

**Урок 5**

**Основы математической статистики**

# На этом уроке

* Поговорим о важности статистики для A/B тестов
* Пройдемся по базовым понятиям статистики
* Посмотрим, что такое Центральная предельная теорема и для чего она применяется в A/B тестировании
* Разберем как оценивать по выборке в каких границах лежат реальные значения ваших метрик
* Рассмотрим как минимизировать выкатку в продакшн неработающих решений и отклонение работающих

# Оглавление

[На этом уроке](#_4wfp8dsrmxhk)

[Оглавление](#_xh2q54mujgl5)

[Важность статистики в A/B тестах](#_gg9rv8vvpiy8)

[Базовые понятия статистики](#_l26xwgyobep3)

[Законы распределения](#_jqsy9kwnm0m0)

[Центральная предельная теорема](#_xwzv8vex1igj)

[Доверительные интервалы](#_dl617upju9tu)

[Ошибки первого и второго рода](#_5e6dkg28xw3b)

[Дополнительные материалы](#_73p6n4irf3ri)

[Используемые источники](#_z337ya)

[Практическое задание](#_trdj6w4lltjg)

# Важность статистики в A/B тестах



Как вы думаете возможно ли построить хороший дом без фундамента? Тут очевидно напрашивается единственно верный ответ.

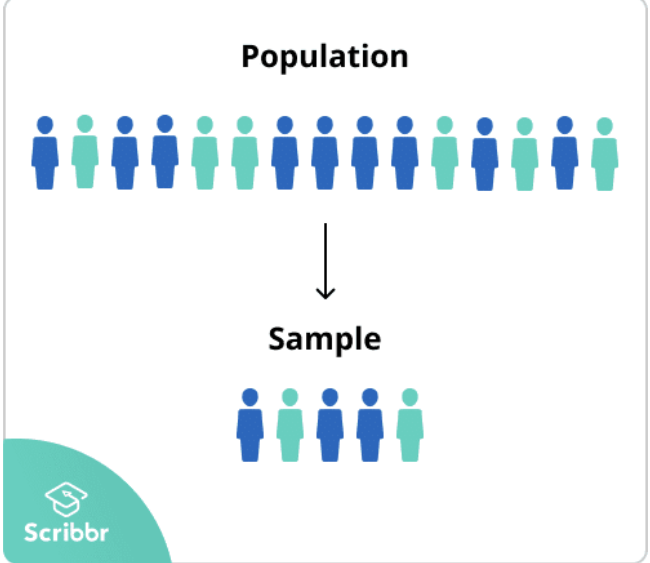
Аналогично, и в A/B тестах нельзя обойтись без базовых знаний по математической статистике, если вы хотите проводить в своем продукте качественные и полезные для бизнеса эксперименты .

Зная основы математической статистики **вы сможете:**

* Оценивать по выборке в каких границах лежат реальные значения ваших метрик
* Сравнивать между собой выборки по различным параметрам
* Понимать, где вы получили случайную разницу в метриках, а где нет
* Рассчитывать длительность теста
* Минимизировать выкатку в продакшн неработающих решений
* Экономить команде время и деньги :)

# 

# Базовые понятия статистики



Начнем с базовых понятий:

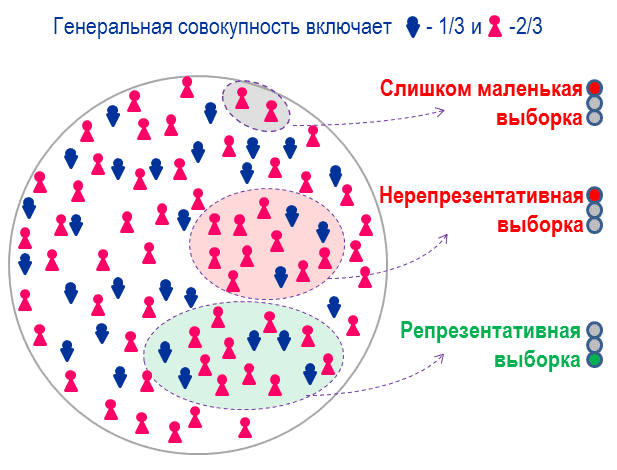
**Генеральная совокупность** - совокупность всех объектов или наблюдений, относительно которых исследователь намерен делать выводы при решении конкретной задачи. В ее состав включаются все объекты, которые подлежат изучению.

**Выборка** - часть генеральной совокупности с помощью определённой процедуры выбранных из генеральной совокупности для участия в исследовании.

Могут возникнуть возникнуть вопросы: зачем же нужна выборка если есть генеральная совокупность? Почему сразу все не считать по ней?

В реальной жизни, у исследователей к сожалению, почти всегда нет доступа ко всем данным генеральной совокупности , это может быть долго, дорого , а иногда и физически невозможно.

Чтобы переносить выводы с выборки на генеральную совокупность, выборка должна быть **репрезентативной,** отражать пропорции и особенности генеральной совокупности.



На репрезентативность сильно влияет метод отбора выборки(их существует довольно много). На практике часто используется [рандомизация.](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D0%BD%D0%B4%D0%BE%D0%BC%D0%B8%D0%B7%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BA%D0%BE%D0%BD%D1%82%D1%80%D0%BE%D0%BB%D0%B8%D1%80%D1%83%D0%B5%D0%BC%D0%BE%D0%B5_%D0%B8%D1%81%D0%BF%D1%8B%D1%82%D0%B0%D0%BD%D0%B8%D0%B5#:~:text=%D0%A0%D0%B0%D0%BD%D0%B4%D0%BE%D0%BC%D0%B8%D0%B7%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%BD%D0%BE%D0%B5%20%D0%BA%D0%BE%D0%BD%D1%82%D1%80%D0%BE%D0%BB%D0%B8%D1%80%D1%83%D0%B5%D0%BC%D0%BE%D0%B5%20%D0%B8%D1%81%D0%BF%D1%8B%D1%82%D0%B0%D0%BD%D0%B8%D0%B5%20(%D1%80%D0%B0%D0%BD%D0%B4%D0%BE%D0%BC%D0%B8%D0%B7%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%BD%D0%BE%D0%B5%20%D0%BA%D0%BE%D0%BD%D1%82%D1%80%D0%BE%D0%BB%D0%B8%D1%80%D1%83%D0%B5%D0%BC%D0%BE%D0%B5,%D0%BF%D1%80%D0%BE%D0%B2%D0%B5%D1%80%D0%BA%D0%B5%20%D1%8D%D1%84%D1%84%D0%B5%D0%BA%D1%82%D0%B8%D0%B2%D0%BD%D0%BE%D1%81%D1%82%D0%B8%20%D0%BD%D0%BE%D0%B2%D1%8B%D1%85%20%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%BE%D0%B2%20%D0%BB%D0%B5%D1%87%D0%B5%D0%BD%D0%B8%D1%8F.))

**Пример:** Допустим в нашем исследовании - мы хотим знать средний рост мужчин в Москве живущих на данный момент .

Генеральная совокупность в данном случае - это все мужчины живущие мужчины в Москве сейчас (будет исчисляться миллионами).

Выборка - например, случайно отобранные 10 к мужчин из разных районов.

