

# **A/Б-тесты**

## **7 занятие**

Игорь Полянский  
Нетология Июнь 2020

# **Ловушки в А/Б- тестировании**

# Проверка связи

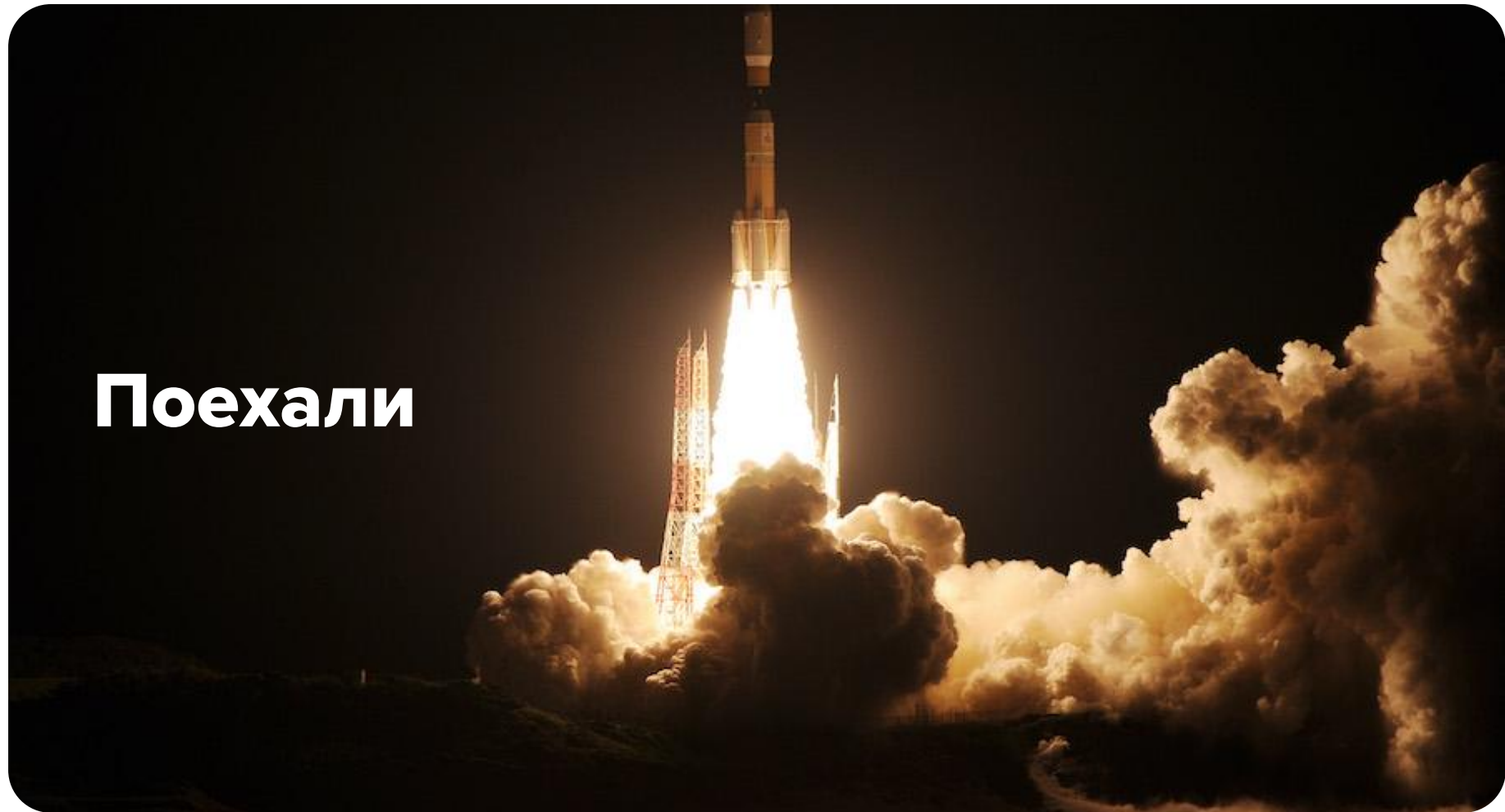


# В этой лекции

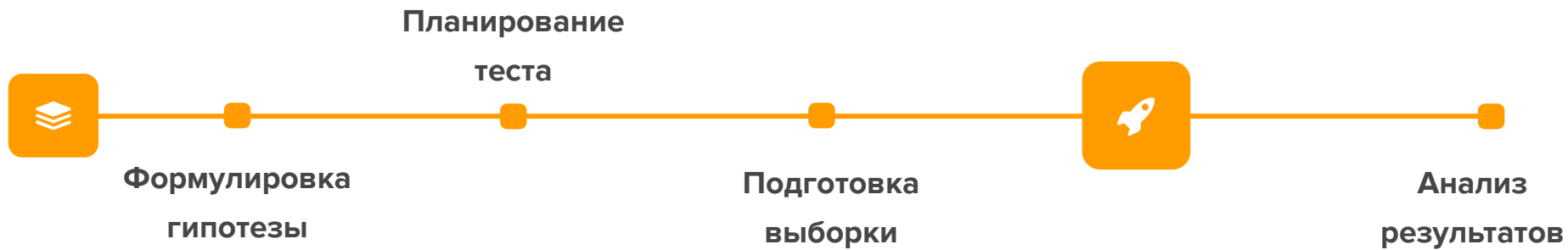
Пробежим весь пройденный материал и разберем типичные **ошибки на этапах:**

- До А/Б-теста
  - Во время А/Б-теста
  - После А/Б-теста
-

**Поехали**



# Структура курса



# О чем говорили на 1 занятии:

- Data-driven -

# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных



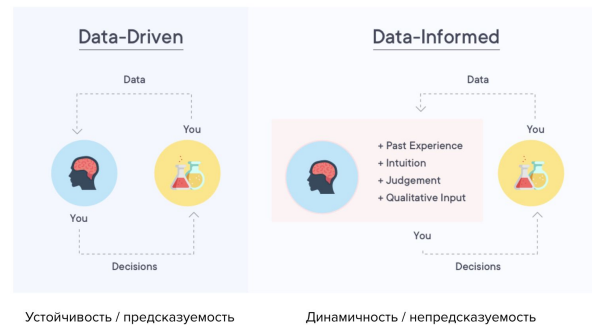


## О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** -

# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие ИСТОЧНИКИ



# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырье data-driven компании -

# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырые data-driven компании - **гипотеза**

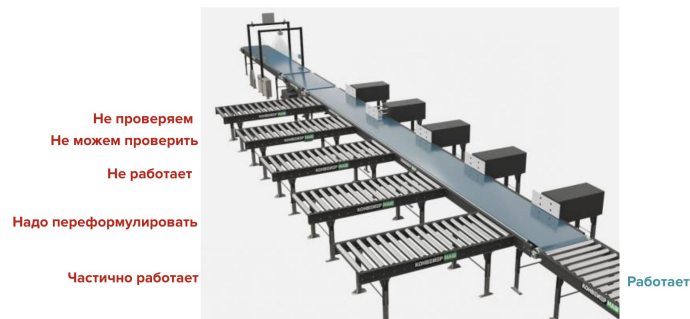


# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырье data-driven компании - **гипотеза**
- **Аналитик** -

# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырые data-driven компании - **гипотеза**
- **Аналитик - конвейер** по проверке гипотез

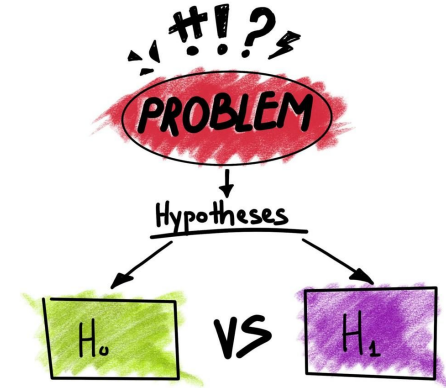


# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырье data-driven компании - **гипотеза**
- **Аналитик - конвейер** по проверке гипотез
- **Источниками гипотез** могут стать:

# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырые data-driven компании - **гипотеза**
- **Аналитик - конвейер** по проверке гипотез
- **Источниками гипотез** могут стать: данные, конкуренты, исследования, feedback



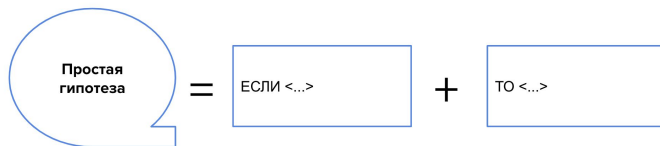


# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырье data-driven компании - **гипотеза**
- **Аналитик - конвейер** по проверке гипотез
- **Источниками гипотез** могут стать: данные, конкуренты, исследования, feedback
- **Простая гипотеза** = Бизнес-гипотеза =

# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырье data-driven компании - **гипотеза**
- **Аналитик - конвейер** по проверке гипотез
- **Источниками гипотез** могут стать: данные, конкуренты, исследования, feedback
- **Простая гипотеза** = Бизнес-гипотеза = **ЕСЛИ ... ТО ...**



