

«Продвинутые» методы ускорения A/B-тестов. CUPED

Цель урока

А теперь обсудим чуть более сложные методы, позволяющие принять решение при той же мощности, но за меньшее количество экспериментов.

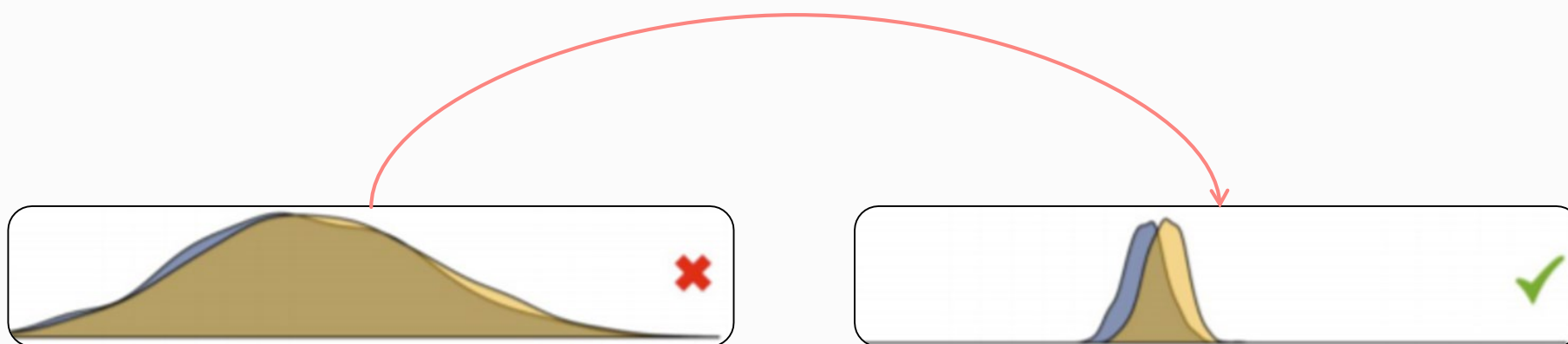
Задачи урока

- ✓ Обсудим, что влияет на количество экспериментов (**спойлер**: дисперсия)
- ✓ Перечислим способы того, как снизить дисперсию метрик
- ✓ Разберём виды линеаризации метрик
- ✓ Насладимся красотой идеи CUPED, про которую все говорят

Ускорение тестов —
это возможность
принять решение
с той же мощностью
при меньшем количестве
наблюдений.

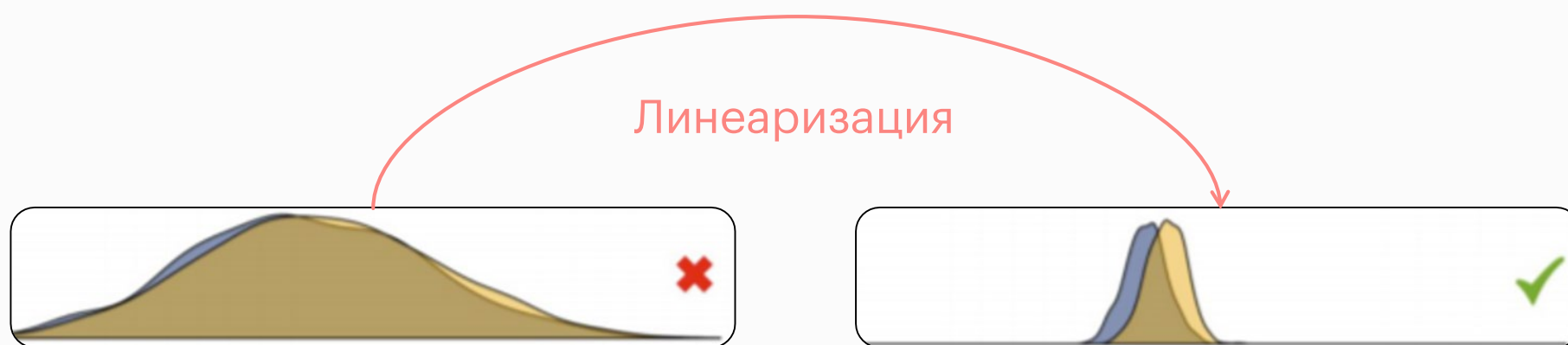
Ускорение тестов

Ускорение тестов — это возможность принять решение с той же мощностью при меньшем количестве наблюдений. Для этого можно **снизить дисперсию**, после чего, как следствие, увеличивается мощность.



