

**Урок 7**

**Документация A/B тестов и основные нюансы**

# На этом уроке

1. Рассмотрим распространенные проблемы/подводные камни на разных этапах эксперимента и их возможные решения.
2. Познакомитесь с шаблоном дизайна эксперимента и сможете на этапе планирования избежать многих ошибок.
3. Познакомимся с чек-листом для экспериментов

**Оглавление**

[На этом уроке](#_17drdfz4hzuf)

[Введение](#_yegt6jl8t3bb)

[Процесс A/B тестирования](#_h6ee38z1maa8)

[Ошибки и нюансы на этапе планирования эксперимента](#_9bcesemxa9ej)

[Неправильно выбранный/посчитанный период эксперимента](#_dpuetu8ba6b0)

[Не учтенное окно конверсии](#_s3rb4yfhsomj)

[Множественные изменения](#_pc01kjmbwxgv)

[Пересечения с другими тестами](#_2mzik4jjrpj8)

[Динамические правила](#_mwbfn3v6badc)

[Множественные сравнения](#_uaya7bpbb0fz)

[Ошибки и нюансы на этапе подготовка продуктовых изменений](#_5s6hcxrogjqe)

[Забывать логировать важную информацию](#_6mwzoab1a2w6)

[Не тестировать изменения перед запуском в достаточной степени](#_9btutxwfa80t)

[Ошибки и нюансы на этапе запуска теста](#_f3qmkq1cjcsy)

[Ждать стат значимости дополнительное время/преждевременно завершать тест.](#_az8qhv48e201)

[Внесение изменений в процессе](#_wfj9h53jqybq)

[Плохое сплитование трафика](#_4e02d0of41ev)

[Ошибки и нюансы на этапе анализа результатов теста](#_sp1n12kb47fx)

[Поиск стат значимости перебором различных сегментов/попытки любыми способами цепляться за гипотезу](#_hp56musgy060)

[Отсутствие критериев принятия решения](#_xfvn42cl6vgl)

[Шаблон дизайна эксперимента](#_g098wigrwoy7)

[Пример](#_ntd8yxe7apzl)

[Гипотеза](#_45hm7u73jfpk)

[Что делаем](#_ww7s902va9mj)

[План действий](#_luthxe3h55xw)

[Чек-лист](#_afbcmac8q71d)

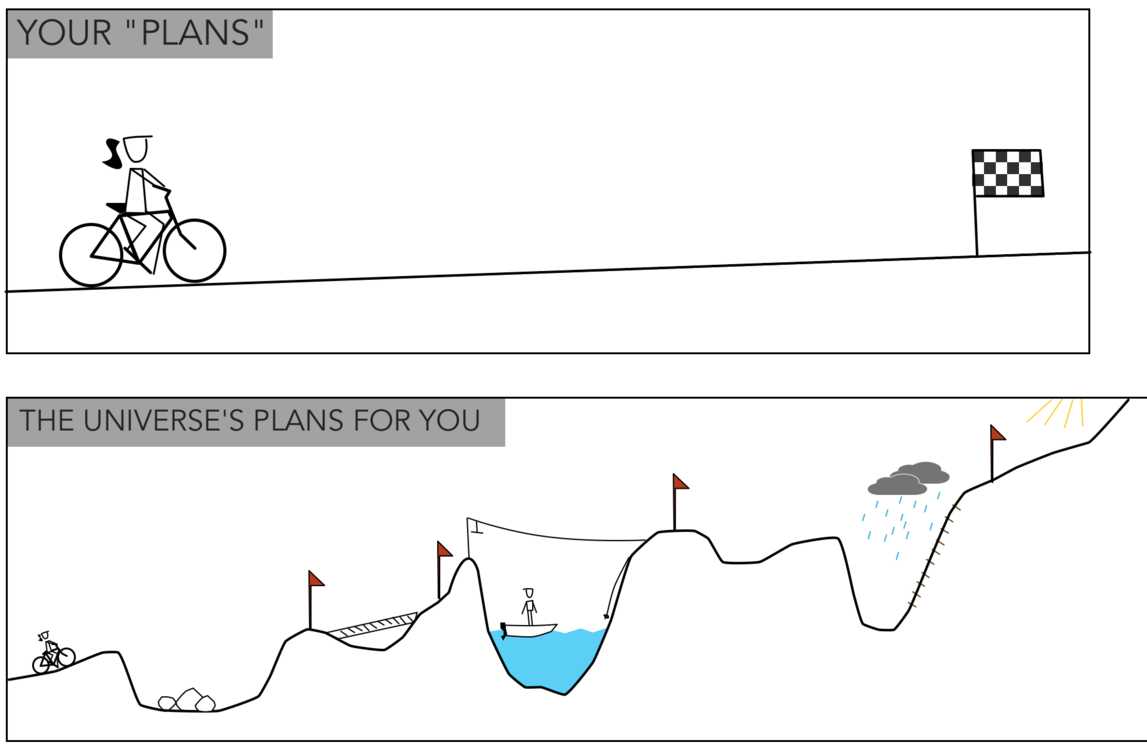
[Практическое задание](#_4t4fygg4zgw3)

[Дополнительные материалы](#_2yv0vvqr61iu)

[Используемые источники](#_z337ya)

# 

# Введение



Ранее мы рассматривали из каких этапов состоит процесс A/B тестирования. На каждом из этапов существуют множество ошибок и нюансов, которые могут приводить к неверной интерпретации теста, раскатке неработающих решений , отклонению хороших гипотез и просто к потраченным впустую ресурсам компании. Правильный дизайн эксперимента позволит не допустить большего числа ошибок.

# Процесс A/B тестирования

Вспомним как может выглядеть процесс a/б тестирования из предыдущего урока.

****

**Возможный процесс:**

**Весь процесс можно разбить на 4-5 этапов:**

**1) Планирование эксперимента**

-Определяем проблему(ы)/ цели и генерируем гипотезы , формируем бэклог

-Приоритезируем их

-Для самых перспективных - продумываем дизайн эксперимента

**2) Подготовка продуктовых изменений**

**3) Запуск теста**

* Запускаем тест, сплитуя юзеров
* Делаем мониторинги(дашборды)

**4) Анализ результатов теста**

**5) Принятие решений**

**Ошибки и нюансы на этапе планирования эксперимента**

* **Неправильно выбранный/посчитанный период эксперимента**
* **Неучтенное окно конверсии**
* **Множественные изменения**
* **Пересечения с другими тестами**
* **Динамические правилав**
* **Множественные сравнения**

**Ошибки и нюансы на этапе подготовка продуктовых изменений**

* **Забывать логировать важную информацию**
* **Не тестировать изменения**

**Ошибки и нюансы на этапе запуска теста**

* **Ждать стат значимости дополнительное время/преждевременно завершать тест**
* **Внесение изменений в процессе**
* **Плохое сплитование трафика**

**Ошибки и нюансы на этапе анализа результатов теста**

* **Поиск стат значимости перебором различных сегментов и метрик /попытки любыми способами цепляться за гипотезу**
* **Отсутствие критериев принятия решения**

Мы собираемся рассмотреть наиболее часто встречающиеся ситуации и понять как можно минимизировать кол-во ошибок при проведении экспериментов.