По этим мужчинам мы замерим рост и сможем оценить его истинное значение с определенной точностью, поговорим дальше как это сделать:)

**Описываем выборку и ГС:**

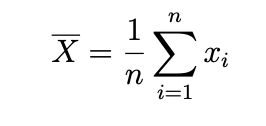
В большинстве случаев как в бизнесе так и в научных исследованиях вы будете считать метрики и делать выводы по выборке, а не по всей генеральной совокупности. Вам нужно уметь описывать и то и другое, в этом вам помогут базовые понятия:

Желающие могут освежить знания ознакомившись со [статьей](https://medium.com/swlh/the-8-basic-statistics-concepts-for-data-science-7b865fca92b9) или посмотреть [видео](https://www.youtube.com/watch?v=4SEX9Bozktw).

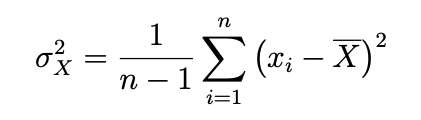
[**Случайная величина (ξ)**](https://ru.wikipedia.org/wiki/%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D0%B0) – это математическое понятие, служащее для представления [случайных явлений](https://ru.wikipedia.org/wiki/%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%BE%D1%81%D1%82%D1%8C), когда для них может быть определена их [вероятность](https://ru.wikipedia.org/wiki/%D0%92%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D1%8C), то есть мера возможности наступления.

По сути это переменная со значениями. Для каждого значения есть своя вероятность исхода

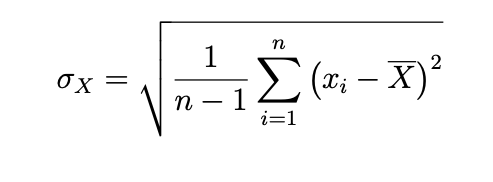
[**Математическое ожидание (*μ, M*)**](https://www.youtube.com/watch?v=4SEX9Bozktw) — это среднее арифметическое значение случайной величины.



**Дисперсия (*S2*)** – рассчитанное расстояние, на которое значения случайной величины находятся вокруг его математического ожидания



**Стандартное отклонение (*SD***)— это квадратный корень от дисперсии



**Пример:** Мы хотим узнать среднюю цену блюда в ресторане 1 и то насколько сильно в нем колеблются цены, для этого можем подсчитать базовые описательные статистики.

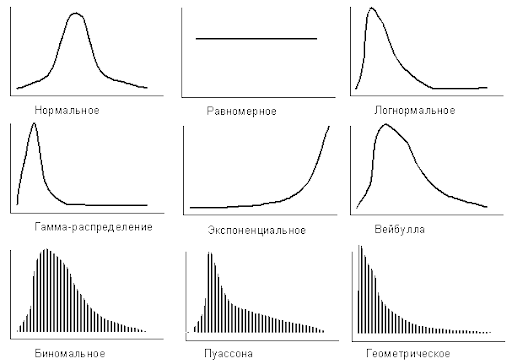
# 

# Законы распределения

Для описания случайных величин используются законы распределения с параметрами.

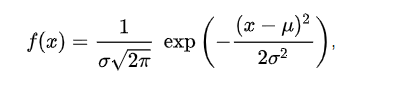
**Распределение вероятностей** — это закон, описывающий область значений [случайной величины](https://ru.wikipedia.org/wiki/%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D0%B0) и соответствующие вероятности появления этих значений.

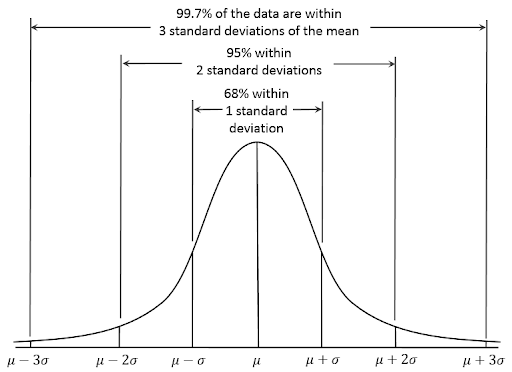
На рисунке ниже визуализированы различные законы распределения:



Нормальное распределение:

Нормальное распределение самый частый вид, который используется в статистике. Закон распределения задается по формуле:





Где, 𝞵 - математическое ожидание и 𝞼 - стандартным отклонением. Эти параметры моделируют, как будет выглядеть распределение. Большая часть тестов при проверке гипотез требует в качестве условий корректности работы- нормальность распределения или принадлежность к какому-либо закону распределения.

# Центральная предельная теорема

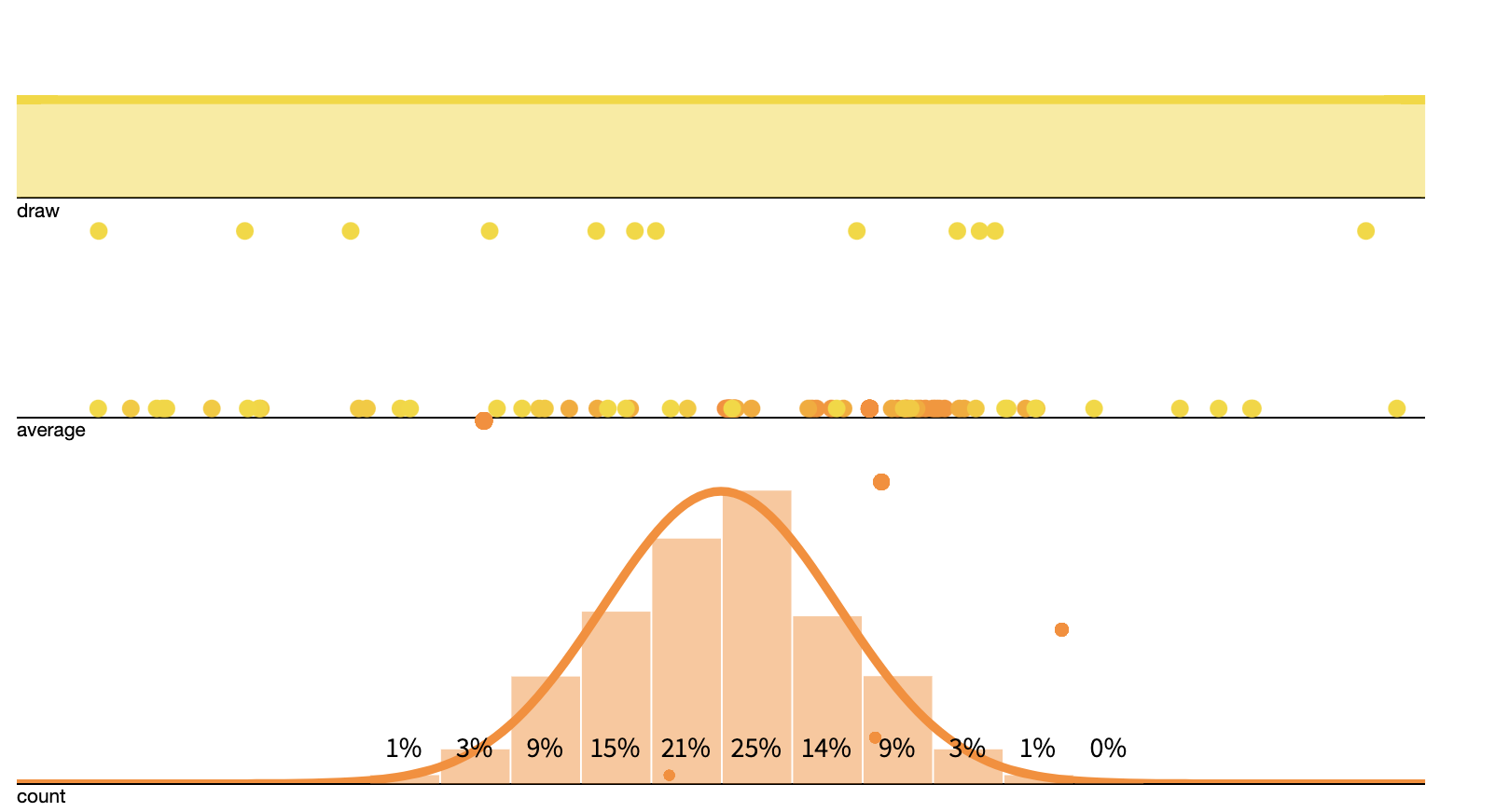
Как можно судить о свойствах генеральной совокупности по выборке?

