

Региональные АВ-тесты. Зачем нужны и как устроены.

Красовский
Игорь
Старший
аналитик

Немного о себе и команде

- Работаю в [АВИТО](#)
- Команда Core Analytics, которая отвечает за:
 - Data Management
 - Strategic Analytics
 - Core Analytics Platform
 - Key Account Analytics



Региональные АВ- тесты. Для каких задач можно применить

Рассмотрим проблемы, которые
требуют особого подхода.



Предположим, что есть задача: измерить эффект от наружной или ТВ-рекламы. Можно ли учесть внутренние и внешние факторы так, чтобы полученный эффект действительно отражал воздействие от рекламы?



Одновременный тест на покупателей и продавцов

Представим, что хотим запустить доставку для заданных категорий объявлений. Какой дизайн АВ-теста вы бы предложили?



Смещенная выборка пользователей

В ценовых и монетизационных тестах приходится часто иметь дело с региональными рынками. Как можно учесть неслучайность тестовой группы в АВ-тесте?

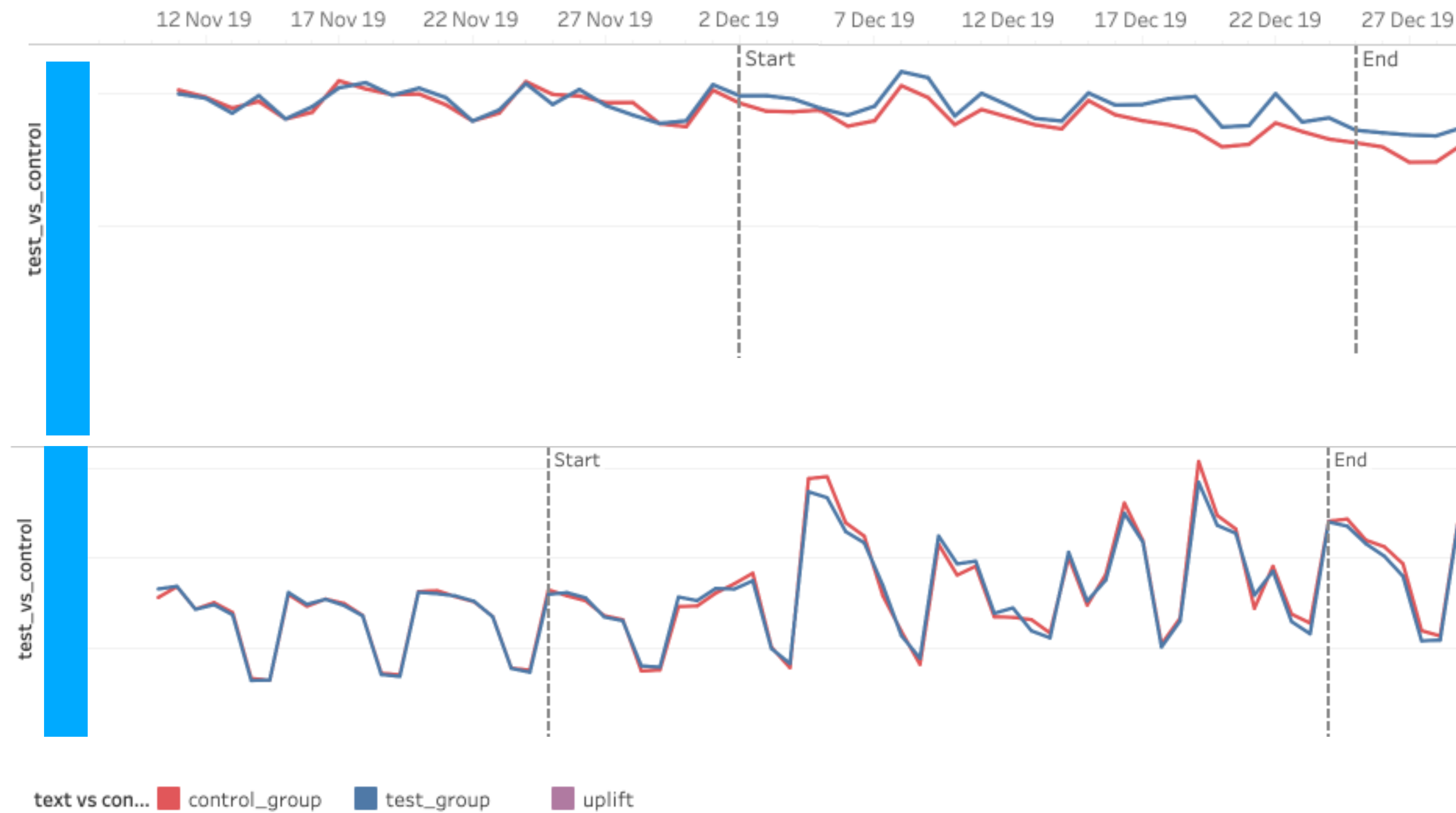


О модели региональных АВ- тестов

“Жадно” подбираем контрольную группу, проводим АА-тест на исторических данных, измеряем точность алгоритма бутстрэпом.

