

BiasConf 2025

Когда классический А/В-тест невозможен: разбираемся в Propensity Score Matching

Терентьев Алексей,
Senior Data Analyst, Yandex



Содержание

- 01** Зачем нужен Propensity Score Matching (PSM)?
- 02** Теория PSM
- 03** Алгоритм исследования с помощью PSM
- 04** Как доказать, что PSM работает?
- 05** Общие рекомендации
- 06** Вопросы

01

Зачем нужен Propensity
Score Matching (PSM)?



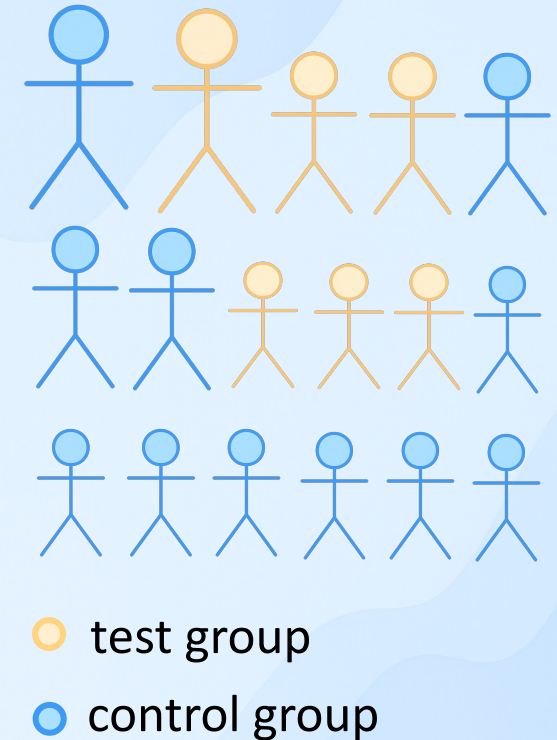
Какую задачу решаем?

Две группы пользователей:

- Тестовая группа — получила воздействие (новая функция, акция).
- Контрольная группа — не получила воздействие.

Задача:

- **Концептуально:** оценить эффект от воздействия на тех, кто его получил.
- **Технически:** понять как воздействие повлияло на бизнес-метрики.



Всегда ли поможет A/B-тест?

Разделение на тест и контроль происходит случайно:

- Пользователи в тесте и контроле НЕ отличаются по своим характеристикам.
- Можно просто сравнить результаты в тесте и контроле.
- Это стандартный A/B-тест.

PSM НЕ нужен

Разделение на тест и контроль происходит НЕ случайно:

- Пользователи в тесте и контроле отличаются по своим характеристикам.
- Эти различия искажают эффект от воздействия (selection bias).
- A/B-тест не применим.

PSM нужен



Поговорим про этот кейс

Когда полезен PSM?

Проведение A/B-теста затруднительно или вовсе невозможно:

- A/B-тест невозможен по техническим, бизнес или этическим причинам.
- Естественное (неслучайное) формирование групп.
- Результат нужен быстро, нет времени на проведение A/B-теста.
- Необходима ретроспективная оценка изменений.

Улучшение точности A/B-теста:

- В проведенном A/B-тесте получилась высокая дисперсия оценок.
- A/B-тест проведен, но есть подозрения, что группы отличаются.

Пример применения PSM

Контекст:

В сервисе существует платная премиум-подписка. Требуется оценить влияние оформления подписки на поведение пользователей.

Проблема:

- Подписку оформляют сами пользователи → провести A/B-тест невозможно.
- Подписчики и неподписчики могут отличаться по характеристикам → сравнение некорректно.
- После оформления подписки поведение пользователей может измениться под влиянием внешних факторов → сравнение до/после некорректно.

Как подойти к задаче:

- Нужно сравнить подписчиков с максимально похожими на них неподписчиками — используем метод Propensity Score Matching (PSM).