[ЦПТ](https://www.marketing.spb.ru/lib-around/stat/Naked_Statistics.htm) говорит, что если мы возьмем достаточно большую выборку из независимых, одинаково распределенных случайных величин( [i.i.d](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D1%8B%D0%B5_%D0%BE%D0%B4%D0%B8%D0%BD%D0%B0%D0%BA%D0%BE%D0%B2%D0%BE_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D1%91%D0%BD%D0%BD%D1%8B%D0%B5_%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B5_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D1%8B#:~:text=%D0%A4%D1%80%D0%B0%D0%B7%D0%B0%20%C2%AB%D0%BD%D0%B5%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D1%8B%D0%B5%20%D0%BE%D0%B4%D0%B8%D0%BD%D0%B0%D0%BA%D0%BE%D0%B2%D0%BE%20%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D1%91%D0%BD%D0%BD%D1%8B%D0%B5%C2%BB%20%D1%87%D0%B0%D1%81%D1%82%D0%BE,%D1%80%C2%BB.)) из генеральной совокупности , то среднее значение будет [нормально распределено](https://ru.wikipedia.org/wiki/%D0%9D%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D1%8C%D0%BD%D0%BE%D0%B5_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5) с ***μ(мат ожидание) и SD(стандартное отклонение) .***

Причем мы можем посчитать оценить параметры *μ и SD по нашей выборке.*

[**Визуализация ЦПТ**](https://seeing-theory.brown.edu/probability-distributions/index.html)

Берем и многократно извлекаем выборки определенного размера и считаем по ним среднее. Распределение средних при соблюдении предпосылок выше будет нормально распределенным и мы сможем оценивать истинное значение в ГС.



(промежуточный скрин из визуализатора показывает что мы постепенно будем приближаться к нормальному распределению при многократном извлечении)

Что нам позволяет делать ЦПТ ?

ЦПТ дает возможность строить доверительные интервалы и проверять статистические гипотезы. Поговорим об этом ниже. При желании вы можете прочитать [хороший разбор ЦПТ тут.](https://www.marketing.spb.ru/lib-around/stat/Naked_Statistics.htm)

# Доверительные интервалы

Допустим у вас e-commerce магазин и вы хотите посчитать какое истинное значение конверсии(по всей ГС) из добавления товаров в корзину в покупку для вашего продукта .

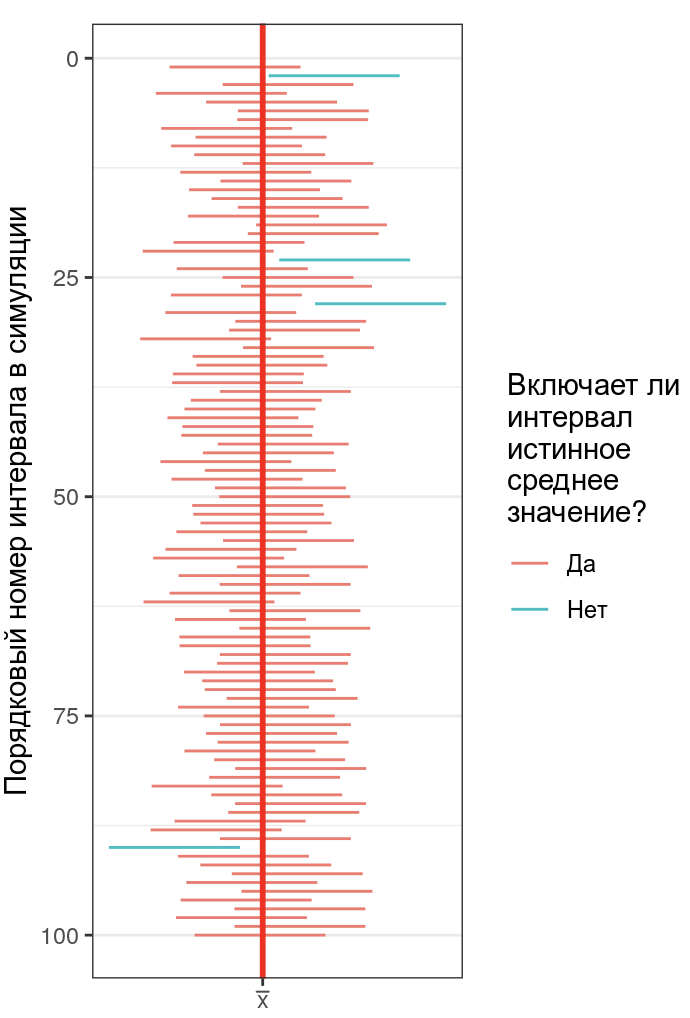
К сожалению, истинное значение конверсии вы не знаете - но зато у вас есть доступ к данным ваших систем аналитики и вы сможете посчитать выборочное значение и по нему оценить в каком диапазоне может лежать истинное значение конверсии. В этом нам поможет доверительный интервал, который мы можем применять благодаря ЦПТ.

**Доверительный интервал для долей**

Что это такое ? Это способ оценки метрики, используя который, мы получим диапазон значений [x,y] , внутри которого будет лежать истинное значение метрики ГС в 95% случаев.

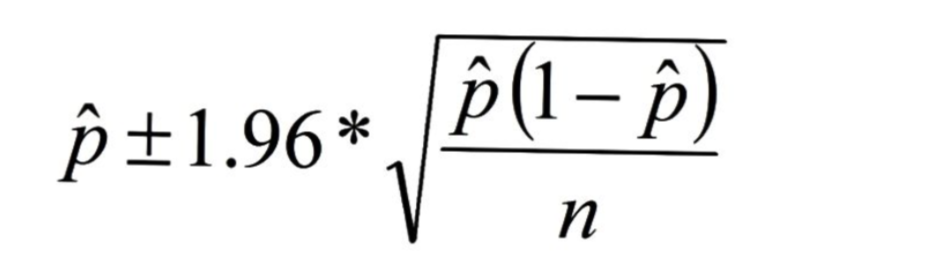
(Если провести очень большое количество независимых экспериментов с аналогичным построением доверительного интервала, то в 95% экспериментов доверительный интервал будет содержать оцениваемый параметр ген совокупности.

В оставшихся 5% экспериментов доверительный интервал не будет содержать параметр ген совокупности.)