Этапы проведения теста

- Выбор тестовой группы, метрик и разрезов данных
- Подбор контрольной группы
- Проведение АВ-теста
- Измерение ошибки

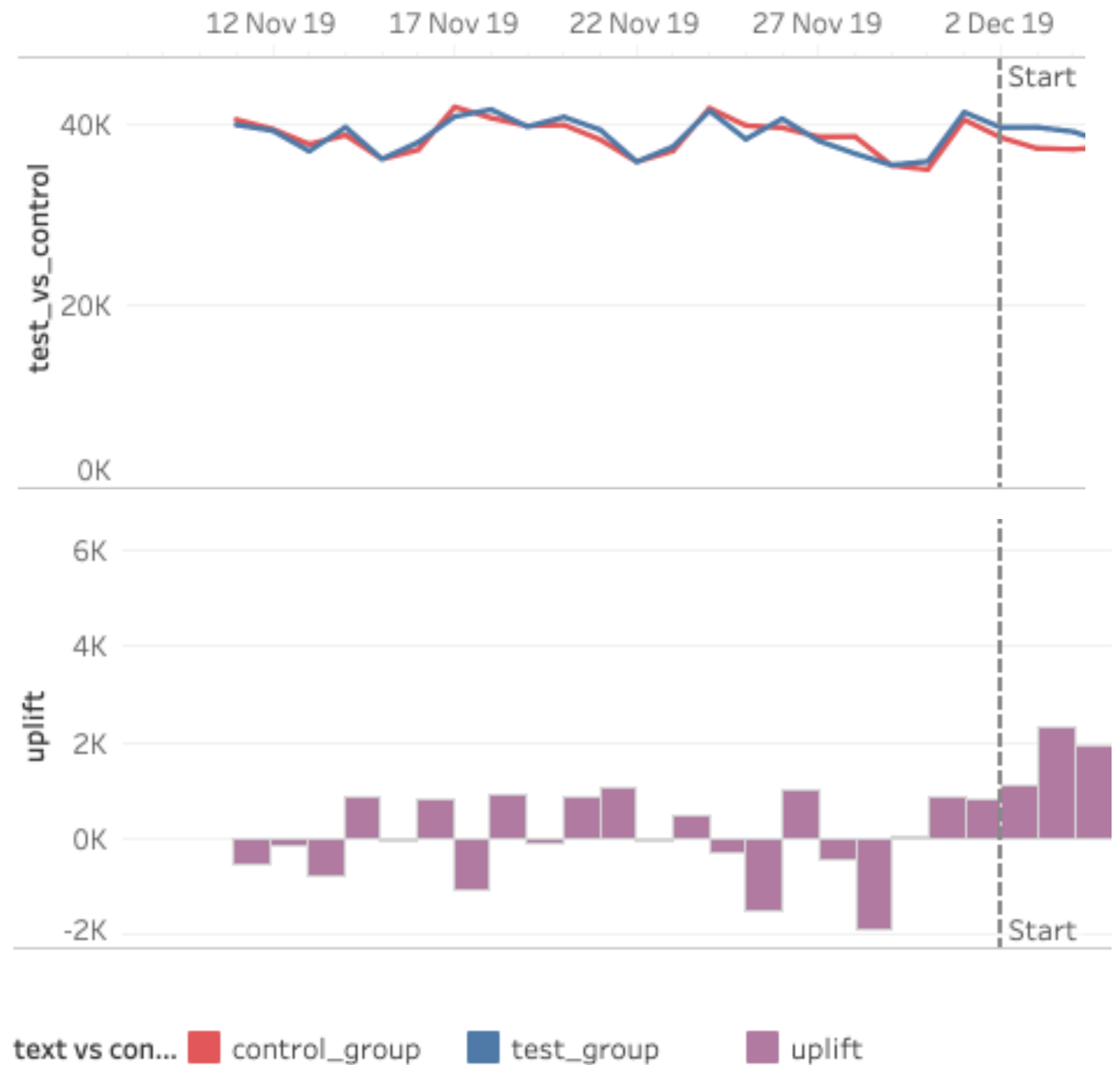


Метрика близости контрольной и тестовой групп

Ожидания старта кампании/теста:

- Схожая динамика тестовой и контрольной групп
- Маленькое подневное отклонение тест vs контроль
- Постоянное отношение значений метрики тест vs контроль

Итого: минимизируем дисперсию подневногo отношения тест vs контроль до начала кампании/теста

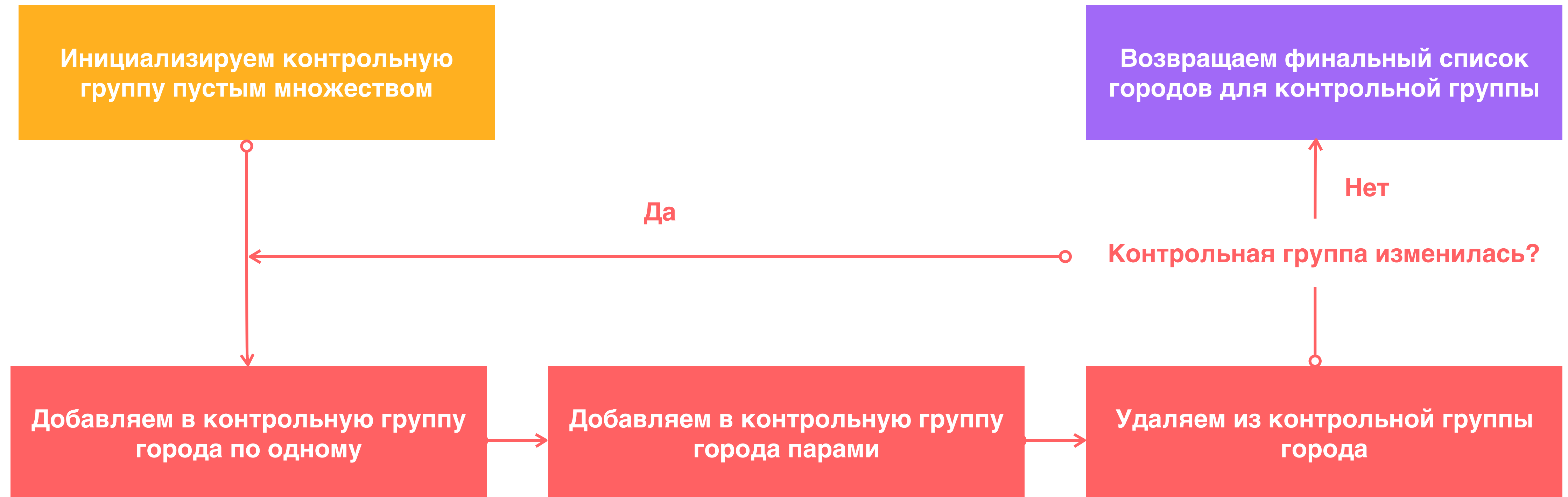


Алгоритм подбора контрольной группы и оценка его точности

Как именно формируем
контрольную группу? Как оценить
точность работы алгоритма? Как
интерпретировать эту точность?



Схема работы алгоритма формирования контрольной группы



На каждом шаге алгоритма контрольная группа последовательно “приближается” к тестовой.



Ретроспективный АА-тест

Алгоритм справляется?

Ключевая проблема:

