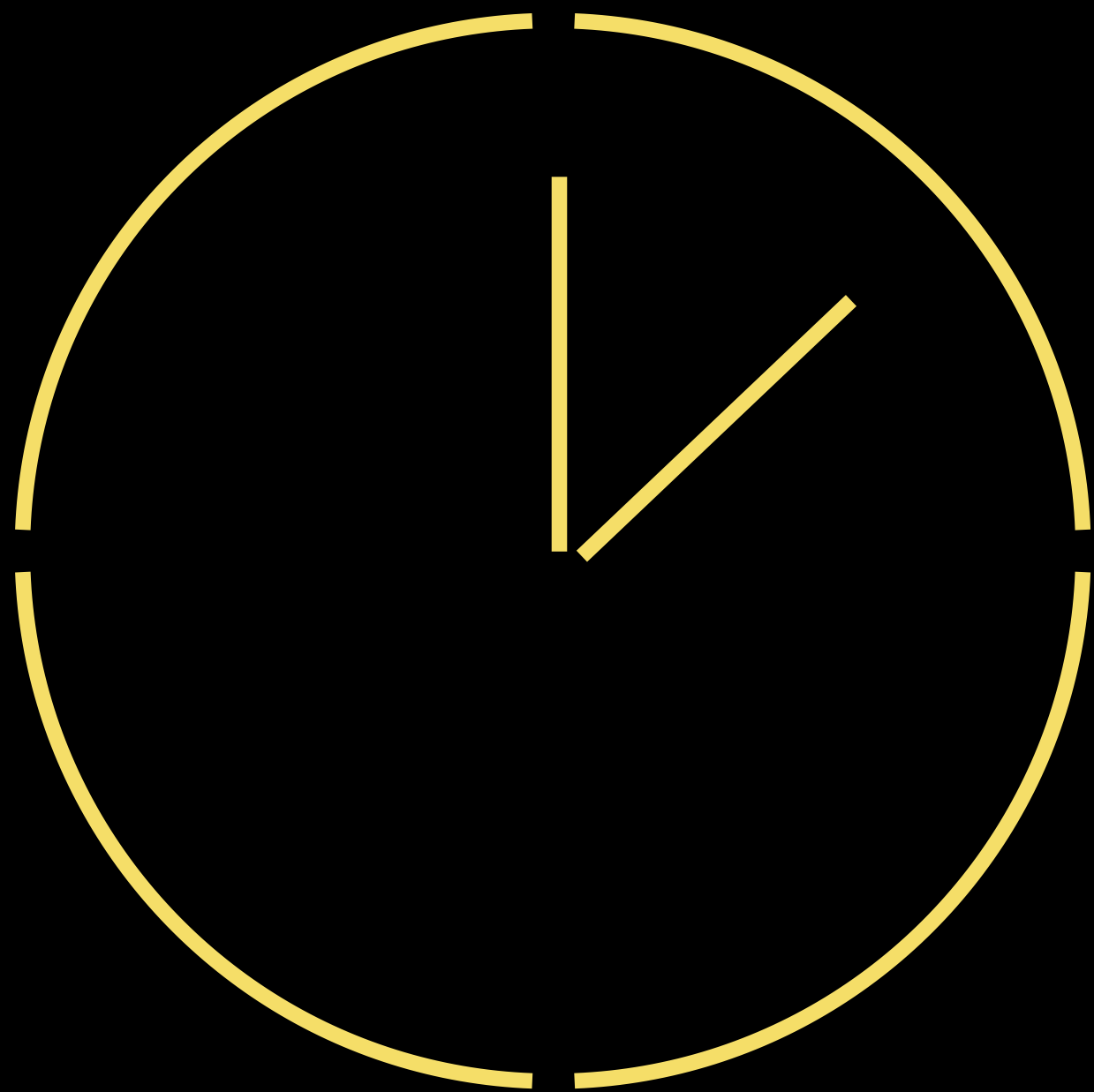


# А/Б тесты и все, что их окружает

Александр Ольферук,  
наставник

Яндекс Практикум

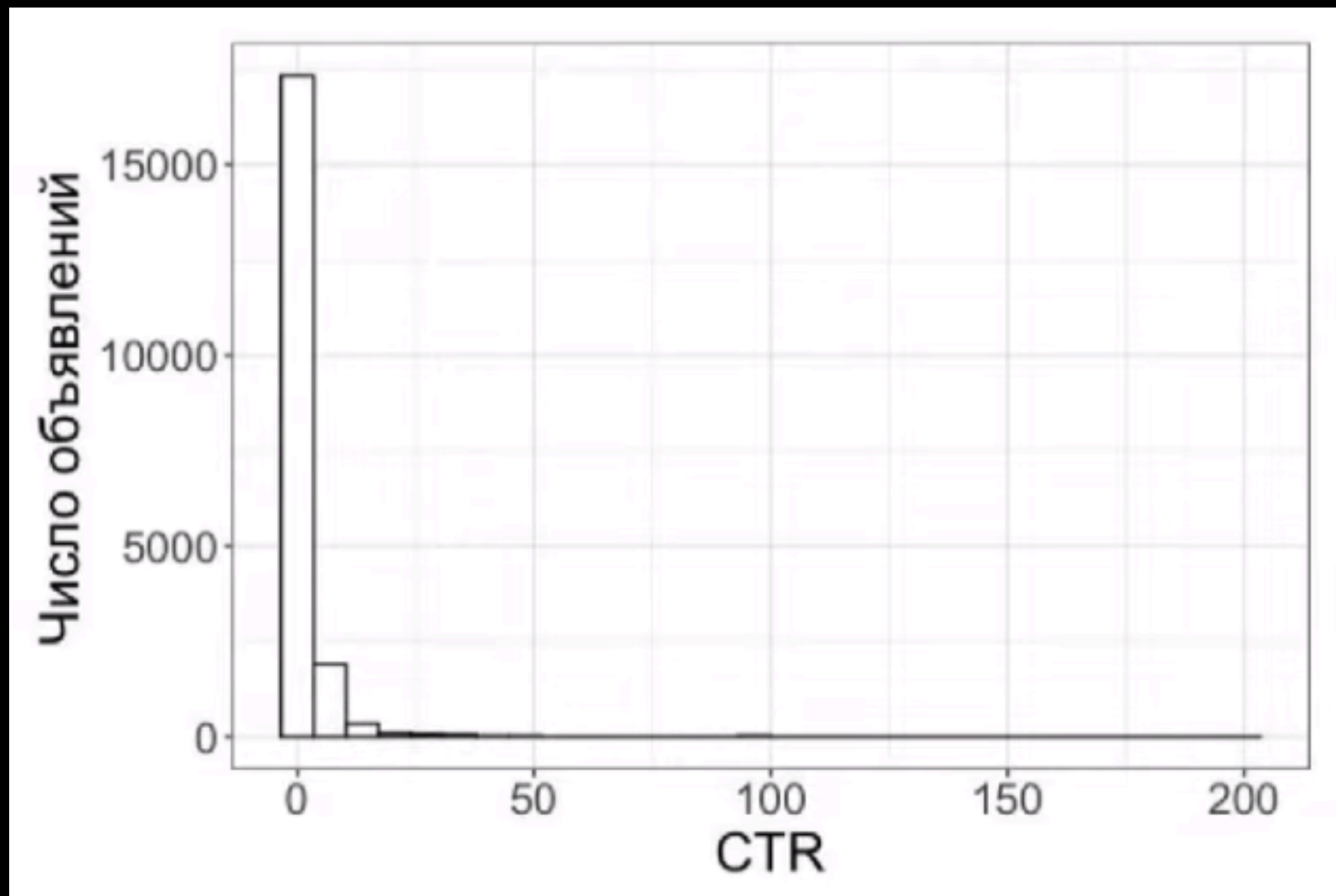
# Цели на консультацию



- Работа с ненормальными распределениями
- Ускорение A/B-тестов
- Работа с A/B тестированием, когда разделение на группы невозможно