(Мы можем регулировать точность и выбирать 99% , 90% и другие диапазоны.)

Для метрик долей\*(ретеншн, конверсия и тп) используется формула



\*Типы метрик которые бывают в экспериментах:

1)Доли - (ретеншн, конверсии) [0,1,0,0,0,1]

2)Непрерывные - таймспент в сек / деньги

3)Отношения - (клики на сессию , поездки на юзера))

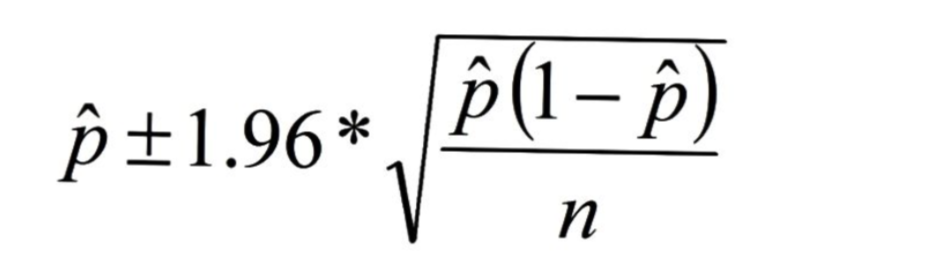
Вы можете рассчитать доверительный интервал на калькуляторе:

[Ссылка на калькулятор](https://www.rotmistrov.com/intrvl)

**Пример расчета доверительного интервала для конверсии:**

Допустим по выборке 1000 юзеров мы посчитали конверсию из добавления товаров в корзину в покупку для нашего e-commerce магазина , конверсия равнялась 10%.

Мы хотим с уровнем доверия 95% оценить истинное значение конверсии. Для этого используем формулу:



0.10 +- 1.96 \* (0.10 \* (1-0.10) /1000) ^ 1/2

Доверительный интервал будет следующий:

[8.14% , 11.86% ]

Давайте попробуем проинтерпретировать полученный результат:

В нашем e-commerce магазина истинная конверсия лежит в диапазоне [8.14% , 11.86% ] с 95% уровнем доверия .

Если бы мы много раз оценивали по выборкам конверсию - ее истинное значение лежало бы в 95% cлучаев в границах доверительного интервала в каждом из расчетов. При желании чуть глубже погрузиться вы можете посмотреть [видео 1](https://www.youtube.com/watch?v=MstzroncW28). + [видео 2](https://www.youtube.com/watch?v=GoK7sod8dhU)

Как вы могли заметить по формуле доверительного интервала: Чем больше выборка, тем точнее оценка.

# Ошибки первого и второго рода

Рассмотрим базовые понятия:

**Нулевая гипотеза** – принимаемое предположение о том, что не существует связи между наблюдениями в двух (или более) событиях (выборках, феноменах, совокупностях). Гипотезу отвергают, если данные показывают разницу между выборками.

**True Positive** = говорим истина, когда по факту истина (факт)**True Negative** = говорим не истина, когда по факту тоже не истина (факт)

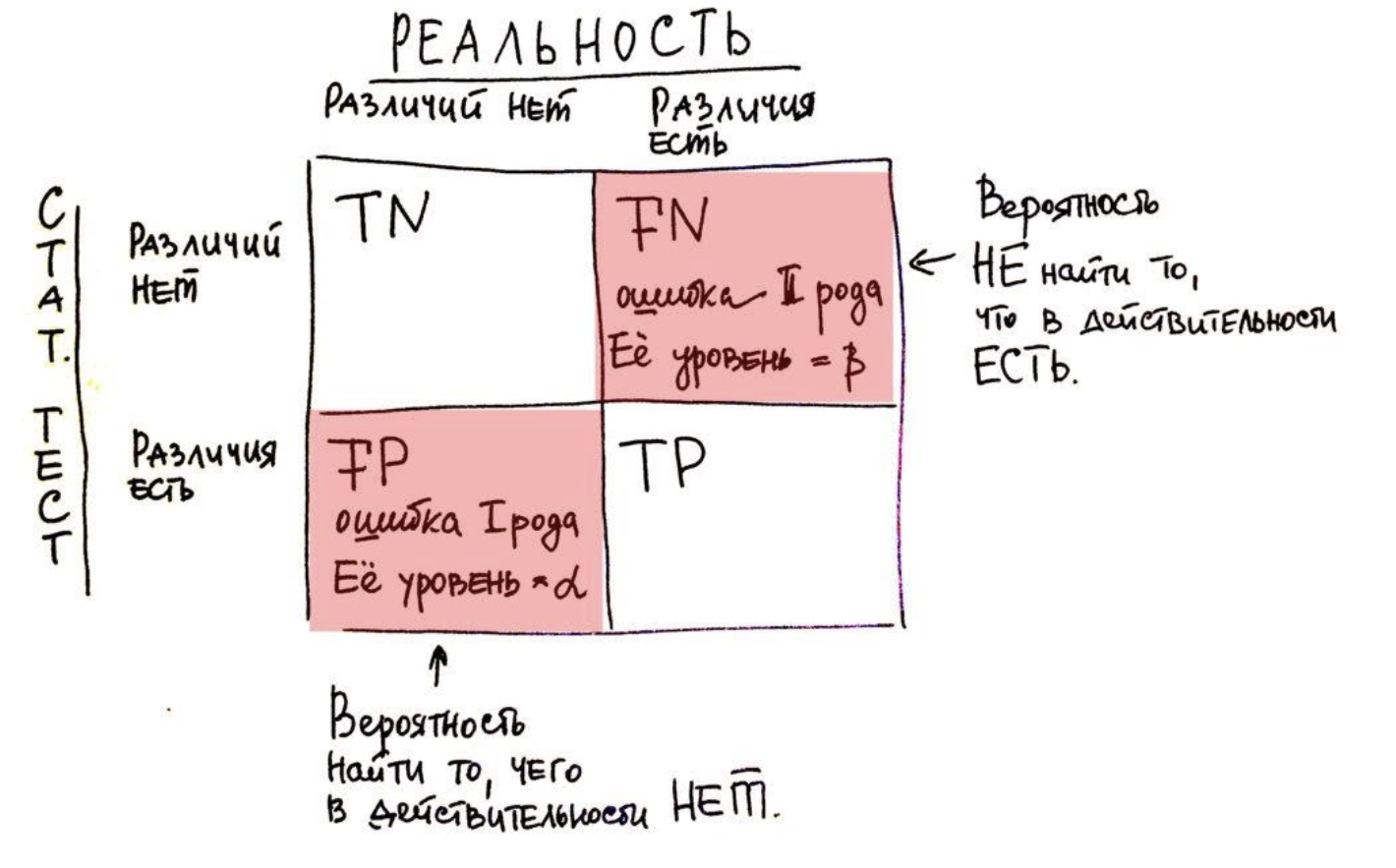
**False Positive (ошибка I рода)** = говорим истина, когда по факту не истина. Отклонение верной нулевой гипотезы. Риск совершить такую ошибку равен выбранному уровню статистической значимости (например, α=0.05) (ошиблись)

**False Negative (ошибка II рода)** = говорим не истина, когда по факту истина (ошиблись). Принятие неверной нулевой гипотезы. Вероятность ошибочно сохранить неверную нулевую гипотезу обозначают буквой β (например, 0.2).

