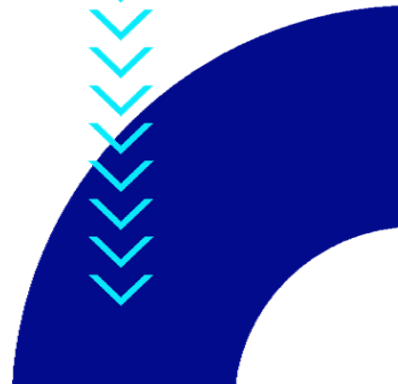


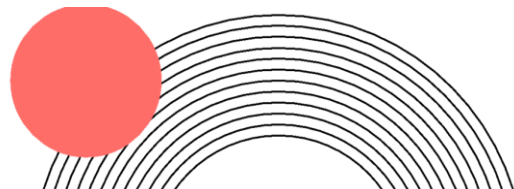
Занятие №4



В ходе четвертого занятия:

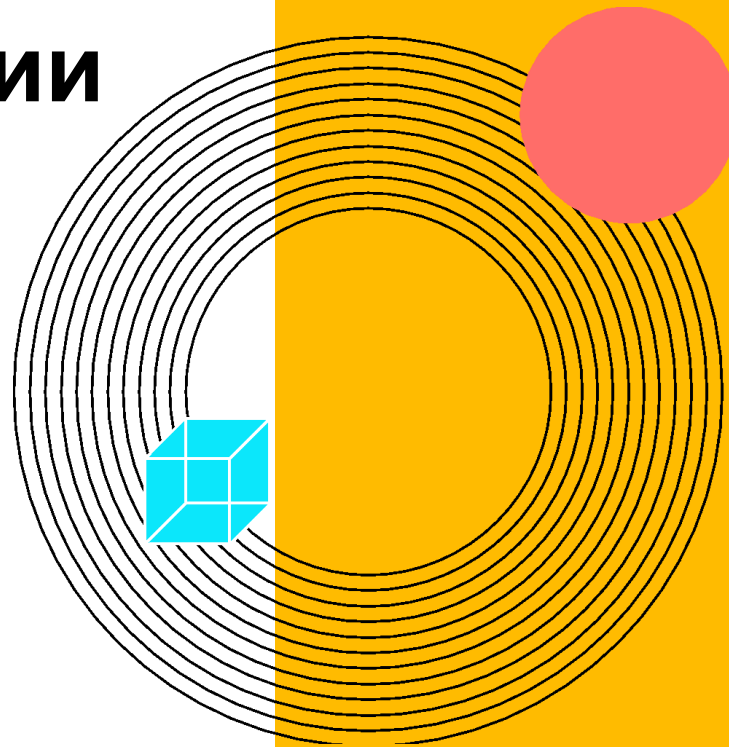
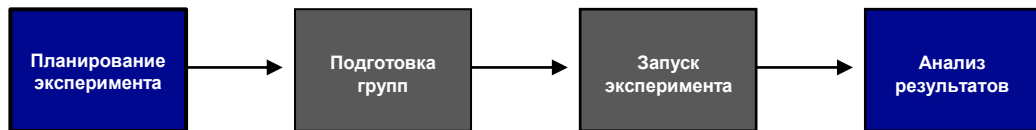


- ▲ обсудим способ снижения ошибок 1-го и 2-го рода: уменьшение дисперсии (CUPED, CUPAC);
- ▲ поймем, что такое прокси метрики и когда их использовать;
- ▲ проговорим, как искать guardrail метрики и какие требования к ним предъявляются;
- ▲ разберемся, как работать с ratio-метриками и какие тесты для них использовать.





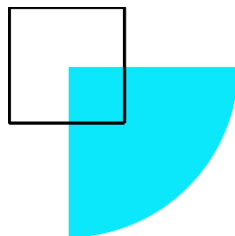
Уменьшение дисперсии CUPED / CUPAC



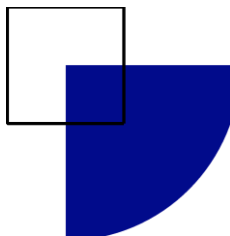
Способ 3. Уменьшение дисперсии. CUPED, CUPAC



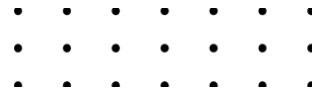
CUPED — это способ уменьшения дисперсии целевой метрики за счет учета ее обычной волатильности.



Как рассчитывается CUPED, CUPAC



Как доказывается CUPED



Способ 3. Уменьшение дисперсии. Как доказывается CUPED



Во-первых, сначала давайте поймем, где мы видим в оригинальной статье формулу, по которой планируем рассчитывать новую метрику.

$$Y_{CUPED} = Y - \theta \hat{X} = Y - \theta(X - \bar{X}) = Y - \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - \bar{X})$$



Deng, A., Xu, Y., Kohavi, R., & Walker, T. (n.d.). Improving the Sensitivity of Online Controlled Experiments by Utilizing Pre-Experiment Data.

