

Что в реальности даст новая модель?

Думбай А.Д.

ММП ВМК МГУ

20 октября 2020 г.

Задача анализа данных

- Предобработка
- EDA
- Разработка моделей + обучение
- Валидация
- **Тестирование на реальных данных (?)**

Как это делать?

Подходы:

- Отправить в работу и посмотреть что будет
- АВ

No Silver Bullet!

Pros

- Надежность
- Стат. обоснованность (~)
- Обоснованность для принимающих решение

Cons

- Дорого
- Медленно
- Не всегда можно

Область применимости

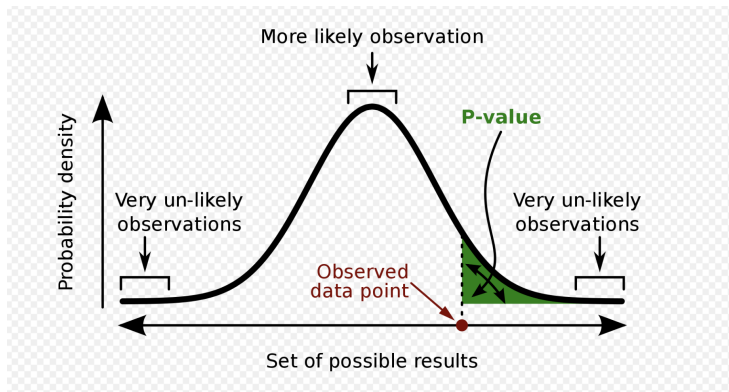
- Есть набор метрик (средние + отклонения)
- Важнее не сломать существующее, чем внедрить новое [Стартап]
- Задача подходит для разбиения [Есть влияние третьей стороны]

Какие этапы возникают:

- Разбиение
- Выкатка функционала
- Проверка стат. значимость (Случайно ли отклонение)
- Анализ эксперимента в целом и принятие решения

Статистическая значимость

- Z-Т-тесты
- U-критерий Манна-Уитни
- p-value
- Bootstrap



Разделение

- Должны быть максимально независимы в рамках цели (!)
- Достаточного объема для значимости
- Могут быть пересекающиеся эксперименты
- Фильтрация и срезы

Интересные статьи на тему

- Offline A/B testing for Recommender Systems
- Adaptive Monte Carlo Multiple Testing via Multi-Armed Bandits (не совсем про АВ, но смежно)
- Lasso adjustments of treatment effect estimates in randomized experiments
- Improving the Sensitivity of Online Controlled Experiments by Utilizing Pre-Experiment Data
- Variance-Weighted Estimators to Improve Sensitivity in Online Experiments