**Статистический критерий(тест)** — это математическое правило, которое помогает нам понять, отвергнута или нет наша гипотеза. Мы поговорим о них подробно на следующем занятии.

**Мощность критерия/теста** - это вероятность, что тест правильно засечёт эффект там, где он и правда есть. (т.е. 1-β)

Допустим мы сформулировали нулевую гипотезу для эксперимента(поговорим о них подробнее на следующих занятиях). У нас могут быть следующие исходы:



Рассмотрим пример :



Однажды воскресным вечером через пару дней после веселых посиделок в баре вы замечаете что у вас пропало обоняние и появилась легкая усталость . Вы пошли и сдали тест на COVID и получили положительный результат. Но не спешите грустить и лучше по возможности сдайте его повторно:)

В реальности могут быть следующие исходы :

Вы реально больны и тест показал что вы больны **True positive**

Вы реально не больны и тест показал что вы больны **False positive**

Вы реально не больны и тест показал что вы не больны **True negative**

Вы реально больны и тест показал что вы не больны **False negative**

У всех тестов разное качество что может влиять на False Positive и False negative. Поэтому лучше перепроверяться

еще один наглядный пример:



Мы в продуктах не хотим принимать ложноотрицательные(отклонять крутые гипотезы) и ложноположительные(раскатывать плохие гипотезы) решения и хотим минимизировать ошибку первого и второго рода .Выбираем alpha и ищем самый мощный критерий среди корректных.

Мощность и точность зависят от размера выборки - чем большую мощность и точность мы хотим - тем больше выборка нам нужна - поговорим об этом в следующей главе. Вам сейчас важно понять суть различных ошибок - поскольку в дальнейшем они будут использоваться регулярно.

# Концепция проверки статистических гипотез

В этой главе поговорим о том, как устроен процесс проверки статистических гипотез и какие нюансы существуют .

Начнем с определения статистической гипотезы.

**Статистическая гипотеза** - это выдвигаемое предположение о свойствах и характеристиках генеральной совокупности(ей).

После выдвижения гипотезы собираются реальные данные выборок(выборки) и к ним применяются статистические методы(критерии) например t test или z test, о которых детальнее мы поговорим позже.

После применения критериев мы либо отвергаем, либо не отвергаем гипотезу. Эта гипотеза называется нулевой и мы дадим ей подробное определение ниже.

# 

**Примеры гипотез:**

Среднее время звонка для всех наших клиентов равно 3 минуты

Уровень брака в текущей партии продукта равен 1%

Retention 5 дня в контрольной группе не отличается от Retention 5 дня в тестовой группе.

**Важно!** : Корректно говорить не отвергаем нулевую гипотезу, а не принимаем .

Проверка гипотез в статистике строится по методу доказательства от противного. Сначала выдвигают нулевую гипотезу и предполагают что она верна. Ее придерживаются пока наблюдаемые данные на докажут обратное.

**Нулевая гипотеза** – принимаемое предположение о том, что не существует связи между наблюдениями в двух (или более) событиях (выборках, феноменах, совокупностях). Гипотезу отвергают, если данные показывают разницу между выборками.

Нулевой гипотезе сопутствует альтернативная гипотеза - принимаемая в случае отвержения нулевой гипотезы.

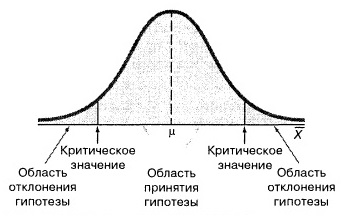
Если предположение противоречит наблюдаемым данным, то гипотезу отклоняют, как ложную в пользу альтернативной, если не противоречит, то говорят что оснований отвергнуть нулевую гипотезу нет.

Сила противоречия зависит от того, того насколько далеко выборочная статистика(например для выб.среднего) отклоняется от гипотезы.

Благодаря [ЦПТ,](https://www.marketing.spb.ru/lib-around/stat/Naked_Statistics.htm) которая рассматривалась на предыдущем занятии знаем что выборочные средние при соблюдении определенных предпосылок имеют нормальное распределение.

Чтобы погрузиться тему чуть глубже /освежить знания можно почитать:

[статья](https://towardsdatascience.com/statistics-is-the-grammar-of-data-science-part-2-8be5685065b5) 1 + [статья 2 +](https://www.marketing.spb.ru/lib-around/stat/Naked_Statistics.htm) [видео 1](https://www.youtube.com/watch?v=2tuBREK_mgE) или [видео 2](https://www.youtube.com/watch?v=9jL5JifQ1WI)



Можно рассматривать площадь под кривой в определенном диапазоне как вероятность получить тот или иной диапазон значений.

**Поизучать по теме статья про** [**z score**](https://medium.com/@ashishtuteja87/what-is-z-score-and-p-value-b7daf7db008f) № 1 и [статья № 2](https://ru.wikihow.com/%D0%B2%D1%8B%D1%87%D0%B8%D1%81%D0%BB%D0%B8%D1%82%D1%8C-Z-%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D1%83)

На графике выше можно заметить что существуют область принятия и область отклонения гипотезы.

**Областью принятия гипотезы** называют совокупность значений статистики критерия, при которых гипотезу H0(нулевую) принимают.

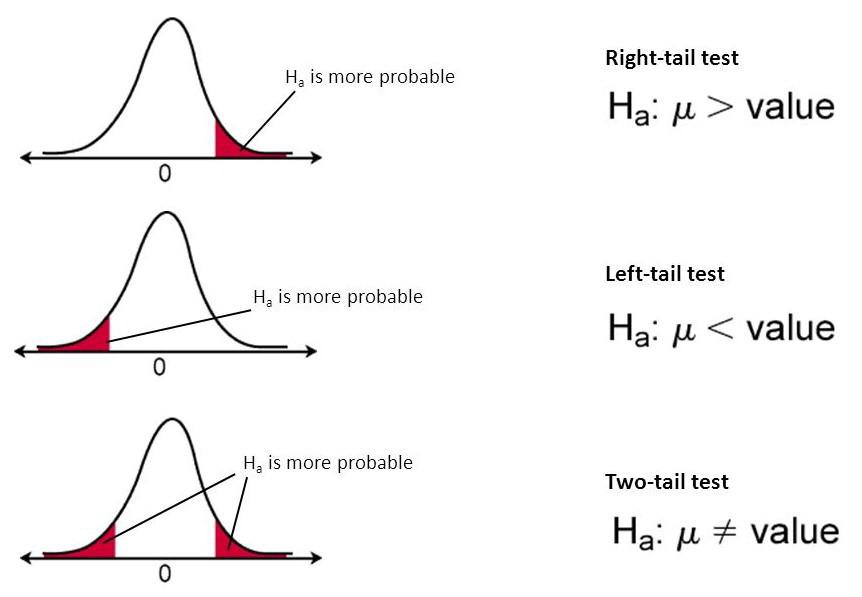
**Областью отклонения гипотезы(критической областью)** - называется область значений статистики критерия, при которых отвергается H0(нулевую).

**Критические значения -** это граница критической области, они зависят от выбранного alpha. Вид критической области определяется видом альтернативной.

Мы смотрим куда у нас попадает выборочное значение. Если рассчитанное значение окажется в зоне близкой к центру(в области принятия гипотезы), то оно не противоречит гипотезе(могли быть просто небольшие колебания в силу случайности).