3.2.1 Control Variates in Simulation

The idea of variance reduction through control variates stems from the following observation. Assume we can simulate another random variable X in addition to Y with known expectation $\mathbb{E}(X)$. In other words, we have independent pairs of $(Y_i, X_i), i = 1, \dots, n$. Define

$$\hat{Y}_{cv} = \bar{Y} - \theta \bar{X} + \theta \mathbb{E}X, \quad (3)$$

where θ is any constant. \hat{Y}_{cv} is an unbiased estimator of $\mathbb{E}(Y)$ since $-\theta \mathbb{E}(\bar{X}) + \theta \mathbb{E}(X) = 0$. The variance of \hat{Y}_{cv} is

$$\begin{aligned} \text{var}(\hat{Y}_{cv}) &= \text{var}(\bar{Y} - \theta \bar{X}) = \text{var}(Y - \theta X)/n \\ &= \frac{1}{n}(\text{var}(Y) + \theta^2 \text{var}(X) - 2\theta \text{cov}(Y, X)). \end{aligned}$$

Note that $\text{var}(\hat{Y}_{cv})$ is minimized when we choose

$$\theta = \text{cov}(Y, X)/\text{var}(X) \quad (4)$$

and with this optimal choice of θ , we have

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - \rho^2), \quad (5)$$

where $\rho = \text{cor}(Y, X)$ is the correlation between Y and X . Compare (5) to the variance of \bar{Y} , the variance is reduced by a factor of ρ^2 . The larger ρ , the better the variance reduction. The single control variate case can be easily generalized to include multiple variables.

It is interesting to point out the connection with linear regression. The optimal θ turns out to be the ordinary least square (OLS) solution of regressing (centered) Y on (centered) X , which gives variance

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - R^2),$$

with R^2 being the proportion of variance explained coefficient from the linear regression. It is also possible to use nonlinear adjustment. i.e., instead of allowing only linear adjustment as in (3), we can minimize variance in a more general functional space. Define

$$\hat{Y}_{cv} = \bar{Y} - \overline{f(X)} + \mathbb{E}(f(X)), \quad (6)$$

and then try to minimize the variance of (6). It can be shown that the regression function $\mathbb{E}(Y|X)$ gives the optimal $f(X)$.

Способ 3. Уменьшение дисперсии. Как доказывается CUPED

Во-первых, сначала давайте поймем, где мы видим в оригинальной статье формулу, по которой планируем рассчитывать новую метрику.

$$Y_{CUPED} = Y - \theta \hat{X} = Y - \theta(X - \bar{X}) = Y - \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - \bar{X})$$

Во-вторых, посмотрим, почему там есть среднее X . Вычитая матожидание X , мы получаем компонент для оценки дисперсии Y , ведь разность между ними должна была бы быть равна нулю.

3.2.1 Control Variates in Simulation

The idea of variance reduction through control variates stems from the following observation. Assume we can simulate another random variable X in addition to Y with known expectation $\mathbb{E}(X)$. In other words, we have independent pairs of (Y_i, X_i) , $i = 1, \dots, n$. Define

$$\hat{Y}_{cv} = \bar{Y} - \theta \bar{X} + \theta \mathbb{E}X, \quad (3)$$

where θ is any constant. \hat{Y}_{cv} is an unbiased estimator of $\mathbb{E}(Y)$ since $-\theta \mathbb{E}(\bar{X}) + \theta \mathbb{E}(X) = 0$. The variance of \hat{Y}_{cv} is

$$\begin{aligned} \text{var}(\hat{Y}_{cv}) &= \text{var}(\bar{Y} - \theta \bar{X}) = \text{var}(Y - \theta X)/n \\ &= \frac{1}{n}(\text{var}(Y) + \theta^2 \text{var}(X) - 2\theta \text{cov}(Y, X)). \end{aligned}$$

Note that $\text{var}(\hat{Y}_{cv})$ is minimized when we choose

$$\theta = \text{cov}(Y, X)/\text{var}(X) \quad (4)$$

and with this optimal choice of θ , we have

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - \rho^2), \quad (5)$$

where $\rho = \text{cor}(Y, X)$ is the correlation between Y and X . Compare (5) to the variance of \bar{Y} , the variance is reduced by a factor of ρ^2 . The larger ρ , the better the variance reduction. The single control variate case can be easily generalized to include multiple variables.

It is interesting to point out the connection with linear regression. The optimal θ turns out to be the ordinary least square (OLS) solution of regressing (centered) Y on (centered) X , which gives variance

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - R^2),$$

with R^2 being the proportion of variance explained coefficient from the linear regression. It is also possible to use nonlinear adjustment. i.e., instead of allowing only linear adjustment as in (3), we can minimize variance in a more general functional space. Define

$$\hat{Y}_{cv} = \bar{Y} - \overline{f(X)} + \mathbb{E}(f(X)), \quad (6)$$

and then try to minimize the variance of (6). It can be shown that the regression function $\mathbb{E}(Y|X)$ gives the optimal $f(X)$.

Способ 3. Уменьшение дисперсии. Как доказывается CUPED

Во-первых, сначала давайте поймем, где мы видим в оригинальной статье формулу, по которой планируем рассчитывать новую метрику.

$$Y_{CUPED} = Y - \theta \hat{X} = Y - \theta(X - \bar{X}) = \underbrace{Y}_{\text{середнее}} - \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - \bar{X})$$

Во-вторых, посмотрим, почему там есть среднее \bar{X} . Вычитая матожидание \bar{X} , мы получаем компонент для оценки дисперсии Y , ведь разность между ними должна была бы быть равна нулю.

В-третьих, с учетом того, что у нас два слагаемых в формуле, посмотрим какой должна быть тета, чтобы минимизировать дисперсию Y_{CUPED} . Для этого, сначала посмотрим на то, во что раскладывается дисперсия Y_{CUPED} .