# Ошибки и нюансы на этапе планирования эксперимента

Начнем с ошибок и нюансов на этапе планирования эксперимента. Как мы уже говорили на предыдущих занятиях любой эксперимент правильно начинать с планирования - это поможет уберечь вас от множества ошибок в будущем и сэкономить время.

Некоторые команды, желая сэкономить время - игнорируют этот этап или подходят к нему формально, делают для галочки.

Эта экономия при планировании может обернуться впустую потраченными ресурсами компании/некачественно проведенным экспериментом.

Какие ошибки нас могут здесь ожидать?

* **Неправильно выбранный/посчитанный период эксперимента**
* **Не учтенное окно конверсии**
* **Множественные изменения**
* **Пересечения с другими тестами**
* **Динамические правилав**
* **Множественные сравнения**

Давайте рассмотрим каждую из них подробнее.

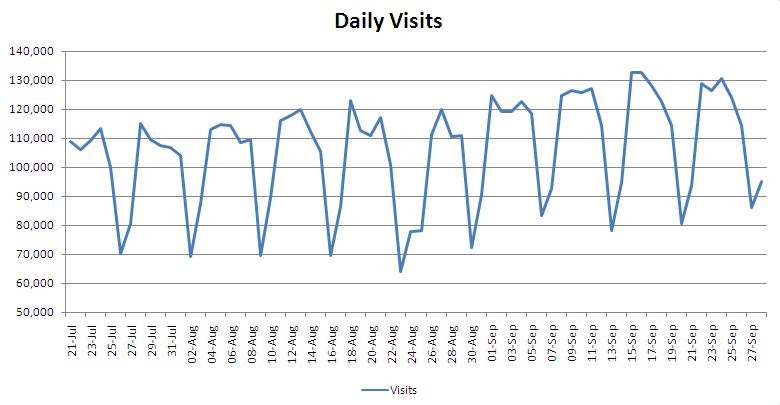
## Неправильно выбранный/посчитанный период эксперимента



Многие в принципе не рассчитывают длительность теста или делают это с ошибками. Это может привести к невалидным результатам(например, к принятию ложноположительных, либо ложноотрицательных решений по продукту).

Здесь можно руководствоваться правилами по расчету выборки из уроков по статистике.

Также часть людей забывают про понятие **сезонность метрики.**

****

Например если в вашем продукте высокая сезонность в рамках недели на выходных - пропуск нескольких дней с высокой/низкой сезонностью может исказить результаты эксперимента. Важно ее учитывать при планировании и по возможности добирать дни до полных n недель, чтобы каждый день недели был представлен равномерно и в тесте не было дисбалансов по дням.

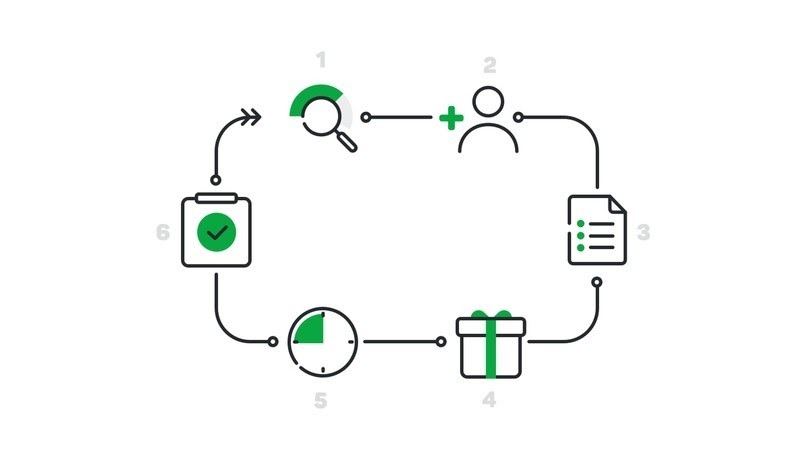
Некоторые любят **запускаться в праздники.**

****

В гендерные, национальные и другие праздники паттерны поведения людей могут очень сильно отличаться от типичных.

При возможности стоит избегать запуска экспериментов в праздники, потому что вы можете получить искаженную картину, которая не будет воспроизводиться в обычные дни. Перенос результатов и выводов эксперимента с праздничных дней на оставшиеся периоды часто оказывается не валидным.

## Не учтенное окно конверсии



Этот пункт перекликается с предыдущим, но я решил его выделить отдельно поскольку он очень важен для получения чистых результатов эксперимента.

**Окно конверсии** — это период времени, в котором мы смотрим за нашими юзерами, ожидая, что они совершат целевое действие.

Соответственно, выгружать статистику и анализировать результаты эксперимента можно только после закрытия окна конверсий для всех пользователей, попавших в эксперимент.

Например, в играх решение может приниматься за неделю/день, а в недвижимости решение о покупке может приниматься полгода.

Допустим, окно конверсии в нашем продукте 5 дней - мы знаем что большинство юзеров совершают целевое действие в течение 5 дней.

Мысленно представим себе когорты пользователей по дате их первого захода в продукт во время эксперимента.

Как вы думаете, если эксперимент идет 14 дней - будет ли у пользователей, которые зашли в продукт только на 13-14 дни столько же времени на совершение конверсии как у пользователей, которые пришли, например, на 5-й день?

Ответ нет:) Часть конверсий у людей из когорт последних дней срежется и это может повлиять на итоговый результат. А что если окно конверсии в вашем продукте 14 дней, а вы крутите эксперимент всего дней 10...?(риторический вопрос).

Очень желательно учитывать этот нюанс в своих продуктах, давая каждой когорте пройти окно конверсии полностью, либо при анализе обрезать когорты, у которых не прошло достаточное время для окна конверсии.