# О чем говорили на 1 занятии:

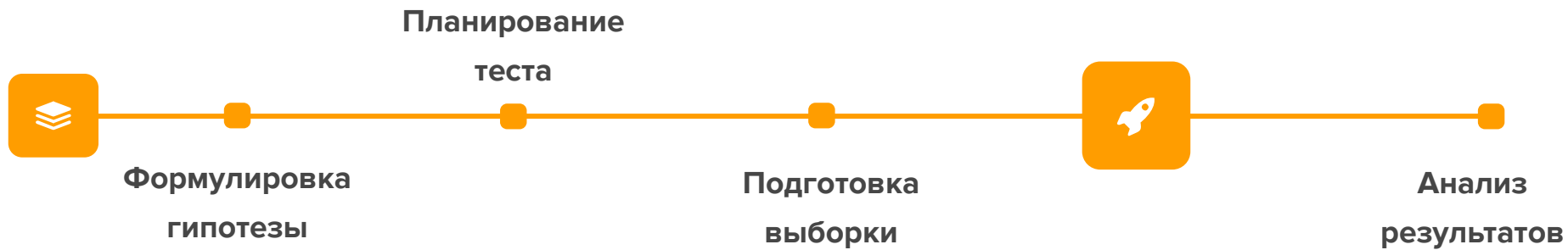
- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырье data-driven компании - **гипотеза**
- **Аналитик - конвейер** по проверке гипотез
- **Источниками гипотез** могут стать: данные, конкуренты, исследования, feedback
- **Простая гипотеза** = Бизнес-гипотеза = **ЕСЛИ ... ТО ...**
- **Маркетинговая/продуктовая гипотеза** =

# О чем говорили на 1 занятии:

- **Data-driven** - компания, принимающая решения на данных
- **Data-informed** - использует не только данные, но и другие источники
- Сырье data-driven компании - **гипотеза**
- **Аналитик - конвейер** по проверке гипотез
- **Источниками гипотез** могут стать: данные, конкуренты, исследования, feedback
- **Простая гипотеза** = Бизнес-гипотеза = **ЕСЛИ ... ТО ...**
- **Маркетинговая/продуктовая гипотеза** = что делаем + на кого повлияет + какой результат ожидаем + почему ожидаем такой результат



# Структура курса



## О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, (что случится)

# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**



## О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация необходима**, чтобы уложиться в наши ресурсы по



## О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация необходима**, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.

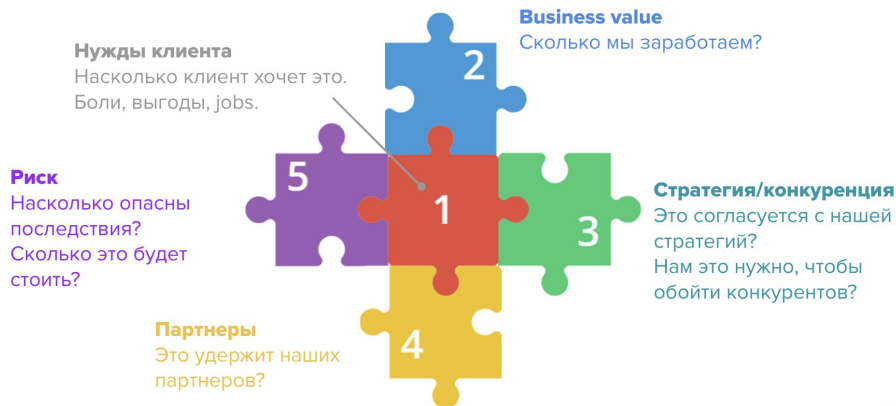


## О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на:

# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**



# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- ICE -

## О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- **ICE** - отличный фреймворк для численной приоритизации гипотез

## ICE framework

ICE = Impact \* Confidence - Effort

# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- **ICE** - отличный фреймворк для численной приоритизации гипотез
- **Можно влиять на effort и confidence**, но ключевая идея ICE:

# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- **ICE** - отличный фреймворк для численной приоритизации гипотез
- **Можно влиять на effort и confidence**, но ключевая идея ICE: довольно сложно понижать effort, но можно довольно легко повысить confidence.



## MVP

Вкладываем минимум ресурсов, чтобы провести пользовательский тест (минимальный работающий продукт)



## Урезаем функционал

Делаем только то, что реально нужно пользователям (80-20) - можно по результатам MVP



## Технологии / каналы

Ищем более дешевые технологии / каналы для маркетинга



## UX-тест прототипа

Нарисуй на бумажке и покажи пользователю - effort близок к 0 (+повышаем confidence)

# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- **ICE** - отличный фреймворк для численной приоритизации гипотез
- **Можно влиять на effort и confidence**, но ключевая идея ICE: довольно сложно понижать effort, но можно довольно легко повысить confidence.
- **Для infra/legal -**





# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- **ICE** - отличный фреймворк для численной приоритизации гипотез
- **Можно влиять на effort и confidence**, но ключевая идея ICE: довольно сложно понижать effort, но можно довольно легко повысить confidence.
- **Для infra/legal - confidence = 100%**



# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- **ICE** - отличный фреймворк для численной приоритизации гипотез
- **Можно влиять на effort и confidence**, но ключевая идея ICE: довольно сложно понижать effort, но можно довольно легко повысить confidence.
- **Для infra/legal - confidence = 100%**
- **RICE** -

# О чем еще говорили на 1 занятии:

- Продуктовая команда может генерить гипотезы из разных источников. При правильной настройке, **гипотез будет сильно больше чем наша возможность их проверить.**
- **Приоритизация** необходима, чтобы уложиться в наши ресурсы по разработчикам, аналитикам, клиентской базе.
- Чтобы приоритизировать гипотезу надо смотреть на: **нужды клиента, business value, соответствие стратегии, партнеров, риски**
- **ICE** - отличный фреймворк для численной приоритизации гипотез
- **Можно влиять на effort и confidence**, но ключевая идея ICE: довольно сложно понижать effort, но можно довольно легко повысить confidence.
- **Для infra/legal - confidence = 100%**
- **RICE** - также учитывает ROI фичей, но имеет склонность делать более маленькие фичи с высоким ROI

## О чем еще говорили на 1 занятии:

- Для приоритизации лучше всего использовать

## О чем еще говорили на 1 занятии:

- Для приоритизации лучше всего использовать **ICE + RICE**

## О чем еще говорили на 1 занятии:

- Для приоритизации лучше всего использовать **ICE + RICE**
- **PIE -**

## О чем еще говорили на 1 занятии:

- Для приоритизации лучше всего использовать **ICE + RICE**
- **PIE** - фреймворк для качественной приоритизации гипотез, хорошо подходит для приоритизации WEB.

### **PIE framework**

PIE = AVG ( Potential, Importance, Ease)

Если лень считать цифры

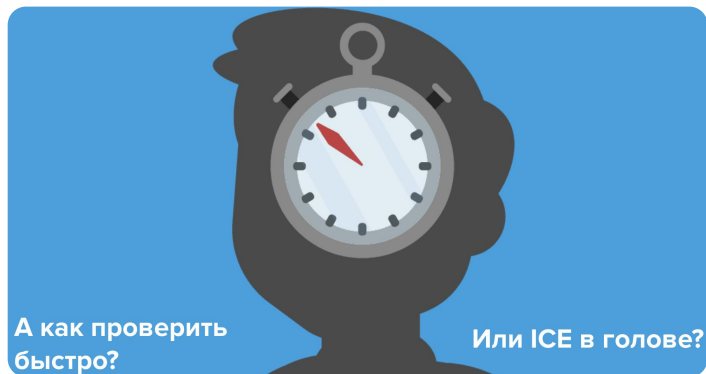
## О чем еще говорили на 1 занятии:

- Для приоритизации лучше всего использовать **ICE + RICE**
- **PIE** - фреймворк для качественной приоритизации гипотез, хорошо подходит для приоритизации WEB.
- **Экспресс-анализ гипотез** используется для того, чтобы не тратить ресурсы на проверку неперспективных гипотез.
- **Экспресс-анализ** проводится путем



# О чем еще говорили на 1 занятии:

- Для приоритизации лучше всего использовать **ICE + RICE**
- **PIE** - фреймворк для качественной приоритизации гипотез, хорошо подходит для приоритизации WEB.
- **Экспресс-анализ гипотез** используется для того, чтобы не тратить ресурсы на проверку неперспективных гипотез.
- **Экспресс-анализ** проводится путем прокручивания в голове процедуры ICE или задаванием себе ряда вопросов



# Кейс на подумать - Gett

Когда только появилась система рейтинга (оценки водителя после поездки), мы наблюдали очень интересную картину.

Рейтинг в России и UK был 4,5+ / 5 (что соответствует нашему позиционированию как сервиса с лучшими водителями).

Рейтинг в UK оказался шокирующим 0.5- / 5.

Бизнес очень всполошился. Было экстренно принято решение создать специальный отдел качества в Израиле, чтобы решить эту проблему.

Но нашелся умный человек, который предложил что-то сделать. **Что он предложил?**

**Из-за чего возникла проблема?**

# Кейс на подумать - Столото

Мы заметили, что клиенты iOS имеют гораздо более высокий средний чек, чем Android.

**Из-за чего так происходило?**

**Как мы это использовали?**

# Кейс на подумать - CRM b2b

Клиенты b2b имеют гораздо более длительный цикл покупки чем b2c.

Мы отправляли клиентам email-ы с предложением лизинга. По рассылке была выделена КГ для оценки влияния рассылки.

Получалась следующая картина:

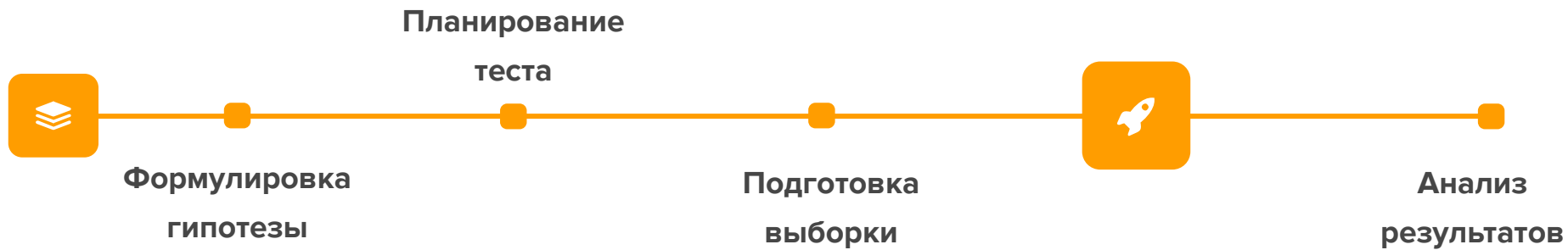
- В первые недели после рассылки конверсия письма была выше на 5-6%
- Но на большом интервале конверсии практически выравнивались.