Но если оно окажется далеко от центра за пределами критических значений(слева или справа) - это значит что вероятность получить такое или еще более высокое отклонение крайне мало и гипотеза скорее всего ложная.

Гипотезы могут быть **односторонними и двухсторонними** (см.график ниже). На практике почти всегда используют двухсторонние тк мы почти никогда не знаем заранее перед запуском теста в какую сторону отклонится метрика в тестовой группе. А после получения данных менять двухсторонний на односторонний критерий некорректно - таким образом мы подгоняем теорию под наши данные.



Для углубленного изучения при желании можно почитать [лекции ВШЭ](http://math-info.hse.ru/f/2017-18/ps-ms/hypo-test.pdf) и [статью.](https://people.richland.edu/james/lecture/m170/ch09-typ.html)

В современном мире используется понятие:

[p-value](https://medium.com/@ashishtuteja87/what-is-z-score-and-p-value-b7daf7db008f) – вероятность при справедливости нулевой гипотезы получить наблюдаемое или еще большее отклонение оценки статистики(например среднего) по сравнению с предполагаемым в гипотезе.

# 

# 

Многие некорректно интерпретируют p-value. Основная проблема - для многих людей это не интуитивная сущность и возникают проблемы с интерпретацией. Я предлагаю вам по желанию ознакомиться с парой интересных статей о нем, чтобы не совершать ошибок при интерпретации данного понятия .

[Статья на русском на хабре](https://habr.com/ru/company/stepic/blog/250527/) и на [английском](http://www.perfendo.org/docs/BayesProbability/twelvePvaluemisconceptions.pdf).

Если p-value < alpha (false poistive), то нулевая гипотеза отвергается в пользу альтернативной!.

Пример: проверяем гипотезу о равенстве конверсии в покупку магазина e-commerce В одной группе CR1 =1 % в другой СR2 = 1,01%.

Выдвигаем нулевую гипотезу что между CR1 и CR2 нет разницы, это двухсторонняя гипотеза, alpha(false positive) возьмем равную 5%.

Допустим , применив критерий мы вычислили p-value =0,7% .

Сравнив p-value и alpha : 0,7 > 0,05 - значит нулевая гипотеза не отвергается.(исторически сложился бенчмарк 5%, но также допустимы и другие отсечки например 1%,10% и другие в зависимости от объема выборки в продукте и готовности получать false positive ошибки.)

К p-value мы еще вернемся в следующих главах. В соседних главах рассмотрим статистические критерии , алгоритм принятия решений и идею статистического вывода по данным.

# Статистические критерии

Чаще всего в A/B тестах проверяют гипотезу о равенстве двух показателей(средних) - используя для этого статистические критерии.

**Статистический критерий** — математическое правило, в соответствии с которым отвергается либо не отвергается та или иная статистическая гипотеза с заданным уровнем значимости.

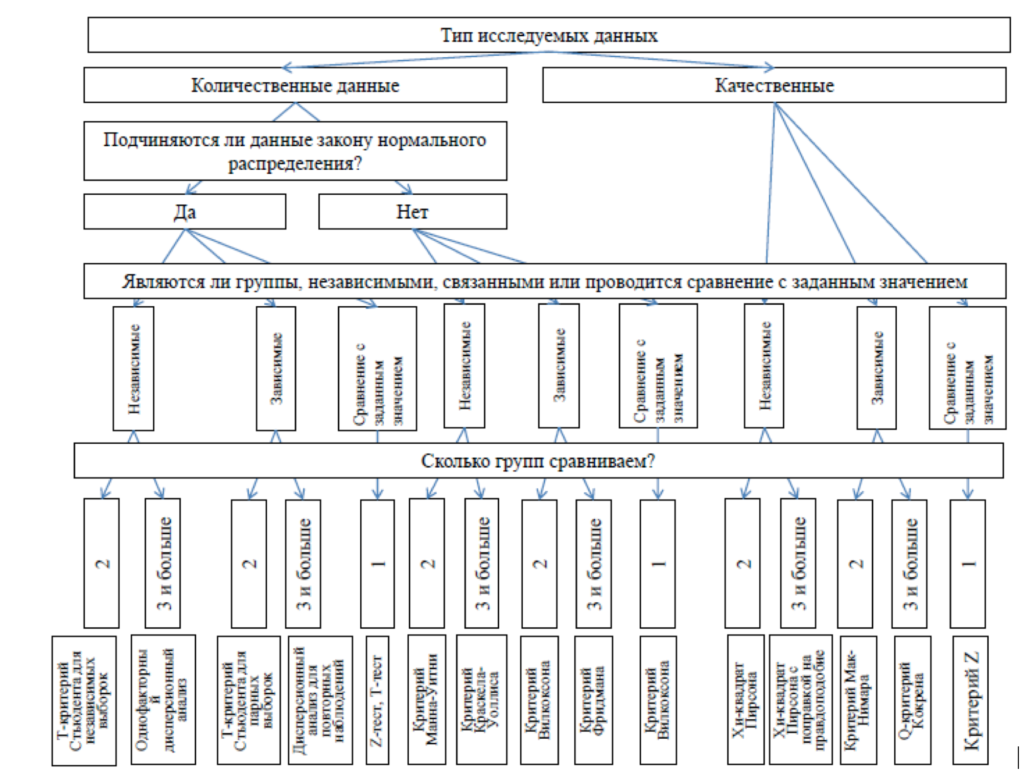
Для того, чтобы проверить гипотезу о равенстве показателей, применяется два типа критериев оценки: *параметрические* и *непараметрические*.

Если просто описывать основную суть их различия - то можно сказать что:

**Параметрическими** называются критерии, в которых мы можем сделать предположение о распределении , относящееся к какой-то выборке. В большинстве случаев в качестве распределения используется нормальное.(t тест)

**Непараметрические** не используют предположения о распределении, а оперируют *рангами* и *частотами.(*[U-критерий Манна — Уитни](https://ru.wikipedia.org/wiki/U-%D0%BA%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D0%B9_%D0%9C%D0%B0%D0%BD%D0%BD%D0%B0_%E2%80%94_%D0%A3%D0%B8%D1%82%D0%BD%D0%B8#:~:text=U%2D%D0%BA%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D0%B9%20%D0%9C%D0%B0%D0%BD%D0%BD%D0%B0%20%E2%80%94%20%D0%A3%D0%B8%D1%82%D0%BD%D0%B8%20(,%D0%B7%D0%BD%D0%B0%D1%87%D0%B5%D0%BD%D0%B8%D0%B8%20%D0%BF%D0%B0%D1%80%D0%B0%D0%BC%D0%B5%D1%82%D1%80%D0%B0%20%D0%BC%D0%B5%D0%B6%D0%B4%D1%83%20%D0%BC%D0%B0%D0%BB%D1%8B%D0%BC%D0%B8%20%D0%B2%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0%D0%BC%D0%B8.)*)*

У каждого критерия есть предпосылки - и можно использовать таблицу ниже для выбора подходящего критерия.