3.2.1 Control Variates in Simulation

The idea of variance reduction through control variates stems from the following observation. Assume we can simulate another random variable X in addition to Y with known expectation $\mathbb{E}(X)$. In other words, we have independent pairs of $(Y_i, X_i), i = 1, \dots, n$. Define

$$\hat{Y}_{cv} = \bar{Y} - \theta \bar{X} + \theta \mathbb{E}X, \quad (3)$$

where θ is any constant. \hat{Y}_{cv} is an unbiased estimator of $\mathbb{E}(Y)$ since $-\theta \mathbb{E}(\bar{X}) + \theta \mathbb{E}(X) = 0$. The variance of \hat{Y}_{cv} is

$$\begin{aligned} \text{var}(Y_{cv}) &= \text{var}(\bar{Y} - \theta \bar{X}) = \text{var}(Y - \theta X)/n \\ &= \frac{1}{n}(\text{var}(Y) + \theta^2 \text{var}(X) - 2\theta \text{cov}(Y, X)). \end{aligned}$$

Note that $\text{var}(\hat{Y}_{cv})$ is minimized when we choose

$$\theta = \text{cov}(Y, X) / \text{var}(X) \quad (4)$$

and with this optimal choice of θ , we have

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - \rho^2), \quad (5)$$

where $\rho = \text{cor}(Y, X)$ is the correlation between Y and X . Compare (5) to the variance of \bar{Y} , the variance is reduced by a factor of ρ^2 . The larger ρ , the better the variance reduction. The single control variate case can be easily generalized to include multiple variables.

It is interesting to point out the connection with linear regression. The optimal θ turns out to be the ordinary least square (OLS) solution of regressing (centered) Y on (centered) X , which gives variance

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - R^2),$$

with R^2 being the proportion of variance explained coefficient from the linear regression. It is also possible to use nonlinear adjustment. i.e., instead of allowing only linear adjustment as in (3), we can minimize variance in a more general functional space. Define

$$\hat{Y}_{cv} = \bar{Y} - \overline{f(X)} + \mathbb{E}(f(X)), \quad (6)$$

and then try to minimize the variance of (6). It can be shown that the regression function $\mathbb{E}(Y|X)$ gives the optimal $f(X)$.

Способ 3. Уменьшение дисперсии. Как доказывается CUPED

Во-первых, сначала давайте поймем, где мы видим в оригинальной статье формулу, по которой планируем рассчитывать новую метрику.

$$Y_{CUPED} = Y - \theta \hat{X} = Y - \theta(X - \bar{X}) = \underbrace{Y}_{\text{Y}} - \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - \bar{X})$$

Во-вторых, посмотрим, почему там есть среднее \bar{X} . Вычитая матожидание \bar{X} , мы получаем компонент для оценки дисперсии Y , ведь разность между ними должна была бы быть равна нулю.

В-третьих, с учетом того, что у нас два слагаемых в формуле, посмотрим какой должна быть тета, чтобы минимизировать дисперсию Y_{CUPED} . Для этого, сначала посмотрим на то, во что раскладывается дисперсия Y_{CUPED} .

В-четвертых, дальше авторы указывают, что такой подход — это линейное уменьшение дисперсии, которое соответствует стандартной OLS регрессии. Но потенциально возможна и более сложная функция теты для уменьшения дисперсии показателя Y_{CUPED} , если нелинейная зависимость между X и Y .

3.2.1 Control Variates in Simulation

The idea of variance reduction through control variates stems from the following observation. Assume we can simulate another random variable X in addition to Y with known expectation $\mathbb{E}(X)$. In other words, we have independent pairs of $(Y_i, X_i), i = 1, \dots, n$. Define

$$\hat{Y}_{cv} = \bar{Y} - \theta \bar{X} + \theta \bar{E}X, \quad (3)$$

where θ is any constant. \hat{Y}_{cv} is an unbiased estimator of $\mathbb{E}(Y)$ since $-\theta \mathbb{E}(\bar{X}) + \theta \mathbb{E}(X) = 0$. The variance of \hat{Y}_{cv} is

$$\begin{aligned} \text{var}(Y_{cv}) &= \text{var}(\bar{Y} - \theta \bar{X}) = \text{var}(Y - \theta X)/n \\ &= \frac{1}{n}(\text{var}(Y) + \theta^2 \text{var}(X) - 2\theta \text{cov}(Y, X)). \end{aligned}$$

Note that $\text{var}(\hat{Y}_{cv})$ is minimized when we choose

$$\theta = \text{cov}(Y, X)/\text{var}(X) \quad (4)$$

and with this optimal choice of θ , we have

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - \rho^2), \quad (5)$$

where $\rho = \text{cor}(Y, X)$ is the correlation between Y and X . Compare (5) to the variance of \bar{Y} , the variance is reduced by a factor of ρ^2 . The larger ρ , the better the variance reduction. The single control variate case can be easily generalized to include multiple variables.

It is interesting to point out the connection with linear regression. The optimal θ turns out to be the ordinary least square (OLS) solution of regressing (centered) Y on (centered) X , which gives variance

$$\text{var}(\hat{Y}_{cv}) = \text{var}(\bar{Y})(1 - R^2),$$

with R^2 being the proportion of variance explained coefficient from the linear regression. It is also possible to use nonlinear adjustment. i.e., instead of allowing only linear adjustment as in (3), we can minimize variance in a more general functional space. Define

$$\hat{Y}_{cv} = \bar{Y} - \overline{f(X)} + \mathbb{E}(f(X)), \quad (6)$$

and then try to minimize the variance of (6). It can be shown that the regression function $\mathbb{E}(Y|X)$ gives the optimal $f(X)$.

Способ 3. Уменьшение дисперсии. Как рассчитывается CUPED




Представим, что

- Y — это расходы пользователя во время эксперимента (целевая метрика теста),
- X — это расходы пользователя на предпериоде.