**Как так получалось?**

**Что мы проанализировали дополнительно?**

**Что изменили в логике?**

# Структура курса



## О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них?

## О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**

## О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать?



## О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать? **(2), но стоит попытаться исключить эффект новизны**

## О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать? **(2), но стоит попытаться исключить эффект новизны**
- Типы метрик для анализа теста:

## О чем говорили на 2 занятии:

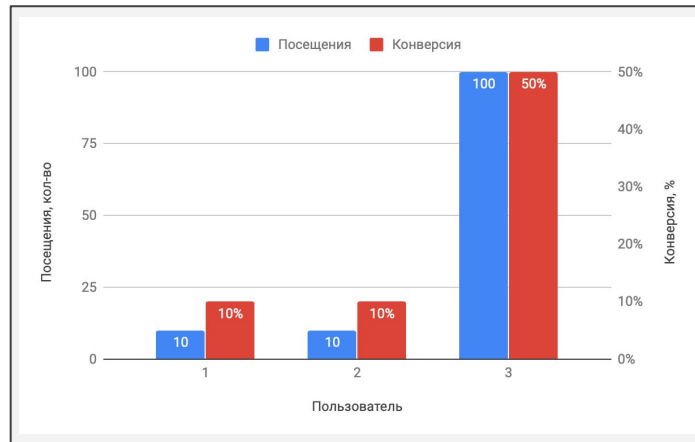
- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать? **(2), но стоит попытаться исключить эффект новизны**
- Типы метрик для анализа теста:  
**Первичные и вторичные**

## О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать? **(2), но стоит попытаться исключить эффект новизны**
- Типы метрик для анализа теста:  
**Первичные и вторичные**
- Методы агрегации значений метрики:

# О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать? **(2), но стоит попытаться исключить эффект новизны**
- Типы метрик для анализа теста:  
**Первичные и вторичные**
- Методы агрегации значений метрики:  
**Среднее по юзерное vs общее**



Общая конверсия:  
$$\frac{(1+1+50)}{(10+10+100)} = 43\%$$

Средняя конверсия:  
$$\frac{(1/10 + 1/10 + 50/100)}{3} = 23\%$$

## О чем говорили на 2 занятии:

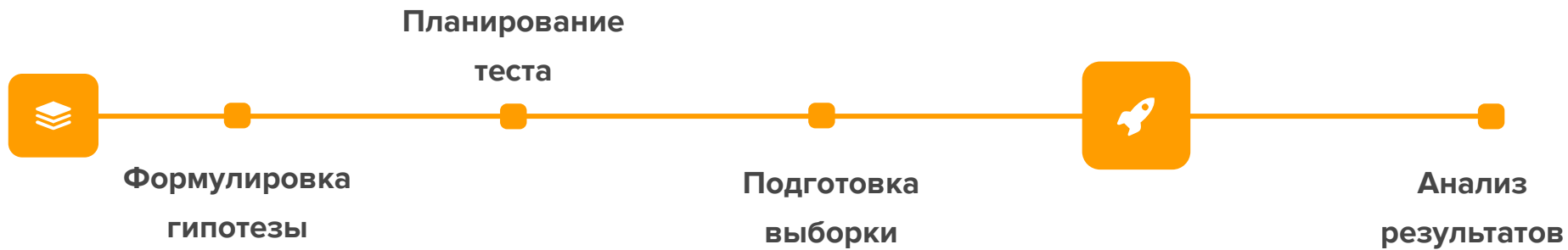
- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать? **(2), но стоит попытаться исключить эффект новизны**
- Типы метрик для анализа теста:  
**Первичные и вторичные**
- Методы агрегации значений метрики:  
**Среднее поюзерное vs общее**
- Стат. Единица и Генеральная совокупность

## О чем говорили на 2 занятии:

- А/Б тесты. Можно ли без них? **Нельзя. Они позволяют делать причинно-следственные выводы**
- Качество продукта (1), принципиально новый опыт (2) и долгосрочное изменение (3) - что можно А/Б тестировать? **(2), но стоит попытаться исключить эффект новизны**
- Типы метрик для анализа теста:  
**Первичные и вторичные**
- Методы агрегации значений метрики:  
**Среднее поюзерное vs общее**
- Стат. Единица и Генеральная совокупность



# Структура курса





## О чем говорили на 3 занятии:

- Описательные статистики:

## О чем говорили на 3 занятии:

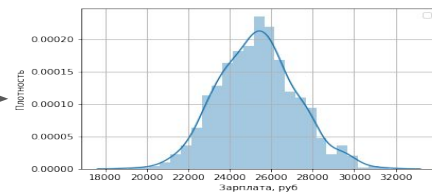
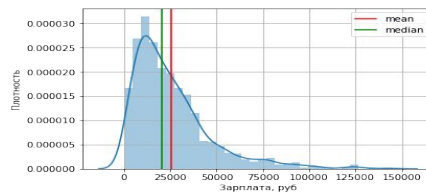
- Описательные статистики:  
**среднее, медиана, персентили**

# О чем говорили на 3 занятии:

- Описательные статистики:  
среднее, медиана, персентили
- ЦПТ

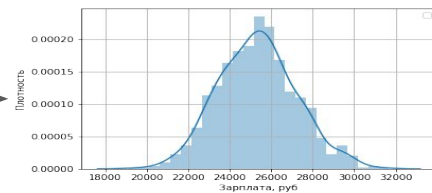
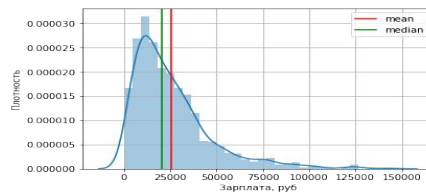
# О чем говорили на 3 занятии:

- Описательные статистики:  
**среднее, медиана, перцентили**
- ЦПТ  
**Среднее распределено нормально**



# О чем говорили на 3 занятии:

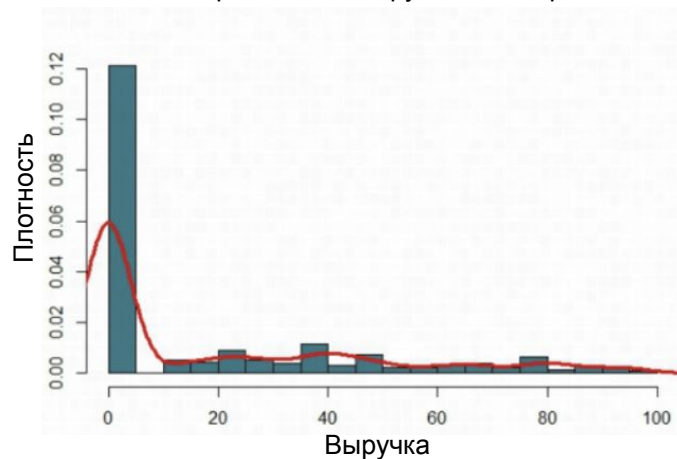
- Описательные статистики:  
**среднее, медиана, персентили**
- ЦПТ  
**Среднее распределено нормально**
- Ограничения ЦПТ:



# О чем говорили на 3 занятии:

- Описательные статистики:  
**среднее, медиана, персентили**
- ЦПТ  
**Среднее распределено нормально**
- Ограничения ЦПТ:  
**Тяжелые хвосты, выбросы**

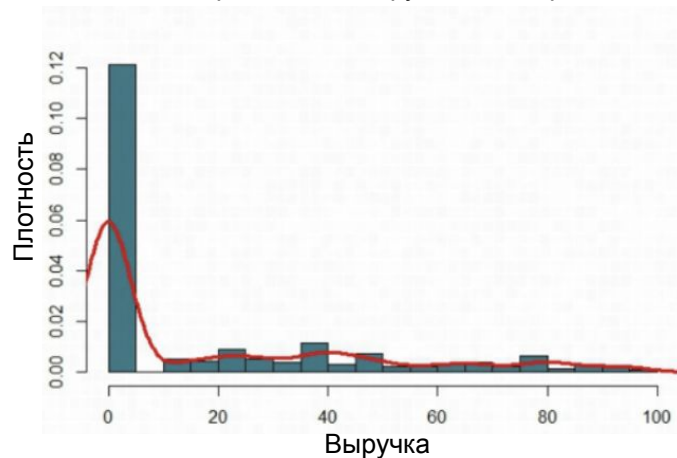
Распределение выручки на юзера



# О чем говорили на 3 занятии:

- Описательные статистики:  
**среднее, медиана, персентили**
- ЦПТ  
**Среднее распределено нормально**
- Ограничения ЦПТ:  
**Тяжелые хвосты, выбросы**
- Параметры А/Б теста:  
Ошибка первого рода -  
Ошибка второго рода -  
Мощность -  
p-value -