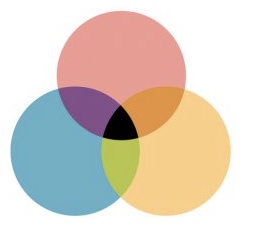
## Множественные изменения



Что же это такое? В чистом A/B тесте группы не должны между собой отличаться больше чем на одно изменение - ведь когда когда мы выкатываем сразу несколько изменений одновременно - мы потом не сможем отделить степень влияния каждой фичи по отдельности и понять что и как повлияло на положительный/негативный результат теста. Результаты могут оказаться бесполезными.

Решение : Делать между группами не более одного отличия . В самых крайних случаях проводить A/B/ ../N тесты, об их нюансах мы поговорим дополнительно далее в пункте про множественные сравнения.

## Пересечения с другими тестами



Этот пункт тесно связан с предыдущим, но причиной могут являться внешние, не зависящие от вашего желания факторы - тесты ваших коллег. Что это такое? Это ситуация, когда один и тот же пользователь может одновременно попадать в несколько тестов.

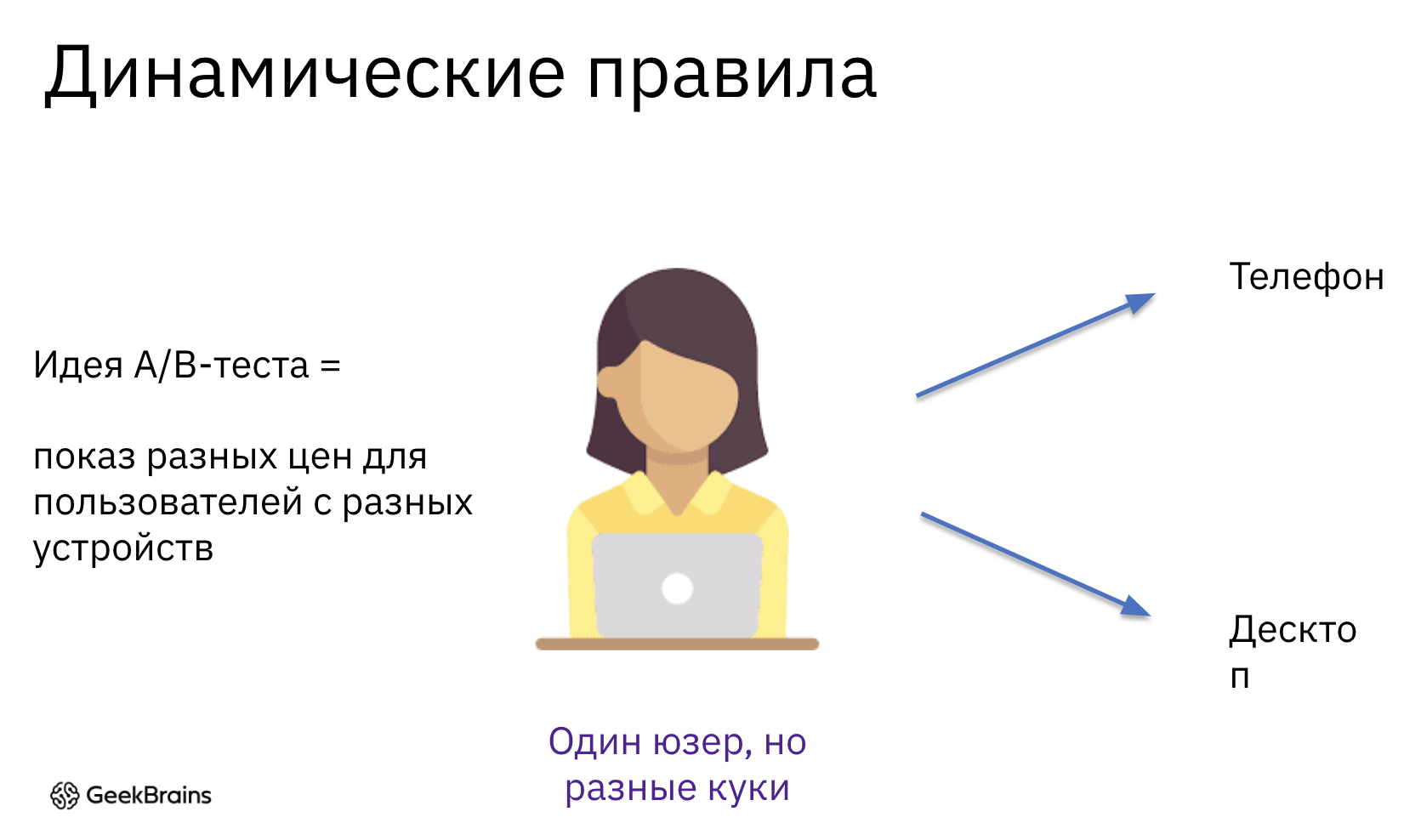
Чем это может быть опасно?

Если на пользователя одновременно влияет множество разных факторов вы потом не сможете определить степень и направление влияния каждого из них.

К сожалению, даже в некоторых известных и сильных продуктах команды забывают про этот нюанс.

Решение здесь довольно простое - вести общий бэклог экспериментов и идей внутри продукта и синхронизироваться с коллегами чтобы не попасть в такую ситуацию.

## Динамические правила

****

Некоторые продукты пытаются показывать разные условия клиентам в зависимости от устройства. Большинство продуктов не имеет устойчивого идентификатора юзера. Чаще всего используются куки, которые могут меняться в разных браузерах + устройствах . Из-за этого один и тот же юзер например может увидеть два разных предложения цены, что может негативно сказаться на его опыте взаимодействия с вашим сервисом. Как итог вы можете получить серьезные репутационные, а в ряде случаев и юридические проблемы. Как правило, потенциальные выгоды от таких действий не перекрывают ущерб для продукта.

