

ТПОЭ - 23/24

Лекция 5

Нерсес Багиян

Мы сформулировали дизайн эксперимента

1. Формулировка гипотезы с сформулированным ожидаемым размером эффекта

Добавление выделенного раздела для экологичных мыл на главной странице увеличит конверсию с главной страницы на страницы товаров на 15% и снизит отказы на главной странице на 10%.

2. Описание аудитории

Посетители онлайн-магазина мыла, включая новых и возвращающихся пользователей.

3. Описание вариантов с размером каждой группы

Контрольная группа (А): Главная страница без выделенного раздела для экологичных мыл. Размер группы – 50% посетителей.

Экспериментальная группа (В): Главная страница с выделенным разделом для экологичных мыл. Размер группы – 50% посетителей.

4. Ожидаемые исходы и метрики

Основная метрика: Увеличение конверсии с главной страницы на страницы товаров.

Guardrail метрика: Конверсия из посещения главной страницы в активную сессию

5. Продолжительность

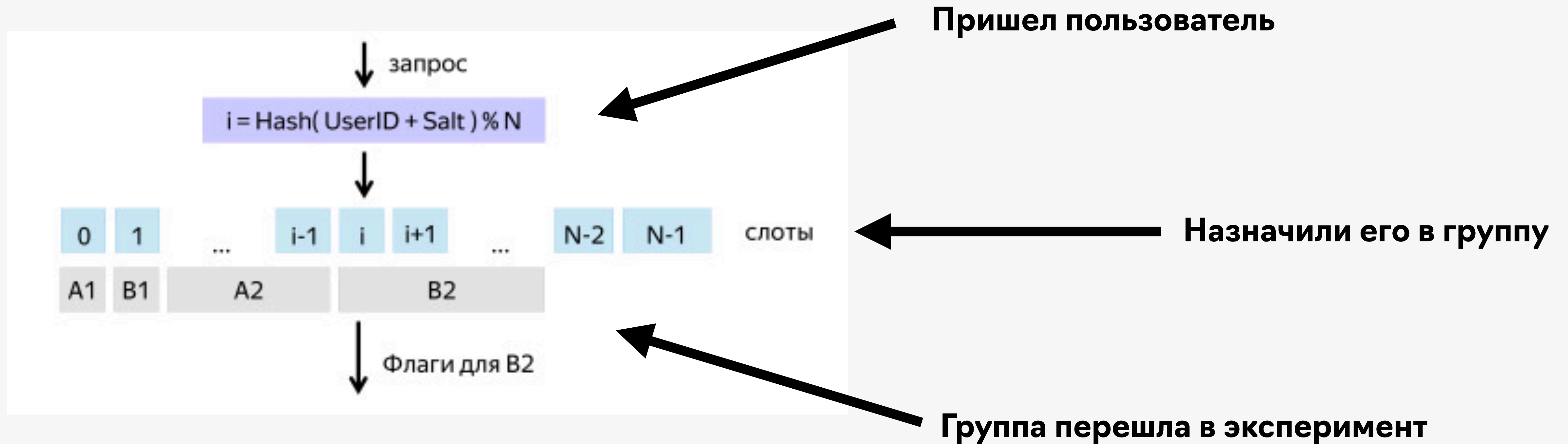
Эксперимент продлится 4 недели, чтобы собрать достаточно данных для статистически значимых результатов, учитывая недельные колебания трафика и поведения покупателей.

6. Результаты

TBD



Как понять сколько нужно брать бакетов?



Как думаете есть ли тут универсальный ответ?

Вопрос аудитории

С какими метриками можно проводить эксперимент?

Чувствительность

Насколько долго нужно ждать, чтобы увидеть изменения в метрике

Достоверность

Можно ли получить точное подтверждение из данных?

Интерпретируемость

Можем ли четко ответить, что улучшили?

Способ подсчета

Среднее, отношение или квантиль



Какой тип метрик мы забыли?