# Работаем с ненормальными распределениями

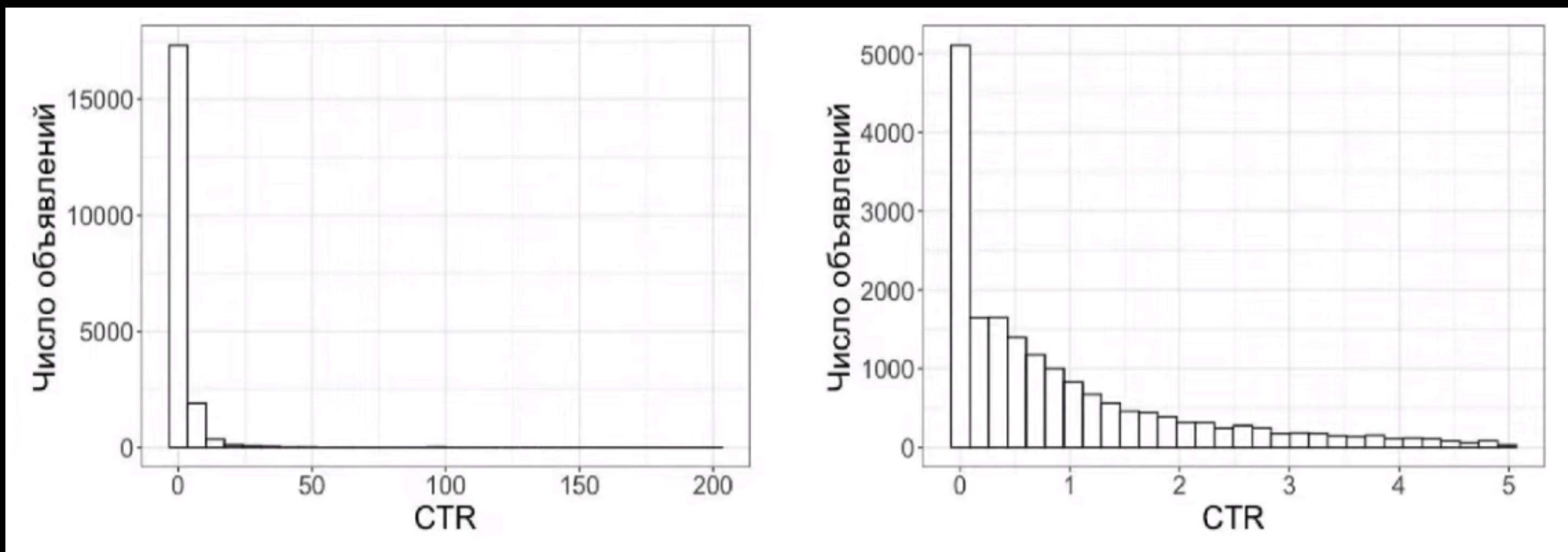
# Ненормальные распределения



# Какие варианты возможны в таком сценарии?

- Удаление выбросов
- Логарифмирование
- Непараметрические тесты
- Bootstrap

# Удаление выбросов

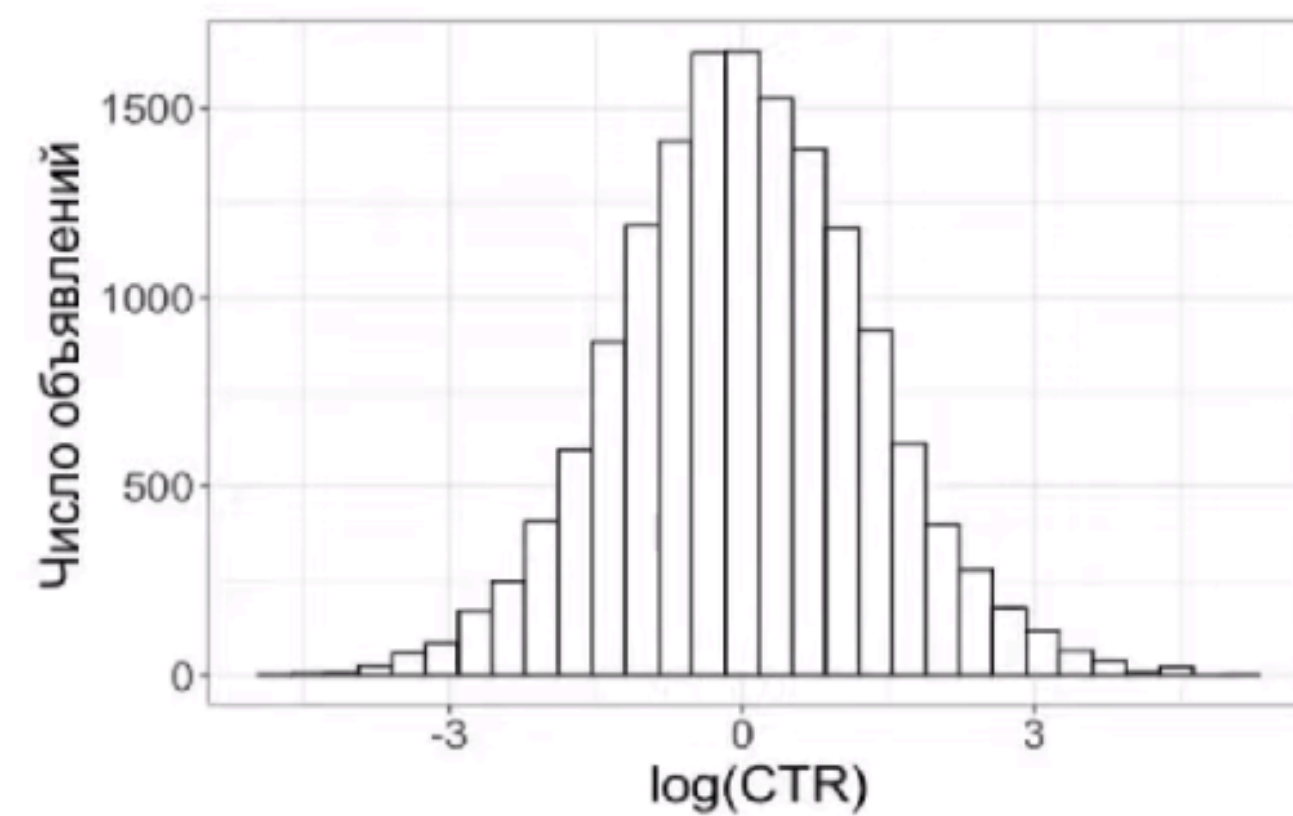
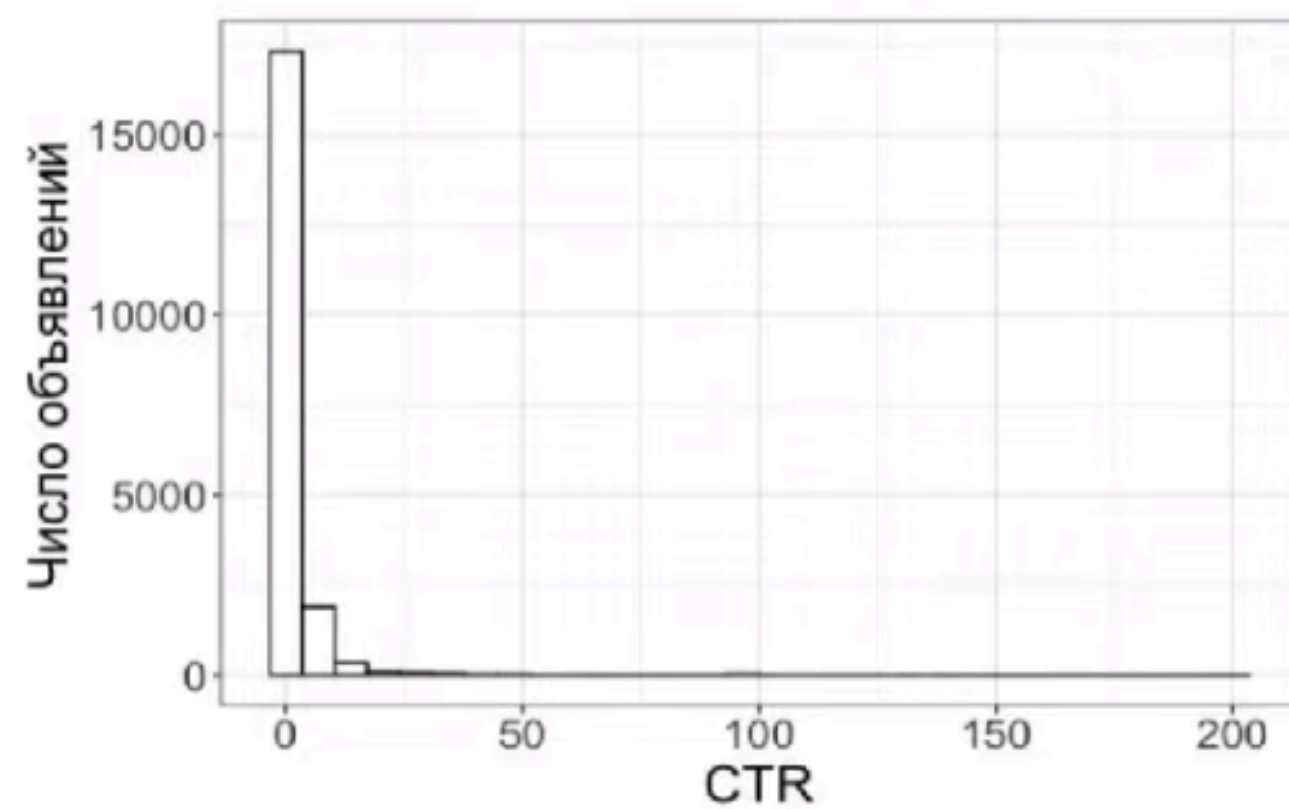


# Удаление выбросов

Однако:

- Однозначного правила, что считать выбросами, не существует
- Если мы работаем с деньгами, то было бы как минимум странно выкидывать самых богатых клиентов