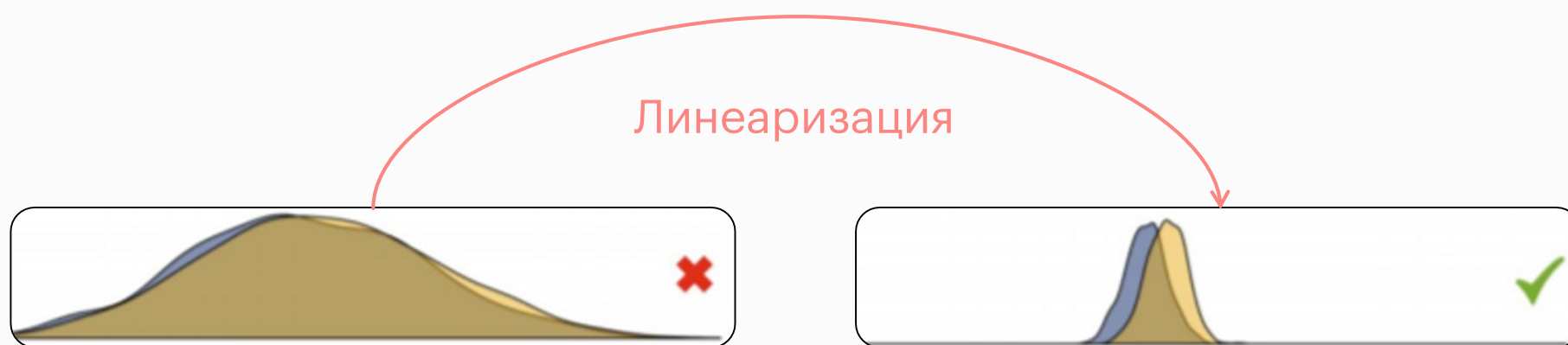
Снижение дисперсии

Снизить дисперсию, чтобы увидеть более тонкие эффекты, можно с помощью **линеаризации**.



Линеаризация

Линеаризацию легко применить к ratio-метрикам — метрикам, где **что-то** делится на **что-то**.
Например, **CTR**.

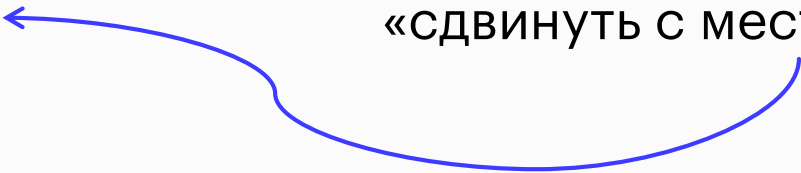


Линеаризация ratio метрик

Ratio-метрика:

$$CTR = \frac{\sum clicks(u)}{\sum views(u)}$$

Обычную ratio-метрику
CTR (глобальный CTR) сложно
«сдвинуть с места»!



Ratio-метрика:

$$linearized_{CTR} = \sum clicks(u) - k \times \sum views(u),$$

$$k = CTR$$

Линеаризация ratio метрик

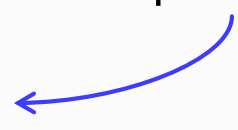
Ratio-метрика:

$$CTR = \frac{\sum clicks(u)}{\sum views(u)}$$

Обычную ratio-метрику
CTR (глобальный CTR) сложно
«сдвинуть с места»!

Ratio-метрика:

А вот линеаризованный CTR
уже сильно проще!

$$linearized_{CTR} = \sum clicks(u) - k \times \sum views(u),$$


$$k = CTR_A$$

Линеаризация ratio метрик

Ratio-метрика:

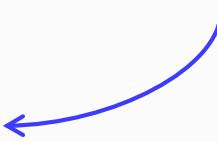
$$CTR = \frac{\sum clicks(u)}{\sum views(u)}$$

Ratio-метрика:

$$linearized_{CTR} = \sum clicks(u) - k \times \sum views(u),$$

$$k = CTR_A$$

Линеаризованный CTR имеет меньшую дисперсию, при этом он сонаправлен с глобальным CTR, чего нельзя сказать о поюзерном CTR.