Вопрос аудитории

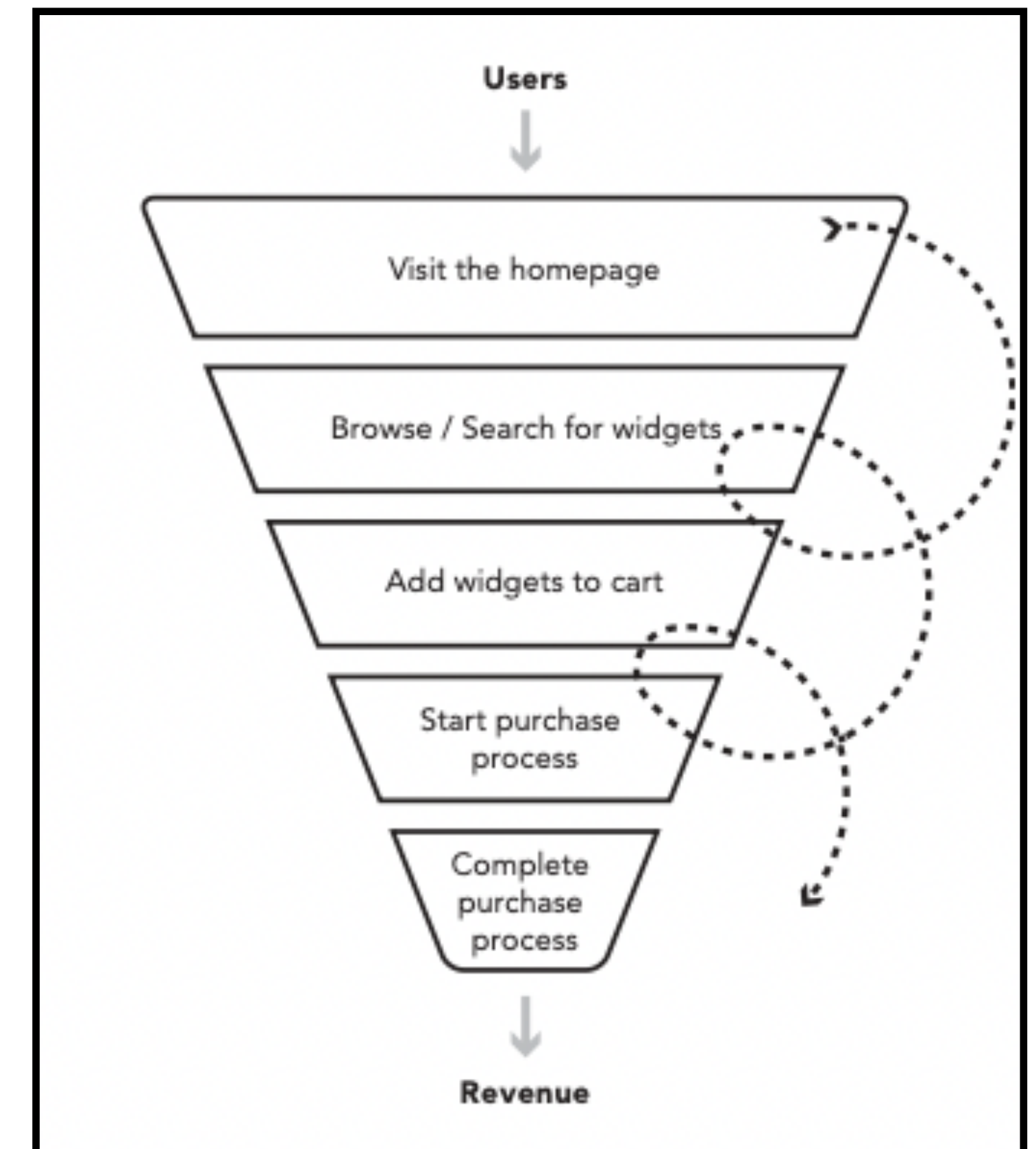
Почему редко смотрят на абсолюты? И нормируют их?

Представим, что у нас есть интернет магазин, в нем пользователи могут покупать мыло. Мы решили, что хотим добавить поле для ввода купона на страницу чекаута

Вопрос: Как правильно нормировать выручку на кол-во юзеров?

Варианты:

1. Все пользователи, которые посещали сайт
2. Только юзеры, которые купили
3. Все, кто приступили к процессу покупки



Почему редко смотрят на абсолюты?

И нормируют их?