Для расчета Y_{CUPED} мы должны:

 Рассчитать ковариацию X, Y — $cov(X, Y)$

 Рассчитать дисперсию X — $var(X)$

 Рассчитать среднее от X — \overline{X}

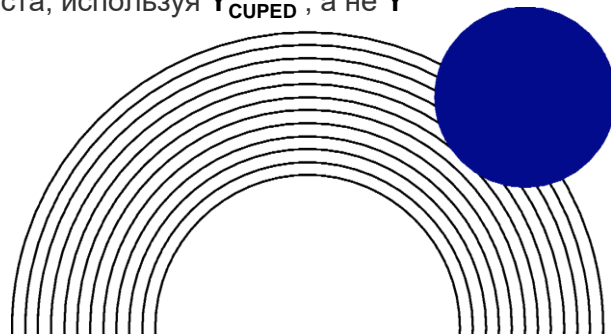


Посчитать для каждого пользователя Y_{CUPED}

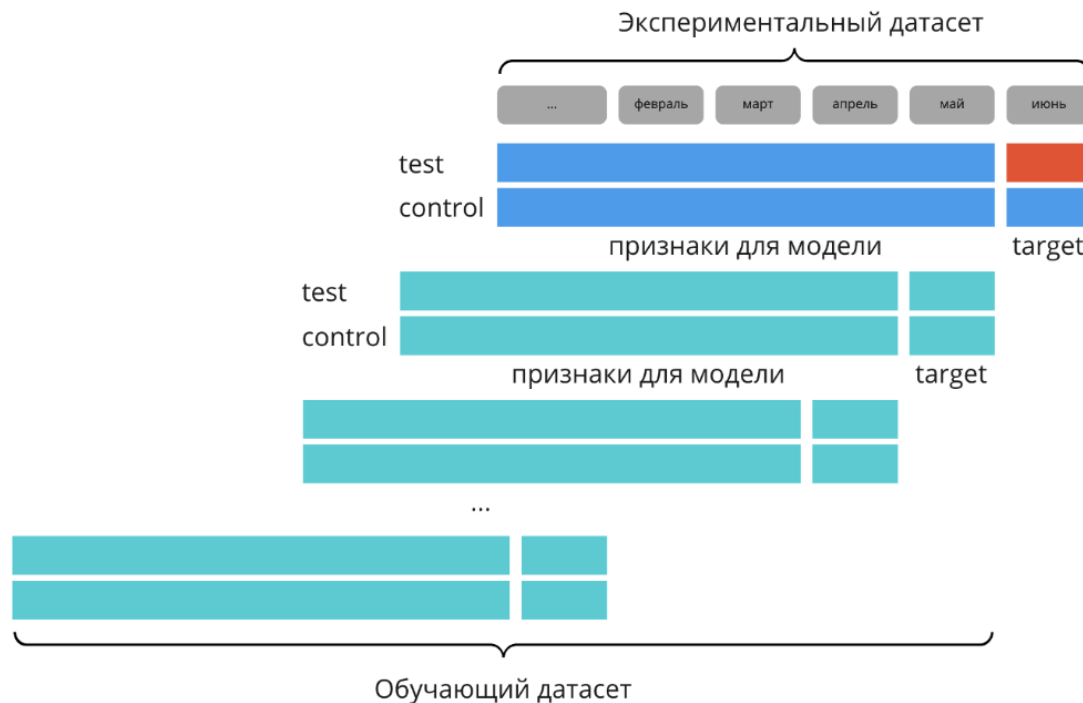
$$Y_{CUPED} = Y - \theta \hat{X} = Y - \theta(X - \overline{X}) = Y - \frac{cov(X, Y)}{var(X)}(X - \overline{X})$$



Подвести итоги теста, используя Y_{CUPED} , а не Y



Способ 3. Уменьшение дисперсии. Как рассчитывается CUPAC



Способ 3. Уменьшение дисперсии. Как рассчитывается CUPED, CUPAC



Чаще всего для CUPED в качестве X берут:

- ▶ Ту же метрику, что и целевая;
- ▶ Берут ее среднее за более длительный период (например, среднее за три предыдущих месяца, если тестируем показатель за месяц).

Следует обратить внимание:

- ▶ Theta рассчитывается одновременно для группы А и группы Б.

Чаще всего для CUPAC в качестве X берут:

- ▶ Прогноз той же метрики, что и целевая.

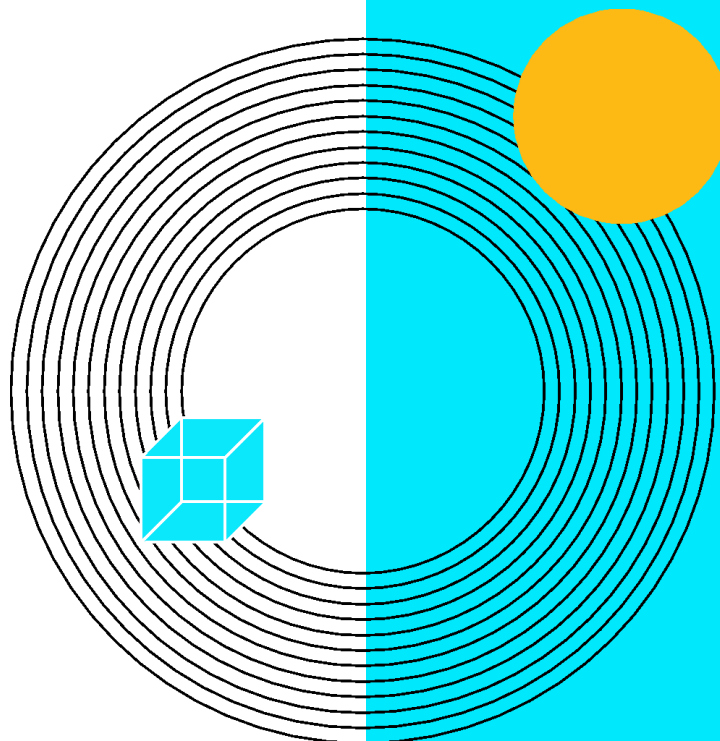
Изначально CUPED был предложен: Deng, A., Xu, Y., Kohavi, R., & Walker, T. (n.d.). Improving the Sensitivity of Online Controlled Experiments by Utilizing Pre-Experiment Data.
URL: <https://www.exp-platform.com/Documents/2013-02-CUPED-ImprovingSensitivityOfControlledExperiments.pdf>