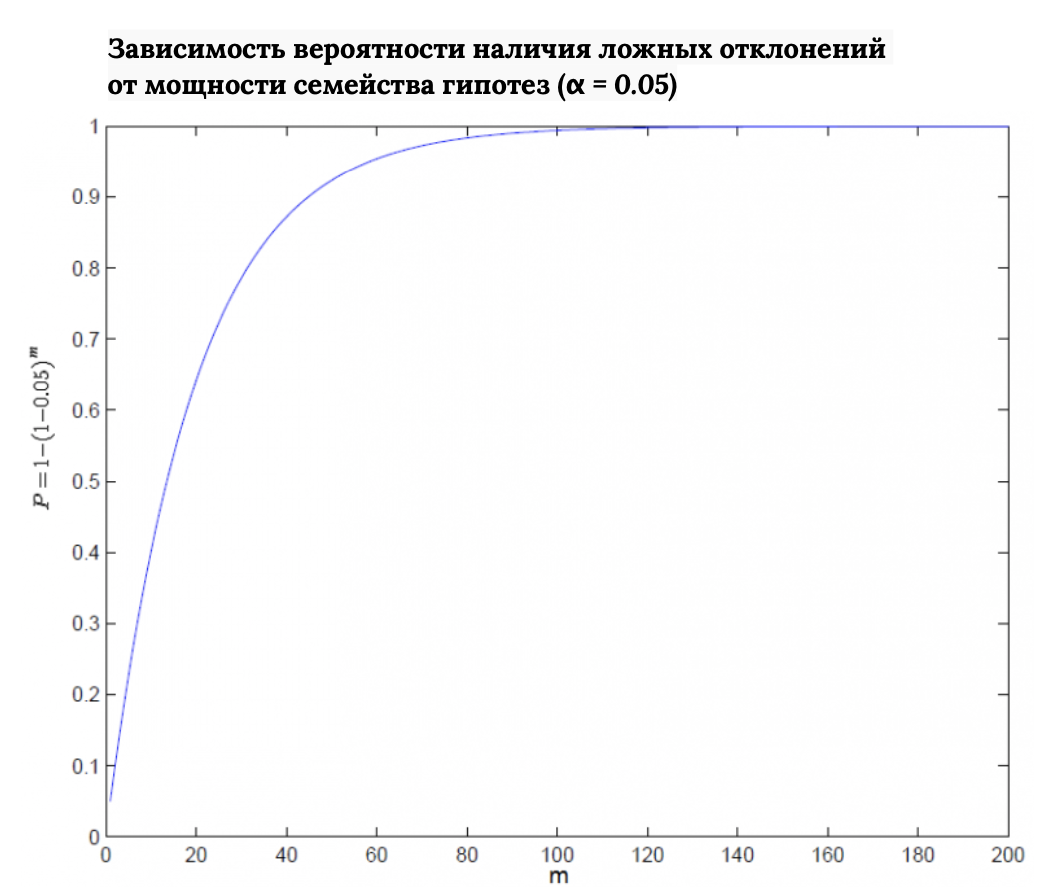
Решение: Не экспериментировать с таким видом тестов. Они могут вызвать негатив и плохую репутацию у сайта (продукта).

## Множественные сравнения

Что же такое множественные сравнения? Это сравнения между собой более 2-х версий чего-либо (каждой версии с каждой).

Например вы запустили A/B/C тест и хотите сравнить 3 версии сайта и понять какая из версий наиболее эффективна.

Какие риски вас могут поджидать?

****

Если сделать N тестов, то вероятность совершить ошибку I рода значительно возрастает по формуле: 1-(1-α)^m

, где m – количество сравнений. В случае, например, A/B/C теста у нас 3 попарных сравнения: A/B, A/C, B/C. Это означает, что при уровне значимости 95%, альфа будет 1- (1-0.05 ) ^ 3 = 0.1426

Как с этим бороться? Есть несколько вариантов вариантов - мы рассмотрим самый простой.

Возьмем для всех наших расчетов ошибку первого рода деленную на кол-во вариантов α/m. Некоторые компании на практике берут заранее маленькую ошибку первого рода(α = 1%),но нужно понимать что у таких компаний зачастую трафик может исчисляться миллионами юзеров в день.

Поправки сильно снижает мощность теста и множественные сравнения не рекомендуется делать без очень острой необходимости. Ведь если вы не Google или Amazon с гигантским объемом трафика - c корректировкой сильно уменьшаете мощность , без корректировки точность эксперимента.

В любом случае вы должны знать о существовании такого нюанса и стараться не тестировать множество вариантов сразу.

# 

# 

# 

# Ошибки и нюансы на этапе подготовка продуктовых изменений

* **Забывать логировать важную информацию**
* **Не тестировать изменения**

## Забывать логировать важную информацию

## 

Часто даже в командах с высоким уровнем культуры экспериментов - в процессе спешки могут забыть либо обвесить метками/целямм/пикселями/ивентами важные куски продукта, либо проверить правильность логирования информации. В результате в лучшем случае команда будет долго думать каким образом оценить собранную информацию, а в худшем ей придется перезапускать тест еще раз.

Как не допустить такое? На этапе планирования стараться синхронизироваться между продактами, разработчиками и аналитиками. Например, под эксперименты можно завести отдельный чат в любом мессенджере. Также можно сделать чек-лист для менеджеров для таких случаев .

## Не тестировать изменения перед запуском в достаточной степени

### 

Часто бывает так, что изменения могут работать не так мы от них ожидаем - не стоит полагаться только на тестировщиков.

Оунер процесса, который лучше всех знает задумку теста должен - проверять новые функции перед запуском.

Это поможет уберечь ваш продукт от лишних репутационных рисков и сделать и провести качественный тест

Именно оунер должен проверить то что сделано по ТЗ на предмет соответствия цели. В случае выкатки серьезных изменений лучше раскатывать их сначала на маленький процент аудитории, наблюдая за тем как все работает в живую.

### 

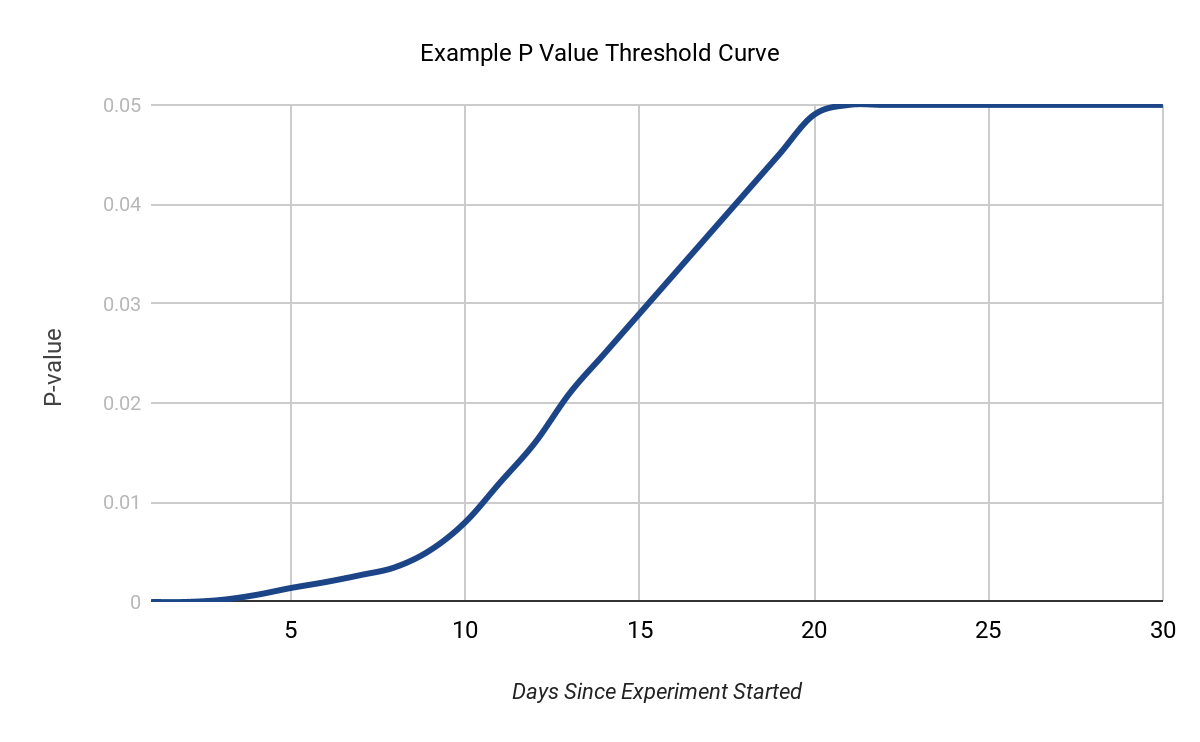
# Ошибки и нюансы на этапе запуска теста