1. Все пользователи, которые посещали сайт

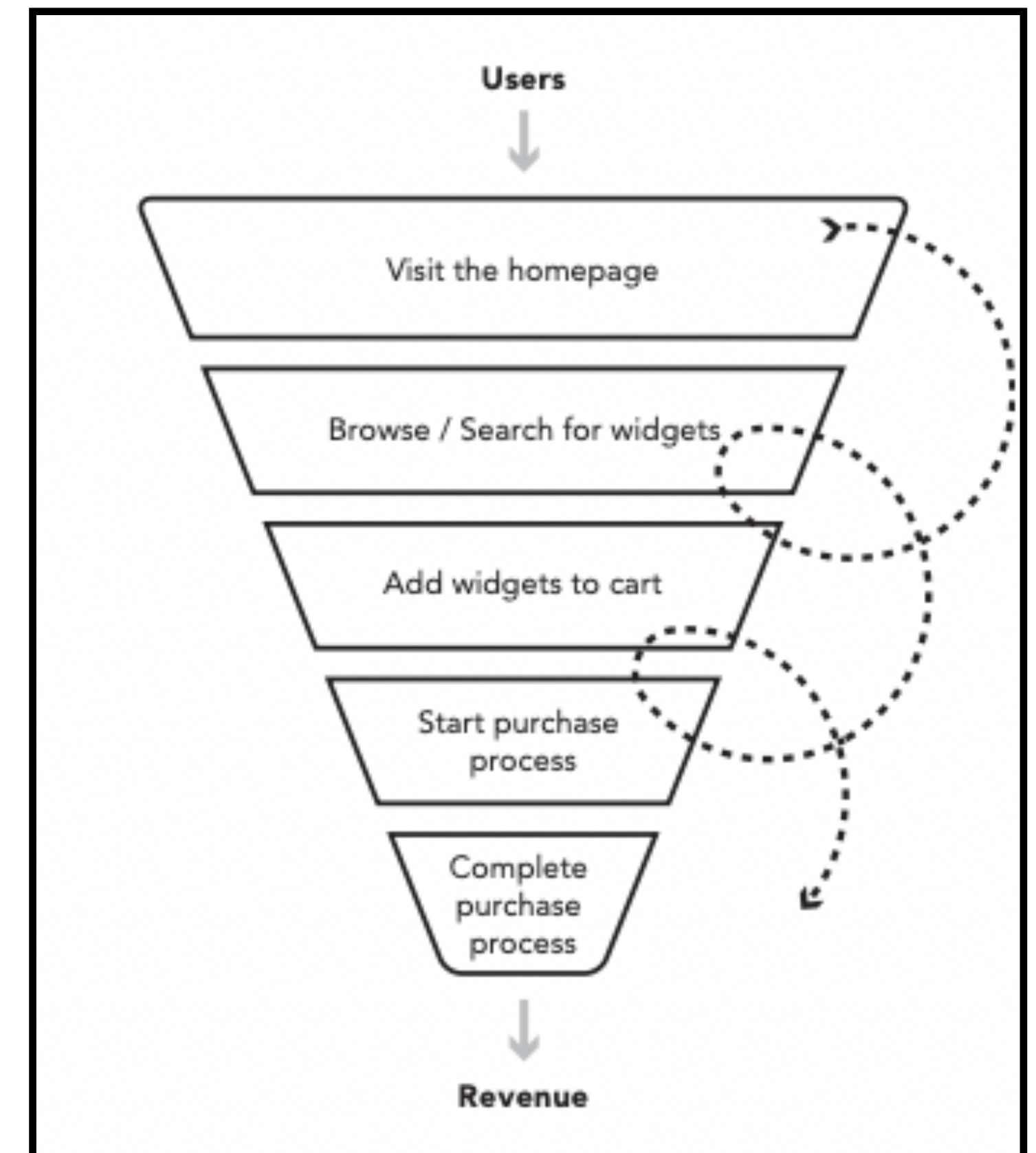
Это приемлемо, хотя и немного усложняется из-за вовлечения пользователей, которые не начинали процесс оформления заказа и куда были внесены изменения. Однако мы уверены, что наши корректировки не окажут влияния на тех, кто не приступал к оформлению заказа.

2. Только юзеры, которые купили

Этот вариант может не подходит, так как он влияет на общую стоимость покупок, а не на долю пользователей, совершающих покупки. Важно учитывать, что при увеличении количества покупателей средний доход на пользователя может сократиться, несмотря на рост общего дохода.

3. Все, кто приступили к процессу покупки

Этот вариант является наиболее предпочтительным, учитывая стадию воронки, на которой происходят изменения. Мы охватываем всех потенциально затронутых пользователей, при этом исключаем тех, кто не вовлечен в процесс (пользователи, не начинавшие оформление заказа), что позволяет нам улучшить результаты.



