

АБТ – часть 1

Михаил Буряков

АБТ: Основы



АБТ: Основы

- система



АБТ: Основы

- система
- состояние



АБТ: Основы

- система
- состояние
- метрика



АБТ: Основы

- система
- состояние
- метрика
- влияние



АБТ: Основы

- система
- состояние
- метрика
- влияние
- эффект



АБТ: Основы

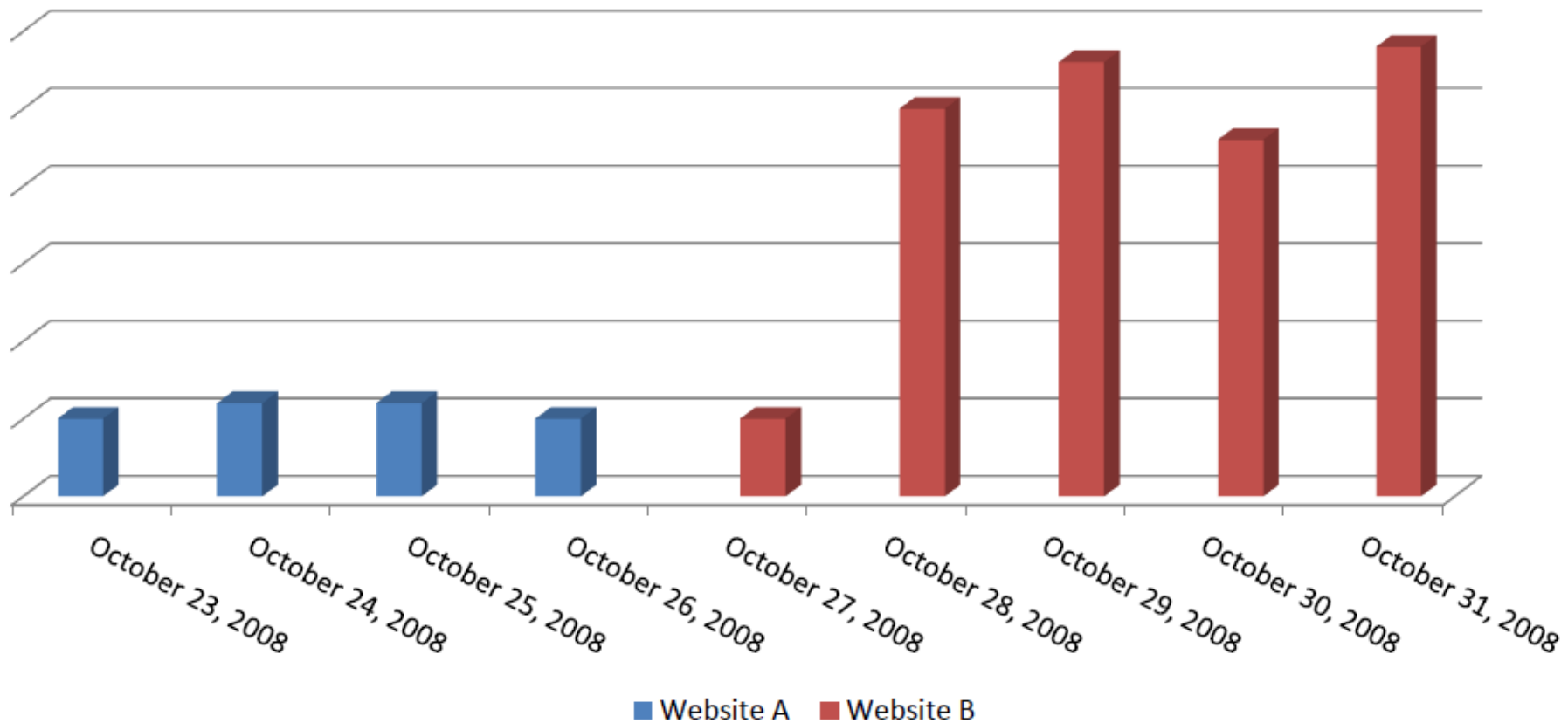
Проблемы:

- корреляция \neq причина
- побочные эффекты & влияние извне

АБТ: Основы

проблемы:

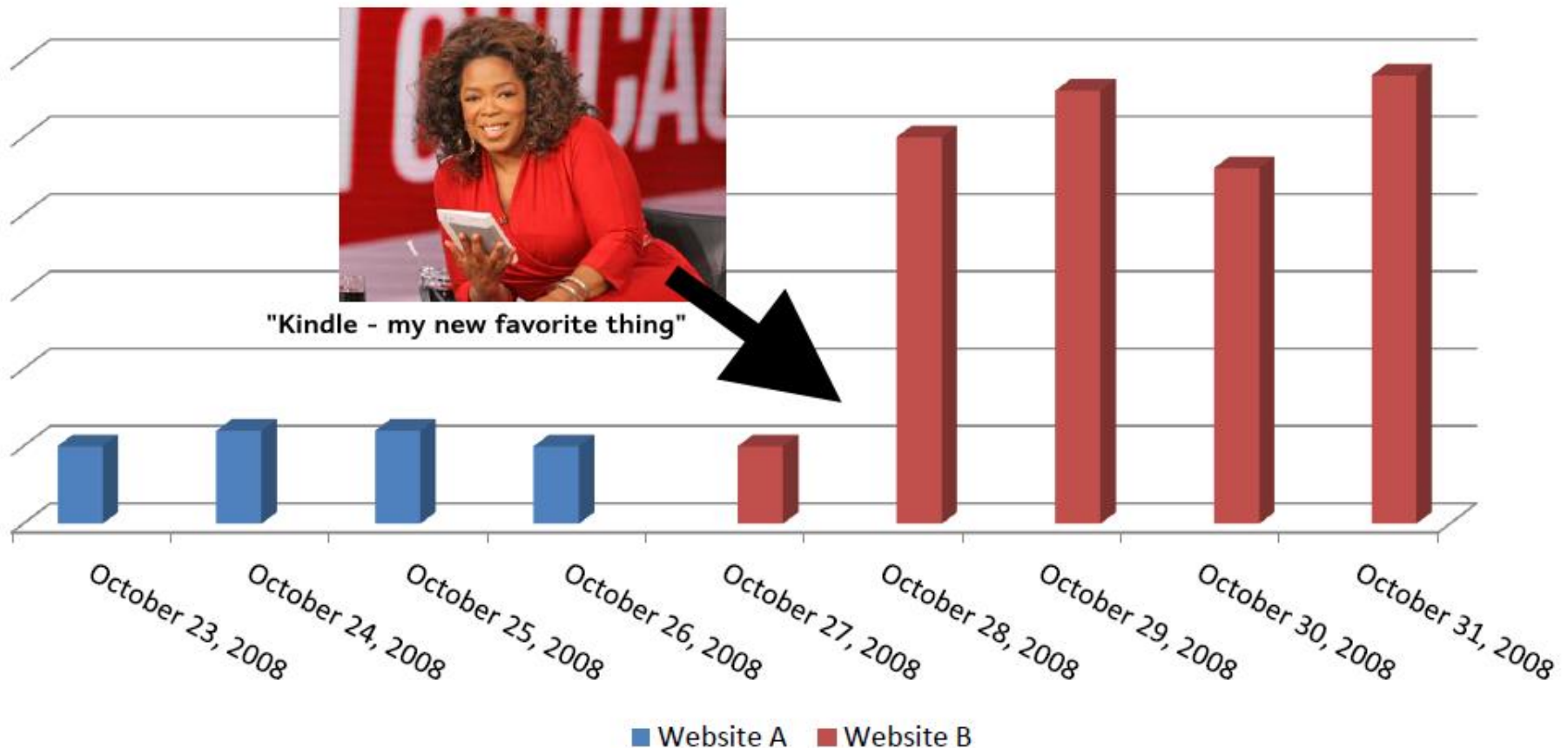
Amazon Kindle Sales



АБТ: Основы

проблемы:

Amazon Kindle Sales



АБТ: Основы

проблемы:

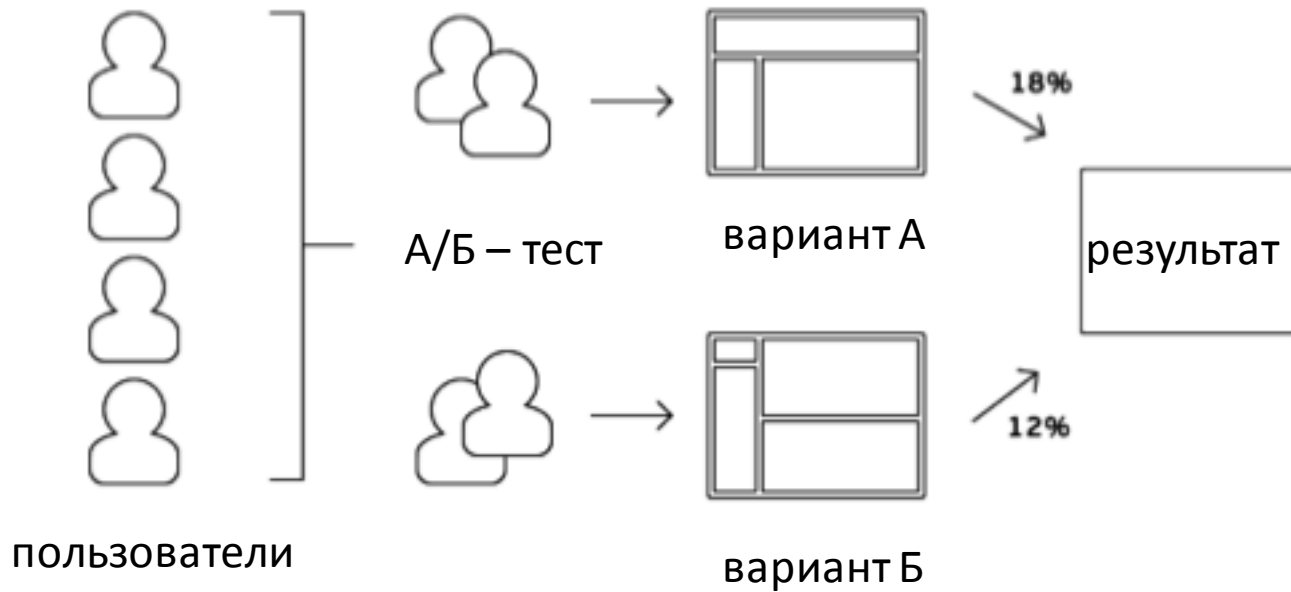
- корреляция \neq причина
- побочные эффекты & влияние извне

решение:

- случайные сплит тесты
- слепые исследования

АБТ: Основы

решение:



АБТ: Основы

особенности для web:

- нет непосредственного контакта
- миллионы наблюдений
- шум

АБТ: Реализация



АБТ: Реализация

- варианты разбиения
- количество информации & шум
- типы экспериментов
- масштабы экспериментов
- технические сложности
- пользовательские сложности
- отладка
- результаты

АБТ: Реализация

Варианты разбиения

АБТ: Реализация

Варианты разбиения

- Разбиение:
 - По пользователям (куки)
 - По визитам (сессии)
 - По действиям (запросы)
 - ...

АБТ: Реализация

Варианты разбиения

- Разбиение
- Особенности:
 - количество
 - независимость

АБТ: Реализация

Количество информации & шум

АБТ: Реализация

Количество информации & шум

- ПЛЮСЫ:
 - меньше дисперсия
 - меньше доверительный интервал
 - выше чувствительность/скорость

АБТ: Реализация

Количество информации & шум

- ПЛЮСЫ
- МИНУСЫ:
 - шум
 - «тяжелые» элементы
 - разнородность

АБТ: Реализация

Типы экспериментов

АБТ: Реализация

Типы экспериментов

- явные / неявные

АБТ: Реализация

Типы экспериментов

- **явные**
 - интерфейсы
 - функциональность
- **неявные**
 - ранжирование
 - персонализация

АБТ: Реализация

Типы экспериментов

- явные / неявные
- сервисы / части сервисов / кросс-сервисные
- улучшения / ухудшения / АА-тесты

АБТ: Реализация

Масштабы экспериментов

АБТ: Реализация

Масштабы экспериментов

Яндекс @ 2017:

- 4778 экспериментов,
- ~400 экспериментов одновременно

АБТ: Реализация

Измерения

АБТ: Реализация

Измерения

- 1-dim - “эксклюзив”

U S E R S	slot1	EXPERIMENT1
	slot2	
	slot3	
	slot4	
	slot5	
	slot6	
	slot7	EXPERIMENT2
	slot8	
	slot9	
	slot10	...
	slot11	
	slot12	
	slot13	
	slot14	
	slot15	
	slot16	
	slot17	
	slot18	
	slot19	
	slot20	
	slot21	
	slot22	
	slot23	
	slot24	CONTROL
	slot25	
	...	
	slotN	