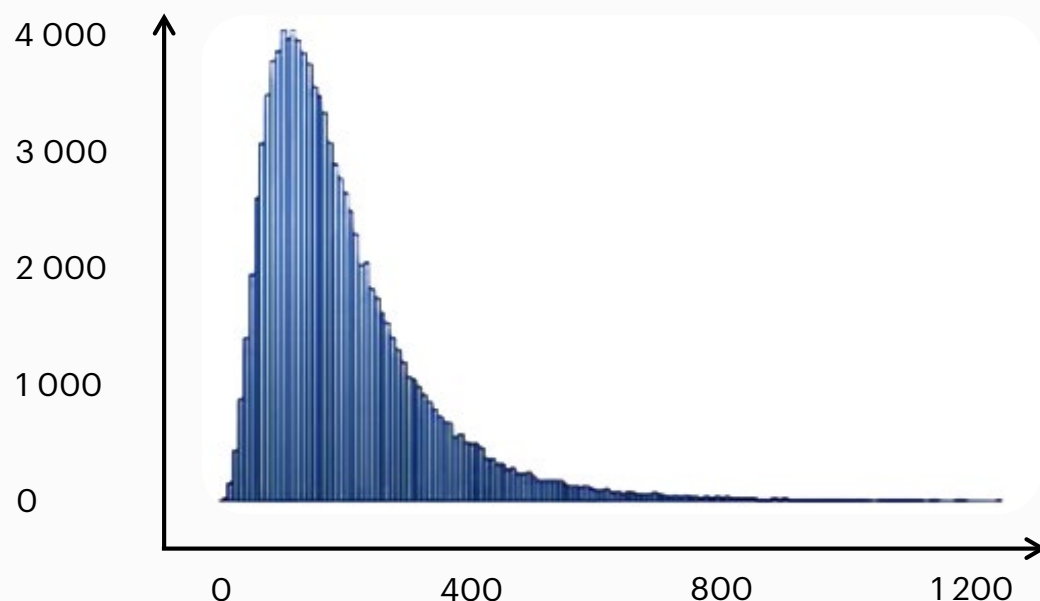
Дисперсия метрик

$$N = N_{groups} \times \frac{\left(\frac{Z_{\alpha/2}}{(C_{N_{groups}}^2 \times N_{metrics})} + Z_{\beta} \right)^2 \times \sigma^2}{\Delta^2}$$



Мы уже видели похожую формулу раньше.

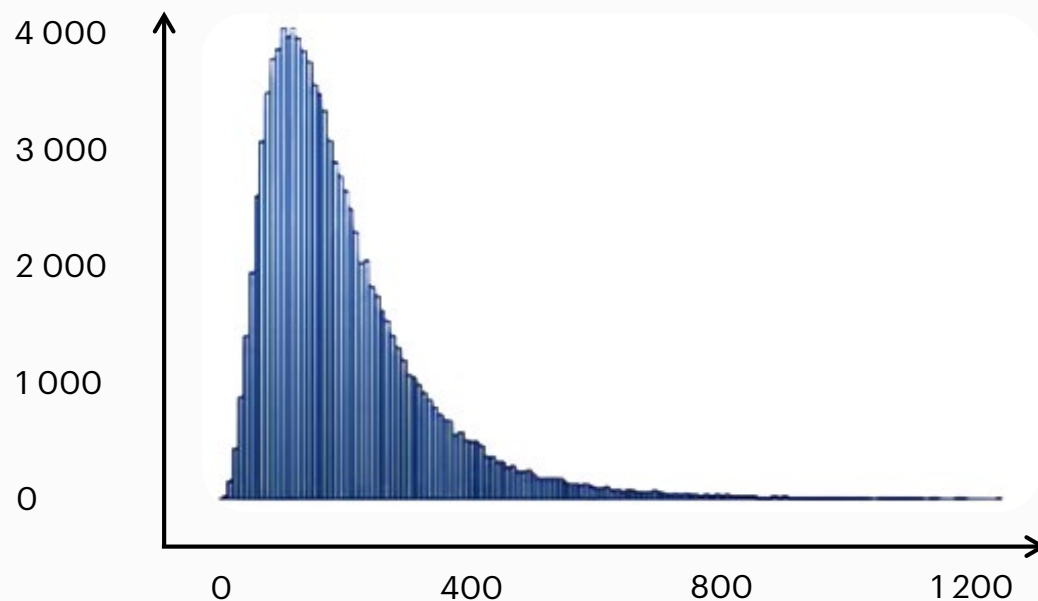
Самое главное: **если снизить дисперсию метрики в два раза, то количество экспериментов снизится тоже в два раза.**