Распределение выручки на юзера



# О чем говорили на 3 занятии:

- Описательные статистики:  
**среднее, медиана, персентили**
- ЦПТ  
**Среднее распределено нормально**
- Ограничения ЦПТ:  
**Тяжелые хвосты, выбросы**
- Параметры А/Б теста:  
Ошибка первого рода - **фиксируем** приемлемую  $P(\text{отклонить } H_0 | H_0 \text{ верна})$   
Ошибка второго рода -  $P(\text{принять } H_0 | H_1 \text{ верна})$   
Мощность -  $P(\text{отклонить } H_0 | H_1 \text{ верна})$   
p-value -  $P(\text{отклонить } H_0 | H_0 \text{ верна})$



# О чем говорили на 3 занятии:

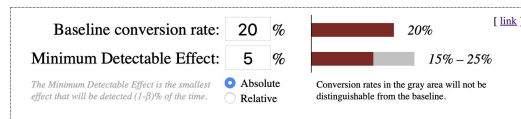
- Описательные статистики:  
**среднее, медиана, персентили**
- ЦПТ  
**Среднее распределено нормально**
- Ограничения ЦПТ:  
**Тяжелые хвосты, выбросы**
- Параметры А/Б теста:  
Ошибка первого рода - **фиксируем** приемлемую  $P(\text{отклонить } H_0 | H_0 \text{ верна})$   
Ошибка второго рода -  $P(\text{принять } H_0 | H_1 \text{ верна})$   
Мощность -  $P(\text{отклонить } H_0 | H_1 \text{ верна})$   
p-value -  $P(\text{отклонить } H_0 | H_0 \text{ верна})$
- Калькуляторы размера выборок А/Б теста

# О чем говорили на 3 занятии:

- Описательные статистики:  
**среднее, медиана, персентили**
- ЦПТ  
**Среднее распределено нормально**
- Ограничения ЦПТ:  
**Тяжелые хвосты, выбросы**
- Параметры А/Б теста:  
Ошибка первого рода - **фиксируем** приемлемую  $P(\text{отклонить } H_0 | H_0 \text{ верна})$   
Ошибка второго рода -  $P(\text{принять } H_0 | H_1 \text{ верна})$   
Мощность -  $P(\text{отклонить } H_0 | H_1 \text{ верна})$   
p-value -  $P(\text{отклонить } H_0 | H_0 \text{ верна})$
- Калькуляторы размера выборок А/Б теста

$P(\text{отклонить } H_0 | H_0 \text{ верна})$

Question: How many subjects are needed for an A/B test?



Sample size:

1,030

per variation

Statistical power  $1-\beta$ :  80% Percent of the time the minimum effect size will be detected, assuming it exists

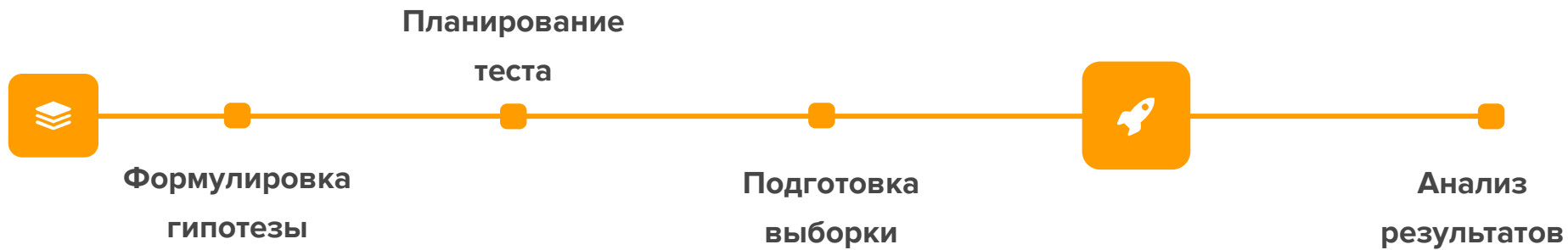
Significance level  $\alpha$ :  5% Percent of the time a difference will be detected, assuming one does NOT exist

# Кейс на подумать - Gett

При анализе данных выявлено, что водители, которые зарабатывают с нами больше, имеют более низкий показатель оттока.

**Что здесь наблюдается? Корреляция / причинно-следственная связь?**

# Структура курса



## О чем говорили на 4 занятии:

- Важность сплита?

## О чем говорили на 4 занятии:

- Важность сплита?

**Разные группы по метрикам => невозможно оценить эффект фичи**

## О чем говорили на 4 занятии:

- Важность сплита?

**Разные группы по метрикам => невозможно оценить эффект фичи**

- Виды сплитов:

# О чем говорили на 4 занятии:

- Важность сплита?

**Разные группы по метрикам => невозможно оценить эффект фичи**

- Виды сплитов:

**Рандом**

**Стратифицированный рандом**



# О чем говорили на 4 занятии:

- Важность сплита?

**Разные группы по метрикам => невозможно оценить эффект фичи**

- Виды сплитов:

**Рандом**

**Стратифицированный рандом**

- Тесты на однородность: зачем?

# О чем говорили на 4 занятии:

- Важность сплита?

**Разные группы по метрикам => невозможно оценить эффект фичи**

- Виды сплитов:

**Рандом**

**Стратифицированный рандом**

- Тесты на однородность: зачем?

**Даже рандом не всегда гарантирует однородность групп :(**

# О чем говорили на 4 занятии:

- Важность сплита?

**Разные группы по метрикам => невозможно оценить эффект фичи**

- Виды сплитов:

**Рандом**

**Стратифицированный рандом**

- Тесты на однородность: зачем?

**Даже рандом не всегда гарантирует однородность групп :(**

- Тесты на однородность: какие?

# О чем говорили на 4 занятии:

- Важность сплита?

**Разные группы по метрикам => невозможно оценить эффект фичи**

- Виды сплитов:

**Рандом**

**Стратифицированный рандом**

- Тесты на однородность: зачем?

**Даже рандом не всегда гарантирует однородность групп :(**

- Тесты на однородность: какие?

**T-тест**

**Ранговые: Манн-Уитни, Крускал-Уоллис**

**Бакетинг**

**Бутстрап**

# Кейс на подумать - Столото

Для оценки эффективности SMS-коммуникации по предотвращению оттока, мы выбрали клиентов-отточников, посчитали необходимый размер выборки, чтобы достичь стат. значимости - 20 000 при заданной вероятности покупки и вариативности. Но результат все равно получился незначимым.

**Из-за чего так могло случиться?**

**Как избежать?**

# Кейс на подумать - CRM b2b

Для одного из экспериментов мы хотели равномерно разделить группу по кол-во авто, взятых в лизинг. Мы сделали это так: в excel отсортировали клиентов по возрастанию кол-ва авто. Далее четным строчкам присвоили А, нечетным - В.

**К какой проблеме это привело?**

**Как этого можно было избежать?**

# Ошибки до теста

---

**“Давайте протестируем как можно  
больше вариаций - одна точно  
сработает!”**

---



# Проблема множественного тестирования

При большом количестве тестов, вероятность ошибки первого рода растет

Велика вероятность найти значимые результаты там, где их нет

**“Два теста одновременно? Нет,  
один тест повлияет на другой и  
результаты будут испорчены”**

## **Один юзер - несколько тестов?**

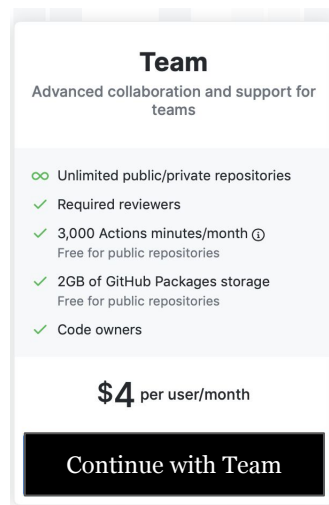
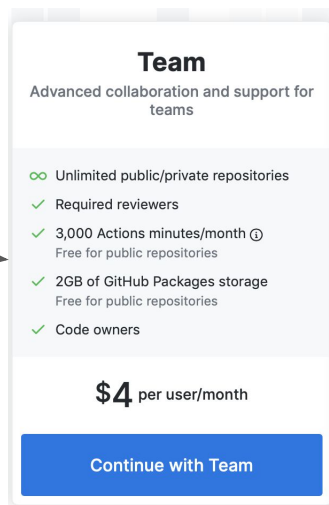
Это возможно, но нужно решать по ситуации

Зависит от того, могут ли фичи оказать влияние на первичные & вторичные метрики друг друга

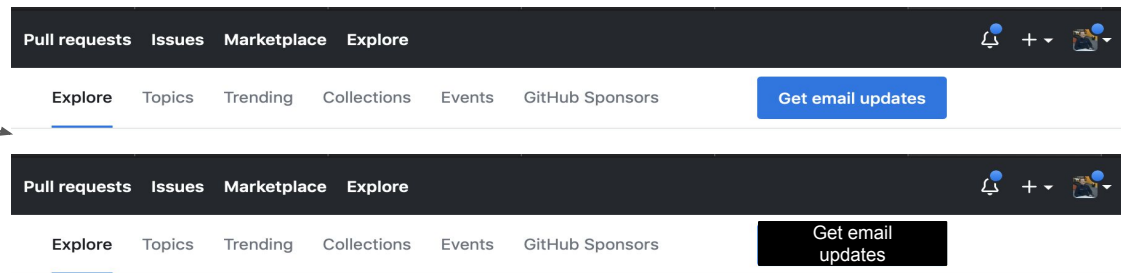
# Один юзер - несколько тестов?: пример 1

**Вопрос:** может ли один юзер участвовать в двух тестах?

Тест 1:



Тест 2:



# Один юзер - несколько тестов?: пример 1

## Тест 1

Первичные метрики: конверсия в покупку  
пакета “Team”

## Тест 2

Первичные метрики: конверсия в подписку на  
обновления GitHub

**Ответ:** можно, так как пакет GitHub подписки юзера не влияет на желание отслеживать изменения в репозиториях!

## Один юзер - несколько тестов?: пример 2

**Предисловие:** допустим, водителю показывается заказ (оффер) на протяжении 5 сек. За это время он может его либо принять, либо отклонить, либо проигнорировать. Если оффер был принят, водитель может потом его отменить.