Итого смотрим на три типа метрик

“Статистическая” классификация

Средняя

- User Average Metrics (ARPU / ARPPU/etc)

Ratio

- User-level Conversion Metrics (Retention / etc)
- Page-level Conversion Metrics (Global CTR / etc)

Квантиль

- Ну тут просто квантиль (.99 latency / перцентиль чека)

Абсолюты

- Метрики (GMV / Выручка / Просмотры)

Какая связь между типами метрик?

Сравниваемся со второй лекцией

Пример метрики продукта

- Retention - user level conversion
- Checkout conversion rate
- ARPU - user level average
- ARPPU - user level average
- LTV - user level average
- Ad CTR - global level conversion

Пример метрики роста

- Выручка - абсолют
- GMV - абсолют
- MAU - абсолют
- WAU - абсолют

Аксессуары для ванной для повышения среднего чека

1. Предпосылка

Аналитика показывает, что средний чек покупки увеличивается на 20%, когда клиенты добавляют в корзину товары из категории "Аксессуары для ванны". Интервью с пользователями выявили, что многие не замечают эту категорию при обычном посещении магазина.

2. Возможность

Ввести функцию персонализированных рекомендаций аксессуаров для ванны на странице оформления заказа мыла.

3. Кого коснется

Покупатели, оформляющие заказ мыла и заинтересованные в улучшении своего опыта ванны.

4. Мотивация

Предложение аксессуаров для ванны на этапе оформления заказа мыла удобно для пользователей, так как позволяет дополнить их покупку, не отвлекаясь на поиск этих товаров отдельно.

5. Эффект, который мы ожидаем

Увеличение среднего чека на 15% за счет добавления аксессуаров для ванны к заказу мыла, что приведет к росту ARPU.

Растим ARPU с помощью рекомендаций аксессуаров

1. Формулировка гипотезы с сформулированным ожидаемым размером эффекта

Предложение комплектов мыла с подборкой популярных ароматов на основе анализа предпочтений покупателей и данных о самых продаваемых товарах увеличит средний доход на пользователя (ARPU) на 20%.

2. Описание аудитории

Покупатели онлайн-магазина мыла, включая как новых, так и возвращающихся пользователей.

3. Описание вариантов с размером каждой группы

Контрольная группа (А): Покупателям предлагается стандартный ассортимент без акцентов на комплекты.

Экспериментальная группа (В): Покупателям активно предлагаются комплекты мыла с популярными ароматами на главной странице и в разделе рекомендаций.

Размер каждой группы составляет 50% от общего числа посетителей в период эксперимента.

4. Ожидаемые исходы и метрики

Основная метрика: Увеличение ARPU в экспериментальной группе по сравнению с контрольной.

Второстепенные метрики: Увеличение среднего размера заказа, конверсии в покупку комплектов.

5. Продолжительность

Эксперимент продлится 4 недели, чтобы собрать достаточно данных для статистически значимых результатов, учитывая недельные колебания трафика и поведения покупателей.

6. Результаты

TBD

Упрощаем ввод адреса для повышения конверсии в заказ

1. Предпосылка

Анализ пользовательского поведения на сайте онлайн-магазина мыла показал, что 30% посетителей отказываются от оформления заказа на этапе ввода адреса доставки. Обратная связь от пользователей указывает на сложности с заполнением формы и недостаточную информацию о способах доставки.

2. Возможность

Упростить процесс оформления заказа, оптимизировав форму ввода адреса и предоставив четкую информацию о способах доставки сразу на странице оформления заказа.

3. Кого коснется

Это коснется всех пользователей, оформляющих заказ в онлайн-магазине мыла, особенно тех, кто ранее сталкивался с трудностями на этапе оформления заказа..

4. Мотивация

Упрощенный процесс оформления заказа сократит время, необходимое для его завершения, и уменьшит количество отказов от покупки, что сделает покупательский опыт более приятным и эффективным.

5. Эффект, который мы ожидаем

Ожидается, что упрощение процесса оформления заказа увеличит конверсию в оформление заказа на 15% за счет снижения числа отказов на этапе ввода адреса доставки.

Увеличиваем релевантность рекламы

Хорошо сформулированная гипотеза:

1. Предпосылка

Анализ данных показывает, что текущие рекламные баннеры на сайте онлайн-магазина мыла имеют низкий CTR (click-through rate), который составляет всего 0.5%. Пользовательские опросы выявили, что многие посетители находят существующие рекламные материалы нерелевантными или недостаточно привлекательными.

2. Возможность

Создать более персонализированные и визуально привлекательные рекламные баннеры, которые будут соответствовать интересам и предпочтениям целевой аудитории, основываясь на их предыдущем поведении на сайте и покупках.

