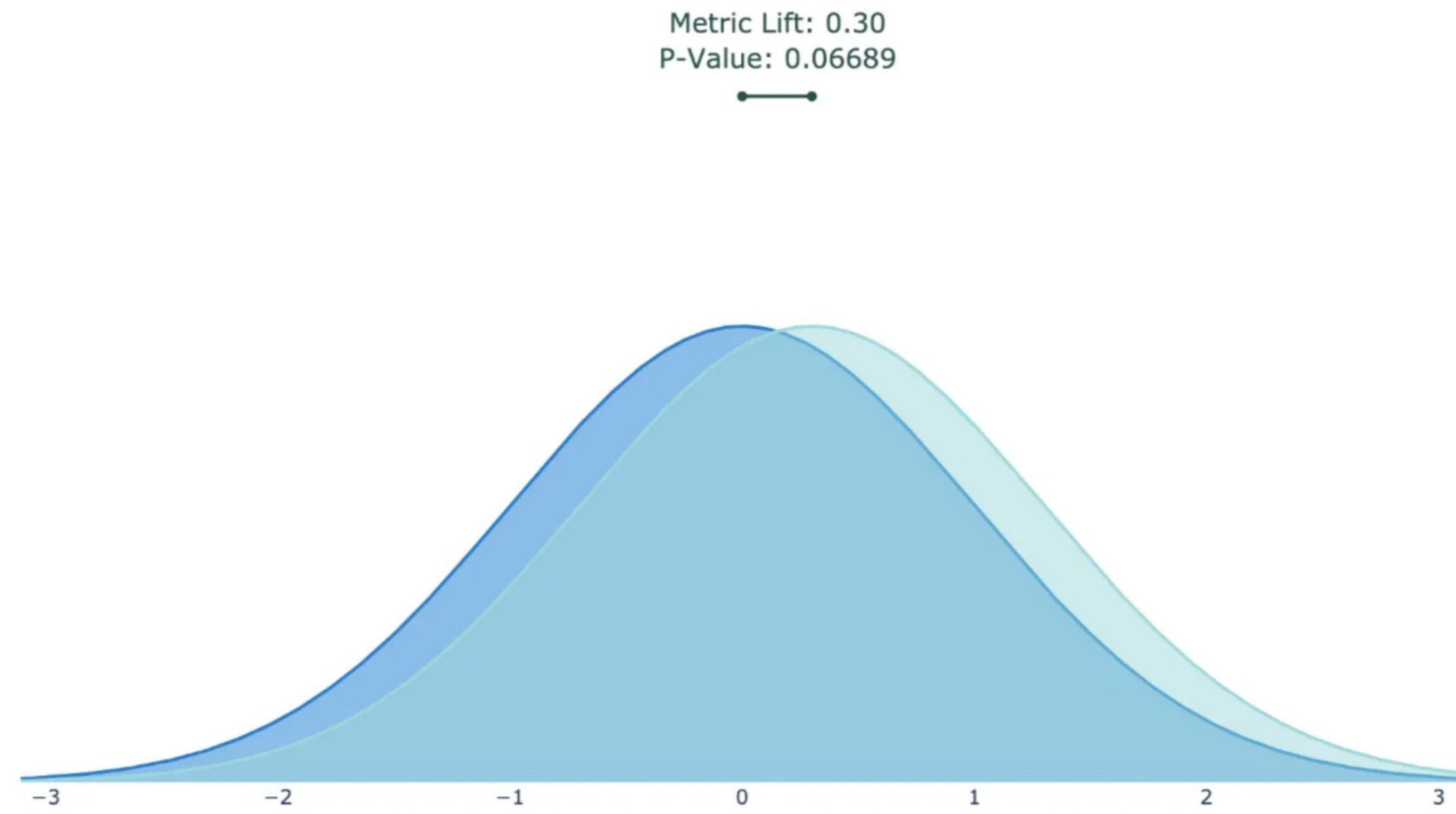


# АВ тестирование

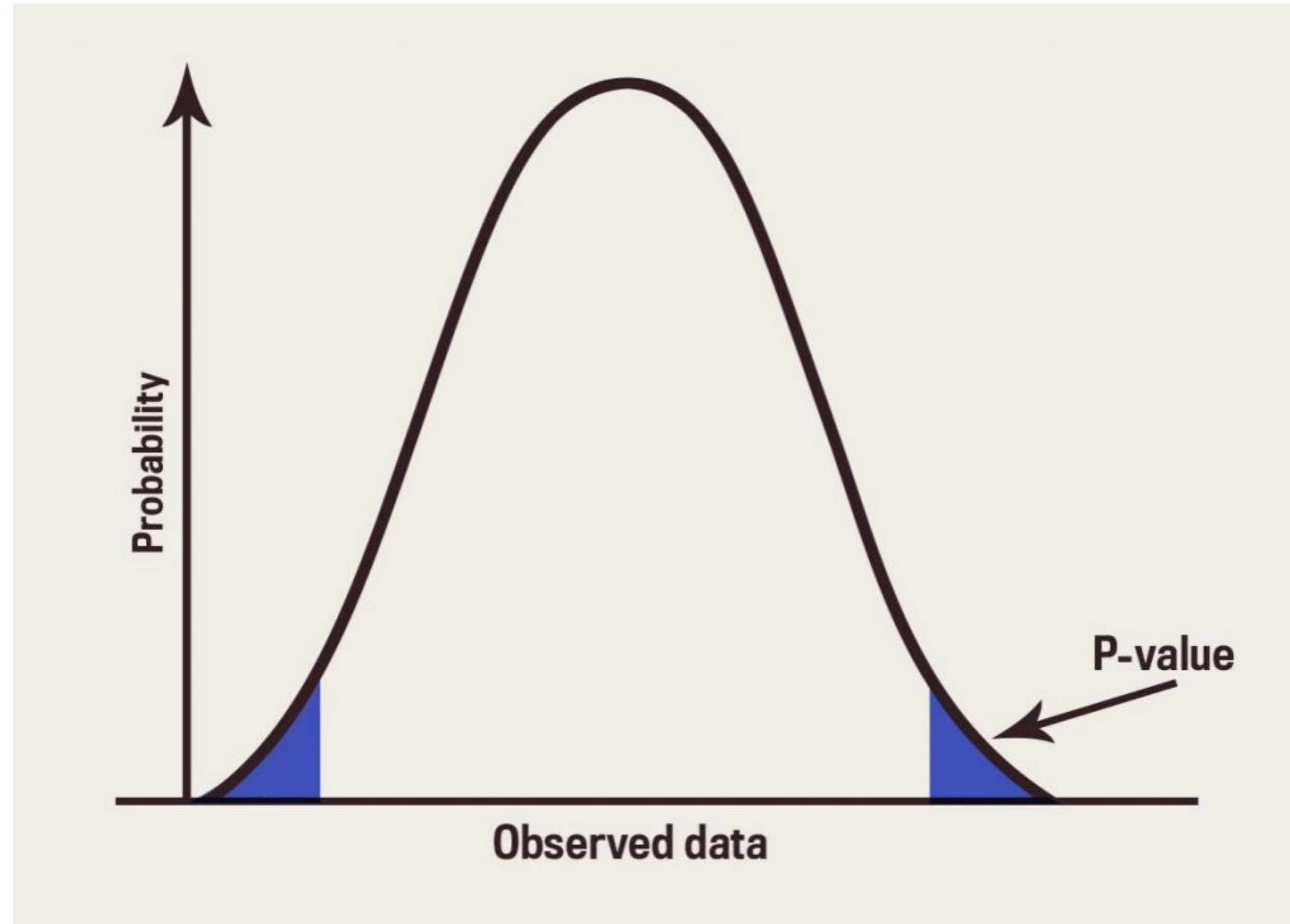


# T-test

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$



# P-value вероятность ошибки первого рода



# Сегментация

## Разбивать желательно на независимые объекты

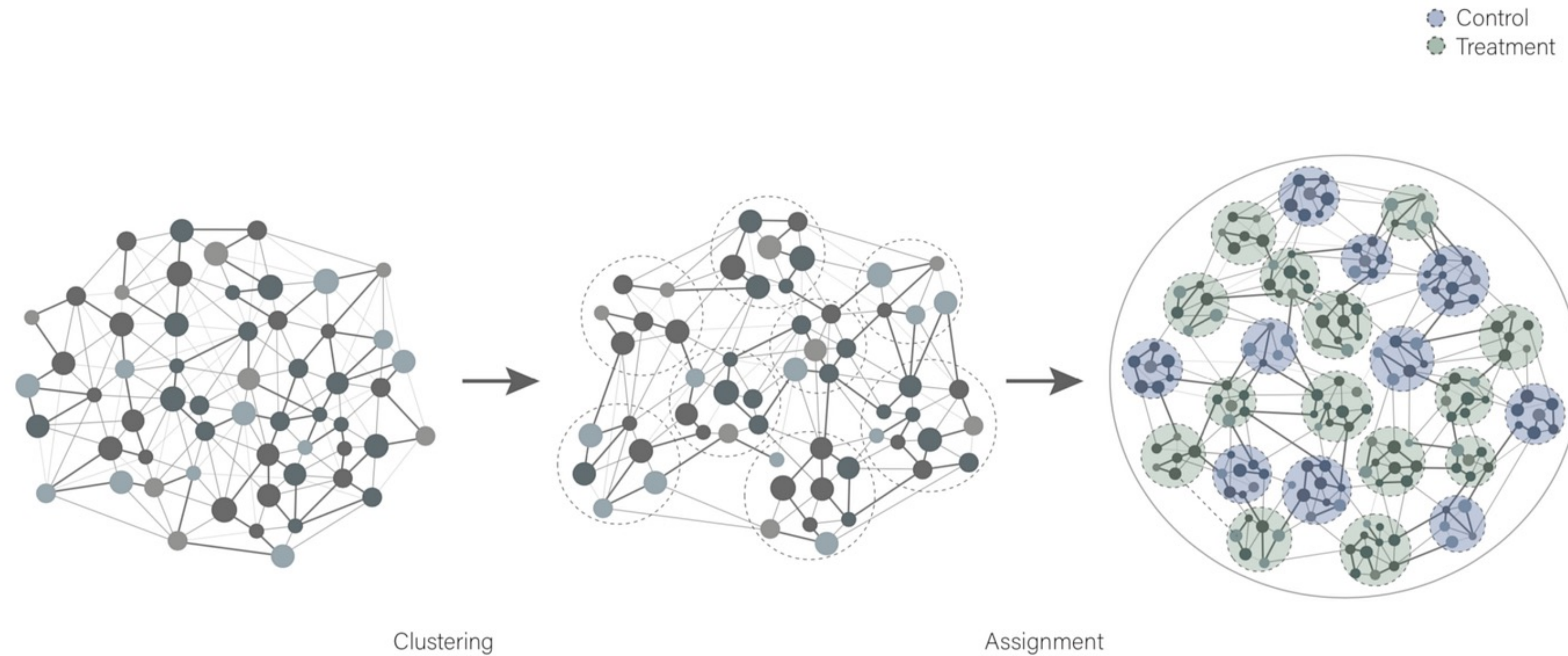
- › По пользователям