* **Ждать стат значимости дополнительное время/преждевременно завершать тест**
* **Внесение изменений в процессе**
* **Плохое сплитование трафика**

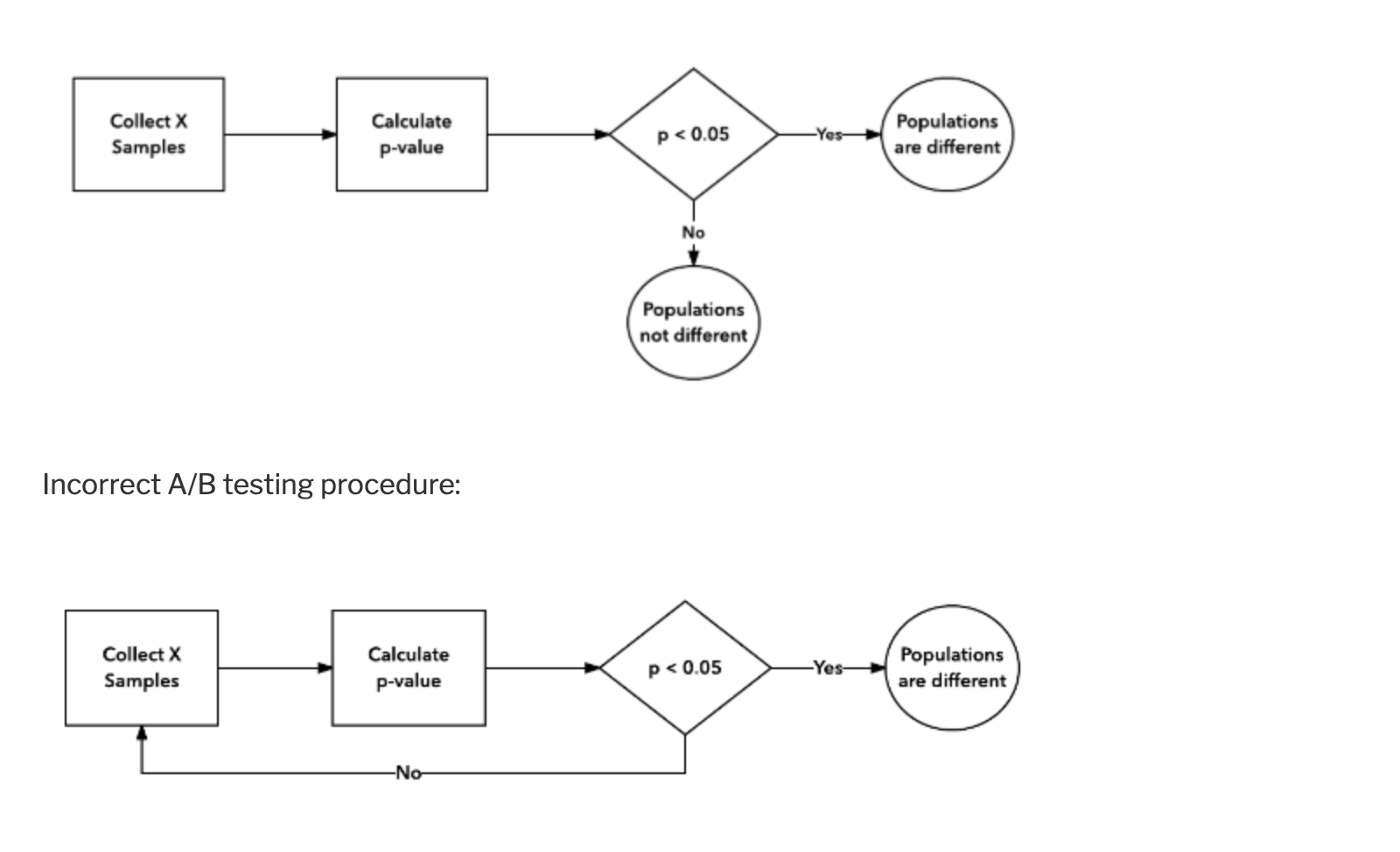
Давайте рассмотрим каждый из них поподробнее.

## Ждать стат значимости дополнительное время/преждевременно завершать тест.

У некоторых команд возникает соблазн остановить эксперимент , когда зафиксирована стат.значимость, не дожидаясь завершения рассчитанного периода. Другие же наоборот ждут достижения стат.значимости уже после завершения эксперимента.



Peeking problem (проблема подглядывания) проявляется, когда вы проверяете промежуточные результаты с готовностью принять решение: раскатить одну из версий, если различие между тестом и контролем окажется .

[](https://gopractice.io/blog/peeking-problem/*)

Это может привести к двум типам последствий:

Преждевременно законченный тест - не валидные результаты.

Продолжение теста в ожидании стат значимости - шанс потратить деньги и время компании не получив результата.

Решение - рассчитывать длительность теста и придерживаться сроков, которые вы посчитали.

Мониторить промежуточные результаты эксперименты можно и нужно, принимать решения на основе неполных данных нельзя.