Дисперсия метрик

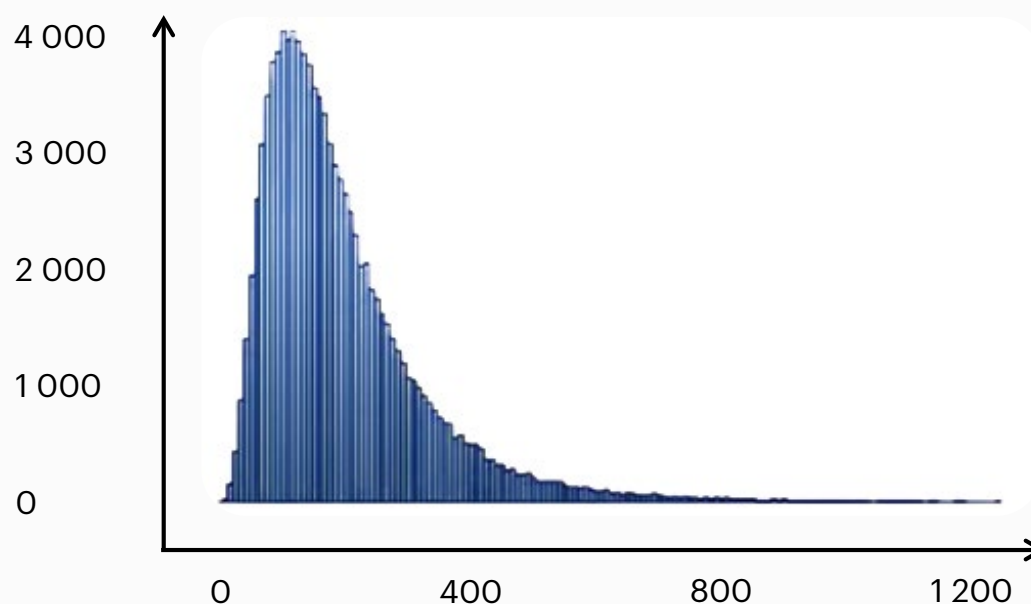
$$N = N_{groups} \times \frac{\left(\frac{Z_{\alpha/2}}{(C_{N_{groups}}^2 \times Nm_{etrics})} + Z_{\beta} \right)^2 \times \sigma^2}{\Delta^2}$$

Как снизить дисперсию???



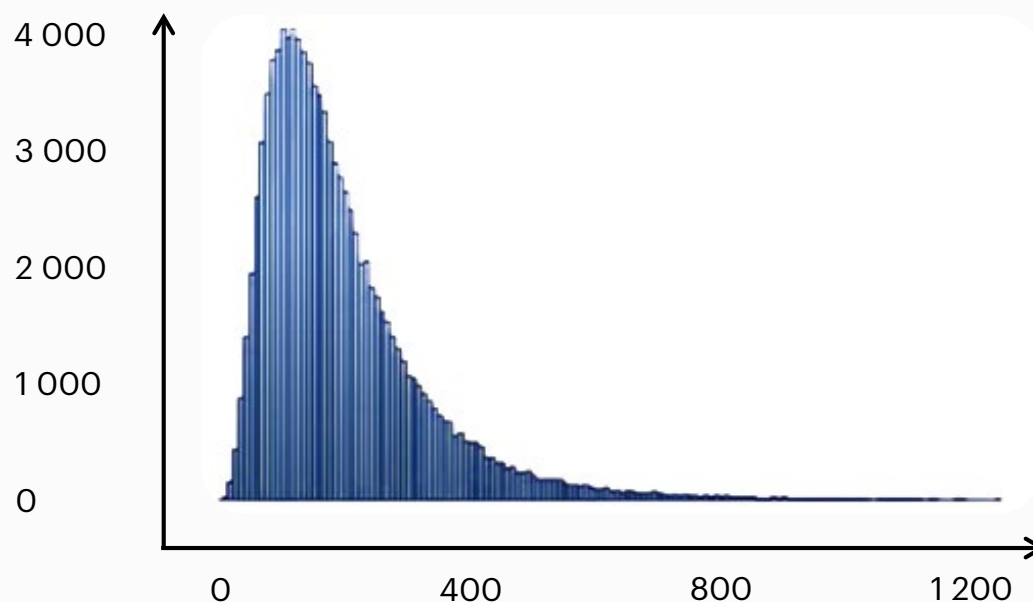
Метрики есть. Снижаем дисперсию метрик!