3. Кого коснется

Эта инициатива коснется как новых, так и возвращающихся пользователей сайта, которые ранее взаимодействовали с рекламными баннерами

4. Мотивация

Персонализированные рекламные баннеры предложат пользователям товары и акции, которые наиболее релевантны их интересам и покупательскому поведению, увеличивая вероятность клика по рекламе.

5. Эффект, который мы ожидаем

Ожидается, что внедрение персонализированных и визуально привлекательных рекламных баннеров увеличит CTR по рекламе на сайте на 20%, что приведет к увеличению трафика на рекламируемые страницы и потенциальному росту продаж

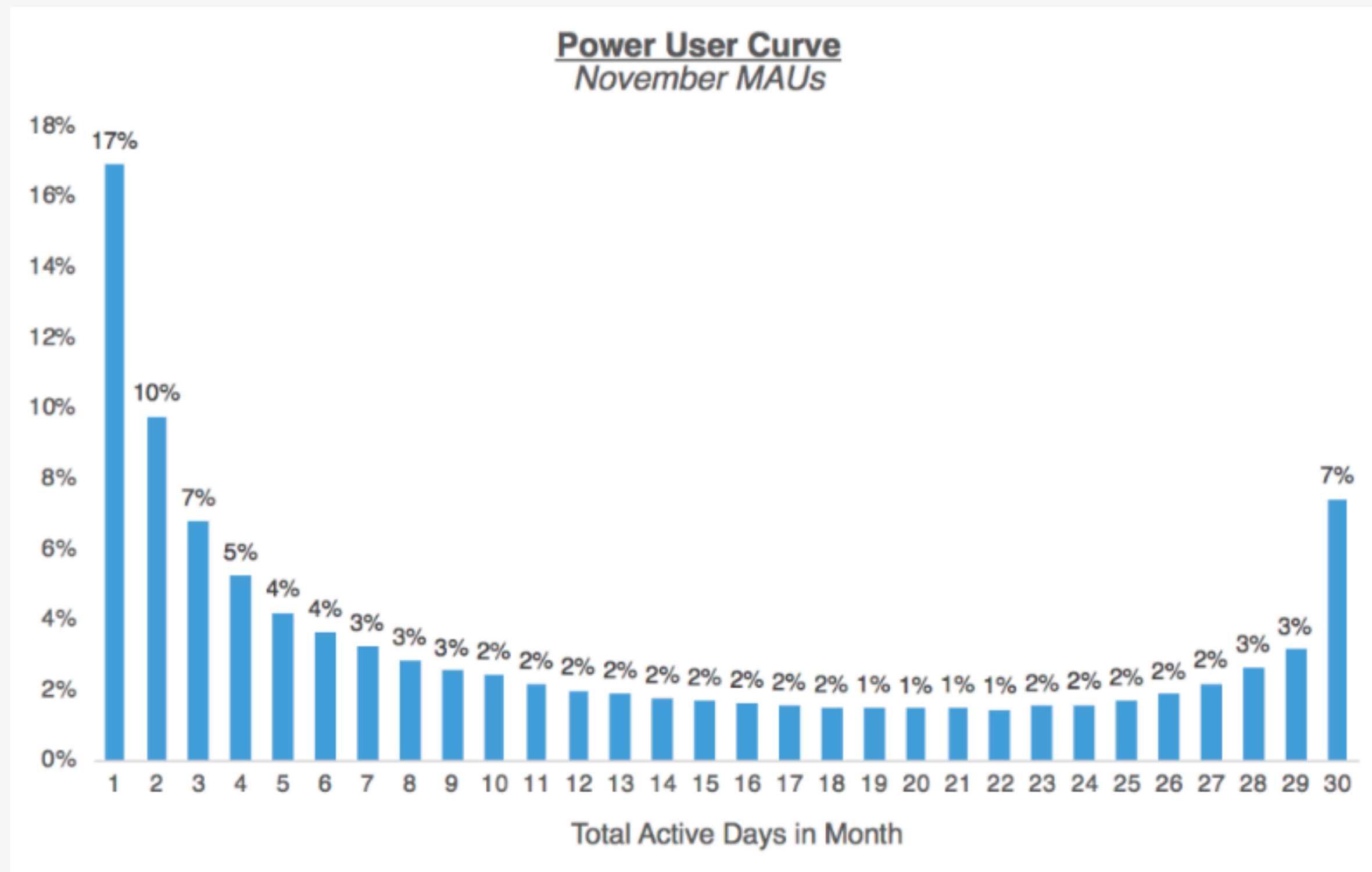
Как думаете для чего тестируют квантиль?