Изначально CUPAC был предложен: Jeff Li, Yixin Tang, and Jared Bauman (2020) Improving Experimental Power through Control Using Predictions as Covariate (CUPAC).
URL: <https://doordash.engineering/2020/06/08/improving-experimental-power-through-control-using-predictions-as-covariate-cupac/>



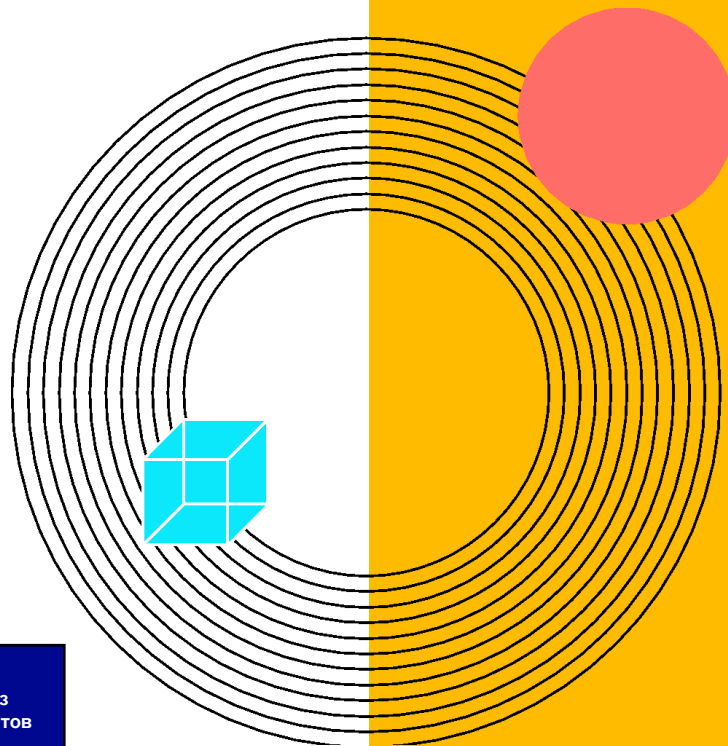
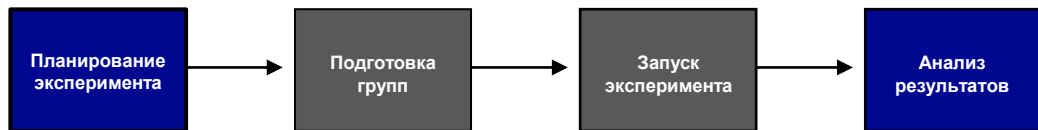
Демонстрация

CUPED, CUPAC





Типы метрик



Типы метрик:



Целевая

- ▼ Основная метрика, на которую ориентируемся, когда принимаем решение о каком-то глобальном изменении
- ▼ Обычно должна быть конвертируема в деньги и понятна бизнесу



Прокси (опережающая)



Guardrail (барьерная)



Прокси (опережающие) метрики



Это метрики, которые позволяют нам раньше или с меньшим числом наблюдений получить статистически значимый результат.



Условие:

Прокси метрика должна быть сонаправлена целевой метрике. То есть должна при воздействии (!) меняться в ту же сторону, что и целевая. Если целевая растет, то и прокси метрика растет.



В каких случаях работает:

Сайты, приложения и любые сервисы, в которых пользователь может совершать много разнообразных действий, а разумная с точки зрения бизнеса переменная не всегда деньги (например, деньги идут не через прямые продажи).



Как находим прокси метрики:

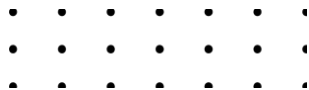
На уже прошедших тестах перебираем метрики и смотрим:

- Коррелированы ли они с целевой;
- Выше ли у них чувствительность;
- Есть ли бизнес логика за тем, что они меняются в унисон с целевой.



Сложности:

Когда целевая метрика - это деньги, бизнес редко соглашается на какую-либо прокси метрику кроме конверсии (но конверсия есть не всегда, и не всегда бизнес согласен и с ней).



Guardrail (барьерные) метрики:



В каких случаях работает:

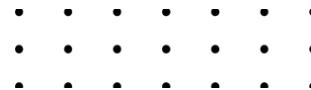
- Когда у бизнеса (продукта) есть много метрик.
- Когда в экосистеме не все продукты направлены на повышение дохода / услуги не монетизируются напрямую.

То есть преимущественно сайты, приложения.