АБТ: Реализация

Измерения

- 1-dim - “эксклюзив”
- N-dim - “мульти-измерения”

USERS salt2		
	EXPERIMENT1d1	
	EXPERIMENT2d1	
U S E R S salt1	CONTROLd2	EXPERIMENT1d2
	CONTROLd1	

АБТ: Реализация

Технические сложности

АБТ: Реализация

Технические сложности

- роботы & шум

АБТ: Реализация

Технические сложности

- роботы & шум
- кэширование

АБТ: Реализация

Технические сложности

- роботы & шум
- кэширование
- логирование множества событий

АБТ: Реализация

Технические сложности

- роботы & шум
- кэширование
- логирование множества событий
- проблемы парсинга логов

АБТ: Реализация

Технические сложности

- роботы & шум
- кэширование
- логирование множества событий
- проблемы парсинга логов
- ...

АБТ: Реализация

Поведение пользователей

АБТ: Реализация

Поведение пользователей

- сезонность
 - использовать 7N периоды
 - исключать крупные праздники

АБТ: Реализация

Поведение пользователей

- сезонность
- спецэффекты
 - изменение множества пользователей
 - изменение поведения пользователей

АБТ: Реализация

Поведение пользователей

- сезонность
- спецэффекты
- поломанные привычки

АБТ: Реализация

Поведение пользователей

- сезонность
- спецэффекты
- поломанные привычки
 - есть странные пользовательские привычки

АБТ: Реализация

Отладка

АБТ: Реализация

Отладка

- мелкие изменения

Пусть ваше изменение затрагивает 1% от всех визитов.

И каждый пользователь становится на 1% счастливее.

Тогда метрика должна заметить 0.01% изменений.

АБТ: Реализация

Отладка

- мелкие изменения
 - выделить релевантный срез
 - не забывать равнозначный контроль

АБТ: Реализация

Отладка

- мелкие изменения
- значимые результаты

Пусть вы получили значимое ухудшение.
Или неожиданное улучшение.
Придется разобраться.

АБТ: Реализация

Отладка

- мелкие изменения
- значимые результаты
 - поиск контрастных срезов
 - браузеры / регионы
 - классы, топ запросов

АБТ: Реализация

Отладка

- мелкие изменения
- значимые результаты
- обмен
 - что-то улучшилось
 - что-то ухудшилось
 - чего больше ?

АБТ: Реализация

Результаты

АБТ: Реализация

Результаты

Яндекс, 2014: 21% **принимаются**, 2017:
28% **принимаются**

Бинг, 201?: ~30% принимаются

Гугл, 201?: ~10% принимаются

АБТ: Реализация

Результаты

Яндекс, 2014: 21% **принимаются**, 2017:
28% **принимаются**

Бинг, 201?: ~30% принимаются

Гугл, 201?: ~10% принимаются

**В среднем: ~20% принимаются, ~50%
непонятных**

АБТ: Реализация

Результаты

Команды делают фичи, в которые верят.

Каждая команда работает над некоторой частью продукта.

Некоторые улучшения части продукта, могут быть ухудшающими для других частей продукта. Или для всего продукта.

Многие эксперименты не показывают улучшения на метриках.

АБТ: Реализация

Результаты

“Our job is to tell clients that their new baby is ugly” © Ronny Kohavi – аналитик в MS

“If you have to kiss a lot of frogs to find a prince, find more frogs and kiss them faster and faster” © Mike Moran

АБТ: Метрики



АБТ: Метрики

Данные

АБТ: Метрики

Данные

Логи:

- события (клики / показы)

АБТ: Метрики

Данные

Логи:

- события (клики / показы)
- время

АБТ: Метрики

Типы метрик

АБТ: Метрики

Типы метрик

- абсолютные (клики, временные)
- относительные (ctr, доля некликнутых)

АБТ: Метрики

Типы метрик

- приемочные
 - влияние на решения
 - полное доверие

АБТ: Метрики

Типы метрик

- приемочные
- метрики контроля
 - контроль валидности эксперимента
 - являются блокерами

АБТ: Метрики

Типы метрик

- приемочные
- метрики контроля
- информационные
 - свойства различных объектов
 - межобъектное взаимодействие
 - отладка

АБТ: Метрики

Приемочные метрики

АБТ: Метрики

Приемочные метрики

- мир одной метрики
 - нереально
 - может быть вредным

АБТ: Метрики

Приемочные метрики

- мир одной метрики
- Парето-оптимизация
 - возможно, но сложно улучшать
 - как правило ищем локальные оптимумы

АБТ: Метрики

Приемочные метрики

- мир одной метрики
- Парето-оптимизация
- иерархия метрик
 - согласованность
 - понятность для высокого уровня
 - чувствительность для низкого уровня

АБТ: Метрики

Приемочные метрики

Объединяем событийные и временные метрики, чтобы измерять совокупные изменения.