## Внесение изменений в процессе

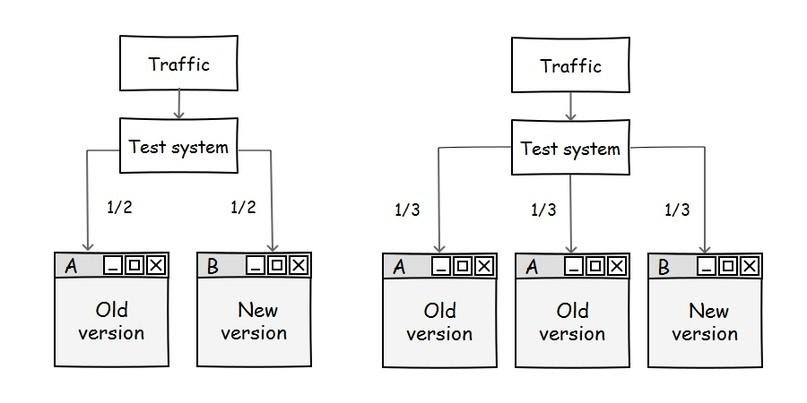
### 

Часть команд практикует внесение дополнительных продуктовых изменений в уже запущенном тесте.

Это не очень хорошая практика, которая может изменить опыт пользователей и затруднить анализ эксперимента.

После внесения изменений вы тестируете другой продукт. Решение здесь довольно простое - не делать хаотичные изменения, когда вы уже запустили тест, лучше потратьте чуть больше времени на качественное планирование эксперимента и четкое понимание его целей.

## Плохое сплитование трафика

****

Неравномерное распределение трафика может сильно исказить результаты. Отличия в метриках которые вы получите в рамках эксперимента могут быть обусловленные не продуктовыми изменениями а разницей в аудитории в двух группах. Например, в одной из групп преобладает аудитория из МСК/СПБ , в другой же наоборот больше людей из регионов.

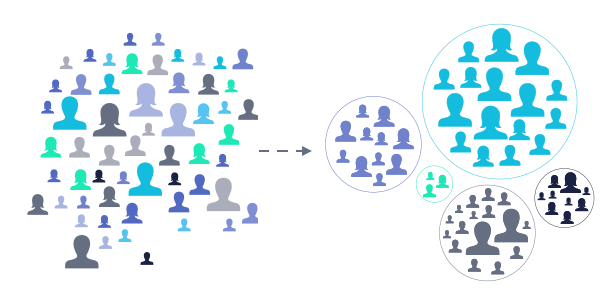
Такое может происходить из-за недоработок или багов инструмента для проведения A/B-тестов.

Решение - проводить A/A или (A/A/B) тесты хотя бы раз в квартал и проверять качество вашей системы сплитования.

# Ошибки и нюансы на этапе анализа результатов теста

* **Поиск стат значимости перебором различных сегментов и метрик /попытки любыми способами цепляться за гипотезу**
* **Отсутствие критериев принятия решения**

## Поиск стат значимости перебором различных сегментов/попытки любыми способами цепляться за гипотезу



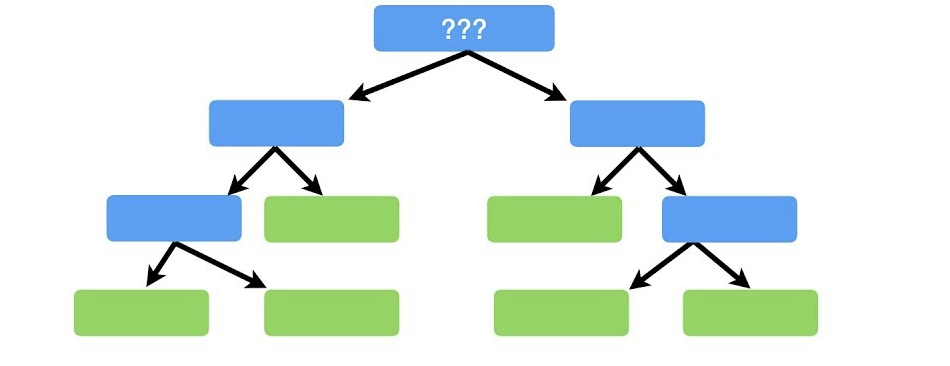
Многие аналитики и продакты при запуске экспериментов очень сильно привязываются к своей гипотезе - и всеми силами пытаются ее подтвердить. В результате когда в основной метрике отличий между группами найти не удалось, у человека может возникнуть соблазн перебирать постфактум разные срезы, сегменты и метрики в поисках стат значимости, пытаясь любыми силами подогнать факты под теорию и пытаться принять решение на основе этого .

Так делать не стоит ! В этом случае мы совершаем множественные сравнения и сильно рискуем случайно в одном из срезов обнаружить желаемую разницу. И как же тут быть?

**Можно** смотреть важные срезы по полу/возрасту (мобайл/десктоп) итд ,если это обосновывается идеей эксперимента. Можно использовать интересные файндинги полученные в срезах для генерации гипотез будущих исследований.

**Нельзя** пытаться использовать рандомные, не обоснованные идеей эксперимента, срезы при принятии решений. Если вы просто перебирали разные комбинации и обнаружили где-то стат значимость - не нужно радостно бежать и обосновывать этим успешность теста.

# Отсутствие критериев принятия решения

****

Тут очень важно понимать с какой целью в принципе проводятся различные виды анализов, включая A/B тесты. А проводятся они для того, чтобы помочь нам принять решение на основе данных.

Без четко определенных критериев принятия решений - вы рискуете зря потратить время на тест и после его завершения плохо понимать, что нужно делать.

Решение: При планировании эксперимента заранее определить для себя критерии принятия решений в зависимости от различных исходов теста и влияния продуктовых фич на метрики.

Можно использовать следующие вопросы:

При каких развитиях событий и получении каких результатов какие решения я принимаю (например, если по результатам эксперимента , конверсия в покупку ниже на 2 п.п., то сразу же вырубаем эксперимент итд**)**

Например : Если основная метрика растет в нужном диапазоне и не падает добавочная, раскатываем на всех.

Если основная растет, а добавочная падает - стопаем тест, разбираемся, пытаемся придумать новые гипотезы.

Если основная падает, добавочная падает - стопаем тест, разбираемся, пытаемся придумать новые гипотезы. итд

Многих ошибок можно избежать используя чек-лист и шаблон дизайна эксперимента.

# Шаблон дизайна эксперимента

Для чего нужен шаблон дизайна эксперимента?