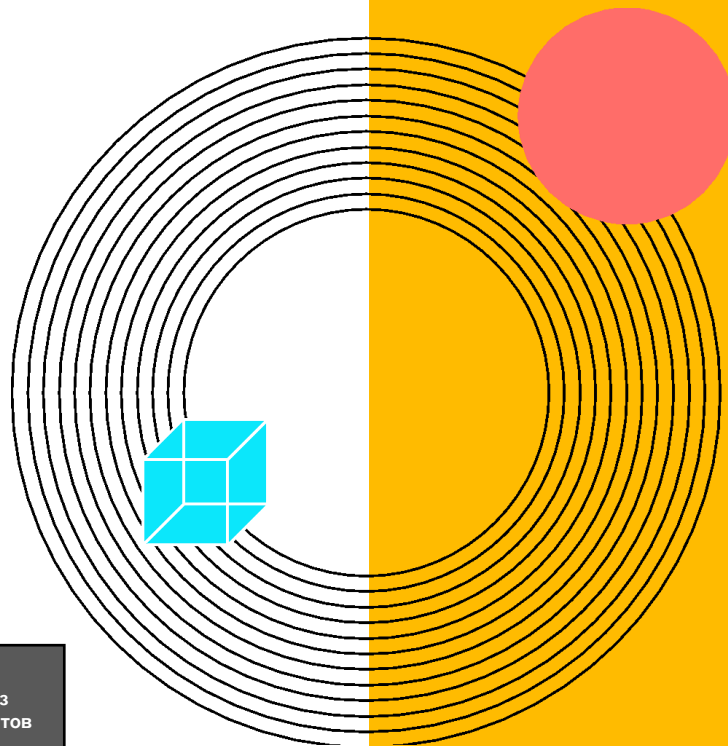
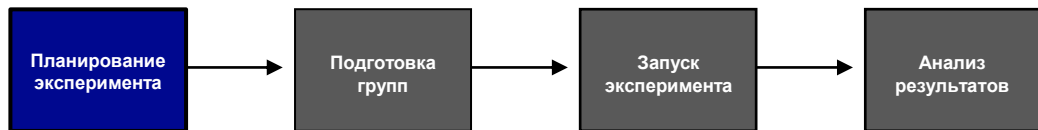
Когда может не работать:

- Если бизнес максимизирует только один показатель (деньги).





Ratio-метрики






Типы метрик по метрической сущности






Тип метрики




Непрерывная

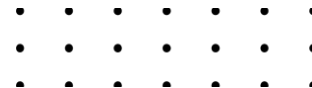
-  Величина чека
-  Время в секундах
-  Рост в сантиметрах

Доля

-  Конверсия в покупку
-  Отклик на кампанию
-  Выход из оттока

Ratio

-  CTR
-  Profit Margin
-  Profit Cost



Ratio-метрики



Ratio-метрика — метрика-отношение двух сущностей.


Чаще всего ratio-метрика поюзерная.

Важно: зависимые непрерывные метрики тоже относятся к ratio и представляют как метрику, деленную на единицу.

Примеры:

 CTR = clicks / session length — кликабельность

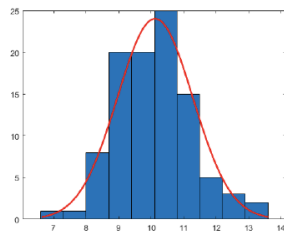
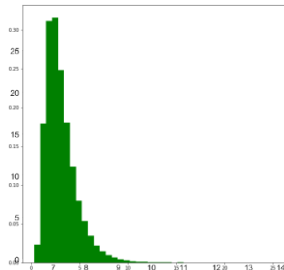
 Profit Margin = Profit / Revenue — маржа

 Profit Percentage = Profit / Cost — наценка

$$\text{CTR}_{total} = \frac{clicks_1 + clicks_2}{app_1 + app_2} \neq \frac{CTR_1 + CTR_2}{2} = CTR_{avg}$$

Свойства:

Ratio =



Ratio-метрики



Дельта-метод — метод приближения среднего и дисперсии ratio-метрики.

$$\text{bias correction}\left[\frac{R}{S}\right] = \frac{\mu_S}{\mu_R^3} * \frac{s_R^2}{n} - \frac{1}{\mu_R^2} * \frac{s_{RS}}{n}$$

$$E\left[\frac{R}{S}\right] = \frac{\mu_S}{\mu_R^3} - 1 + \text{bias correction}$$

$$\text{Var}\left[\frac{R}{S}\right] = \frac{s_S^2}{\mu_R^2} - 2 * \frac{\mu_S}{\mu_R^3} * s_{RS} + \frac{\mu_S^2}{\mu_R^4} * s_R^2$$

} Рассчитать для каждой группы

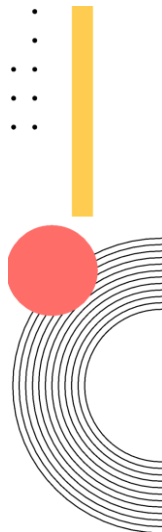
μ_R — Среднее числителя
 μ_S — Среднее знаменателя
 s_R — Ст. откл. числителя
 s_S — Ст. откл. знаменателя
 s_{RS} — Ковариация

Далее можно рассчитать t-статистику и провести t-test:

$$t = \frac{\overline{x_B} - \overline{x_A}}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{E\left[\frac{R}{S}\right]_B - E\left[\frac{R}{S}\right]_A}{\sqrt{\frac{\text{Var}\left[\frac{R}{S}\right]_A}{n_A} + \frac{\text{Var}\left[\frac{R}{S}\right]_B}{n_B}}}$$

[Applying the Delta Method in Metric Analytics: A Practical Guide with Novel Ideas \[paper\]](#)

[Applying Delta Method in A/B Tests Analysis \[article\]](#)



Ratio-метрики

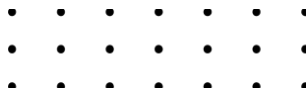


Разложение Тейлора — метод приближения среднего и дисперсии ratio-метрики с помощью разложения Тейлора.

$$\left. \begin{aligned} E\left[\frac{R}{S}\right] &\approx \frac{\mu_R}{\mu_S} - \frac{s_{RS}}{\mu_S^2} + \frac{s_S^2 \mu_R}{\mu_S^3} \\ \text{Var}\left[\frac{R}{S}\right] &\approx \frac{\mu_R^2}{\mu_S^2} \left[\frac{s_R^2}{\mu_R^2} - 2 \frac{s_{RS}}{\mu_R \mu_S} + \frac{s_S^2}{\mu_S^2} \right] \end{aligned} \right\} \begin{array}{l} \text{Рассчитать для} \\ \text{каждой группы} \end{array}$$