АБТ: Метрики

Приемка «вручную»

- [может быть] опасной и субъективной
 - любую метрику можно накрутить
 - модель интерпретации может изменить знак
 - никакой интерпретации метрик a posteriori!

АБТ: Метрики

Приемка «вручную»

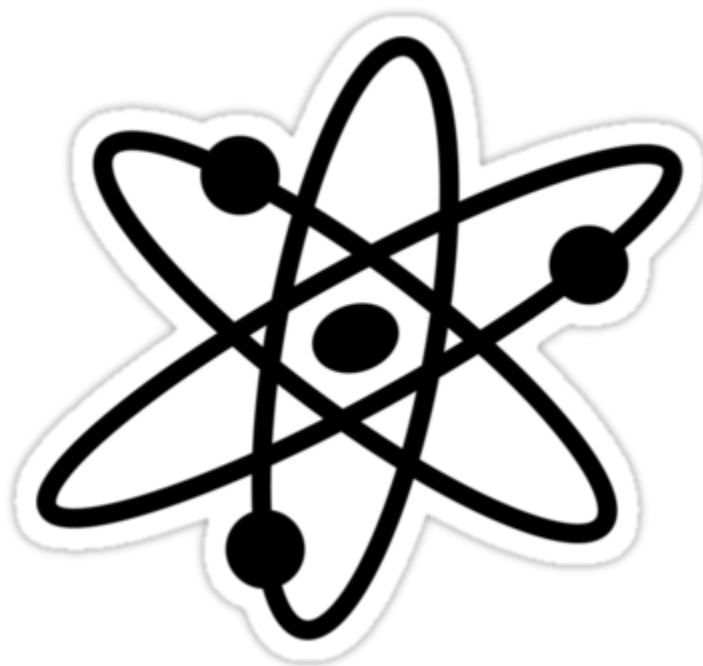
- опасна и субъективна
- когда применяется?
 - сложное изменение
 - абсолютно новое изменение
 - ожидается ухудшение приемочных метрик
 - приемочные метрики серые

АБТ: Метрики

Приемка «вручную»

- опасна и субъективна
- когда применяется?
- как применяется?
 - заранее фиксируем требования к продукту, хорошие и плохие сценарии
 - заранее фиксируем ожидания от известных метрик
 - при необходимости добавляем новую метрику

АБТ: Матчасть



АБТ: Матчасть

- А & В (“контроль” и “эксперимент”)

АБТ: Матчасть

- A & B (“контроль” и “эксперимент”)
- $M(X)$ – **случайная** величина
- $diff = M(A) - M(B)$

АБТ: Матчасть

- A & B (“контроль” и “эксперимент”)
- $M(X)$ – **случайная** величина
- $diff = M(A) - M(B)$
- $diff$ **значима?** (significance)
- насколько **достоверно?** (confidence)

АБТ: Матчасть

- $diff$ = эффект воздействия изменения + случайность
- чтобы измерить эффект, используем стат. тесты

АБТ: Матчасть

- *diff* = эффект воздействия изменения + случайность
- чтобы измерить эффект, используем стат. тесты:
 - Z-test
 - Student t-test
 - Mann–Whitney / U test / Wilcoxon
 - Wald test
 - ...

АБТ: Матчасть

- $diff$ = эффект воздействия изменения + случайность
- чтобы измерить эффект, используем стат. тесты:
 - Z-test
 - Student t-test
 - Mann–Whitney
 - Wald test
 - ...

АБТ: Матчасть

Как получить распределение (если требуется)?

- предположить класс, подобрать параметры
 - $N(\mu, \sigma^2)$
 - $B(n, p)$
 - $\ln N(\mu, \sigma^2)$

АБТ: Матчасть

Как получить распределение (если требуется)?

- предположить класс, подобрать параметры
- разбиение и бутстрап

АБТ: Вопросы?