Вопрос аудитории

А если взять такую метрику как сумма заказа?

Вопрос аудитории

Теория power users



Power users - people who love their product, are highly engaged, and contribute a ton of value to the network. In ecommerce marketplaces it's power sellers, in ridesharing platforms it's power riders, and in social networks it's influencers.

Andrew Chen, author of The Cold Start Problem

Эксклюзив для самых лояльных

Хорошо сформулированная гипотеза:

1. Предпосылка

Аналитические данные онлайн-магазина мыла показывают, что 75-й квантиль суммы заказов составляет \$50. Опросы лояльных клиентов выявили интерес к эксклюзивным продуктам и персонализированным предложениям. Возможность Создать более персонализированные и визуально привлекательные рекламные баннеры, которые будут соответствовать интересам и предпочтениям целевой аудитории, основываясь на их предыдущем поведении на сайте и покупках.

2. Кого коснется

Программа коснется лояльных клиентов магазина, которые совершают покупки регулярно и имеют историю покупок выше среднего.

3. Мотивация

Лояльные клиенты получают доступ к эксклюзивным продуктам и персонализированным предложениям, что сделает их покупательский опыт более индивидуализированным и премиальным.

4. Эффект, который мы ожидаем

Ожидается, что внедрение программы лояльности увеличит 75-й квантиль суммы заказов с \$50 до \$60, улучшая показатели среднего чека среди лояльной аудитории.

Какой тест обычно применяется для средних?

Вопрос аудитории

Расчет размера выборки

Формулируем нулевую гипотезу

Общий вид:

$$H_0 : \theta \in \Theta$$

$$H_1 : \theta \notin \Theta$$

Для средних:

$$H_0 : \theta = \mu_C - \mu_T = 0$$

$$H_1 : \theta = \mu_C - \mu_T \neq 0 = \Delta$$

Расчет размера выборки

Формулируем нулевую гипотезу

Пусть есть тест статистика T нашего теста:

$$T = \frac{\overline{X}_C - \overline{X}_T}{\sqrt{\frac{s_C^2}{n_1} + \frac{s_T^2}{n_2}}} \sim F(x), t(2n - 2)$$

Для простоты предположим, что $n_1 = n_2$ и у выборок одинаковые дисперсии:

$$T = \frac{\overline{X}_C - \overline{X}_T}{\sqrt{\frac{2s^2}{n}}} = \frac{\bar{\theta}}{\sqrt{\frac{2s^2}{n}}}$$

Расчет размера выборки

Что такое мощность?

Мы во время теста хотим контролировать мощность:

$$\beta \geq P(\text{принять } H_0, \text{ если } H_1 \text{ верна})$$

Перепишем это на определение:

$$\beta \geq P(|T| < F^{-1}(1 - \frac{\alpha}{2}) \mid H_1 \text{ is true})$$

Усложняем:

$$\beta \geq P(-F^{-1}(1 - \frac{\alpha}{2}) < \frac{\bar{\theta}}{\sqrt{\frac{2s^2}{n}}} < F^{-1}(1 - \frac{\alpha}{2}) \mid H_1 \text{ is true})$$

Расчет размера выборки

Какое распределение у T?

Если альтернативная гипотеза верна, то:

$$T = \frac{\bar{\theta}}{\sqrt{\frac{2s^2}{n}}} \approx t(2n - 2)$$

Это так, потому что мы предполагаем, что в разнице средних присутствует некоторое смещение Δ , но его можно устранить:

$$\frac{\bar{\theta} - \Delta}{\sqrt{\frac{2s^2}{n}}} \sim t(2n - 2)$$

Воспользуемся любым трюком математиков, добавим и вычтем Δ

Расчет размера выборки

Добавляем и вычитаем дельту

Простите на прошлый слайд не поместилось:

$$\beta \geq P\left(-F^{-1}\left(1 - \frac{\alpha}{2}\right) < \frac{\bar{\theta} + \Delta - \Delta}{\sqrt{\frac{2s^2}{n}}} < F^{-1}\left(1 - \frac{\alpha}{2}\right)\right)$$

Давайте оставим по середине только, то что распределено по Стьюденту

$$\beta \geq P\left(-F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}} < \frac{\bar{\theta} - \Delta}{\sqrt{\frac{2s^2}{n}}} < F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right)$$

Расчет размера выборки

Внимательно смотрим на формулу

Штука посередине у нас распределена по Стьюденту, давайте распишем неравенство в виде двух функций распределения:

$$\beta \geq F\left(F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right) - F\left(-F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right)$$

А как это упростить?