μ_R — Среднее числителя
 μ_S — Среднее знаменателя
 s_R — Ст. откл. числителя
 s_S — Ст. откл. знаменателя
 s_{RS} — Ковариация

Далее можно рассчитать t-статистику и провести t-test:
$$t = \frac{\overline{x_B} - \overline{x_A}}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{E\left[\frac{R}{S}\right]_B - E\left[\frac{R}{S}\right]_A}{\sqrt{\frac{\text{Var}\left[\frac{R}{S}\right]_A}{n_A} + \frac{\text{Var}\left[\frac{R}{S}\right]_B}{n_B}}}$$



Ratio-метрики



Линеаризация — способ перевода ratio-метрики в обычную дискретную.

$$CTR = \frac{\sum_{u \in U} C(u)}{\sum_{u \in U} S(u)}, \text{ где } C \text{ — clicks, } S \text{ — session length}$$

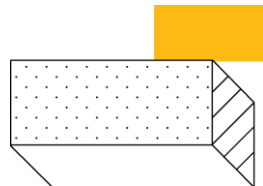
$L(u) = C(u) - K * S(u)$, где K — CTR в контрольной группе

Новая метрика:
$$LCTR = \frac{\sum_{u \in U} L(u)}{|U|}$$

С линеаризованной метрикой можно работать как с обычной непрерывной, т.е. применять методы понижения размерности вида CUPED/CUPAC, использовать тесты для непрерывных метрик.



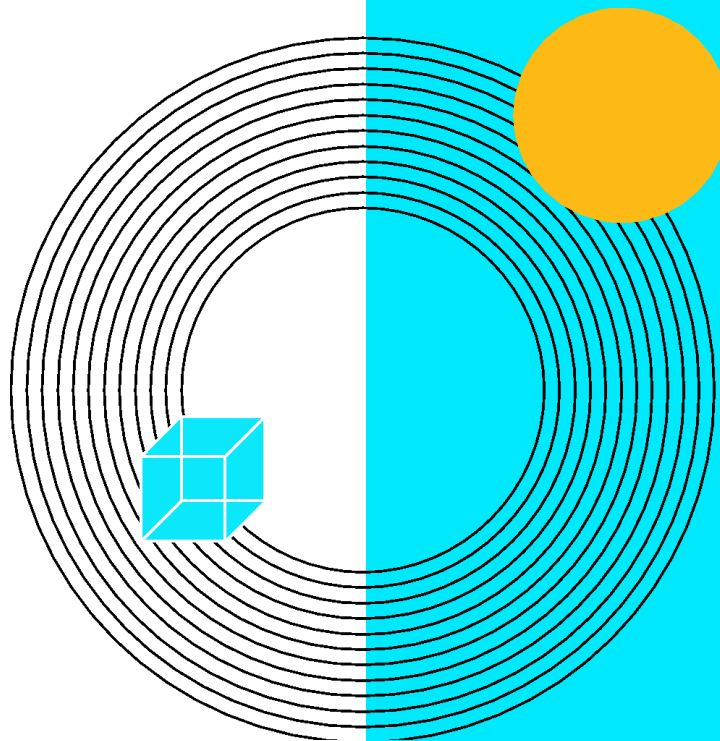
[Как ускорить сотни А/Б-тестов в десятки раз \[YouTube\]](#)





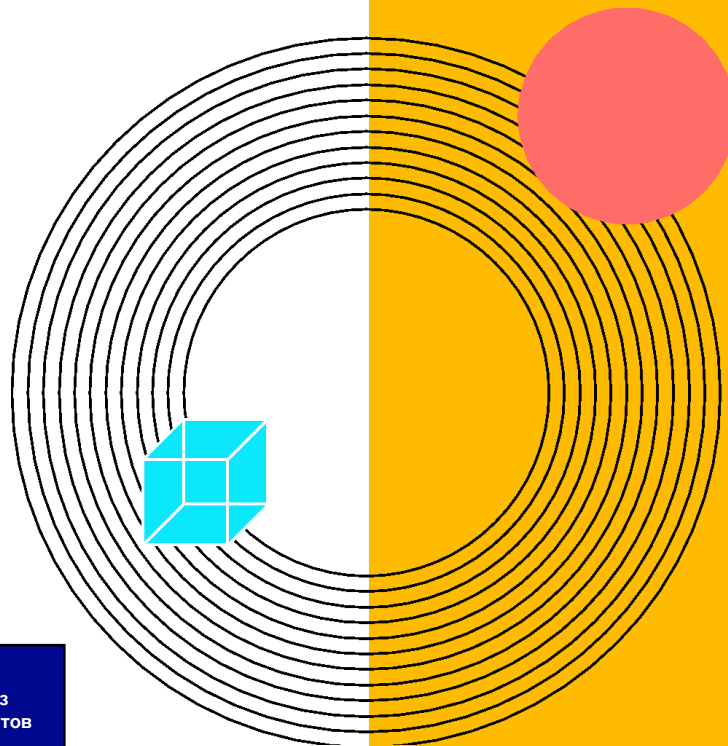
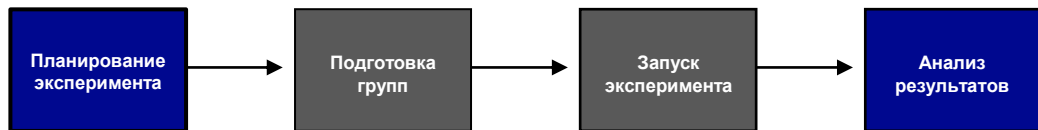
Демонстрация

Линеаризация

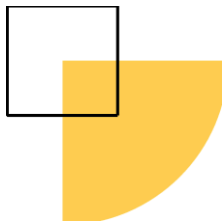




Длительность эксперимента

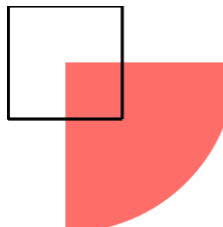


Длительность эксперимента



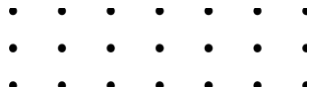
Fixed horizon test

Эксперимент с фиксированным временем проведения и конкретной датой окончания.