1. Это полноценный документ.
2. Это исторические данные.
3. Это хорошая практика, чтобы не замыкать на себе процессы, если вы решите сменить место работы.
4. Это может помочь задать себе правильные вопросы перед запуском и избежать ненужных ошибок.

Недостатки, которые можно продумать за час планирования и проектирования, с легкостью компенсируются месяцем переработок по ночам :)



**Краткая версия шаблона с вопросами для самопроверки:**

1. **Гипотеза** (Какую проблему решаем? Какая сейчас ситуация? Зачем что-то менять?).
2. **Что делаем** (Что будут видеть пользователи в каждой из групп?).
3. **На каких пользователях тестируем** (Почему именно на этих группах пользователей?).
4. **Ключевые метрики для оценки эксперимента.** (Основные метрики (что хотим улучшить) ,

Добавочные метрики (что хотим не уронить):

1. **Выводы** (Какие результаты получили, какие выводы сделали, какие решения приняли).

## Пример

### Гипотеза

Если на странице лендинга факультета веб-аналитики GeekBrains поднять блок «На курсе вы научитесь» на второй экран, то % показателя отказов снизится на 10% а СR% подачи заявки увеличится на 1%, потому что по результатам интервью, навыки, которым обучатся студенты, для них важнее, чем объяснение, кто такой веб-аналитик.

### Что делаем

*Контрольная версия:* оставляем текущий вид страницы. *Тестовая версия:* передвигаем блок «На курсе вы научитесь» на первый экран под кнопку «Начать обучение».

**На каких пользователях тестируем:** только на новых пользователях.

**Метрики:** Процент конверсий в заявку (%CR) — primary. Процент показателя отказов — secondary.

### План действий

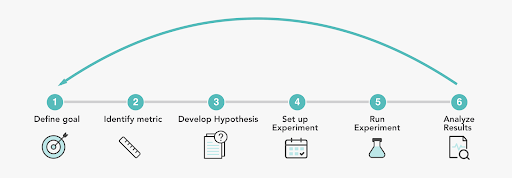
**Если** наш эксперимент будет положительным и мы зафиксируем ожидаемое улучшение в в ключевых метриках и не посадим добавочные, то масштабируем изменение и «выкатываем» его на всех пользователей.

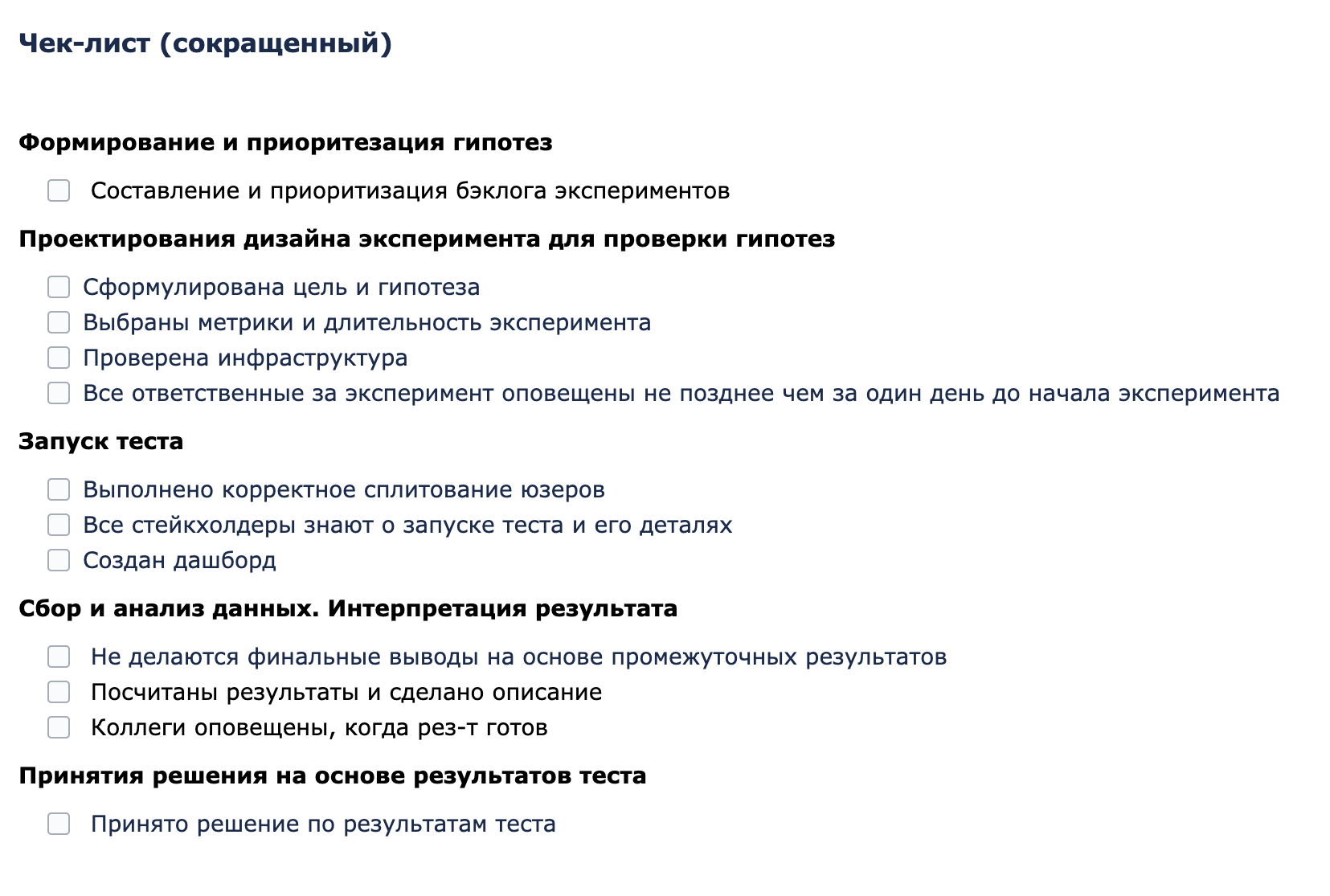
**Если** основные метрики падают или растут недостаточно либо добавочные метрики падают, откатываем эксперимент.

В процессе проведения эксперимента может поджидать большое число нюансов. Важно понимать эти нюансы, а также способы минимизации негативного эффекта от них.

# Чек-лист

Многих ошибок можно избежать используя чек-лист и шаблон дизайна эксперимента.





# Практическое задание

Составьте шаблон дизайна эксперимента для гипотезы, которая набрала больше всего баллов в практическом задании предыдущего урока

Формат: gdocs c возможностью редактирования/комментирования.