# Логарифмирование переменных



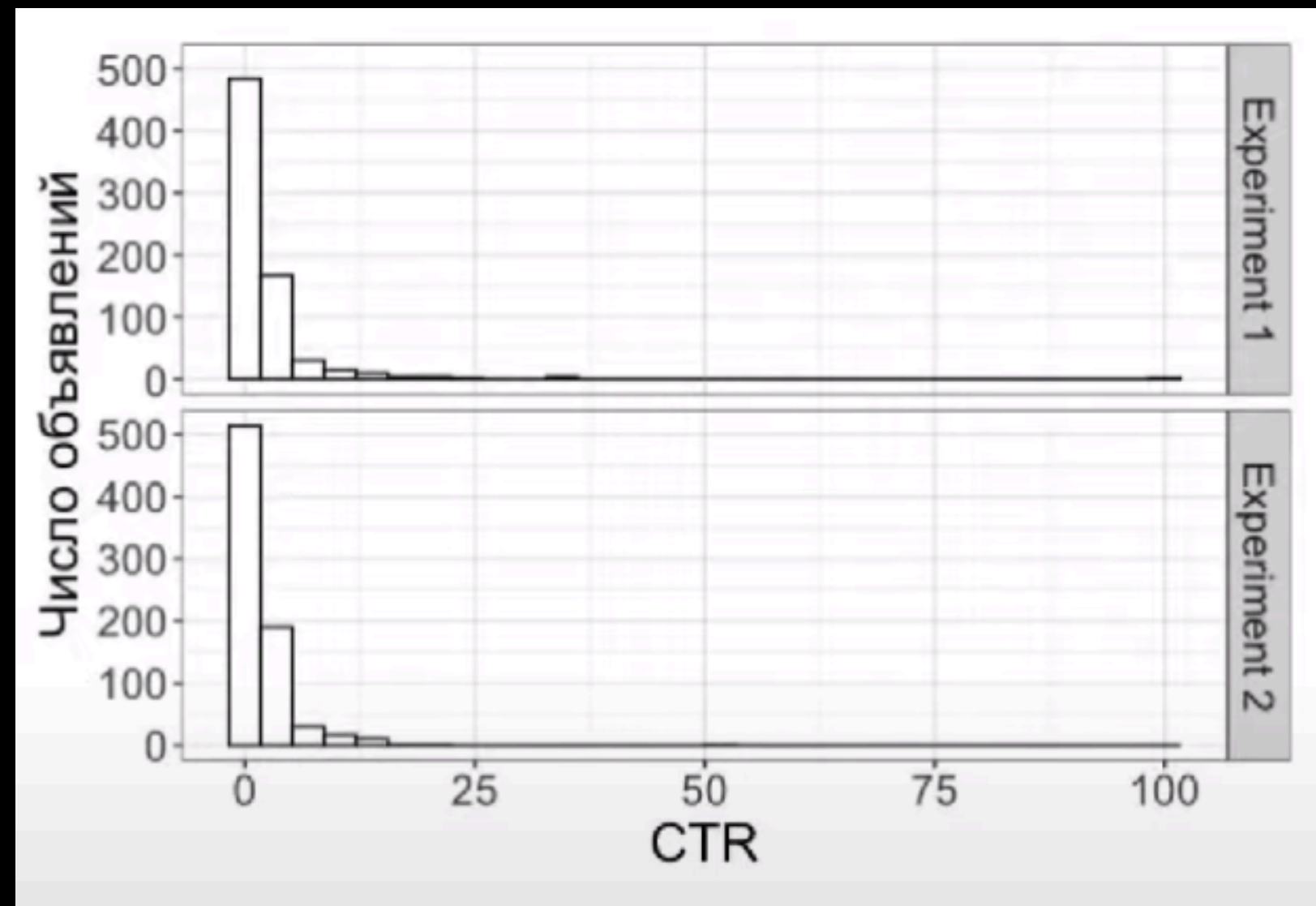


# Почему это работает?

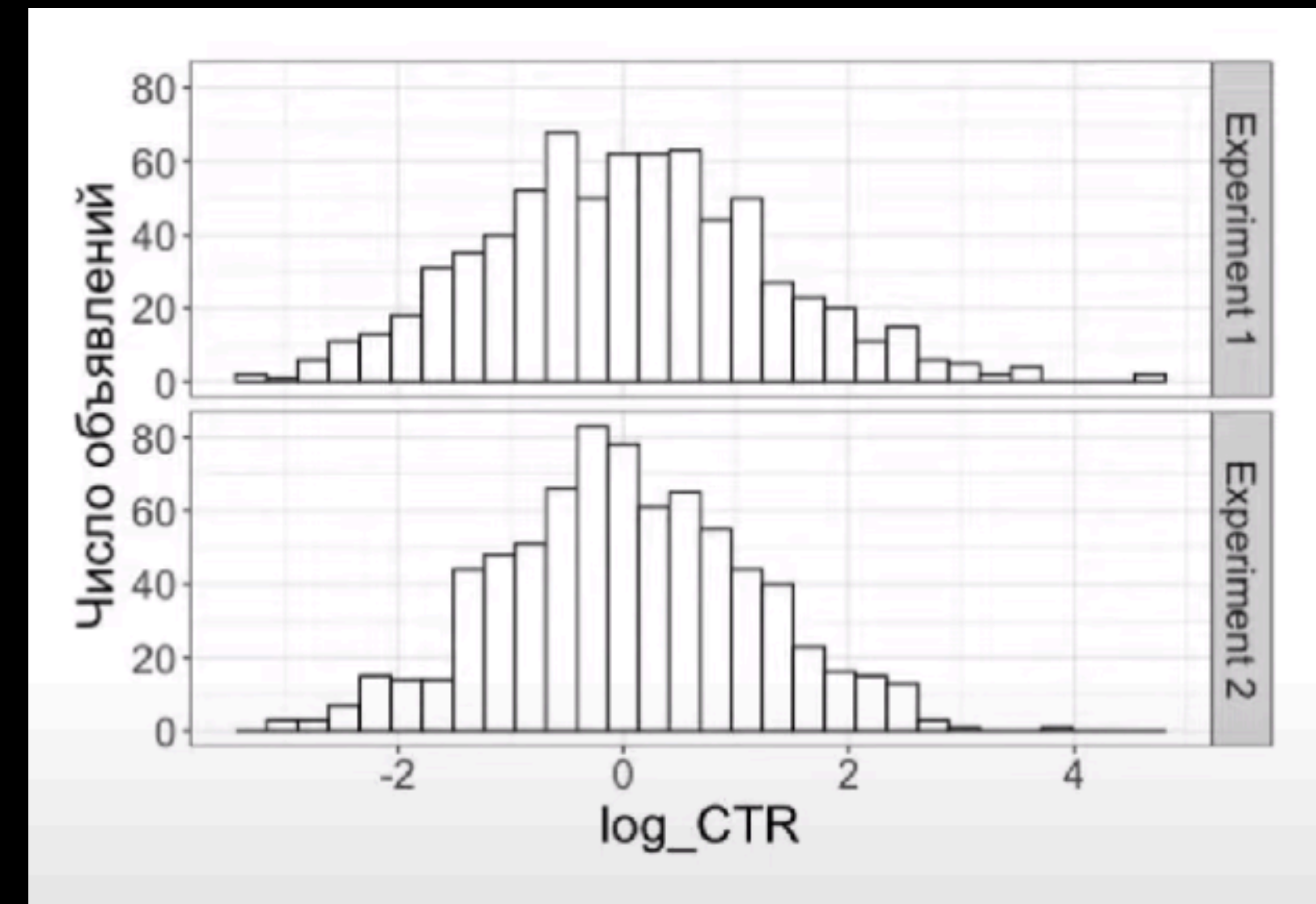
Потому что логарифм в куда меньшей степени штрафует выбросы:

X	log(x)
1	0
2	0.69
3	1.10
100	4.61
1000	6.91

# Применяем параметрические статистические тесты

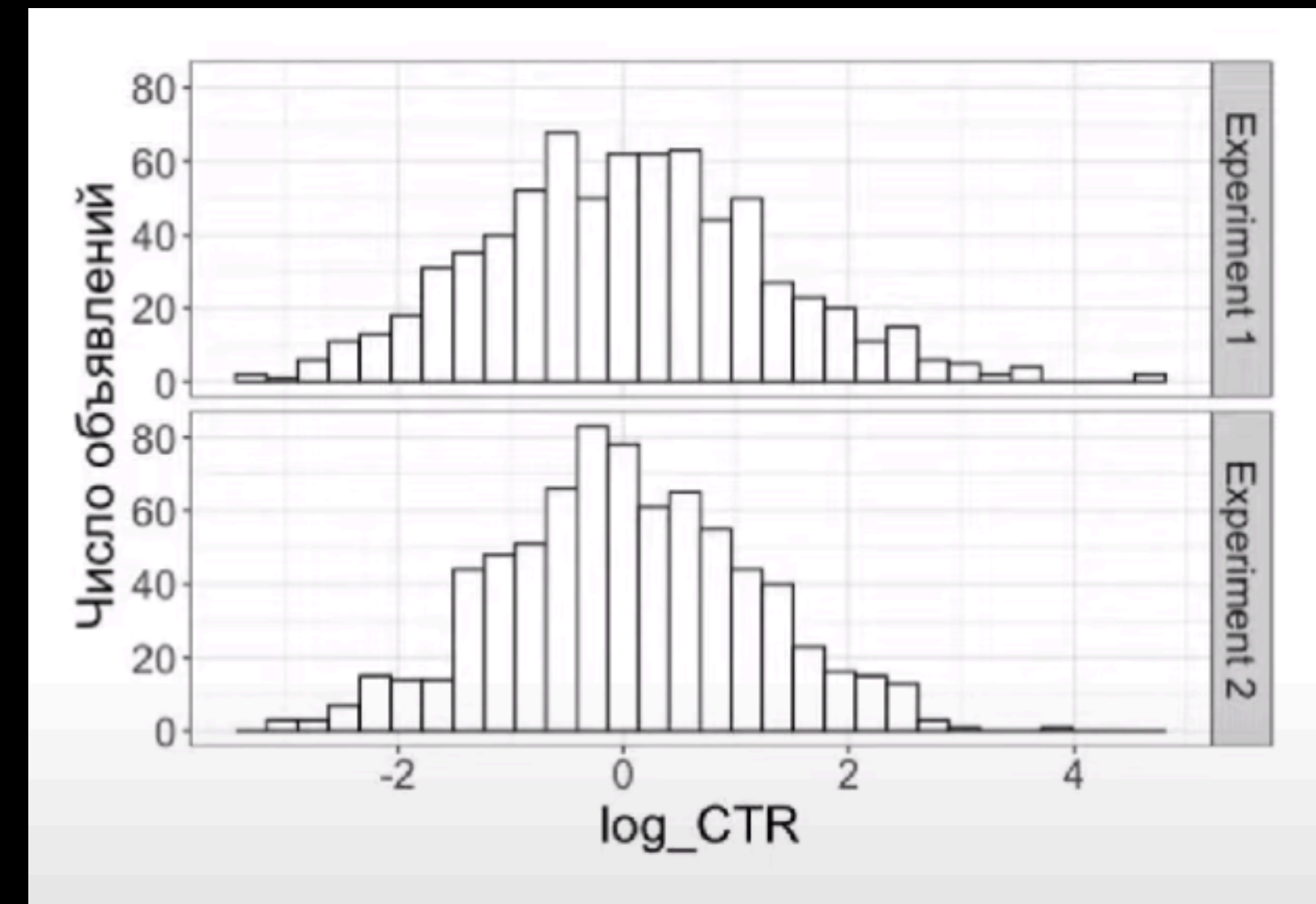
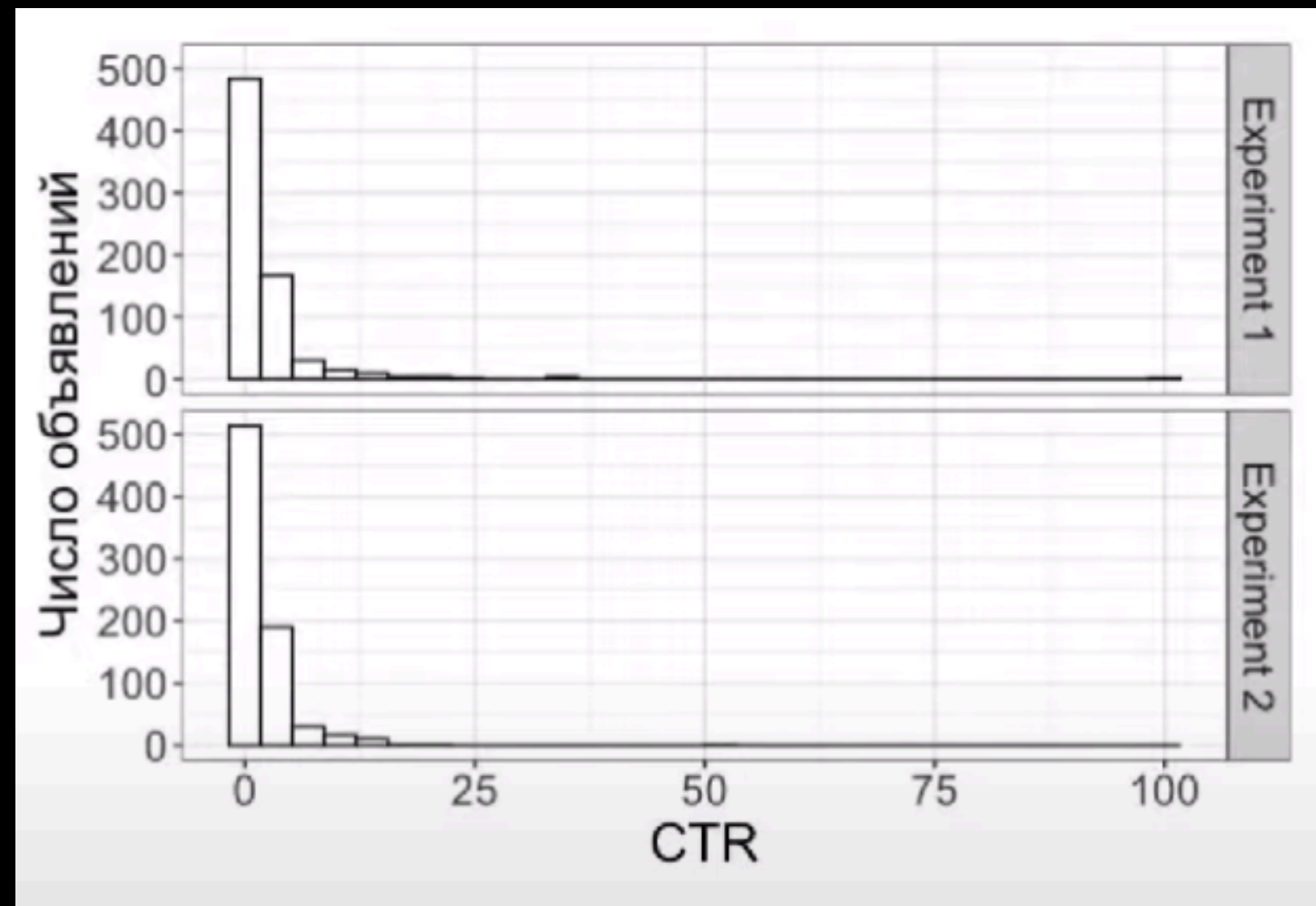