02

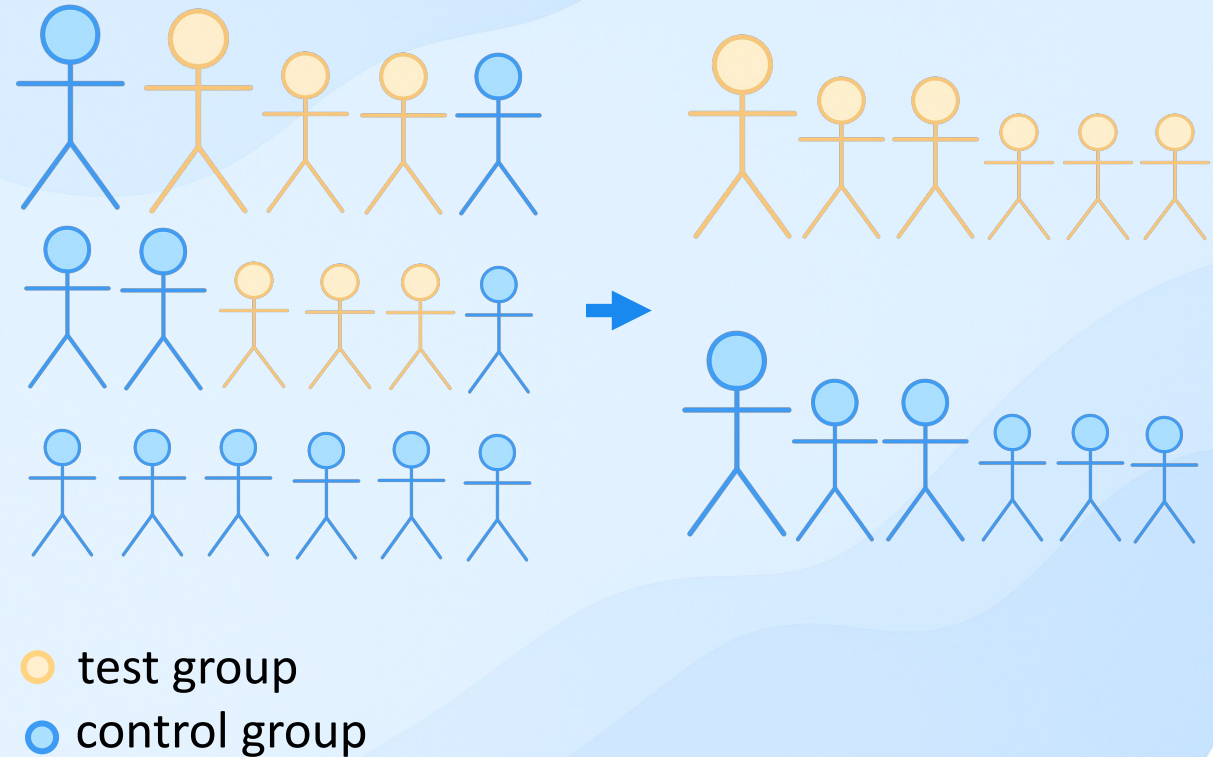
Теория PSM



Propensity Score Matching (PSM)

Propensity Score Matching: статистический метод для сравнения групп, уменьшающий систематические различия между ними.

Идея метода: сравнивать пользователей из тестовой группы только с максимально похожими пользователями из контрольной группы.



Что анализируем?

Есть данные:

- Есть выборка пользователей, часть получила воздействие (Тест), другая — нет (Контроль).
- Для всех пользователей известны исходные характеристики (пол, возраст, регион, история покупок и т.д.).
- Есть бизнес-метрика, эффект на которую мы хотим измерить в рамках нашего исследования.

Проблема:

- Пользователи в тесте и контроле отличаются по своим исходным характеристикам.

Задача:

- Измерить эффект от воздействия на бизнес-метрику, минимизировав различия между группами.

Метод PSM

Цель:

- Подобрать среди контрольной группы пользователей, которые были бы похожи на пользователей из тестовой группы по своим исходным характеристикам.

Что делает метод?

- Формирует “псевдо-контрольную” группу: для каждого объекта из теста подберет наиболее похожий объект из контроля.