**Фича 1:** даем водителям из тестовой группы 10 сек (вместо 5) на раздумья.

**Ожидаем 1:** увеличение доли принятых заказов (больше времени прочитать детали)

## Один юзер - несколько тестов?: пример 2

**Предисловие:** допустим, водителю показывается заказ (оффер) на протяжении 5 сек. За это время он может его либо принять, либо отклонить, либо проигнорировать. Если оффер был принят, водитель может потом его отменить.

**Фича 1:** даем водителям из тестовой группы 10 сек (вместо 5) на раздумья.

**Ожидаем 1:** увеличение доли принятых заказов (больше времени прочитать детали)

**Фича 2:** увеличиваем штраф за отмену заказа со 100 до 200 руб (в реальности штрафов нет)

**Ожидание 2:** уменьшение отмен, лучше юзер experience

**Вопрос:** может ли один юзер участвовать в двух тестах?

## Один юзер - несколько тестов?: пример 2

**Предисловие:** допустим, водителю показывается заказ (оффер) на протяжении 5 сек. За это время он может его либо принять, либо отклонить, либо проигнорировать. Если оффер был принят, водитель может потом его отменить.

**Фича 1:** даем водителям из тестовой группы 10 сек (вместо 5) на раздумья.

**Ожидаем 1:** увеличение доли принятых заказов (больше времени прочитать детали)

**Фича 2:** увеличиваем штраф за отмену заказа со 100 до 200 руб (в реальности штрафов нет)

**Ожидание 2:** уменьшение отмен, лучше юзер experience

**Вопрос:** может ли один юзер участвовать в двух тестах?

---

**Ответ:** нет.

Фича 1 увеличивает долю принятых заказов, а фича 2 - понижает, так как водители принимают осторожнее => **эффект от первой фичи определить невозможно**



## Один юзер - несколько тестов?: пример 2: решения

Не допускать пересекающихся тестовых  
выборок в тестах



## Один юзер - несколько тестов?: пример 2: решения

Не допускать пересекающихся тестовых  
выборок в тестах



Многомерное тестирование

(как будто у нас один тест с несколькими  
вариациями)

# Один юзер - несколько тестов?: пример 2: решения

Не допускать пересекающихся тестовых  
выборок в тестах

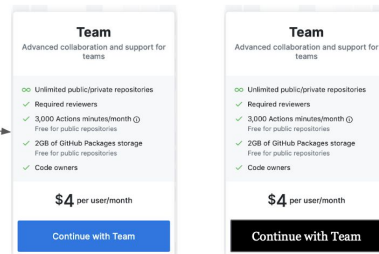


Многомерное тестирование

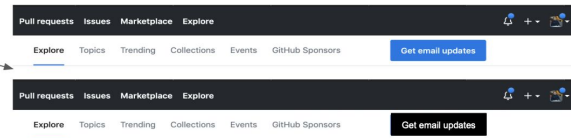
(как будто у нас один тест с несколькими  
вариациями)

Вопрос: сколько будет вариаций в примере  
про GitHub?

Тест 1:

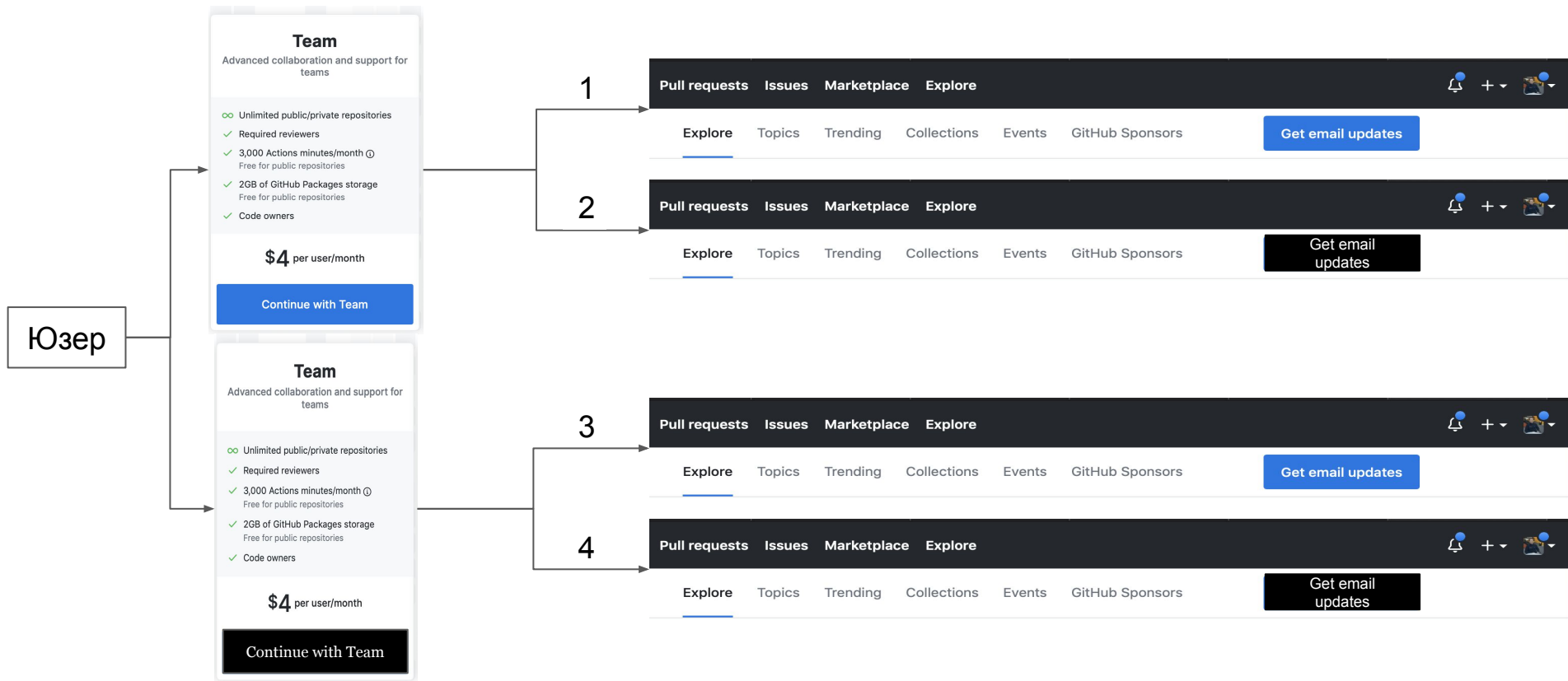


Тест 2:



# Один юзер - несколько тестов?: пример 1

Ответ: 4.



**“Растут клики, следовательно,  
растет конверсия!”**

---

## Клики vs конверсия: частая ошибка в ДЗ!

Увеличение кликов не обязательно значит увеличение конверсии!

Поэтому **не выбирайте клики в качестве первичных метрик**

Посещения страницы  
“категории товаров”



+ 12,7 % (significant)

## Клики vs конверсия: частая ошибка в ДЗ!

Увеличение кликов не обязательно значит увеличение конверсии!

Поэтому **не выбирайте клики в качестве первичных метрик**

Посещения страницы “категории товаров”	↑	+ 12,7 % (significant)
Посещения страницы товара	↑	+ 18,3 % (significant)

## Клики vs конверсия: частая ошибка в ДЗ!

Увеличение кликов не обязательно значит увеличение конверсии!

Поэтому **не выбирайте клики в качестве первичных метрик**

Посещения страницы “категории товаров”	↑	+ 12,7 % (significant)
Посещения страницы товара	↑	+ 18,3 % (significant)
Добавления в корзину	↑	+ 6,4 % (significant)



## Клики vs конверсия: частая ошибка в ДЗ!

Увеличение кликов не обязательно значит увеличение конверсии!

Поэтому **не выбирайте клики в качестве первичных метрик**



## Клики vs конверсия: частая ошибка в ДЗ!

Увеличение кликов не обязательно значит увеличение конверсии!

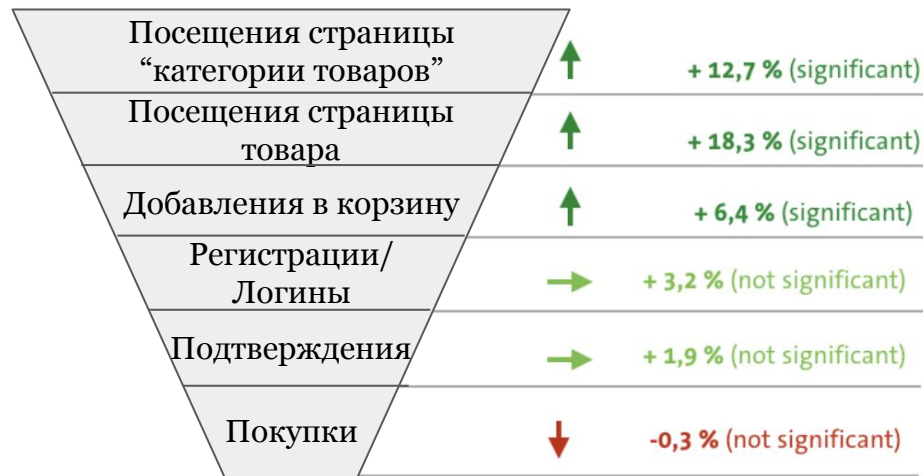
Поэтому **не выбирайте клики в качестве первичных метрик**



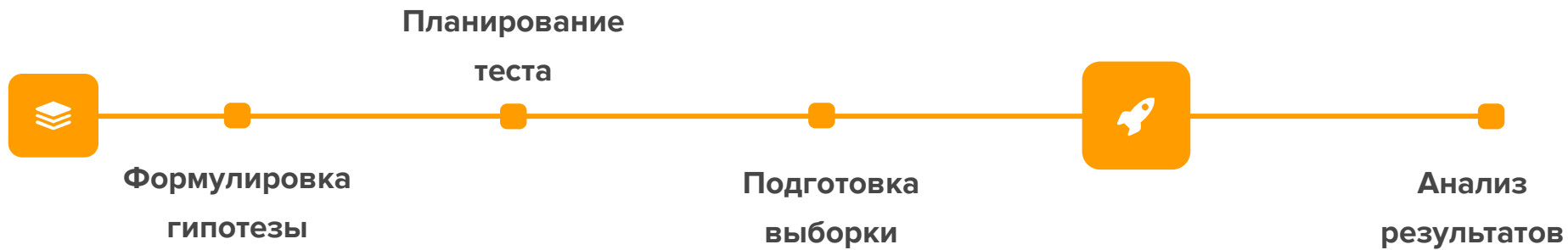
## Клики vs конверсия: частая ошибка в ДЗ!