- › По водителям
- › По запросам
- › ...

## Разбиение случайное

- ›  $\text{Hash}(\text{id} + \text{salt}) \% n$



# Не всегда можно сегментировать



# Статистические тесты

## **T-test**

- › Хорошо интерпретируется, но не всегда применим

## ***Mann–Whitney U test***

- › *Плохо интерпретируется, но почти всегда применим*

## ***Bootstrap***

- › *Вычислительно дороже, можно применять даже для экзотических статистик*

# Срезы

**Фича может влиять только на часть пользователей**

› Например, только на пользователей android

**На разных срезах влияние может быть разным**



02



# Платформа АВ тестирования

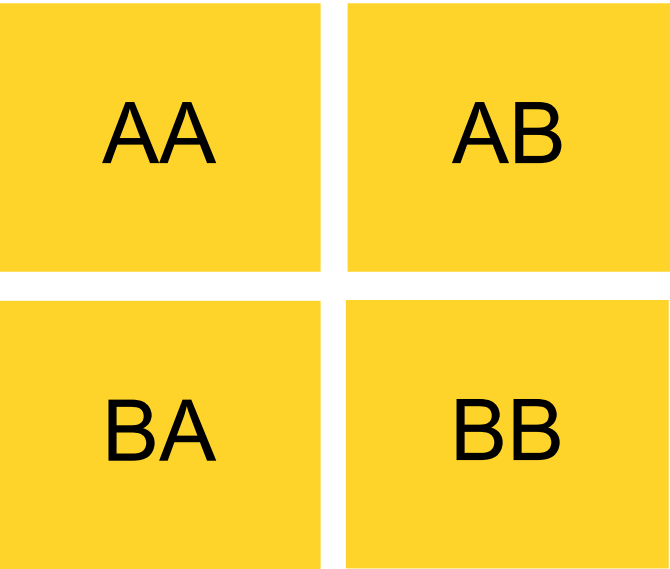
**Задача – ускорение проведения экспериментов**