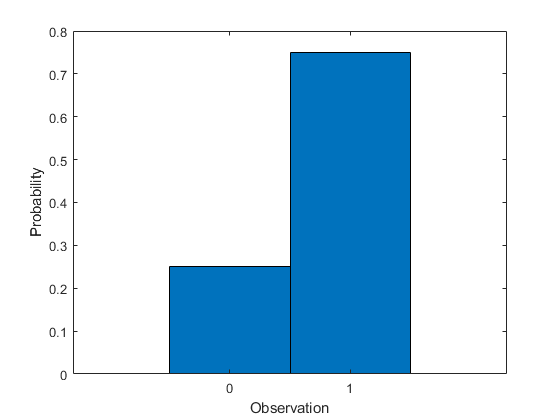


Необязательно знать все критерии - на практике чаще всего используется [лишь их часть.](https://lit-review.ru/biostatistika/vybor-statisticheskogo-kriteriya/)

Часто используются [t-test Уэлча](https://ru.wikipedia.org/wiki/T-%D0%BA%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D0%B9_%D0%A3%D1%8D%D0%BB%D1%87%D0%B0) для сравнения средних , z - критерий , для сравнения метрик долей(конверсии/ретеншн) , критерий Хи-квадрат тоже подойдет для долей + Критерий Манна Уитни для сравнения ненормально распределенных величин.

Рассмотрим z - критерий для метрик долей как с помощью него проверять гипотезы для конверсий /ретеншн итд используя p-value и доверительный интервал.

***Критерии долей*** — это критерии, которые работают с [распределениями Бернулли](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%91%D0%B5%D1%80%D0%BD%D1%83%D0%BB%D0%BB%D0%B8).

Они принимают на вход выборки из нулей и единиц и проверяют гипотезы о вероятности появления единицы в выборке

Гипотезы можно проверять с помощью доверительного интервала или p-value.

Допустим мы запустили эксперимент направленный на улучшение конверсии на каком-нибудь участке продукта , получили результаты конверсий в обеих группах и хотим понять эта разница получена не случайно, а действительно существует закономерность.

# 

У каждой из метрик в двух выборок есть сигнал(разница средних) и шум (дисперсия). Мы хотим понять несмотря на наличие шума а есть ли действительная разница между средними ? В этом нам помогают статистические критерии.

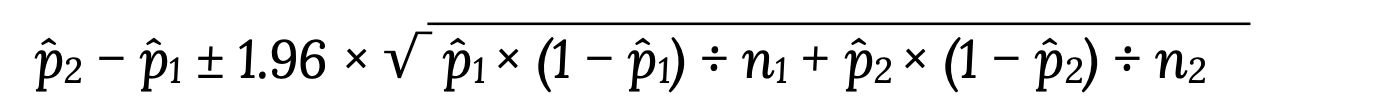
В случае с критериями долей мы можем использовать следующий принцип:

Пусть p̂ 1 и p̂ 2 посчитанные в A/Б-тесте конверсии, в контроле p̂ 1 , а в тесте p̂ 2.

*n1 и n2 —* количество элементов в каждой

1) Используя доверительный интервал: выберем уровень значимости alpha = 5%

Тогда 95% доверительный интервал для их разницы будет считаться следующим образом:



Мы получим диапазон значений [х;y]

Если 0 входит в интервал значит на основе собранных данных статистически значимой разницы между группами не обнаружено.

Если не входит, то между группами есть стат.значимая разница.

2) Альтернатива - можем также по нашим данным посчитать p-value:

Выберем уровень значимости alpha = 5%

Посчитаем p-value на полученных в эксперименте данных.

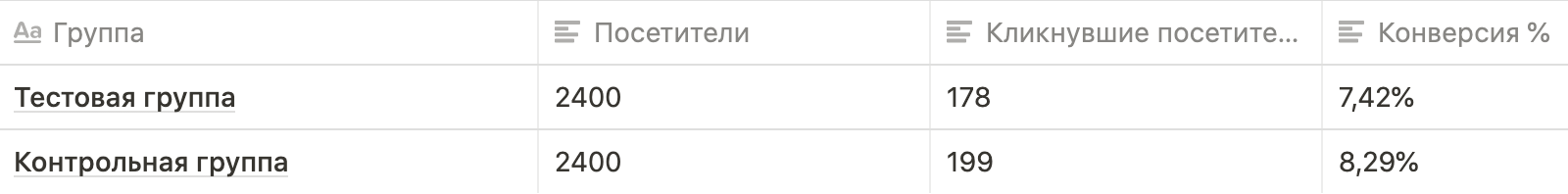
Помним про нулевую гипотезу:(Разницы между средними в двух группах нет.)

Если p-value > 5% значит на основе собранных данных статистически значимой разницы между группами не обнаружено.

Если p-value < 5% , то между группами есть стат.значимая разница

Рассмотрим пример:

Допустим мы запустили эксперимент направленный на улучшение конверсии на одном из кусков сайта:



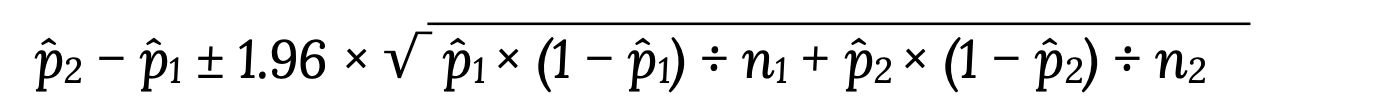
Хотим понять есть ли стат.значимая разница между группами с достоверностью 95%:

Пусть p̂ 1 и p̂ 2 посчитанные в A/Б-тесте конверсии, в контроле p̂ 1 = 8,29%

, а в тесте p̂ 2. = 7,41%.

*n1 и n2 —* количество посетителей в каждой из них.

Посчитаем доверительный интервал по формуле:



0.0742-0.0829 ± 1.96 \* ( 0.0742 × (1 − 0.0742) ÷ 2400 + 0.0829 × (1 − 0.0829) ÷ 2400 ) ^1/2 = [-2.392% ; 0.652%]

0 входит в интервал - значит на основе собранных данных статистически значимой разницы между группами не обнаружено.

Также можно было бы посчитать p-value воспользовавшись например калькуляторами:

[Калькулятор 1](https://abtestguide.com/calc/)

[Калькулятор 2](https://www.surveymonkey.ru/mp/ab-testing-significance-calculator/)

p value = 0.1299 > 0.05 - на основе собранных данных статистически значимой разницы между группами не обнаружено.

Похожий принцип принятия решений действует и для других критериев.

# Алгоритм проверки гипотез

Давайте рассмотрим алгоритм проверки гипотез, который можно использовать при проверке гипотез.

1. Выбираем метрику и формулируем нулевую и альтернативную гипотезы
2. Выбираем параметр alpha (например 5% ) равный вероятности допустить ошибку первого рода
3. Выбираем критерий, подходящий под наши условия
4. Считаем p - value и(или) доверительный интервал и делаем вывод:

Если p-value < alpha - разница между группами стат.значима . Либо если доверительный интервал для разницы не включает 0.