$p = 0.0217$



$p = 0.8981$

# Применяем непараметрические статистические тесты

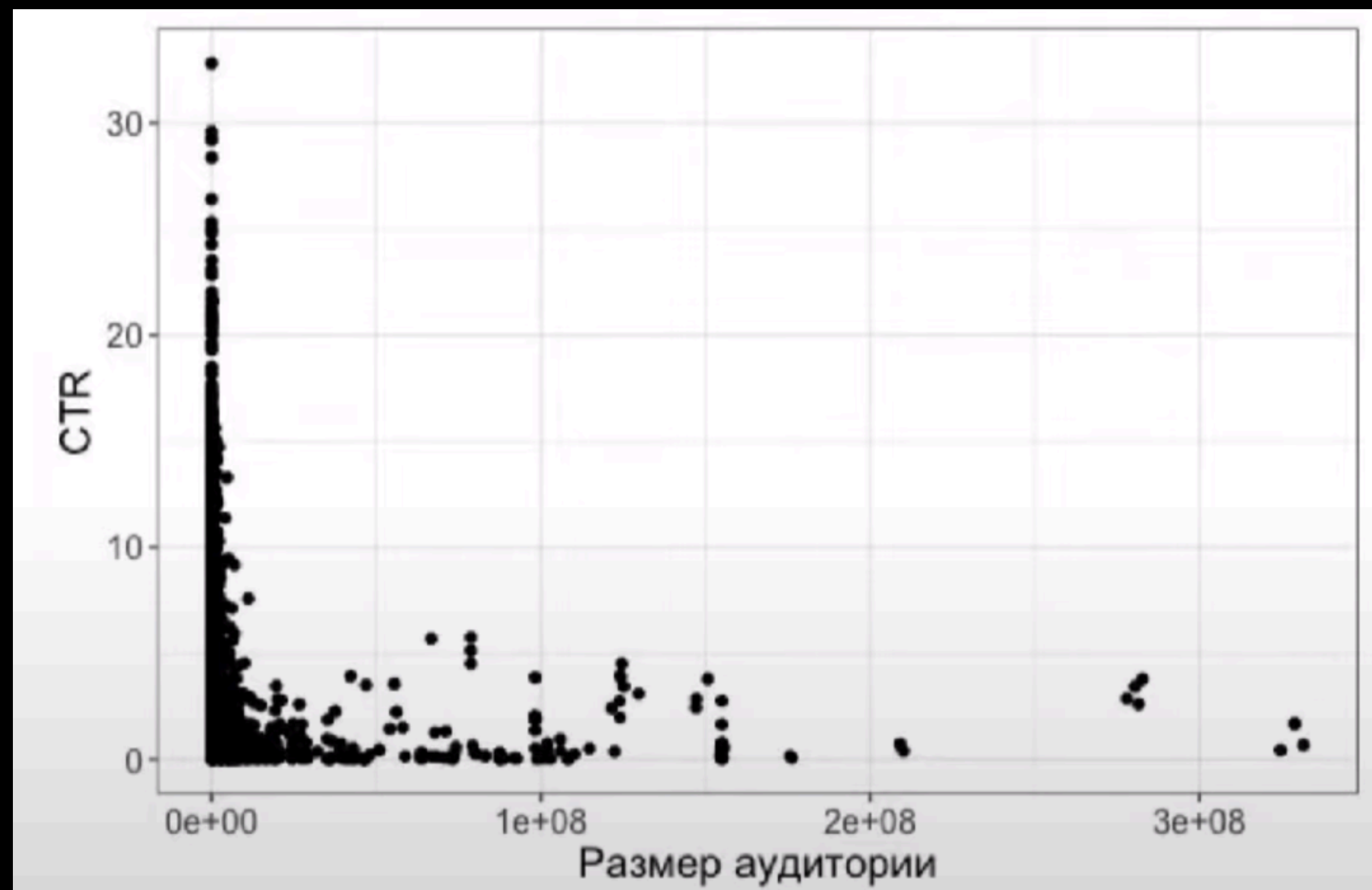


Манн-Уитни:  $p = 0.8531$

# Выводы

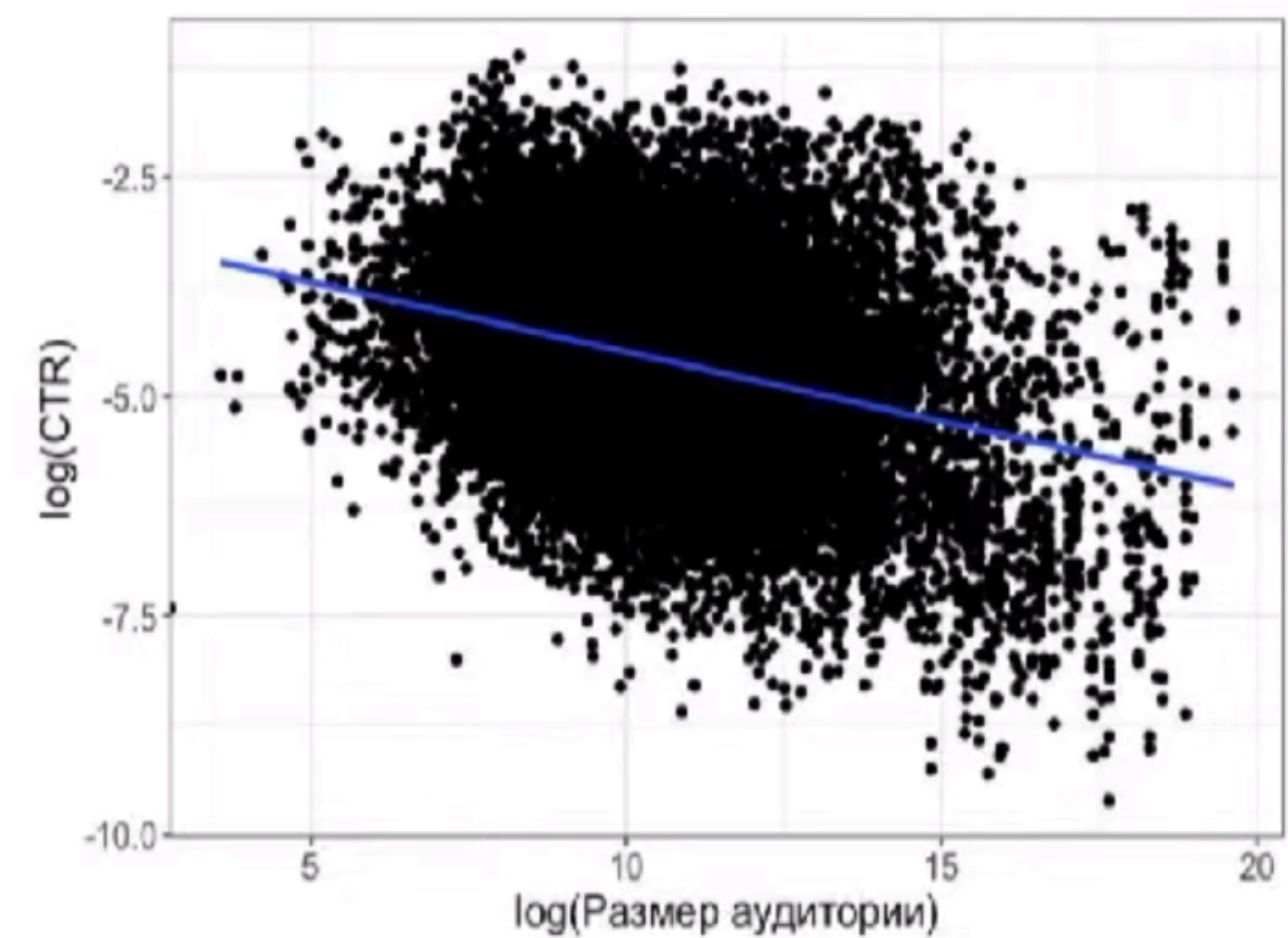
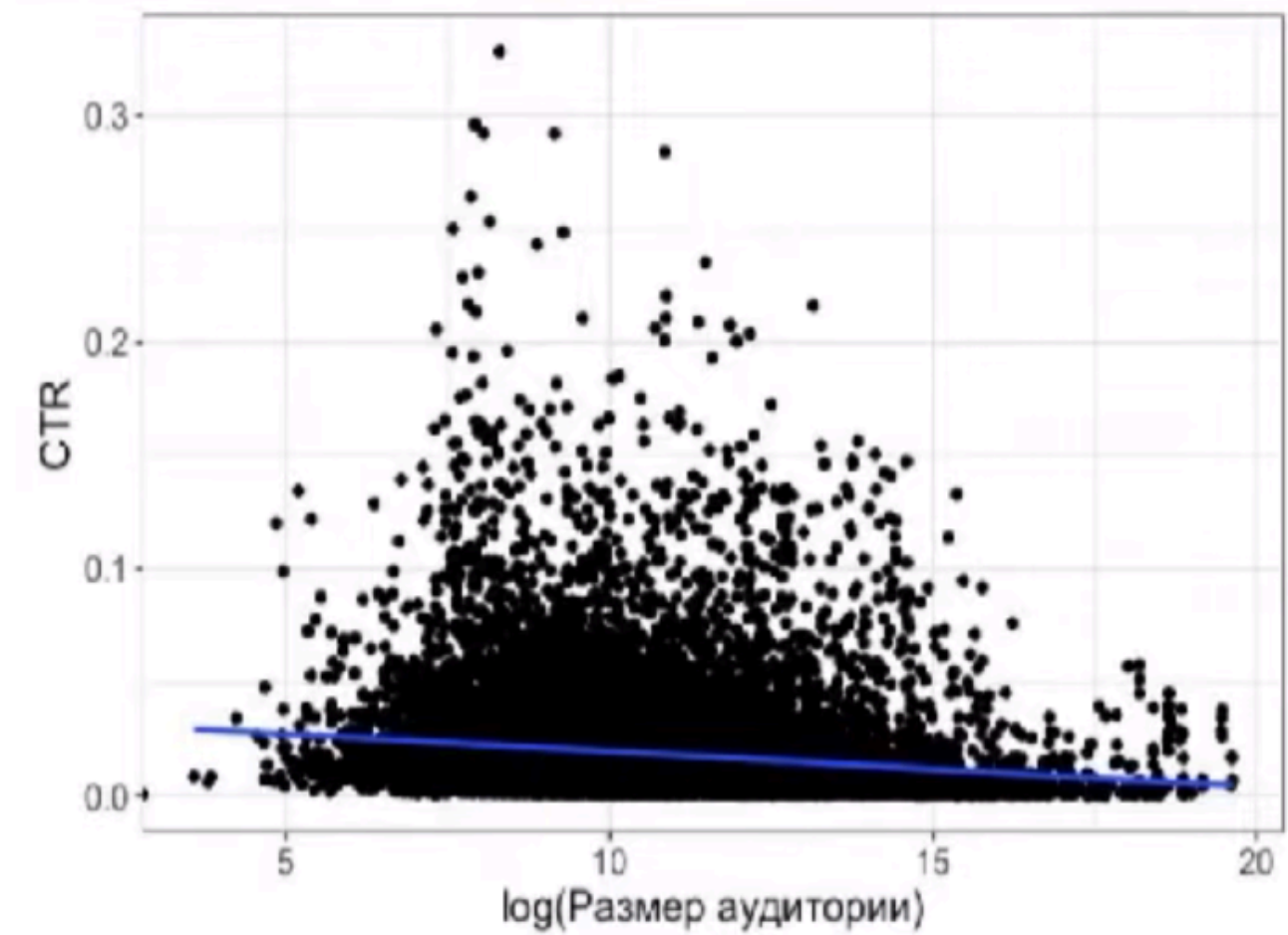
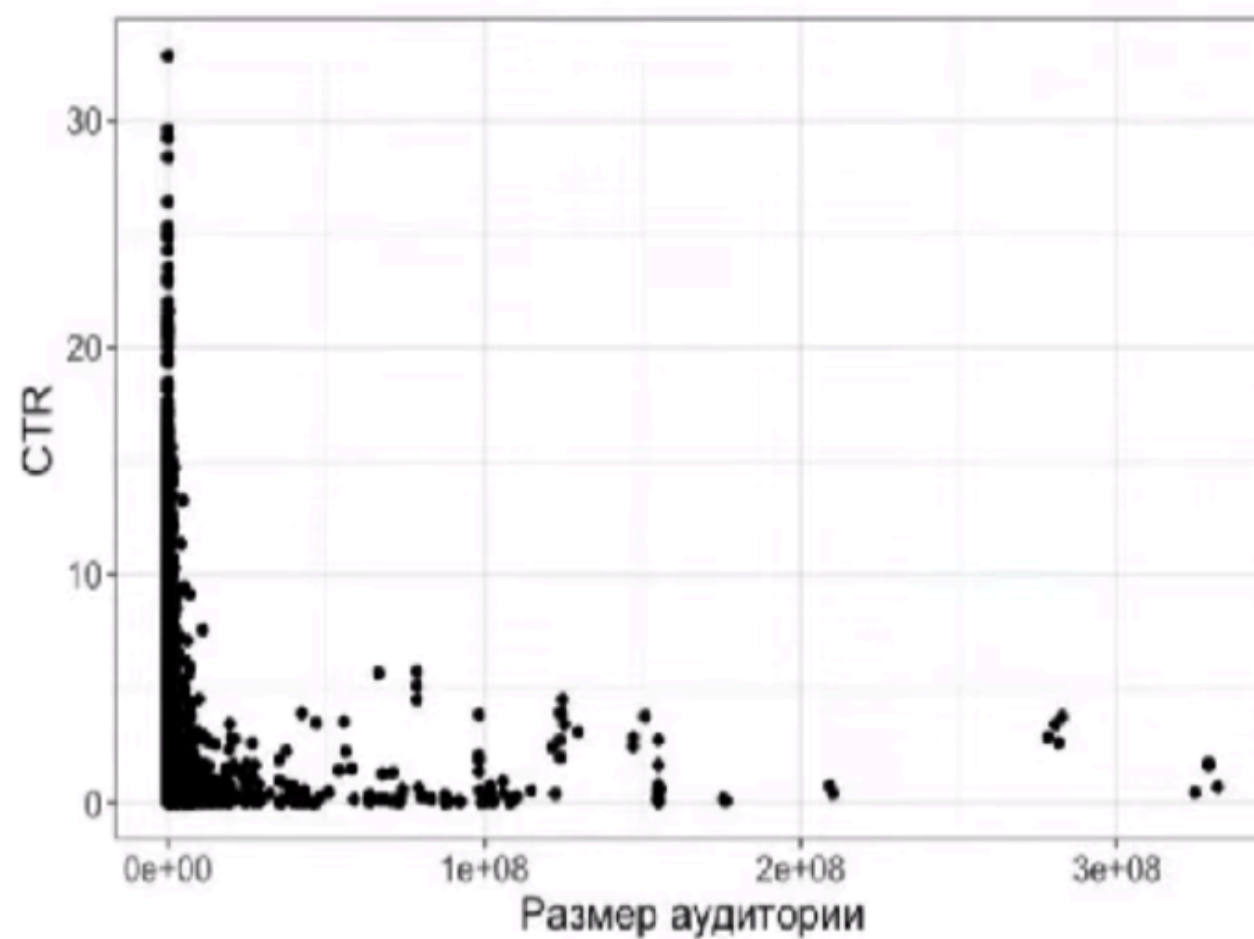
- После такого трюка можно применять даже параметрические тесты
- Хорошо работает на реальных данных
- Не забудьте сдвинуть к положительным значениям (**`np.log1p`** или **`np.log1p(x + np.min(x))`**)