# Схемы проведения экспериментов

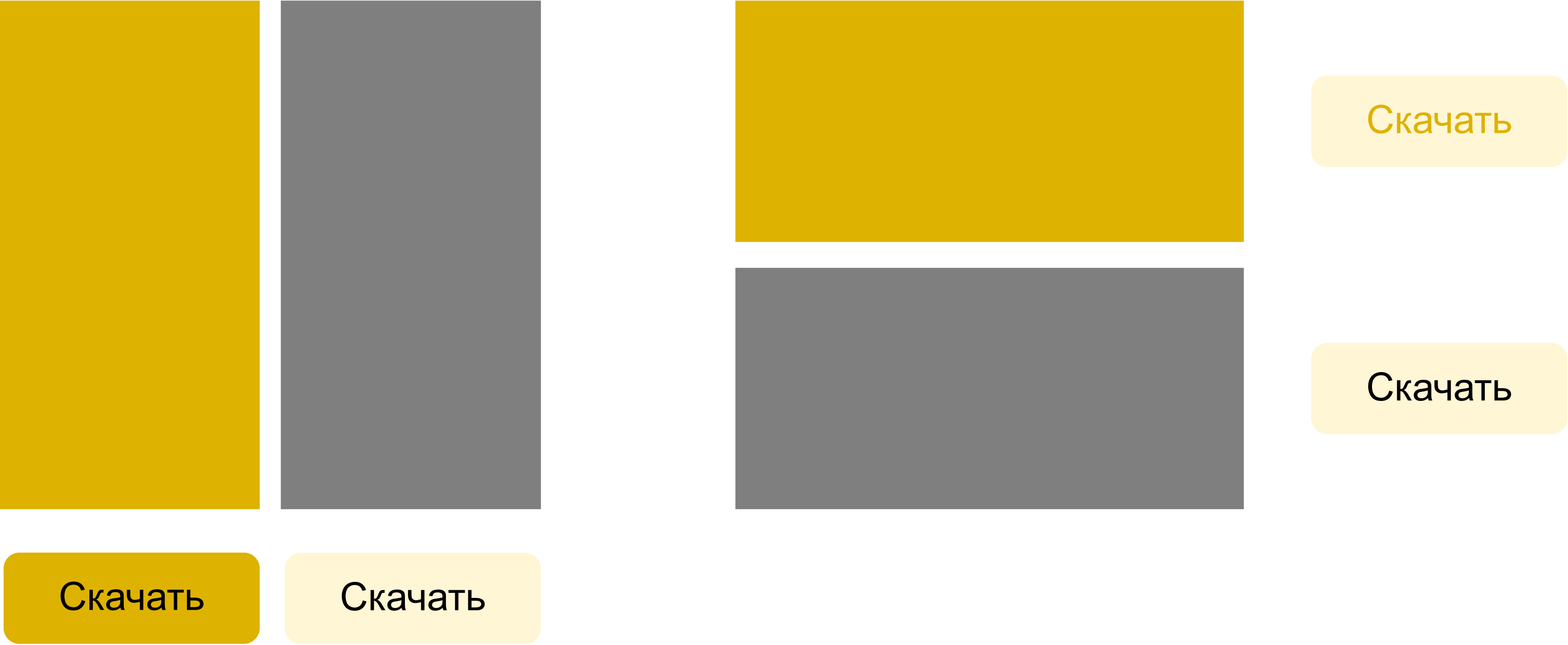
**Одномерная**



**Многомерная**