- Заменяем выбросы максимальным значением
- Критерий Манна — Уитни



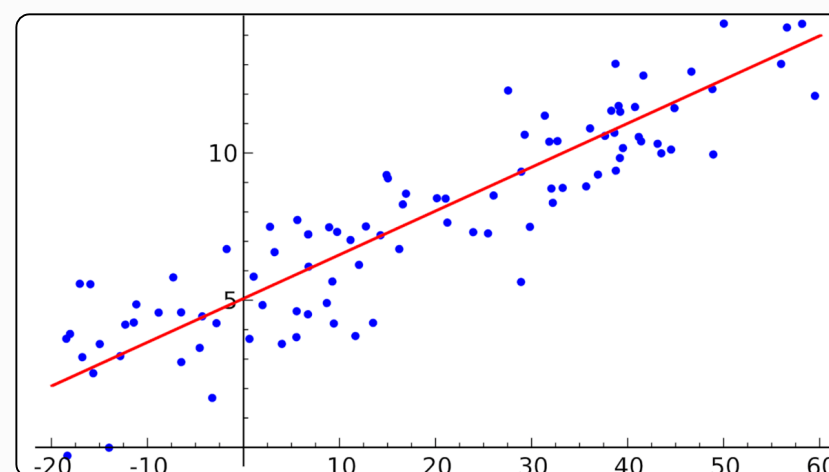
Метрики есть. Снижаем дисперсию метрик!

- Заменяем выбросы максимальным значением
- Критерий Манна — Уитни
- Применяем линеаризацию
 - Постстратификация
 - CUPED
 - CUPAC



Линейная регрессия

Прежде чем мы будем говорить про виды линеаризации, необходимо вспомнить, что по сути это способ моделирования линейной зависимости, подсчёт некоторой метрики, которая линейно зависит от каких-то параметров.



Линейная регрессия

$$metric = a_0 + a_1 \times is_test$$

(Обычный A/B с T-критерием)

Индикаторы группы
A или B (OneHotEncod).

Индикатор тестовой группы.

На линеаризованной
метрике используем
обычный T-критерий.

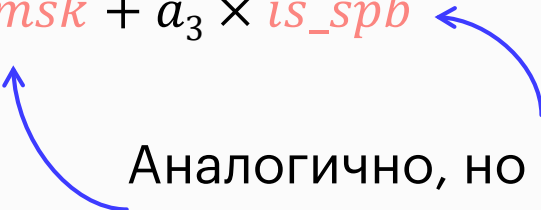
Линейная регрессия

$$metric = a_0 + a_1 \times is_test$$

(Обычный A/B с T-критерием)

$$metric = a_0 + a_1 \times is_test + a_2 \times is_msk + a_3 \times is_spb$$

(Стратификация)



Аналогично, но добавляются ещё индикаторы принадлежности к какой-то категории, например, того, что человек из Москвы или Санкт-Петербурга.

Линейная регрессия

$$metric = a_0 + a_1 \times is_test$$


(Обычный A/B с T-критерием)

$$metric = a_0 + a_1 \times is_test + a_2 \times is_msk + a_3 \times is_spb$$

(Стратификация)

$$metric = a_0 + a_1 \times is_test + a_2 \times metric_{t_21}$$

(CUPED)



Ничего нам не мешает
использовать значения
метрики до эксперимента
(исторических данных)!