# Помогает обнаружить зависимости на графике





# Помогает обнаружить зависимости на графике



# Но и это еще не все!

Обрати внимание на семейство функций Бокса-Кокса:

- Корень
- Возведение в степень (дробную)

В результате это помогает визуализировать зависимости, но усложняет интерпретацию результатов

# Интерпретация результатов

Некоторая положительная функция  $Y$  зависит от размера аудитории  $X$  и подчиняется следующему закону:

$$Y = -0.05X + 30$$

Как можно это трактовать?



# Интерпретация результатов

Некоторая положительная функция  $Y$  зависит от размера аудитории  $X$  и подчиняется следующему закону:

$$Y = -0.05X + 30$$

При изменении аудитории на одного человека функция поменяется на 0.05.

# Интерпретация результатов

Некоторая положительная функция  $Y$  зависит от размера аудитории  $X$  и подчиняется следующему закону:

$$Y = -0.05 \log(X) + 30$$

Как можно это трактовать?

# Интерпретация результатов

Некоторая положительная функция  $Y$  зависит от размера аудитории  $X$  и подчиняется следующему закону:

$$Y = -0.05 \log(X) + 30$$

При **изменении размера аудитории в 10 раз** функция изменяется на 0.05.

# Интерпретация результатов

Некоторая положительная функция  $Y$  зависит от размера аудитории  $X$  и подчиняется следующему закону:

$$Y = -0.05 \log(X) + 30$$

При **изменении размера аудитории в 10 раз** функция изменяется на 0.05.

При возрастании  $X$  с 1 до 10 – изменение будет 0.05.

А с 100 до 110 – изменение будет 0.005.

# В заключении

Если хочется сравнивать медианы, 95-ые перцентили и другие странные значения, то вам в помощь:

- Метод Монте-Карло
- Бутстрап
- Непараметрические тесты

# Ускорение А/Б-тестов

# От чего зависит время проведения теста?

- От метрики
- От размера выборки
- От дисперсии выборочных средних

# Метрики

Меняя метрику, можно подменить задачу и упростить себе жизнь.

Например, **Netflix** заметил, что с месячным ретеншном очень сильно коррелирует дневной просмотр контента (в часах).

Это и логично: те люди, кто больше смотрит ежедневно, с большой вероятностью продлят подписку еще на месяц.

Заменив метрику, они здорово сократили время теста.



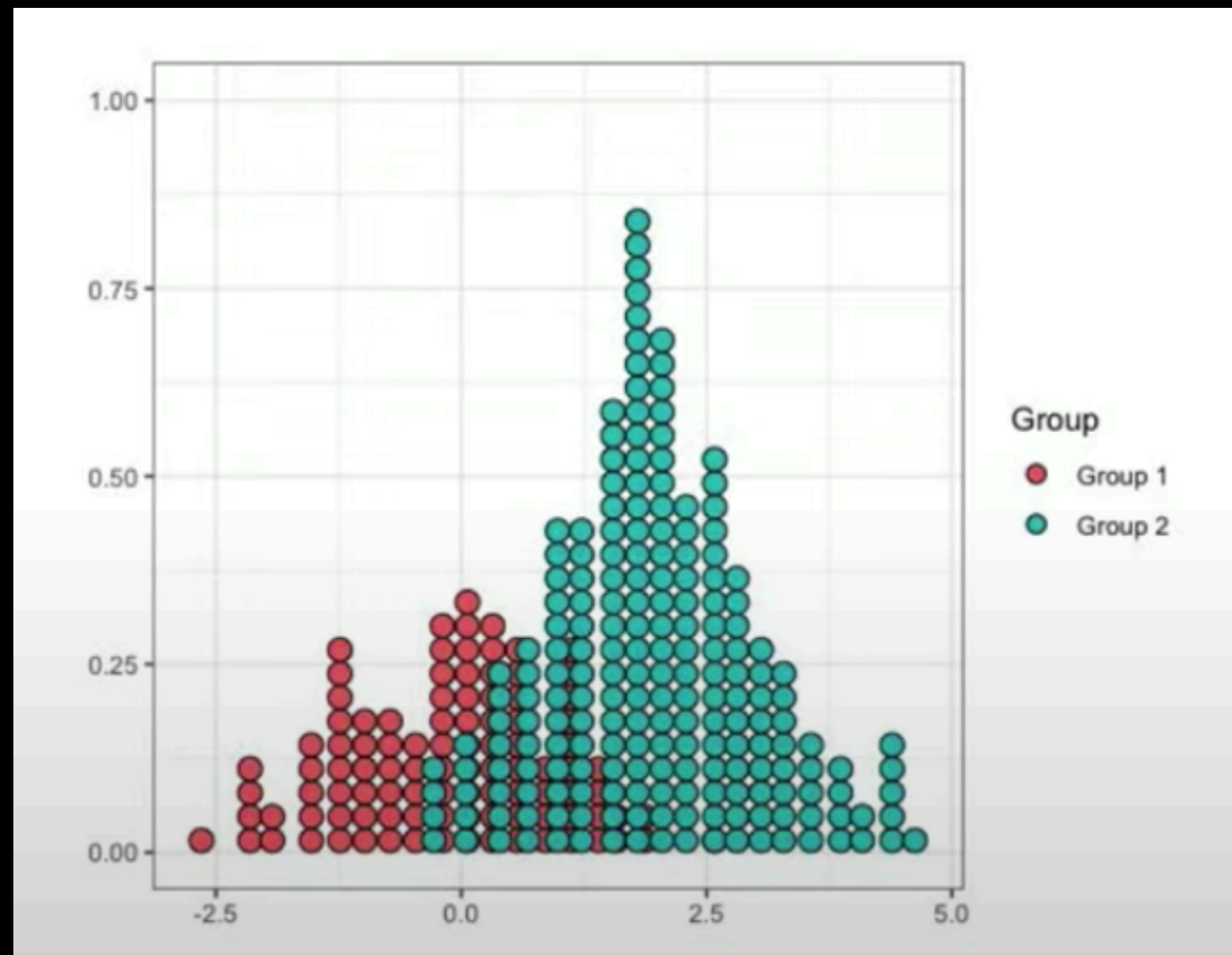
# Сокращаем дисперсию

Суть идеи **стратификации**: делаем деление на А/В не случайно, а согласно некоторой пропорции, которую знаем сами из данных.