# Проблема многомерной схемы



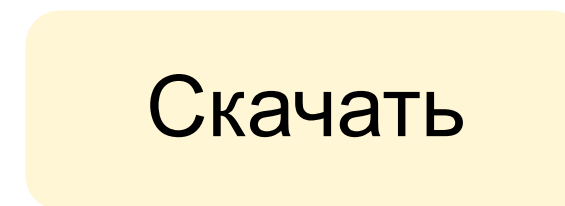
# Проблема многомерной схемы



Скачать



Скачать

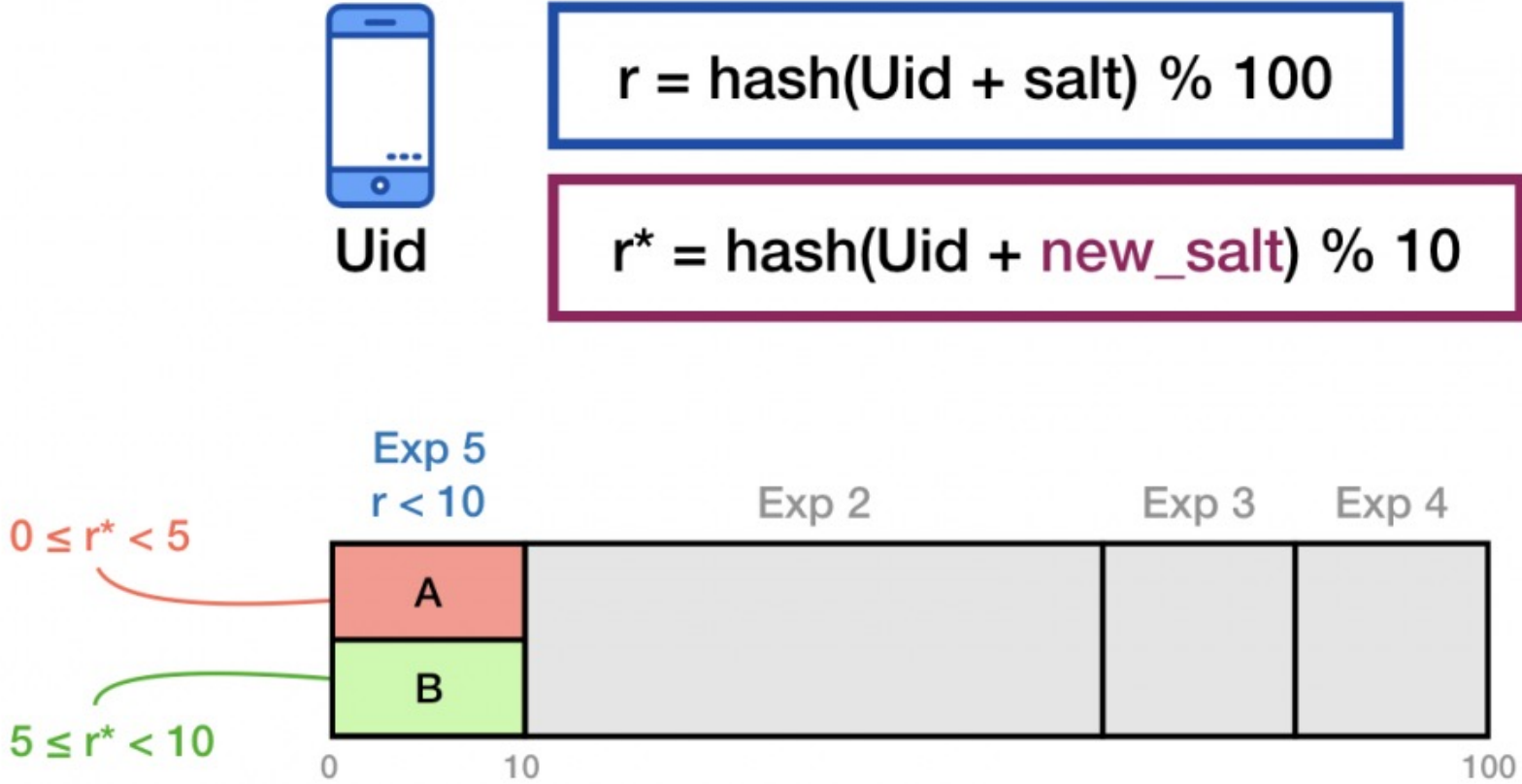
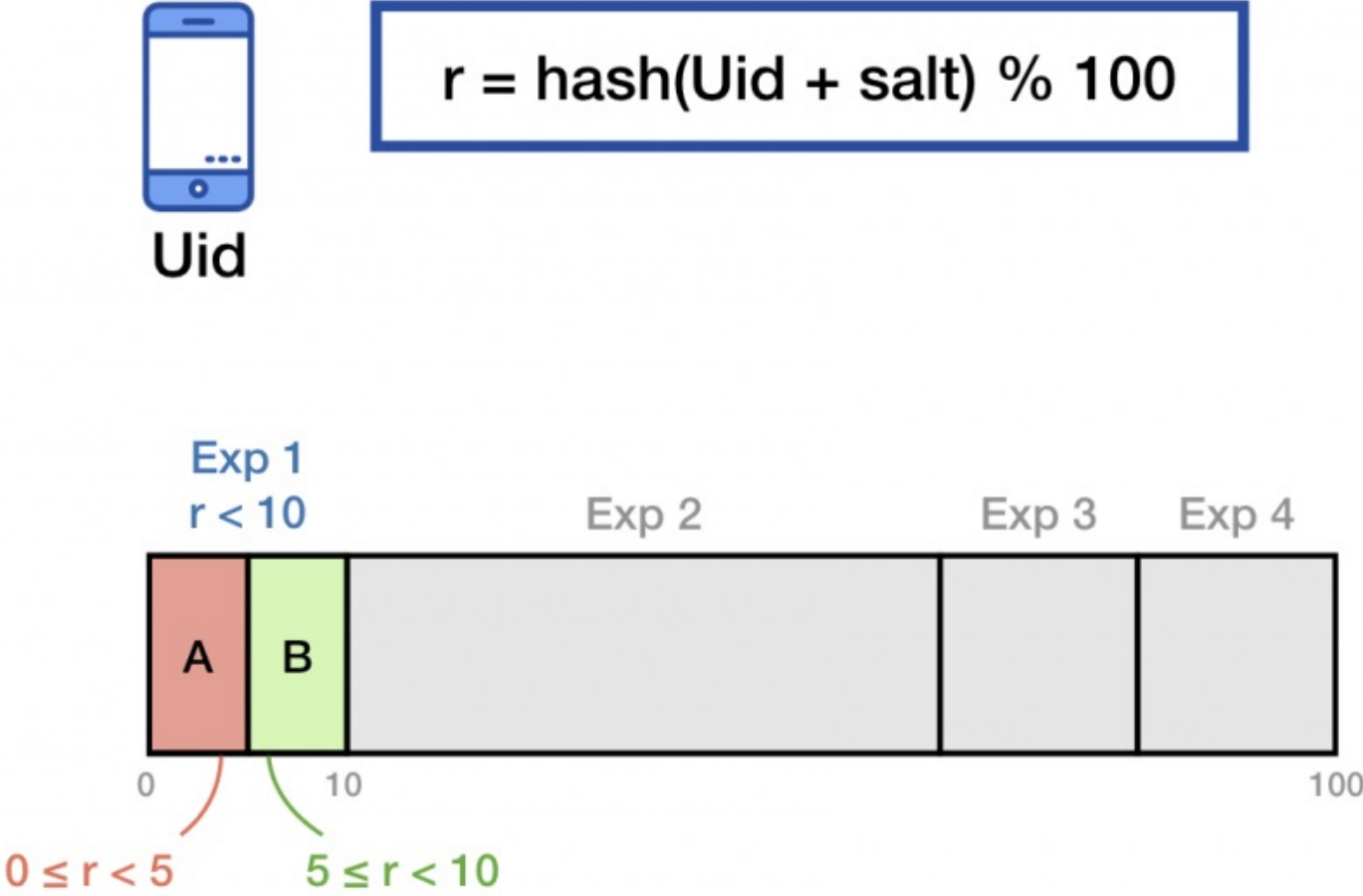