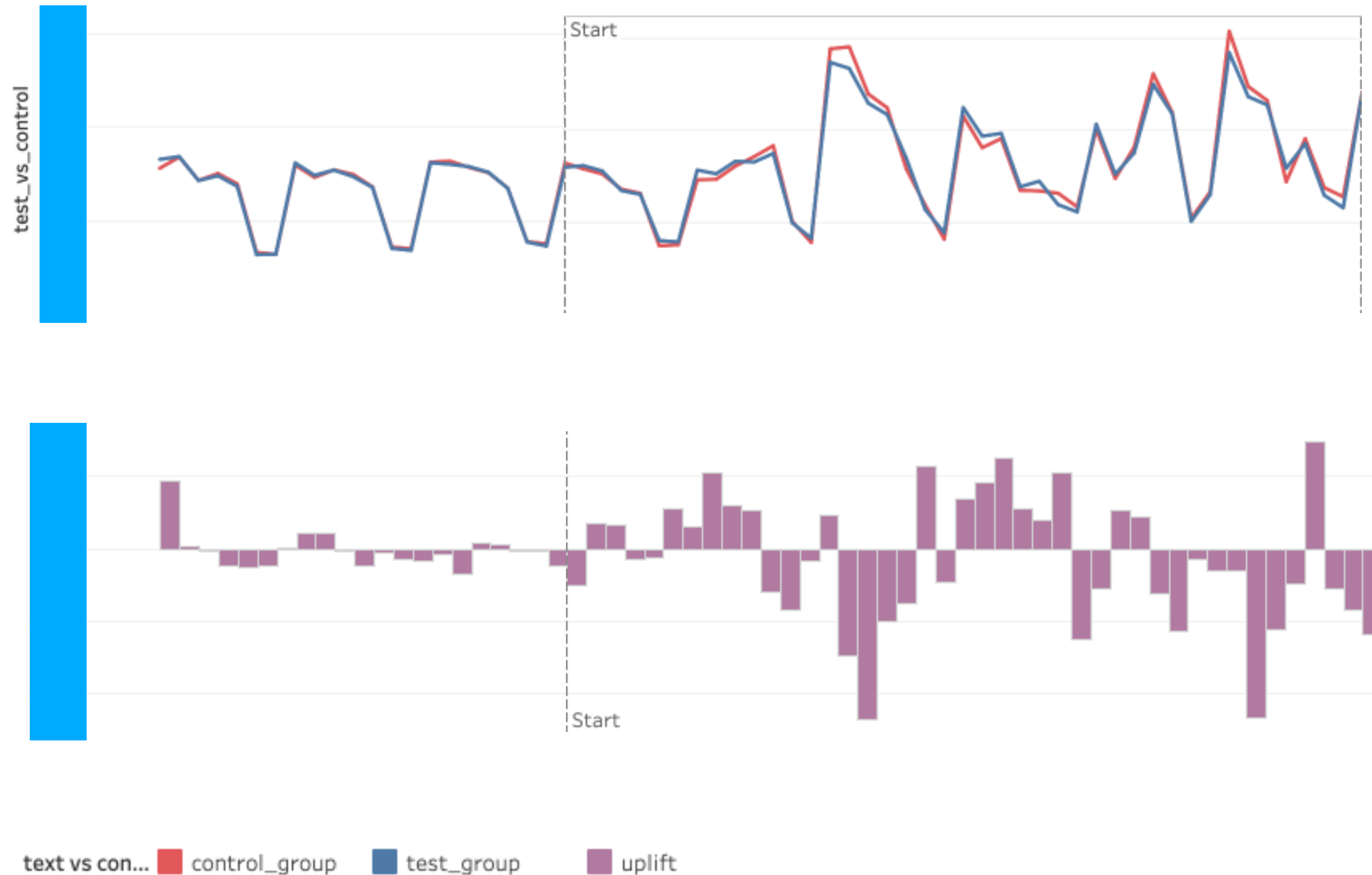
Метрики в группах в АА-тесте на фейковом АВ-периоде отличаются. Сильно или слабо – как понять?

Этапы АА-теста:

- Обучение в АА-тесте
- Фейковый АВ-период

Разница (аплицт):