Например, мы знаем, что пользователи, пользующиеся мобильным приложением, приносят нам больше денег, чем те, что сидят через web, но составляют 30% от общего числа аудитории.

Тогда будем делить пользователей на А и В, сохраняя пропорцию 30%/70%.

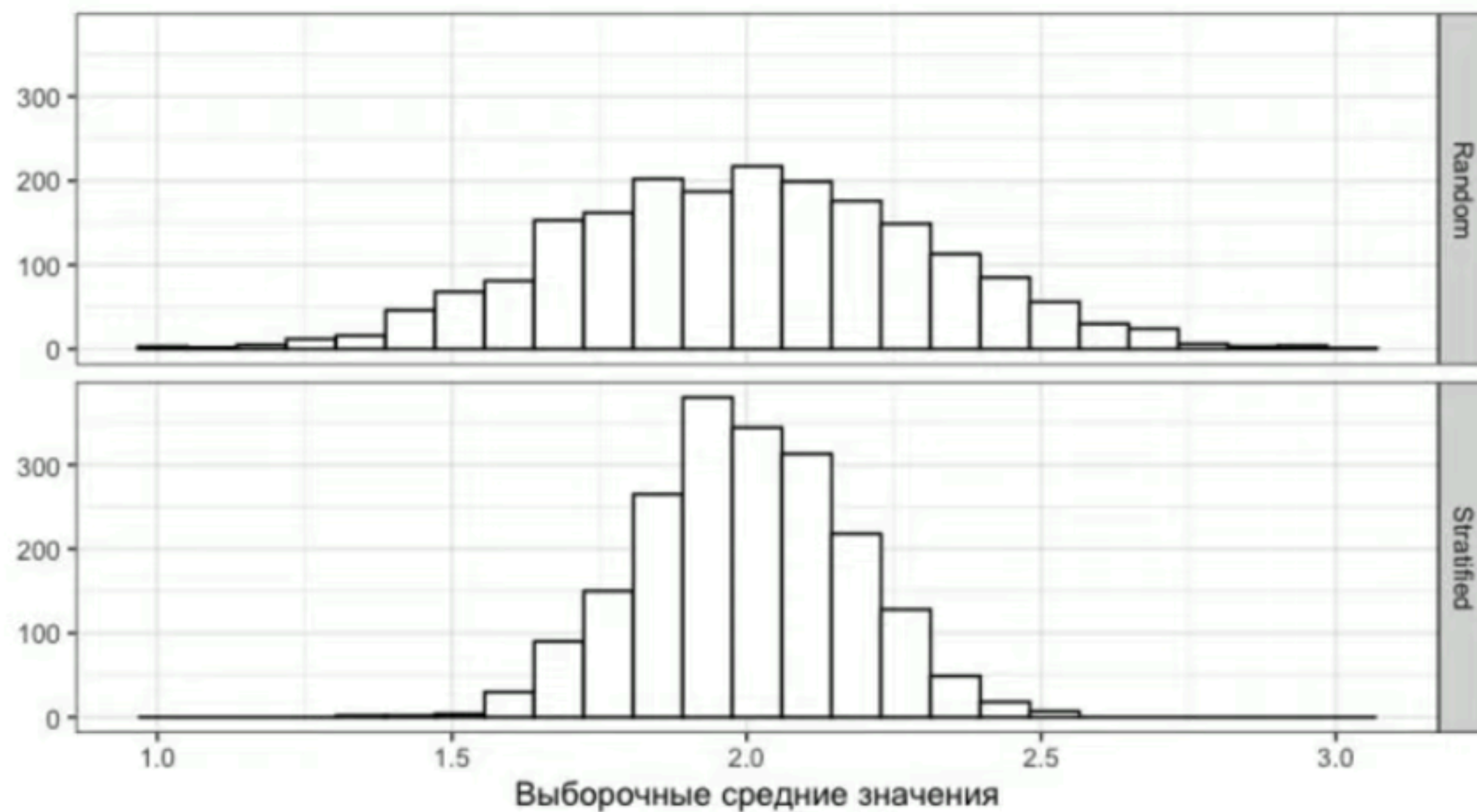
# Сокращаем дисперсию



$$\hat{Y}_{strat} = \sum_{k=1}^K p_k \bar{Y}_k,$$

, здесь **K** – количество страт.

# Результаты иногда хорошие



# Особенности

Разделение на страты можно делать и после проведения A/B-теста, т.н. пост-стратификация.

Идея та же: ищем фичу, аудитория по ней делится в некоторой пропорции, считаем по формуле с предыдущего слайда, учитывая эту пропорцию  $p_k$ , и *ВОЗМОЖНО*, это сократит нам дисперсию.

# **А/В тестирование при невозможности разделения на группы А и В**

# Когда такое возможно?

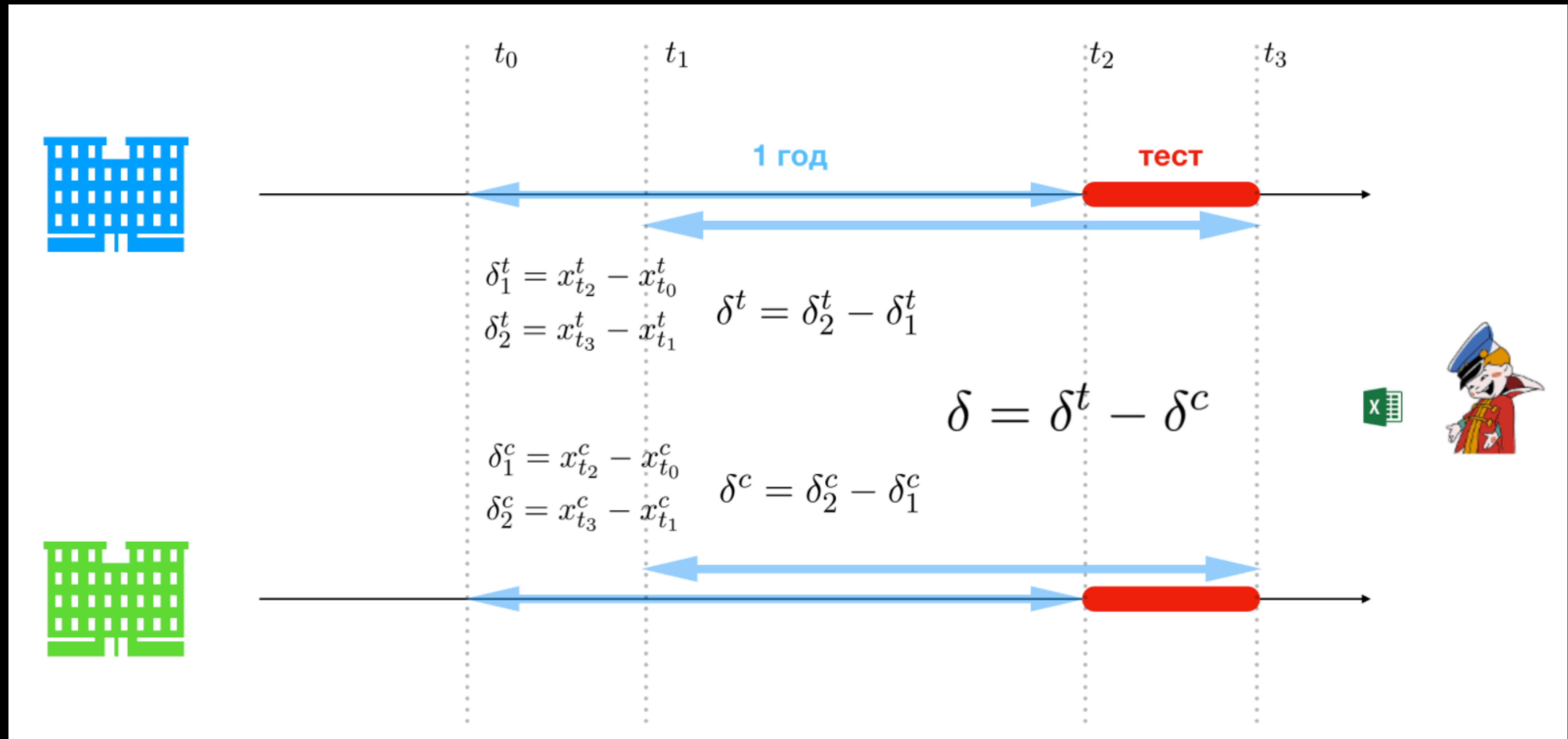
При проведении некоторого глобального события, например, при проведении чемпионата мира по футболу. Мы ведь не можем посмотреть на то, каким бы был мир, в котором одновременно провели и не провели чемпионат.

При проведении оффлайн-теста, когда вы не можете в один магазин расставить товары для одной аудитории так, а для другой – по-другому.

# Что делать?

- Методология тройной разности
- Causal impact
- ML

# Методология тройной разности





# Когда такое возможно?

При проведении некоторого глобального события, например, при проведении чемпионата мира по футболу. Мы ведь не можем посмотреть на то, каким бы был мир, в котором одновременно провели и не провели чемпионат.

На этом примере разберем Causal Impact

При проведении оффлайн-теста, когда вы не можете в один магазин расставить товары для одной аудитории так, а для другой – по-другому.