Скачать

# Измерения

- | Разделение экспериментов на измерения
  - › Например: дизайн, ранжирование, контент, ...
- | Несовместимые эксперименты проводятся только в одном измерении

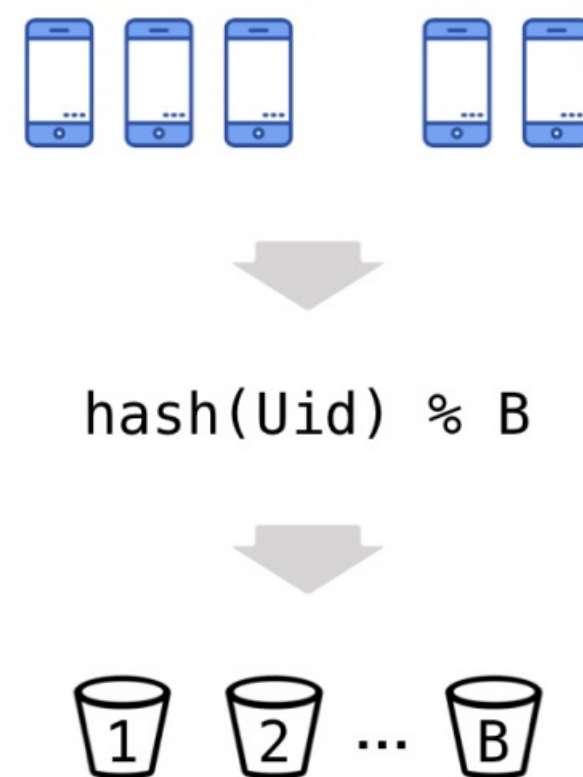
# Память эксперимента



# Вычислительная оптимизация

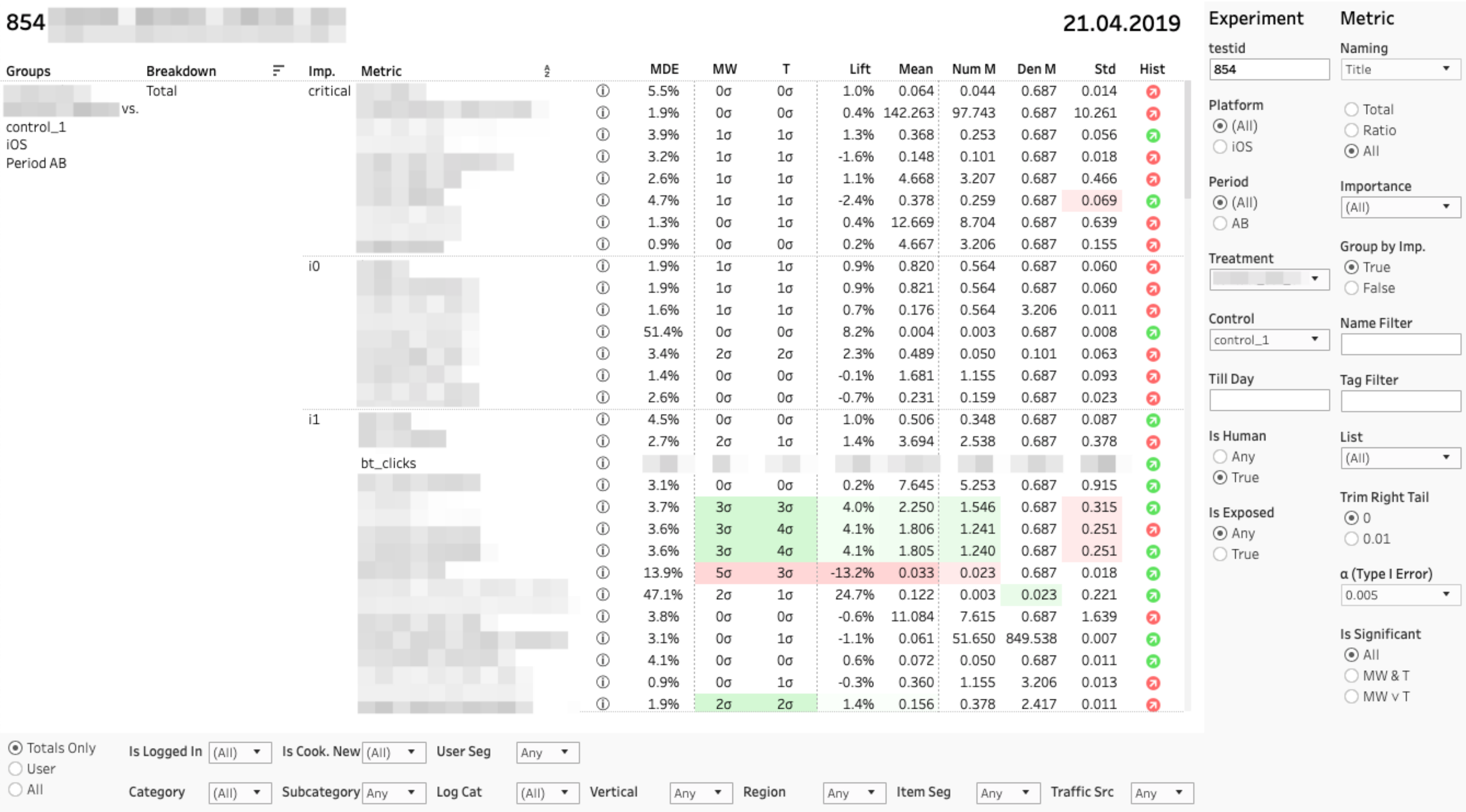
Большое количество метрик, экспериментов, срезов