# Дополнительные материалы

1. [Структура гипотезы](https://help.optimizely.com/Ideate_and_Hypothesize/Design_an_effective_hypothesis).
2. [Как используют когнитивные искажения в бизнесе](https://vc.ru/marketing/154908-22-kognitivnyh-iskazheniya-i-kak-ih-ispolzovat-v-marketinge)
3. [Как Uber улучшает пользовательский опыт с помощью психологии](https://rb.ru/story/how-uber-uses-psychology/).
4. [Как построить непрерывный процесс growth hacking](https://vc.ru/marketing/96436-growth-hacking-kak-postroit-nepreryvnyy-process-po-vzlamyvaniyu-rosta).
5. [Обзорная статья по методам приоритизации](https://cxl.com/blog/better-way-prioritize-ab-tests/).
6. [Неплохая статья по ICE](https://skillsetter.io/blog/how-to-prioritize-hypothesis).
7. [ICE/RICE](https://habr.com/ru/company/hygger/blog/422131/).
8. [Довольно подробно про уверенность в идее](https://itamargilad.com/the-tool-that-will-help-you-choose-better-product-ideas/).
9. [Yet another article](https://geekbrains.ru/posts/tasks_scoring).
10. «Принцип пирамиды», Барбара Минто.
11. [Неплохой вводный туториал по A/B тестам](https://cxl.com/blog/ab-testing-guide/)
12. [Ухудшающие эксперименты](https://gopractice.ru/ab-test/)
13. [Интересная статья про то как Flo повысили долю успешных экспериментов](https://gopractice.ru/how_to_increase_the_number_of_successful_experiments/)
14. [Статья для общего развития про проблему множественной проверки гипотез](https://habr.com/ru/company/yandex/blog/476826/)
15. [Сравнение нескольких готовых инструментов для A/Б тестирования](https://insightwhale.com/google-optimize-vs-optimizely-comparing-a-b-testing-tools/)
16. [Курс по Google Optimize](https://www.udemy.com/course/learn-google-optimize-course/)
17. [Туториал по Google Optimize](https://support.google.com/optimize/answer/9340015?hl=en)
18. [Статья мини-гайд по Google Optimize](https://blog.click.ru/growthhacking/google-optimize-kak-testirovat-ux-sajta-bez-programmist)
19. [Google Optimize Beginner’s Guide](https://cxl.com/blog/google-optimize/)
20. Замечательная книга по A/B тестам [Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing](https://www.cambridge.org/core/books/trustworthy-online-controlled-experiments/D97B26382EB0EB2DC2019A7A7B518F59)
21. [Статья на хабре про множественные сравнения](https://habr.com/ru/company/yandex/blog/476826/)
22. .[Статья на хабре про проверку качества системы сплитования](https://habr.com/ru/company/hh/blog/321386/)
23. [Статья про проблему подглядывания](https://gopractice.ru/how-not-to-analyze-abtests/)
24. [Отличная статья на хабре про проблемы в A/B тестах](https://habr.com/ru/company/jugru/blog/358104/)
25. Замечательная книга по A/B тестам [Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing](https://www.cambridge.org/core/books/trustworthy-online-controlled-experiments/D97B26382EB0EB2DC2019A7A7B518F59)
26. [Кладезь знаний от Майкрософт по экспериментам](https://exp-platform.com/)

# Используемые источники

1. [Структура гипотезы](https://help.optimizely.com/Ideate_and_Hypothesize/Design_an_effective_hypothesis).
2. [Как используют когнитивные искажения в бизнесе](https://vc.ru/marketing/154908-22-kognitivnyh-iskazheniya-i-kak-ih-ispolzovat-v-marketinge)
3. [Как Uber улучшает пользовательский опыт с помощью психологии](https://rb.ru/story/how-uber-uses-psychology/).
4. [Как построить непрерывный процесс growth hacking](https://vc.ru/marketing/96436-growth-hacking-kak-postroit-nepreryvnyy-process-po-vzlamyvaniyu-rosta).
5. [Обзорная статья по методам приоритизации](https://cxl.com/blog/better-way-prioritize-ab-tests/).
6. [Неплохая статья по ICE](https://skillsetter.io/blog/how-to-prioritize-hypothesis).
7. [ICE/RICE](https://habr.com/ru/company/hygger/blog/422131/).
8. [Довольно подробно про уверенность в идее](https://itamargilad.com/the-tool-that-will-help-you-choose-better-product-ideas/).
9. [Yet another article](https://geekbrains.ru/posts/tasks_scoring).
10. «Принцип пирамиды», Барбара Минто.
11. [Неплохой вводный туториал по A/B тестам](https://cxl.com/blog/ab-testing-guide/)
12. [Ухудшающие эксперименты](https://gopractice.ru/ab-test/)
13. [Интересная статья про то как Flo повысили долю успешных экспериментов](https://gopractice.ru/how_to_increase_the_number_of_successful_experiments/)
14. [Статья для общего развития про проблему множественной проверки гипотез](https://habr.com/ru/company/yandex/blog/476826/)
15. [Сравнение нескольких готовых инструментов для A/Б тестирования](https://insightwhale.com/google-optimize-vs-optimizely-comparing-a-b-testing-tools/)
16. [Курс по Google Optimize](https://www.udemy.com/course/learn-google-optimize-course/)
17. [Туториал по Google Optimize](https://support.google.com/optimize/answer/9340015?hl=en)
18. [Статья мини-гайд по Google Optimize](https://blog.click.ru/growthhacking/google-optimize-kak-testirovat-ux-sajta-bez-programmist)
19. [Google Optimize Beginner’s Guide](https://cxl.com/blog/google-optimize/)
20. Замечательная книга по A/B тестам [Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing](https://www.cambridge.org/core/books/trustworthy-online-controlled-experiments/D97B26382EB0EB2DC2019A7A7B518F59)
21. [Статья на хабре про множественные сравнения](https://habr.com/ru/company/yandex/blog/476826/)
22. .[Статья на хабре про проверку качества системы сплитования](https://habr.com/ru/company/hh/blog/321386/)
23. [Статья про проблему подглядывания](https://gopractice.ru/how-not-to-analyze-abtests/)
24. [Отличная статья на хабре про проблемы в A/B тестах](https://habr.com/ru/company/jugru/blog/358104/)
25. Замечательная книга по A/B тестам [Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing](https://www.cambridge.org/core/books/trustworthy-online-controlled-experiments/D97B26382EB0EB2DC2019A7A7B518F59)
26. [Кладезь знаний от Майкрософт по экспериментам](https://exp-platform.com/)

# 