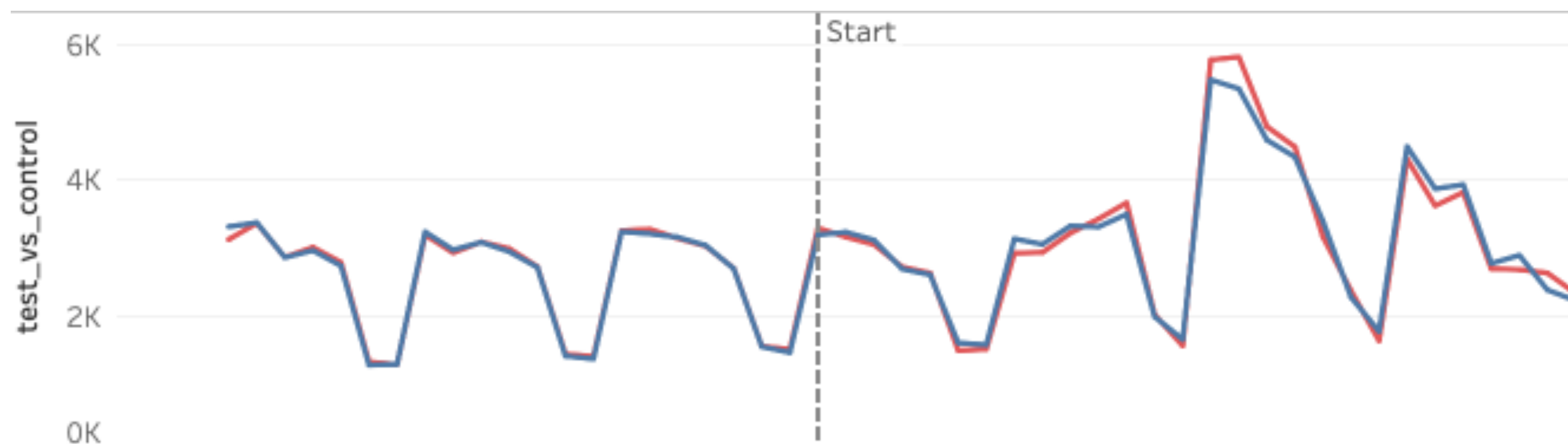
- На обучение в АА-тесте
- На фейковый АВ-периоде



Оценка точности алгоритма

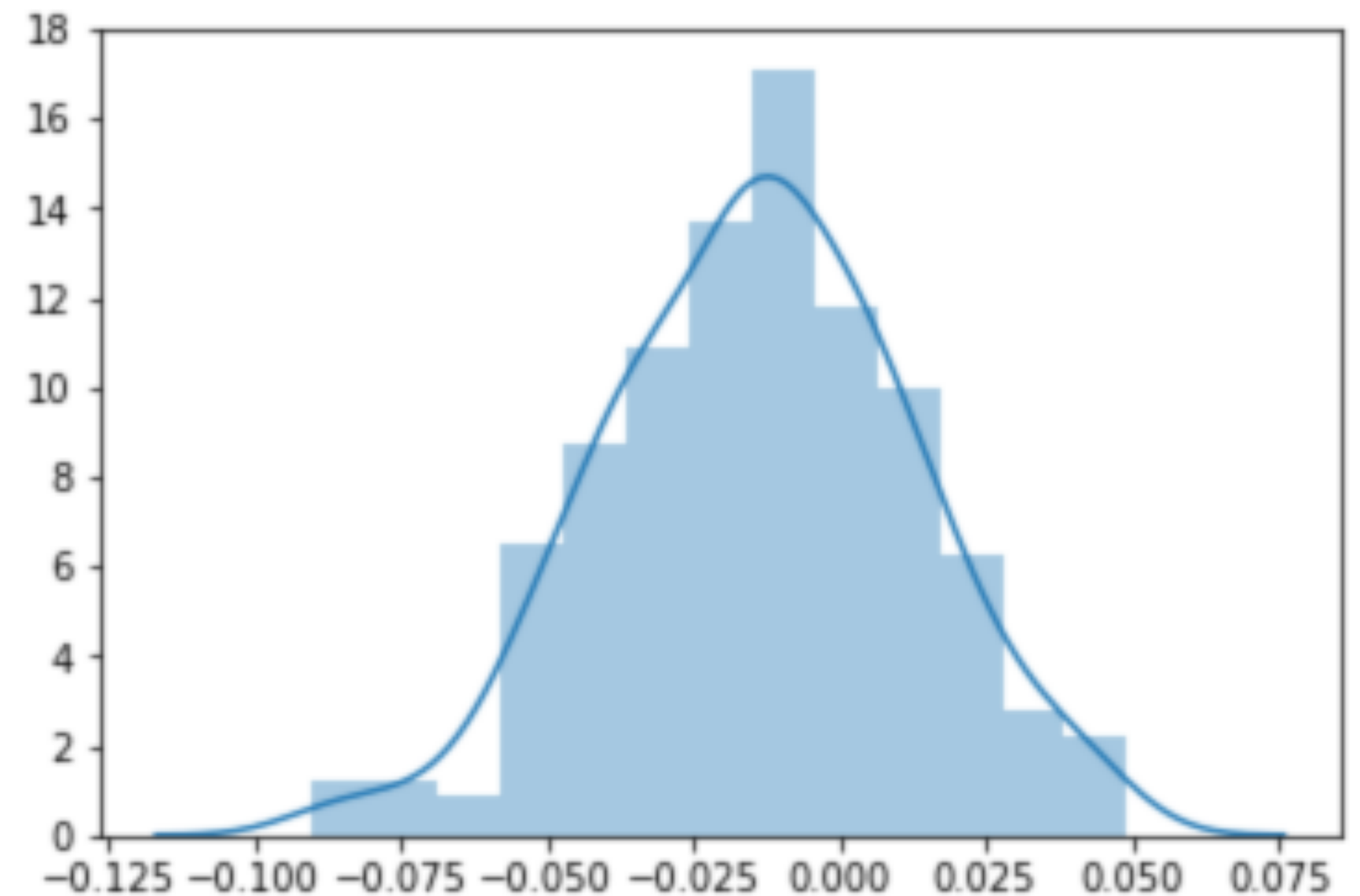
Как получаем распределения ошибок:

- Случайным образом выбираем тестовую группу
- Проводим АА-тест
- Различие в группах на обучающем и фейковом АВ-периодах определяет величину ошибки



Идея:

По историческим данным строим бутстрэп распределение ошибок алгоритма и доверительный интервал.

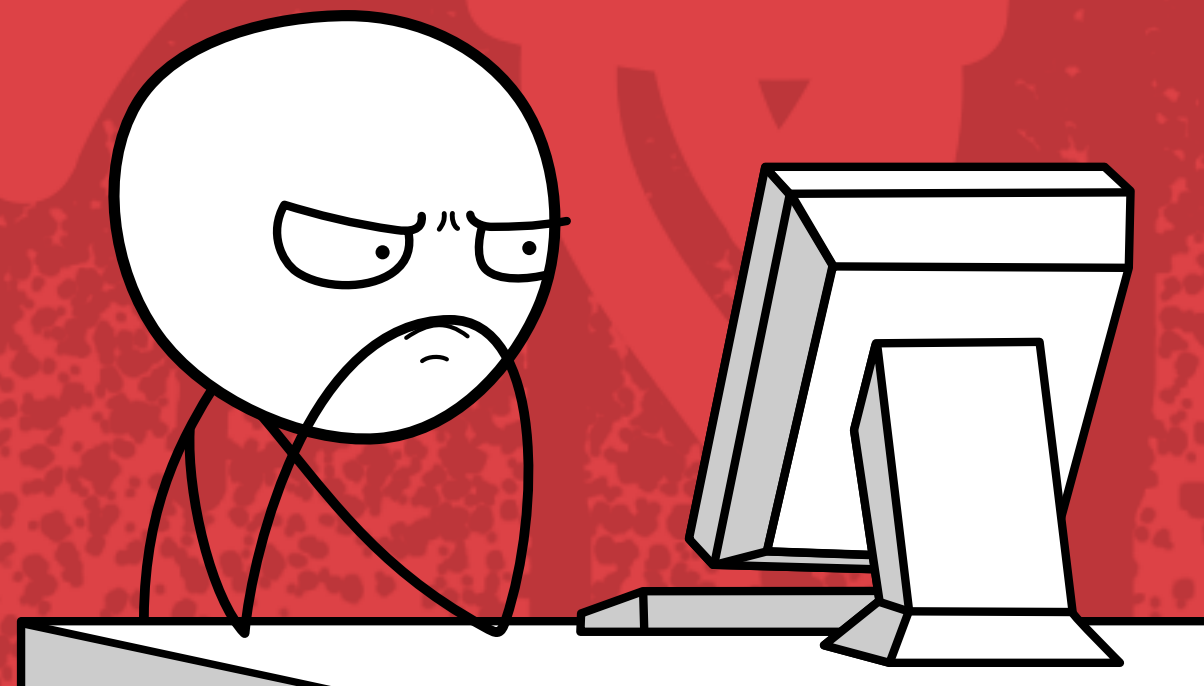


А можно лучше?

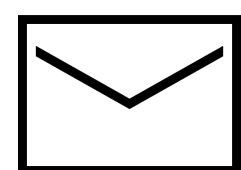
Точки роста:

- Качественно сегментируем данные;
- Фильтруем выбросы;
- Добавляем скоррелированные переменные;
- Присваиваем веса городам;
- Используем алгоритмы ML.

Ещё чуть-чуть...



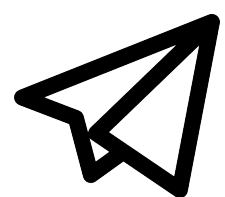
Спасибо



ivkrasovski@avito.ru



krasovskii-igor-378668109



@HagenIK

Красовский Игорь
Старший аналитик