Работает в два шага:

- Шаг 1: Построить propensity score: оценить вероятность попадания в тест.
- Шаг 2: Выполнить matching: подобрать похожие объекты по рассчитанной вероятности.

Шаг 1: Propensity score

Propensity score (PS) — это вероятность, что пользователь попадёт в тестовую группу, на основе его характеристик.

Как получить Propensity Score?

- Строим модель, которая учится отличать пользователей теста и контроля на основе их характеристик.
- Задача бинарной классификации: предсказывается вероятность попадания в тест.
- Методы: градиентный бустинг, логистическая регрессия, случайный лес и т.д.

Что даёт:

- Агрегирует характеристики объекта в одну фичу — оценку склонности попасть в тест.
- Если модель хорошо предсказывает, значит, разделение на тест и контроль не случайно.

Шаг 2: Matching

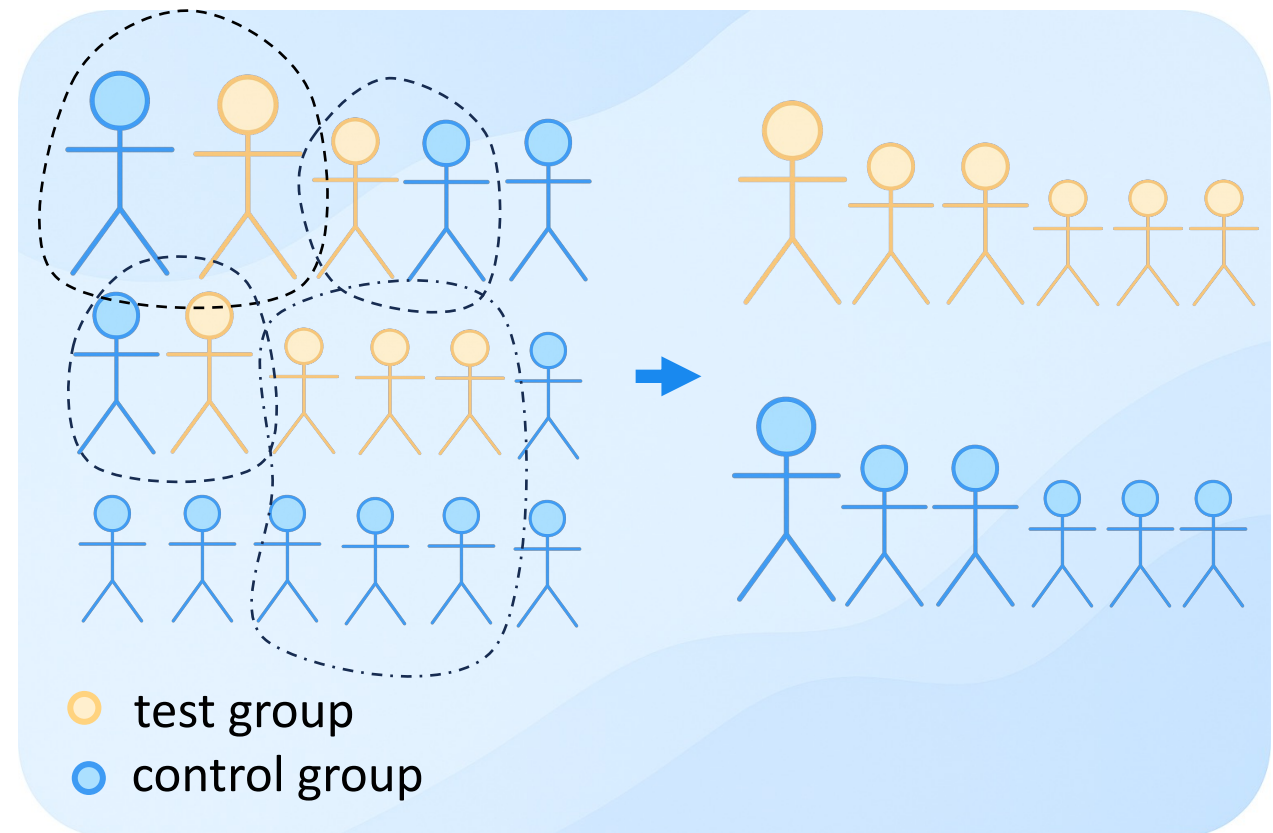
Matching — это процесс подбора объектов из контрольной группы, максимально похожих на объекты из тестовой группы.