Увеличение кликов не обязательно значит увеличение конверсии!

Поэтому **не выбирайте клики в качестве первичных метрик**



# Структура курса



# Ошибки во время теста

---

**“Тест запущен уже 3 дня и  
вариация показывает низкую  
конверсию, останавливаем тест”**

---

# Преждевременная остановка

Тест “угрожает” критически важным бизнес-метрикам

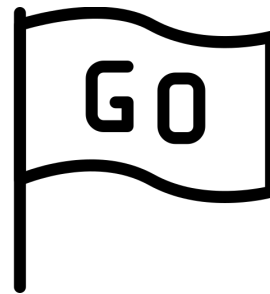


## Преждевременная остановка

Тест “угрожает” критически важным бизнес-метрикам



Ухудшения не критичны для бизнеса и могут быть чистой случайностью

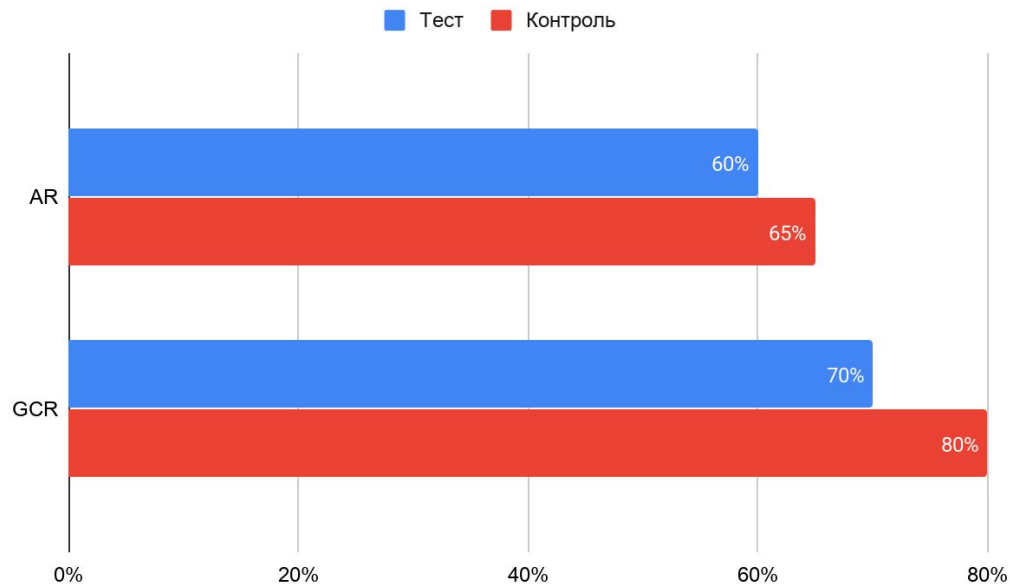


**Помним** про *minimal required sample size!*



## Угроза критически важным метрикам: пример

Проседание в доле принятых заказов на 5 п.п. еще можно пережить, но на 10 п.п. в доле вывезенных заказов - нет

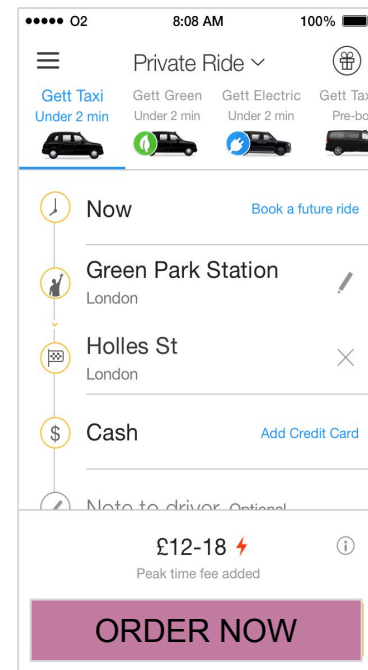
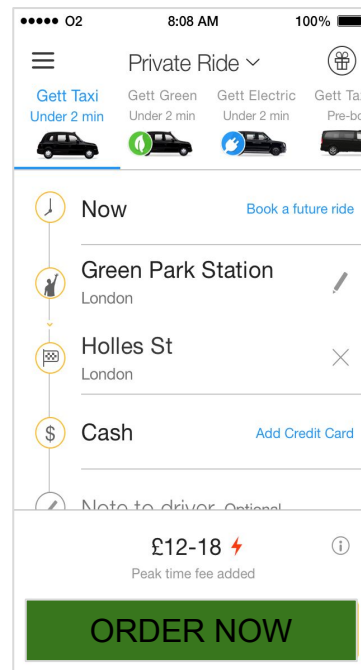
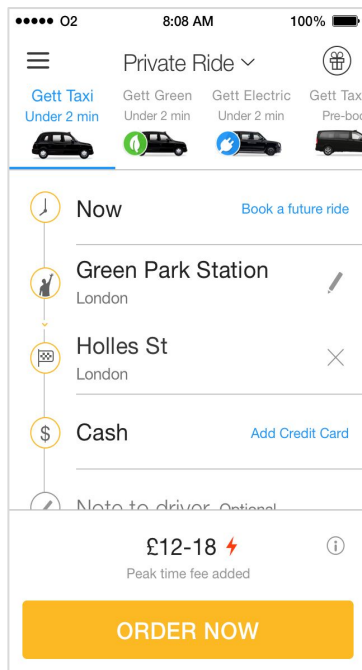


**“Одна из вариаций показывает  
плохие результаты.  
Изменим ее”**

---

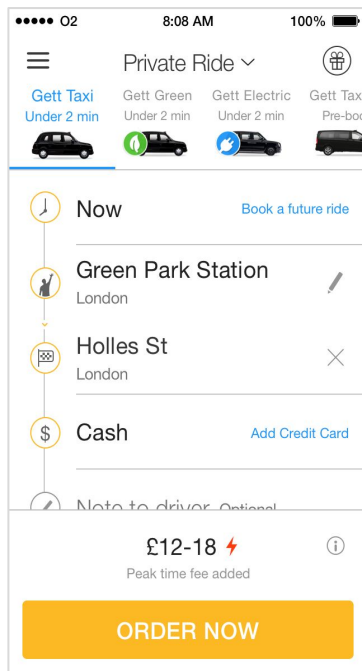
# Изменение вариаций: кейс

1. Запустили тест с тремя вариациями

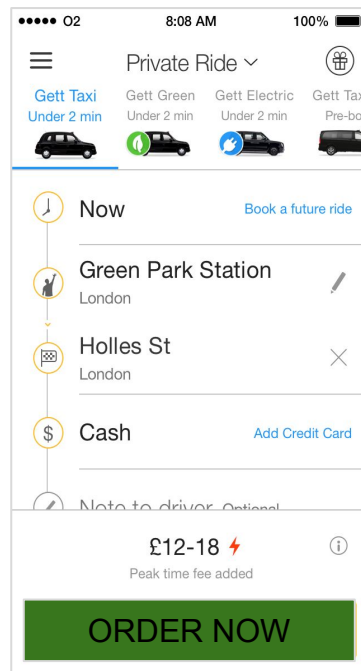


# Изменение вариаций: кейс

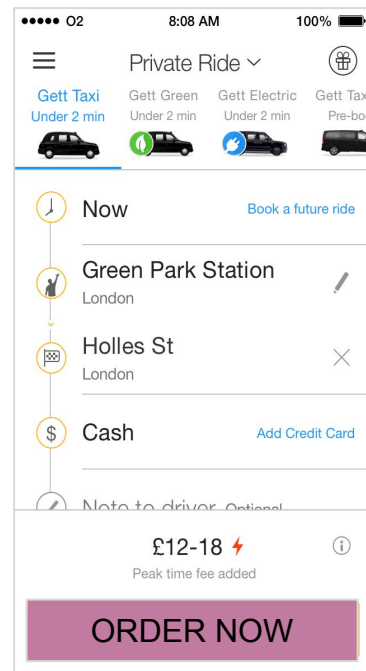
1. Запустили тест с тремя вариациями
2. После первой недели конверсия по **розовому** цвету была ниже всех, и решили заменить его на **коричневый**



Конверсия: 80%



Конверсия: 82%



Конверсия: 77%

# Изменение вариаций: кейс

Длительность запуска каждой вариации:



**Можем ли сравнить три  
вариации?**

---

## Один юзер - несколько тестов?

**Нет!**

Данных по первой неделе коричневого цвета нет

=> результаты этой недели имеют меньший вес в результатах А/Б теста

=> **невалидные выводы**

## Один юзер - несколько тестов?

**Нет!**

Данных по первой неделе коричневого цвета нет

=> результаты этой недели имеют меньший вес в результатах А/Б теста

=> **невалидные выводы**

**Интуитивно:**

Вдруг на первой неделе был аномально высокий спрос

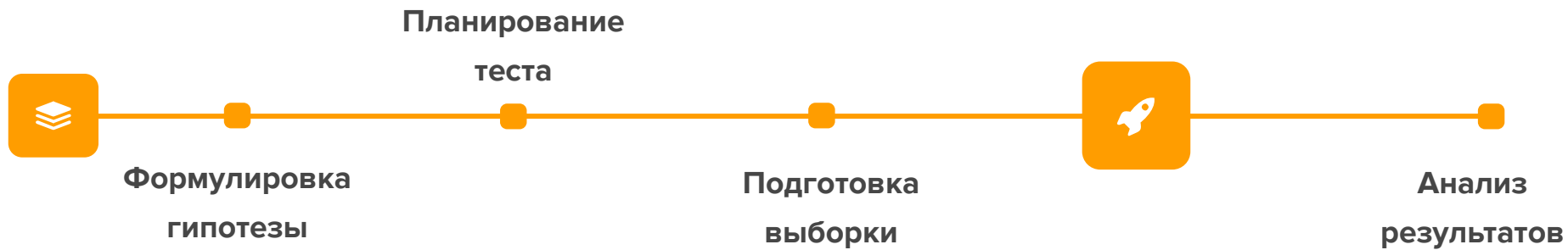
=> цены выше

=> конверсии просели у всех вариантов одинаково?

Коричневый цвет этого проседания не учитывает



# Структура курса



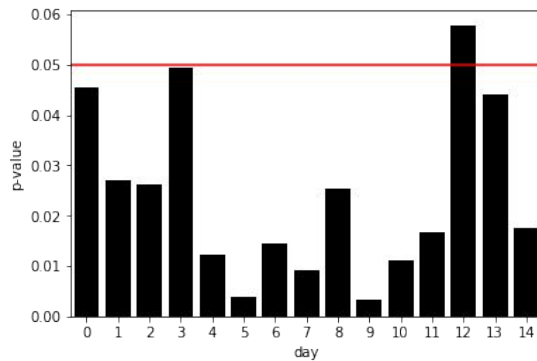
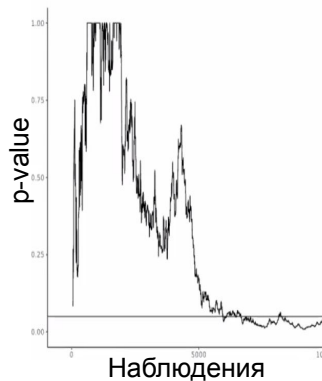
## О чем говорили на 6 занятии:

- Подходы к p-value

# О чем говорили на 6 занятии:

## Подходы к p-value

### Кумулятивный и периодический



## О чем говорили на 6 занятии:

- Подходы к p-value  
**Кумулятивный и периодический**
- Ускорить А/Б тест. Что это и как сделать?

## О чем говорили на 6 занятии:

- Подходы к p-value  
**Кумулятивный и периодический**
- Ускорить А/Б тест. Что это и как сделать?  
**Ускорить - суметь сделать вывод с заданными вероятностями ошибок 1 и 2 рода при меньших выборках**

## О чем говорили на 6 занятии:

- Подходы к p-value  
**Кумулятивный и периодический**
- Ускорить А/Б тест. Что это и как сделать?  
**Ускорить - суметь сделать вывод с заданными вероятностями ошибок 1 и 2 рода при меньших выборках**
- Сегментированный анализ: мораль

# О чем говорили на 6 занятии:

- Подходы к p-value  
**Кумулятивный и периодический**
- Ускорить А/Б тест. Что это и как сделать?  
**Ускорить - суметь сделать вывод с заданными вероятностями ошибок 1 и 2 рода при меньших выборках**
- Сегментированный анализ: мораль  
**Не нужно пытаться найти значимое различие в 100 разбивках, а то найдете (но неверно)**

# Ошибки после теста

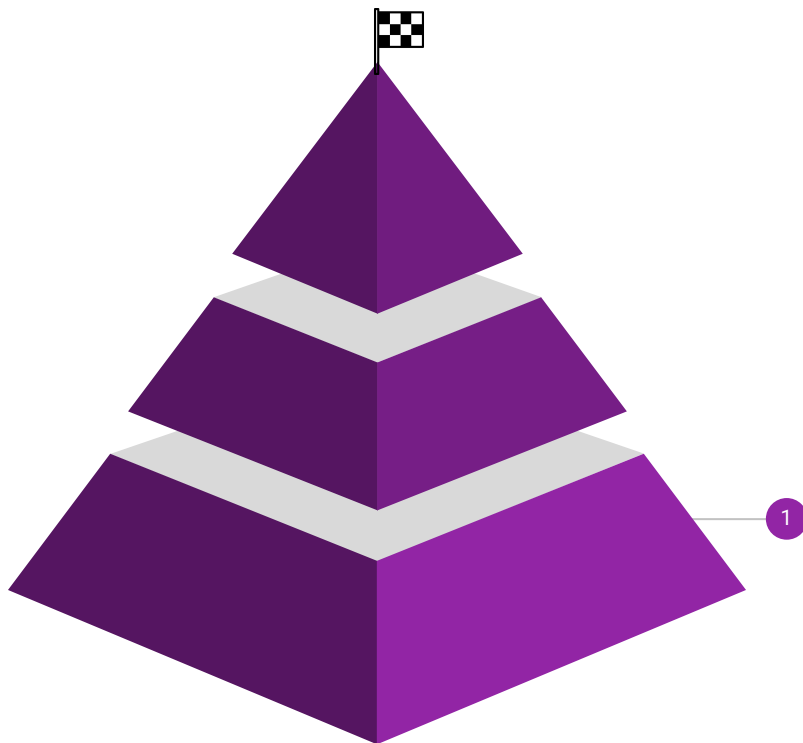
---



**“Wow, какие распределения!  
Сравню t-тестом!”**

---

# Нестандартное распределение?



## Бутстрап

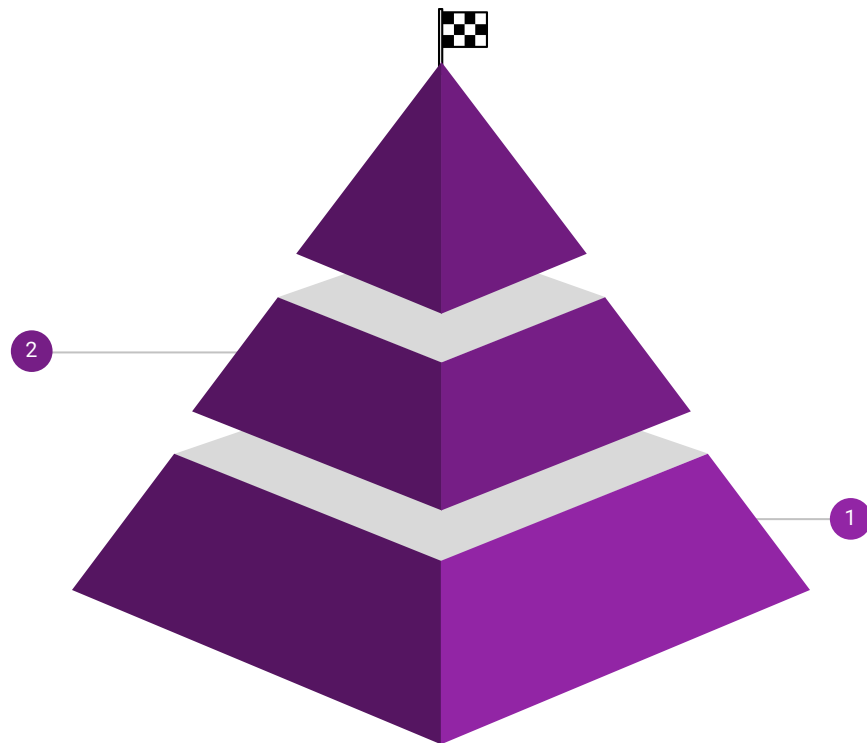
- + применим почти к любым распределениям
- + не требует много данных
- + применим для обоих методов агрегации значений метрики
- требует значительные вычислительные мощности

# Нестандартное распределение?

## Бакетинг

- + применим к любым распределениям
- + сравнительно легко считается
- + применим для обоих методов агрегации значений метрики

- **требуется мин. ~2500 наблюдений в каждой группе**



## Бутстрап

- + применим почти к любым распределениям
- + не требует много данных
- + применим для обоих методов агрегации значений метрики