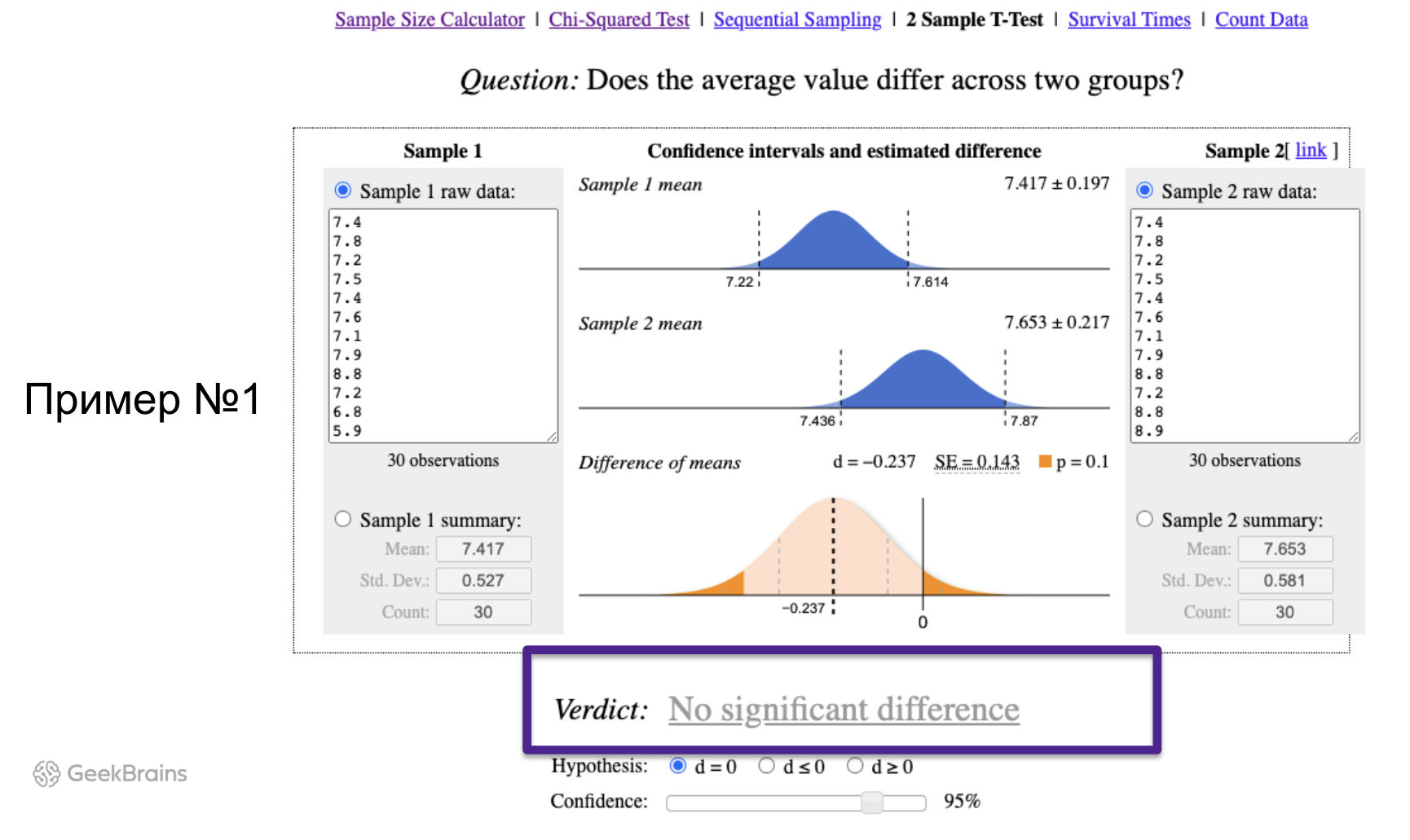
1. Даем (рекомендации лицам принимающим решения /принимаем решение ) выкатывать или не выкатывать новое изменение - если ключевая метрика в группе с новой фичей стат значимо лучше чем в контроле а дополнительные метрики не упали- можно раскатывать , помня про ошибку первого рода).

Критерии принятия решений определяются заранее на этапе планирования эксперимента.

# Калькуляторы, которые вы можете использовать:

1. <https://www.evanmiller.org/ab-testing/t-test.html>
2. <https://abtestguide.com/calc/>

3. <https://www.surveymonkey.ru/mp/ab-testing-significance-calculator/>



# 

# 

# Дополнительные материалы

1. [Видео про описательную статистику](https://www.youtube.com/watch?v=4SEX9Bozktw)
2. [Базовые концепты статистики](https://medium.com/nuances-of-programming/8-%D0%B1%D0%B0%D0%B7%D0%BE%D0%B2%D1%8B%D1%85-%D0%BF%D0%BE%D0%BD%D1%8F%D1%82%D0%B8%D0%B9-%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B8-%D0%B4%D0%BB%D1%8F-%D0%BD%D0%B0%D1%83%D0%BA%D0%B8-%D0%BE-%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85-14a1ed2e31ae)
3. [Хорошее описание ЦПТ](https://www.marketing.spb.ru/lib-around/stat/Naked_Statistics.htm)
4. [Про доверительные интервалы от Демешева](https://www.youtube.com/watch?v=GoK7sod8dhU)
5. [Про доверительные интервалы от Кирилла Мильчакова](https://www.youtube.com/watch?v=MstzroncW28)
6. [Про MDE и ошибки 1 и 2 типа - вводная статья](https://medium.com/@PAPP/significant-power-d9c3bc3dd3c)
7. [Про MDE посложнее](https://rpubs.com/hedbergec/sirc)
8. [Книга Голая статистика](https://www.livelib.ru/book/1001518877-golaya-statistika-samaya-interesnaya-kniga-o-samoj-skuchnoj-nauke-charlz-uilan)
9. [Книга статистика и котики](https://www.livelib.ru/book/1002721779-statistika-i-kotiki-vladimir-savelev)
10. [Книга Statistics in Nutshell](https://www.oreilly.com/library/view/statistics-in-a/9781449361129/)

# Используемые источники

1. [Видео про описательную статистику](https://www.youtube.com/watch?v=4SEX9Bozktw)
2. [Базовые концепты статистики](https://medium.com/nuances-of-programming/8-%D0%B1%D0%B0%D0%B7%D0%BE%D0%B2%D1%8B%D1%85-%D0%BF%D0%BE%D0%BD%D1%8F%D1%82%D0%B8%D0%B9-%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B8-%D0%B4%D0%BB%D1%8F-%D0%BD%D0%B0%D1%83%D0%BA%D0%B8-%D0%BE-%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85-14a1ed2e31ae)
3. [Хорошее описание ЦПТ](https://www.marketing.spb.ru/lib-around/stat/Naked_Statistics.htm)
4. [Про доверительные интервалы от Демешева](https://www.youtube.com/watch?v=GoK7sod8dhU)
5. [Про доверительные интервалы от Кирилла Мильчакова](https://www.youtube.com/watch?v=MstzroncW28)
6. [Про MDE и ошибки 1 и 2 типа - вводная статья](https://medium.com/@PAPP/significant-power-d9c3bc3dd3c)
7. [Про MDE посложнее](https://rpubs.com/hedbergec/sirc)
8. [Книга Голая статистика](https://www.livelib.ru/book/1001518877-golaya-statistika-samaya-interesnaya-kniga-o-samoj-skuchnoj-nauke-charlz-uilan)
9. [Книга статистика и котики](https://www.livelib.ru/book/1002721779-statistika-i-kotiki-vladimir-savelev)

10) [Книга Statistics in Nutshell](https://www.oreilly.com/library/view/statistics-in-a/9781449361129/)