Как выполнить Matching?

- Для каждого пользователя из теста ищем схожих по PS из контроля.
- Методы: K-ближайших соседей, Caliper Matching.
- Разрешены повторы контрольных объектов.

Что делает Matching:

- Формируем “псевдо-контрольную” группу, сопоставимую с тестом.



Пример применения PSM

Контекст:

В сервисе существует платная премиум-подписка. Требуется оценить влияние оформления подписки на поведение пользователей.

Решение через PSM:

- **Собираем данные:** выделяем подписчиков и неподписчиков, фиксируем их характеристики (до подписки).
- **Строим propensity score:** обучаем модель, которая рассчитывает вероятность оформления подписки.
- **Проводим matching:** для каждого подписчика находим максимально похожего неподписчика по propensity score, формируем “псевдо-контрольную” группу.
- **Сравниваем поведение:** анализируем разницу между подписчиками и “псевдо-контрольной” группой.

03

Алгоритм исследования с
помощью PSM



PSM: пошаговая инструкция

✅ Собери данные

- Определи treatment date
дата, после которой тестовая группа получает воздействие
- Собери признаки
пользователей ДО этой даты
- Формируй группы
выдели тест и потенциальный контроль

🧠 примени PSM

- Построй PSM-модель
шаг 1: модель propensity score
шаг 2: выполни matching
результат: “псевдо-контрольная” группа
- Проверь качество модели
насколько хорошо она определяет группы

🎯 Применяй и оценивай

- Измерь эффект
оцени метрики в тестовой и “псевдо-контрольной” группе после и оцени эффект
- Исследование готово!

Алгоритм исследования



04

Как доказать, что PSM
работает?



Как доказать, что PSM работает?

1. Проверка на отсутствие изначальных различий

- Исследуемая метрика до treatment date в тесте и псевдо-контроле — не различается.

2. Проверка модели Propensity Score

- Propensity-модель отличает тест от контроля ($\text{ROC-AUC} > 0.6$).

3. Проверка баланса после matching

- Распределения propensity score в тесте и псевдо-контроле не различаются.
- Ключевые признаки в двух группах максимально близки.

4. Проверка размера и структуры групп

- Размер псевдо-контроля \approx тестовая группа.

05

Общие рекомендации



Рекомендации по проведению исследования

1. Дизайн исследования – залог успеха

- Пробуйте разные варианты: контрольные группы, наборы признаков, дату тритмента.

2. Помогайте модели на старте

- Формируйте контрольную группу так, чтобы она была больше похожа на тест.
- Хороший вариант: на шаг позади по воронке.

3. Правильно собирайте данные

- Контрольная группа должна быть больше теста в 5+ раз, чтобы алгоритм мог подобрать подходящие пары для matching.
- Используйте только признаки до воздействия — чтобы избежать утечки и смещения.

4. Не гонитесь за высоким ROC-AUC

Рекомендации по использованию метода

1. Если можно провести A/B-тест — делайте его

- PSM — не замена A/B-тестированию, а способ анализа, когда рандомизация невозможна.

2. PSM требует больше аналитических ресурсов

- Нужно собрать признаки, построить модель, валидировать и интерпретировать результаты.

3. Корректно интерпретируйте выводы

- PSM показывает эффект на пользователях, похожих на тех, кто был в тесте — это не всегда обобщаемо на всю аудиторию.

06

Вопросы



BiasConf 2025

Спасибо
за внимание!

Терентьев Алексей,
Senior Data Analyst, Yandex
Linkedin: aleksei-terentevv