В этом и есть весь CUPED.

Линейная регрессия

$$metric = a_0 + a_1 \times is_test$$

(Обычный A/B с T-критерием)

$$metric = a_0 + a_1 \times is_test + a_2 \times is_msk + a_3 \times is_spb$$


(Стратификация)

$$metric = a_0 + a_1 \times is_test + a_2 \times metric_{t-21}$$

(CUPED)

$$metric = a_0 + a_1 \times is_test + a_2 \times metric_{t-21} + a_3 \times Age + a_4 \times is_msk$$

(CUPAC)



Мы можем пойти дальше
и использовать ещё какие-то
данные про пользователей.
Например, их возраст.
С некоторыми коэффициентами.

Линейная регрессия

$$metric = a_0 + a_1 \times is_test$$

(Обычный A/B с T-критерием)

$$metric = a_0 + a_1 \times is_test + a_2 \times is_msk + a_3 \times is_spb$$

(Стратификация)

$$metric = a_0 + a_1 \times is_test + a_2 \times metric_{t-21}$$

(CUPED)

$$metric = a_0 + a_1 \times is_test + a_2 \times metric_{t-21} + a_3 \times Age + a_4 \times is_msk$$

(CUPAC)

Вот и всё! Если думать обо всех этих методах в таком ключе, то всё становится понятно.

Теперь и вы сможете ускорять свои тесты в десятки раз! Пользуйтесь!

Линейная регрессия

$$metric = a_0 + a_1 \times is_test$$

(Обычный A/B с T-критерием)

$$metric = a_0 + a_1 \times is_test + a_2 \times is_msk + a_3 \times is_spb$$

(Стратификация)

$$metric = a_0 + a_1 \times is_test + a_2 \times metric_{t-21}$$

(CUPED)

$$metric = a_0 + a_1 \times is_test + a_2 \times metric_{t-21} + a_3 \times Age + a_4 \times is_msk$$

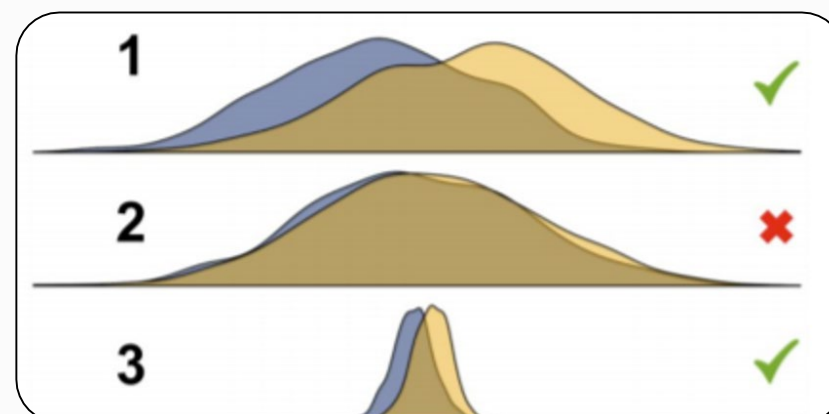
(CUPAC)

Ускорение тестов: ещё раз про идею CUPED

Ускорение тестов — это возможность принять решение с той же мощностью при меньшем количестве наблюдений. Для этого можно снизить дисперсию, после чего, как следствие, увеличивается мощность. Смысл метода CUPED: для уменьшения срока учитываем поведение пользователей до эксперимента.

Обычно дисперсия сокращается на 30–40 % и больше.