Бакетное преобразование позволяет уменьшить объем хранимых данных





# Пример платформы АВ тестов



# Полезные свойства АВ платформы

**Удобство и скорость проведения экспериментов**

**Переиспользование метрик других команд**

**Автоматически откатывать эксперименты в критических ситуациях**

**Алерты ответственных за метрики в случае прокрасов**

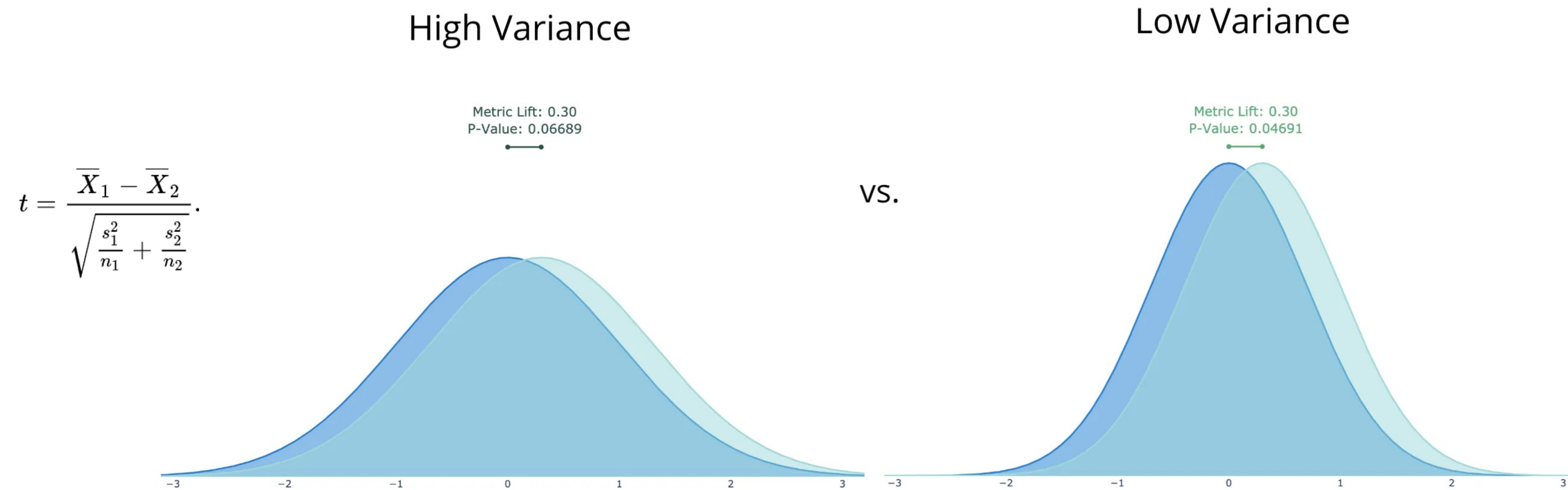
**Единообразное логирование экспериментов**

03



# Разработка метрик

# Уменьшение дисперсии метрики



# Вычитание прогноза

User	Actions (before exp)	Actions	Diff
1	10	12	2
2	1	1	0
3	20	19	-1
4	3	5	2

**avg = 9.25 std = 6.87    avg = 0.75 std = 1.30**

# Увеличение чувствительности

Вычитание прогноза и CUPED (Controlled-experiment Using Pre-Experiment Data)

›  $\hat{Y}_i = Y_i - \theta X_i$

Стратификация

› Перевзвешиваем объекты в тесте так, чтобы распределение определенного параметра совпадало в тесте и контроле

Линеаризация и Дельта метод

›  $\mathcal{R}(U) = \frac{\sum_{u \in U} X(u)}{\sum_{u \in U} Y(u)} \quad L_{X,Y,\alpha}(u) = X(u) - \alpha Y(u) \quad \forall u \in U$

# Разработка метрик

- Собираем пул АВ экспериментов с известным вердиктом

- Подбираем метрику, которая правильно упорядочивает все АВ эксперименты

- И как можно более чувствительная

# Ухудшающие эксперименты

- Для накопления пула экспериментов с известными вердиктами

- Проверка что отключение полезных фичей все еще роняет метрики

- Обратный с отключением всех фичей за длительный период, чтобы убедиться, что суммарно внедрения приносят пользу



# Резюме

- 1 | АВ тесты
- 2 | Платформа АВ тестирования
- 3 | Разработка метрик

# Полезные ссылки

## **АВ тесты в Яндекс (2017)**

› <https://habr.com/ru/companies/yandex/articles/342704>

## **АВ тесты в Авито (2019)**

› <https://habr.com/ru/companies/avito/articles/454164>

## **Про CUPEd (2020)**

› <https://habr.com/ru/companies/yandex/articles/497804>



# Вопросы