Sequential testing

Эксперимент без фиксированного количества требуемых клиентов и без конкретной даты окончания.



Fixed horizon test. Проблема подглядывания

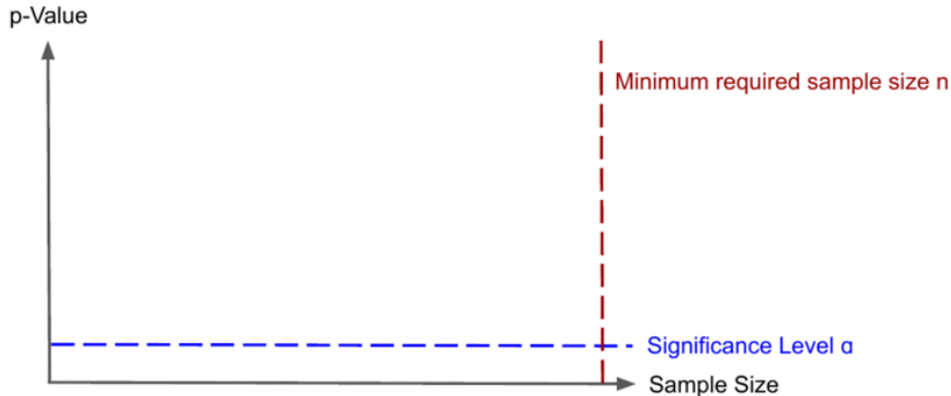


Подглядывание в A/B-экспериментах — слежение за результатами эксперимента по мере его проведения.



Основное правило: смотреть можно, вносить изменения в дизайн — нет.

Единственное исключение: если в ходе эксперимента выяснилось, что он сломан или явно ухудшает целевую и/или барьерную метрику.




Sequential testing Overview




Sequential probability ratio test (SPRT) — sequential test для бинарной метрики.

Область решений определяется неравенством $a < S_i < b$

где:

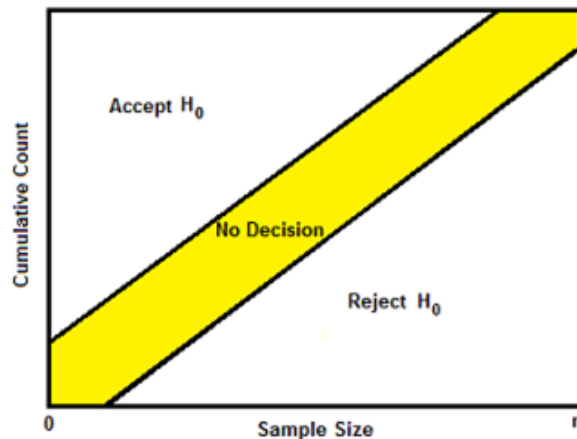
 $S_i = S_{i-1} + \log \Lambda_i, S_0 = 0$

 Λ — cumulative sum of log-likelihood ratio

 $a \approx \log \frac{\beta}{1 - \alpha}$ and $b \approx \log \frac{1 - \beta}{\alpha}$

Прочие подходы:

- Modified SPRT
- Group Sequential Methods
- Sequential Bayes Factors
- Sequential t-test

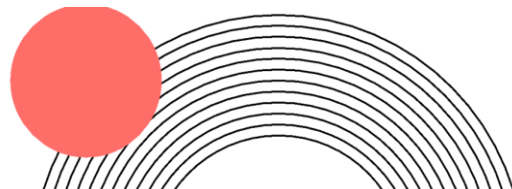


Sequential Probability Ratio Test: Definition & Overview
Sequential Probability Ratio Test for Reliability Demonstration

Выводы по четвертому занятию



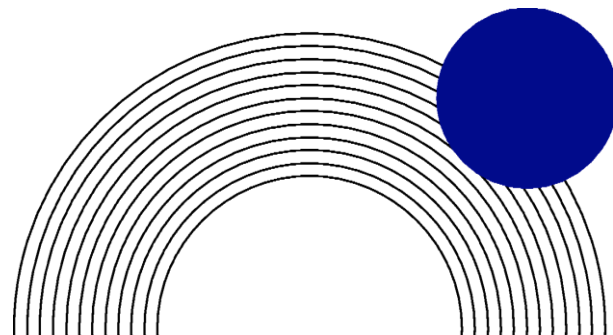
- ▲ На сложных продуктах, где есть серьезный интерес бизнеса к метрикам кроме “денег”, вводят (разработают) опережающие и guardrail метрики, которые позволяют лучше оценивать результат тестирования.
- ▲ Нельзя работать с ratio-метриками, как с обычными непрерывными.
- ▲ Существует способ приведения ratio-метрики к обычной дискретной.
- ▲ Существуют способы приближения среднего и дисперсии ratio-метрики для использования стандартных статистических тестов.
- ▲ Альтернатива проведению фиксированного по времени эксперимента — sequential testing.



Дополнительная литература



- ▶ Sequential Probability Ratio Test: Definition & Overview
- ▶ Sequential Probability Ratio Test for Reliability Demonstration





ВОПРОСЫ