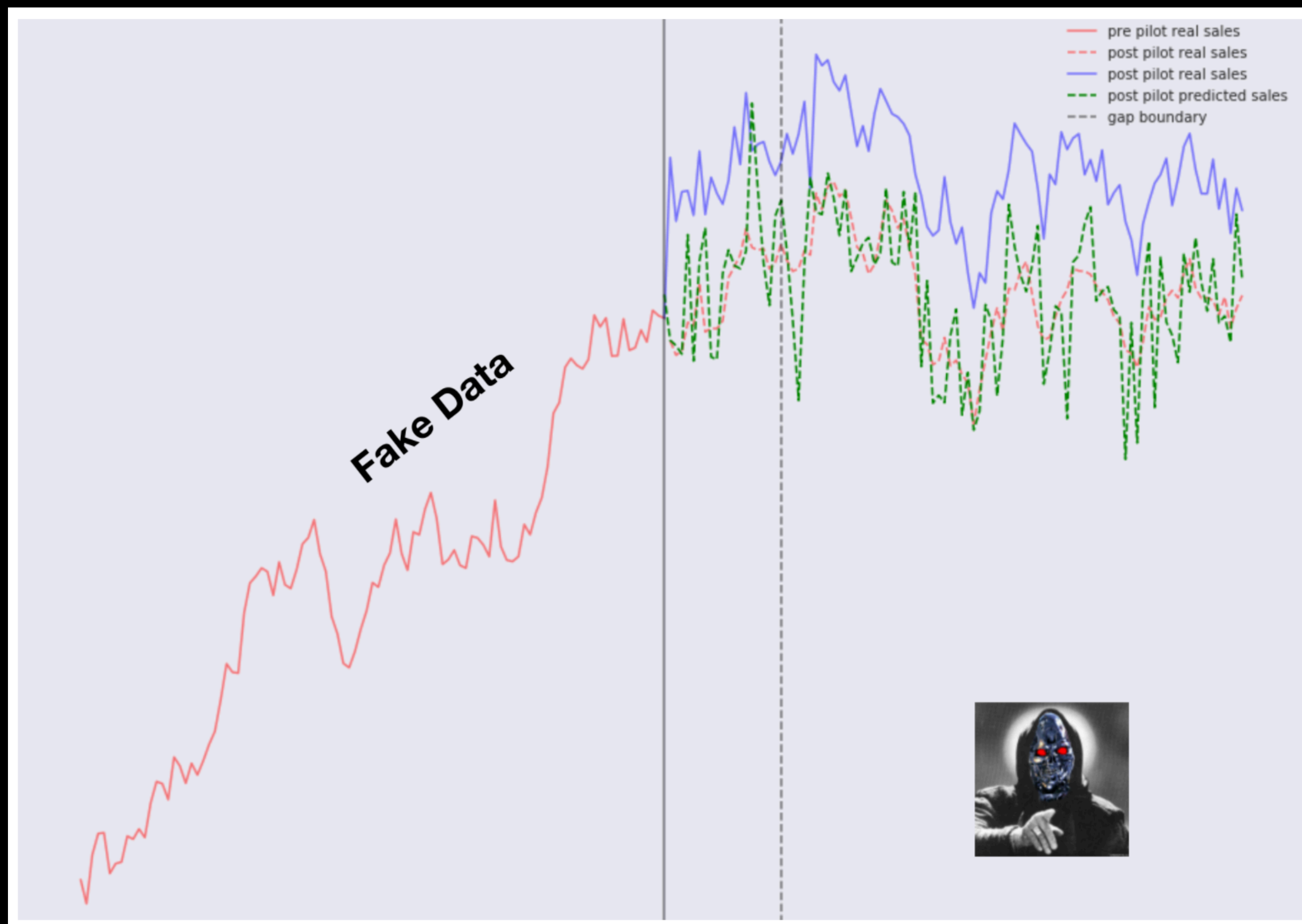
На этом примере разберем применение ML-моделей

# Байесовские модели пространства-состояния

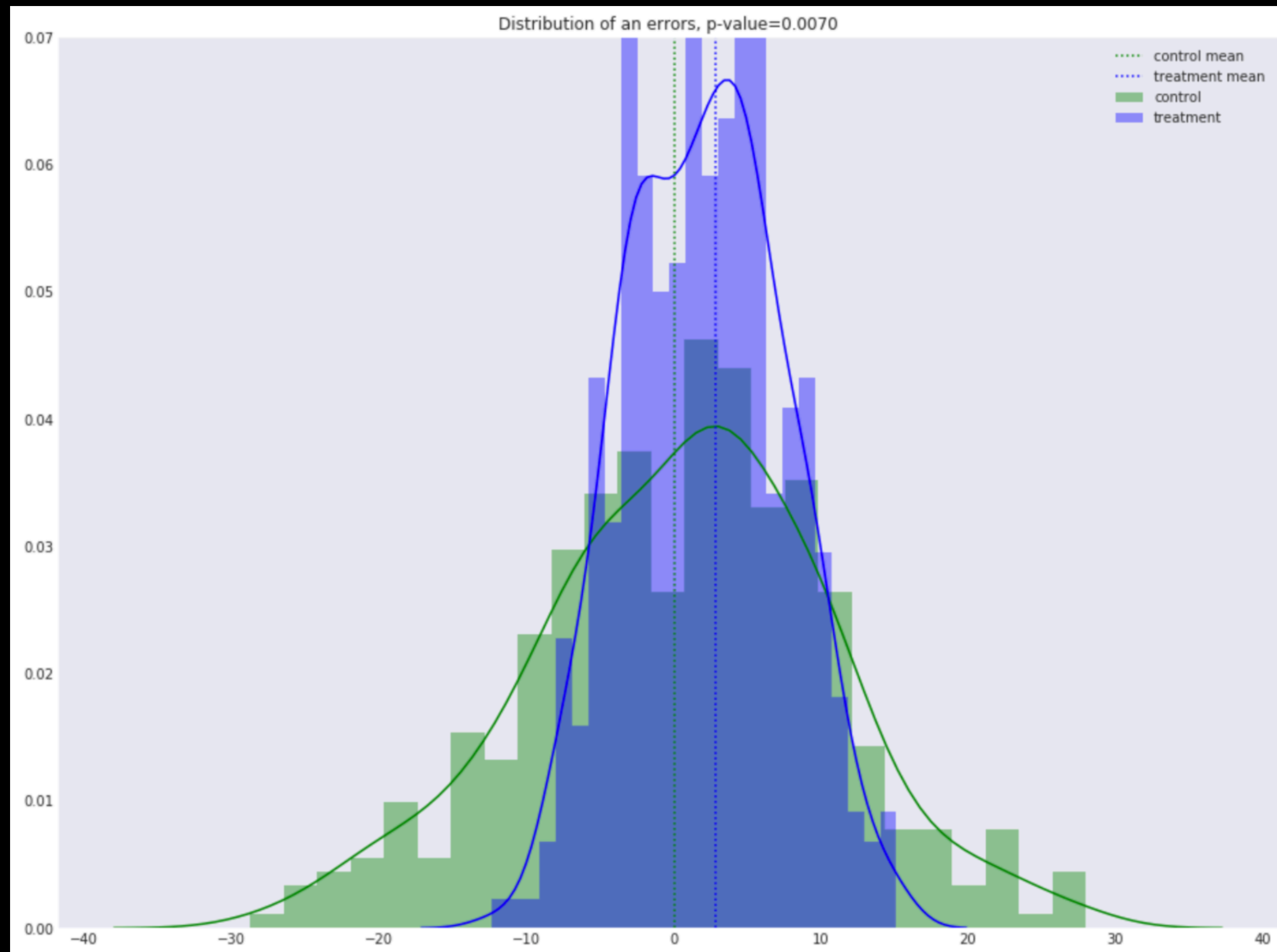
- Берем нашу задачу и рассматриваем ее как временной ряд: например, продажи пива день ото дня во время чемпионата.
- Далее находим некоторый коррелирующий временной ряд. Например, продажи пива в другом городе.
- На основе этих двух рядов будет сформировано предсказание.
- Если фактические цифры выйдут за доверительных интервал предсказанного продолжения ряда, то мы уверены, что некоторое воздействие было оказано.

[https://nbviewer.jupyter.org/github/dafiti/causalimpact/blob/master/examples/getting\\_started.ipynb](https://nbviewer.jupyter.org/github/dafiti/causalimpact/blob/master/examples/getting_started.ipynb)

# Использование описательных ML-моделей



# Использование описательных ML-моделей



# Немного подробнее про gap boundary

, или о том, как бороться с эффектом новизны.

Методология **Facebook**:

- Выкатываем новый дизайн, группа А видит старый, группа В – новый
- Делим группу В еще раз, но хитрее: будем сравнивать тех, кто видел старый и тех, кто зарегистрировался **после** того, как новый дизайн был предложен.

# Почитать

- Статья [на медиуме](#)
- <https://www.amazon.com/Trustworthy-Online-Controlled-Experiments-Practical-ebook/dp/B0845Y3DJV>
- <https://towardsdatascience.com/cuped-r-shiny-tool-7a4869d77f0a>
- <https://booking.ai/how-booking-com-increases-the-power-of-online-experiments-with-cuped-995d186fff1d>
- <https://habr.com/ru/company/ods/blog/325416/>

# А/Б тесты и все, что их окружает

Александр Ольферук,  
наставник

Яндекс Практикум