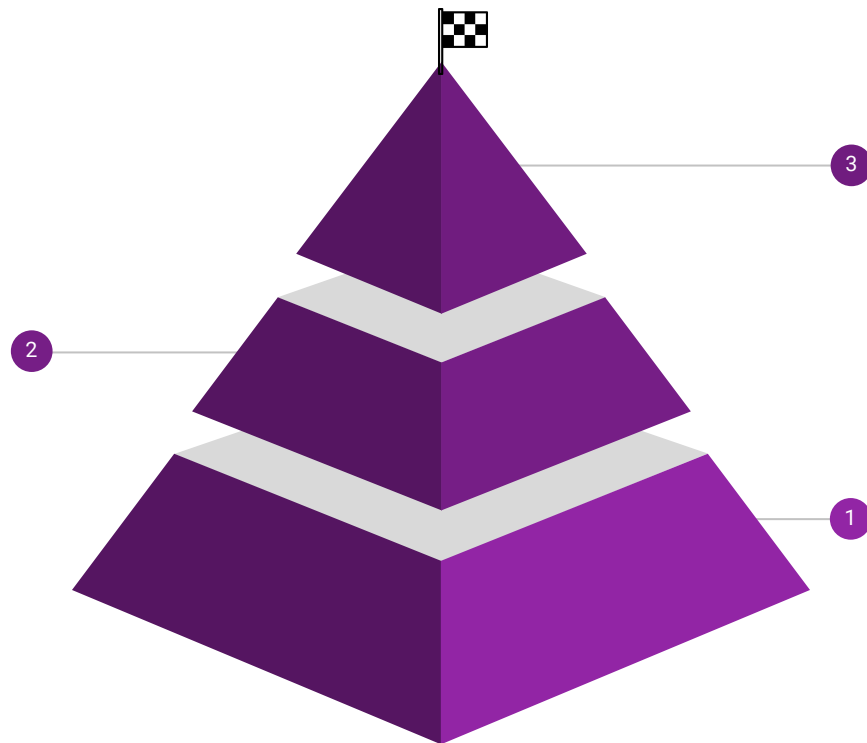
- **требуется значительные вычислительные мощности**

# Нестандартное распределение?

## Бакетинг

- + применим к любым распределениям
- + сравнительно легко считается
- + применим для обоих методов агрегации значений метрики

- **требует мин. ~2500 наблюдений в каждой группе**



## Ранговые критерии

(Mann-Whitney/Kruskal-Wallis)

- + применимы к любым распределениям
- + не требуют много данных

- **не применим к категориальным данным**

- **не применимы для общего значения метрики**

## Бутстрап

- + применим почти к любым распределениям
- + не требует много данных
- + применим для обоих методов агрегации значений метрики

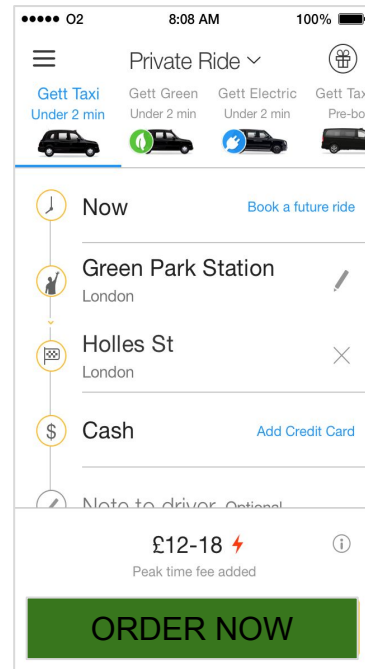
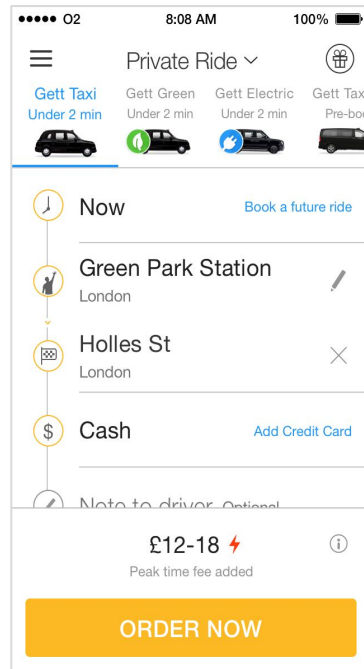
- **требует значительных вычислительных мощностей**

**“Похоже, тест был не таким и  
успешным. Фичу выкатили, а  
конверсия не выросла...”**

---

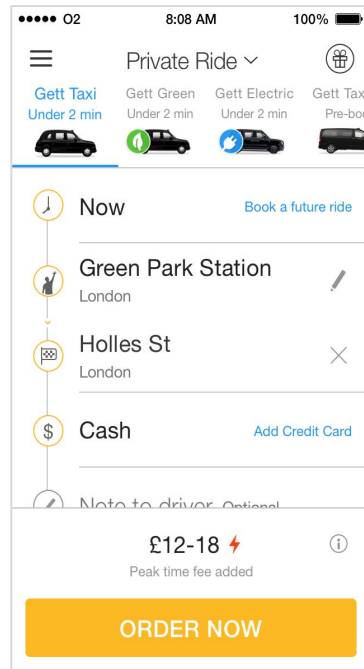
# Экстраполяция результатов: кейс

## 1. Запустили тест

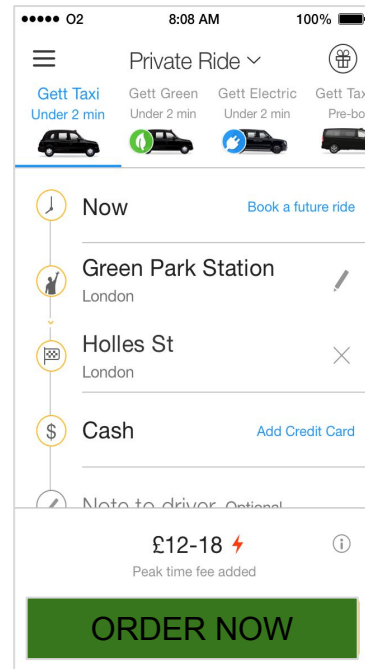


# Экстраполяция результатов: кейс

1. Запустили тест
2. Получили результаты



Конверсия: 60%

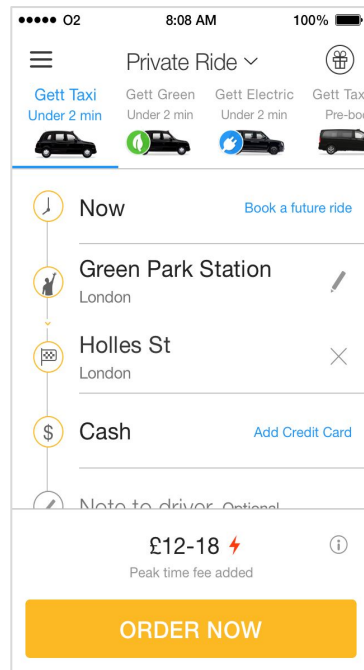


Конверсия: 84%

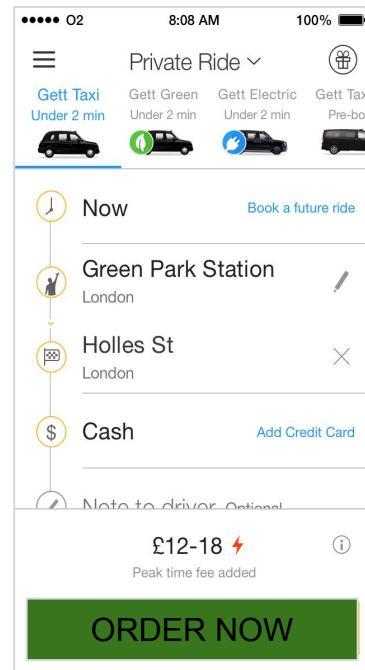
# Экстраполяция результатов: кейс

1. Запустили тест
2. Получили результаты
3. Конверсия значительно выросла на 40%.

То есть эта фича будет генерить +40% к выручке в течение, например, следующих двух лет. Так?



Конверсия: 60%



Конверсия: 84%



## Экстраполяция результатов: кейс

**Нет!**

**Причина:** поведение юзера может измениться. **А/Б тесты** - это про **краткосрочные** изменения!

# Любопытные кейсы

---

**Угадайте более успешную  
вариацию**

---

# Сертификаты безопасности

Форма подписки сервиса:

First Name: \*

Last Name: \*

Email: \*

ZIP/Postal Code: \*

Primary Telephone: \*

submit

First Name: \*

Last Name: \*

Email: \*

ZIP/Postal Code: \*

Primary Telephone: \*



**TRUSTe**  
CERTIFIED PRIVACY

submit

# Сертификаты безопасности

Форму без логотипа заполнили на 12.6% больше пользователей!

В данном случае, **логотип** TRUSTe был опубликован **в неподходящем месте в неподходящее время**

Посетители привыкли видеть этот элемент именно в корзине, а не при заполнении формы.

Многие из них на подсознательном уровне стали подозревать, что сейчас с них спишут деньги.

First Name:*	<input type="text"/>
Last Name:*	<input type="text"/>
Email:*	<input type="text"/>
ZIP/Postal Code:*	<input type="text"/>
Primary Telephone:*	<input type="text"/>

submit

First Name:*	<input type="text"/>
Last Name:*	<input type="text"/>
Email:*	<input type="text"/>
ZIP/Postal Code:*	<input type="text"/>
Primary Telephone:*	<input type="text"/>



**TRUSTe**  
CERTIFIED PRIVACY

submit

# Социальное одобрение

**Get Email Updates (it's Free)**

Join **14,752** others and get free updates!

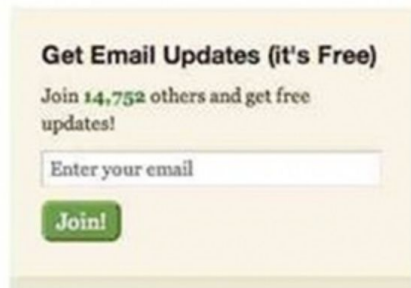
  

**Get Email Updates (it's Free)**

# Сертификаты безопасности

Правый вариант заполнили на 122% больше юзеров!

В мире, где все пресытились огромным количеством аккаунтов в Facebook или Twitter, 14 тысяч подписчиков не выглядит достаточно убедительно, чтобы мотивировать потенциальных клиентов совершить действие.



Get Email Updates (it's Free)

Join 14,752 others and get free updates!

Join!



Get Email Updates (it's Free)

Join!

**Спасибо за  
внимание!  
Вы молодцы!**

**Вопросы?**