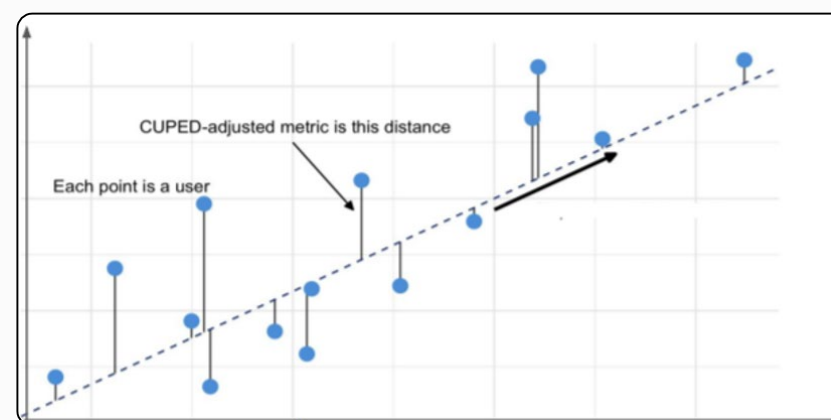
- 1 Большой эффект = большая мощность
- 2 Большая дисперсия + маленький эффект = маленькая мощность
- 3 Маленькая дисперсия = большая мощность



CUPED: метод работы

Дисперсия метрики:

- 1 Объясняемая данными до тестирования
- 2 Не объясняемая данными до тестирования CUPED сокращает дисперсию, опираясь на 1



$$CUPED = metric - (covariate - mean(covariate)) \times theta$$

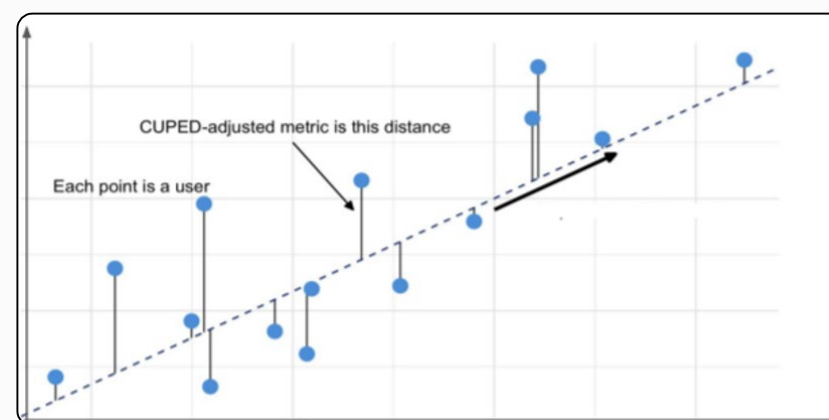
- **covariate** — метрика до эксперимента
- **metric** — метрика после эксперимента

Коэффициент b
линейной регрессии
 $y_t = a + bxt + \varepsilon_t$

CUPED: метод работы

Дисперсия метрики:

- 1 Объясняемая данными до тестирования
- 2 Не объясняемая данными до тестирования CUPED сокращает дисперсию, опираясь на 1



- **theta** вычисляется как
$$\frac{\text{covariance}(\text{metric}, \text{covariate})}{\text{variance}(\text{covariate})}$$

Варианты равнозначны, но первый сохраняет среднее значение метрики.

$$\text{Metric}_{\text{CUPED}} = \text{metric} - \text{covariate} \times \text{theta}$$

CUPED: кейс

Первичная метрика: среднее число покупок в приложении за неделю.

Изменение: дизайн.

По результатам тестирования: покупки могут обладать большей дисперсией. Известно:

- 1 Количество покупок каждого покупателя до теста
- 2 Дисперсия покупок в результате тестирования объясняется
 - a) количеством поездок до теста
 - b) и ещё чем-то

CUPED: кейс

Откажемся от A (количество поездок до теста).

Поэтому можем тестировать, будет ли покупатель делать больше покупок.

У разницы значения, вероятно, дисперсия значительно уменьшится.

Итоги и выводы урока

Если вы уже задействовали все простые техники ускорения тестирования, то самое время использовать что-то посерьёзнее, чтобы ускорить тестирование гипотез ещё в несколько раз.

Итоги и выводы урока

- ✓ Методы линеаризации могут сильно снизить дисперсию в ваших данных, что поможет детектировать более тонкие эффекты
- ✓ Выбор вида зависит от того, насколько разнообразны ваши данные
- ✓ Помните, что ускорение самого тестирования / сбора данных — это только часть всего теста. Пытаться ускорить этот этап надо, но лучше делать это не в первую очередь, а когда у вас уже запускаются периодически более 10 A/B-тестов в месяц. То есть когда речь идёт про настройку массового тестирования