Куда на графике попадет $-F^{-1}\left(\frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}$?

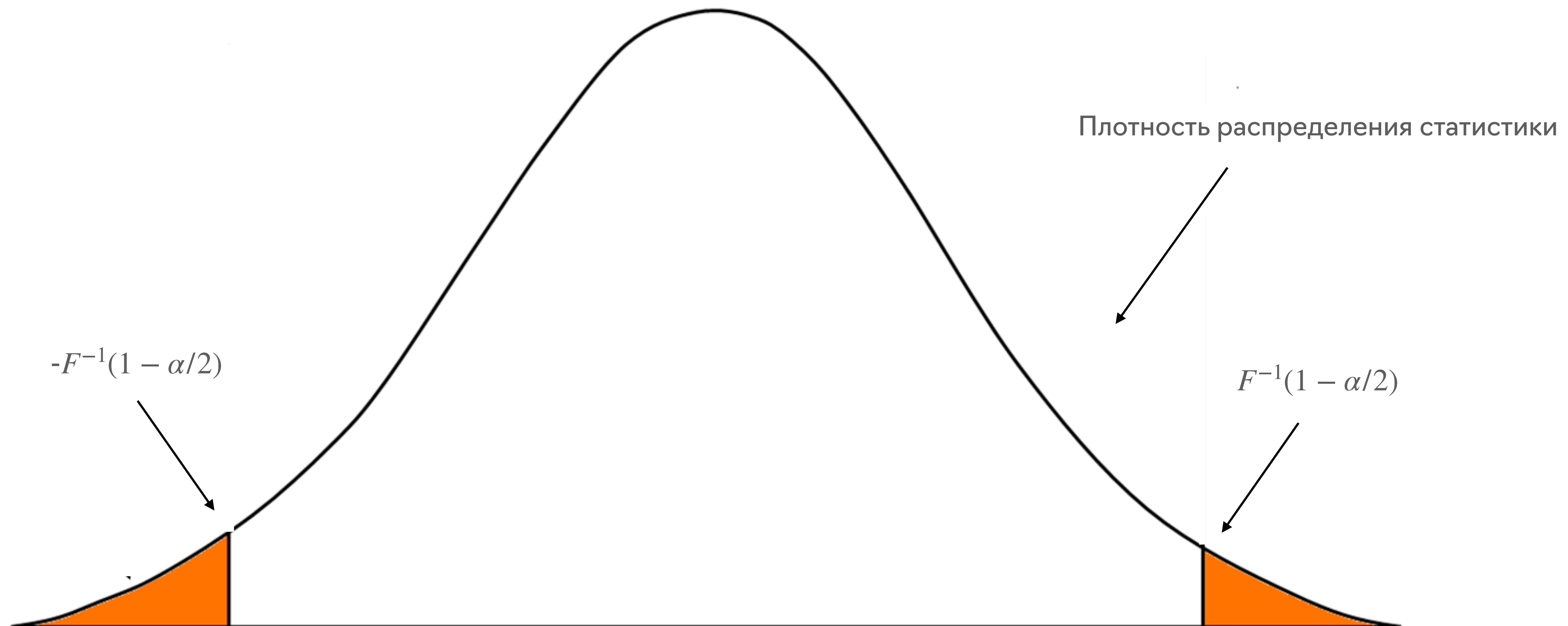
При условии $\Delta > 0$

Вопрос аудитории

Вспомним наши стулья с предыдущих лекций

Отметим необходимые значения из формулы:

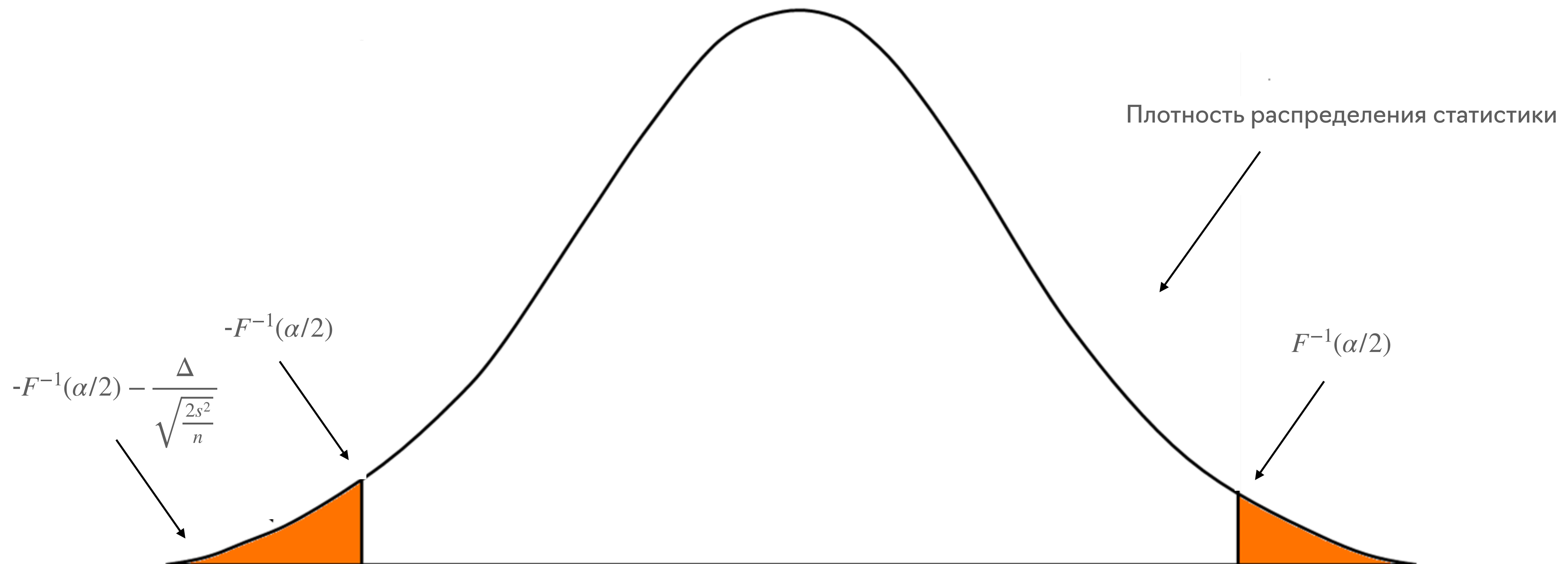
$$\beta \geq F\left(F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right) - F\left(-F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right)$$



Вспомним наши стулья с предыдущих лекций

Отметим необходимые значения из формулы:

$$\beta \geq F\left(F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right) - F\left(-F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right)$$



**Может ли Δ быть одновременно больше и
и меньше нуля?**

Вопрос аудитории

Расчет размера выборки

Δ принимает только один знак

Из-за этого мы можем предположить, что одно из слагаемых очень близко к нулю и записать:

$$\beta \geq F\left(F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}\right)$$

Упрощаем применяя обратную функцию F^{-1} :

$$F^{-1}(\beta) \geq F^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\Delta}{\sqrt{\frac{2s^2}{n}}}$$

Расчет размера выборки

Ну все, можем выразить n

Выражаем n:

$$\frac{\Delta}{\sqrt{\frac{2s^2}{n}}} \geq (F^{-1}(1 - \frac{\alpha}{2}) - F^{-1}(\beta))$$

Упрощаем еще:

$$\frac{\sqrt{n}}{\sqrt{2s^2}} \geq \frac{(F^{-1}(1 - \frac{\alpha}{2}) - F^{-1}(\beta))}{\Delta}$$

И еще:

$$n \geq \frac{2(F^{-1}(1 - \frac{\alpha}{2}) - F^{-1}(\beta))^2 s^2}{\Delta^2}$$

Будут ли полностью одинаковыми два стандартных нормальных распределения из μ и σ ?

Вопрос аудитории

MDE - минимально-детектируемый эффект

MDE - значение эффекта, меньше которого мы предполагаем, что выборки равны.

Тогда гипотезу можно сформулировать так:

$$H_0 : \theta = |\mu_C - \mu_T| \leq MDE$$
$$H_1 : \theta = |\mu_C - \mu_T| \geq MDE = \Delta$$

Теоретическая формула для размера выборки

С учетом MDE

Так как MDE всегда меньше или равна Δ , то

$$n \geq \frac{2(F^{-1}(1 - \frac{\alpha}{2}) - F^{-1}(\beta))^2 s^2}{MDE^2}$$

**Как думаете единственный ли это способ
посчитать размер эффекта?**

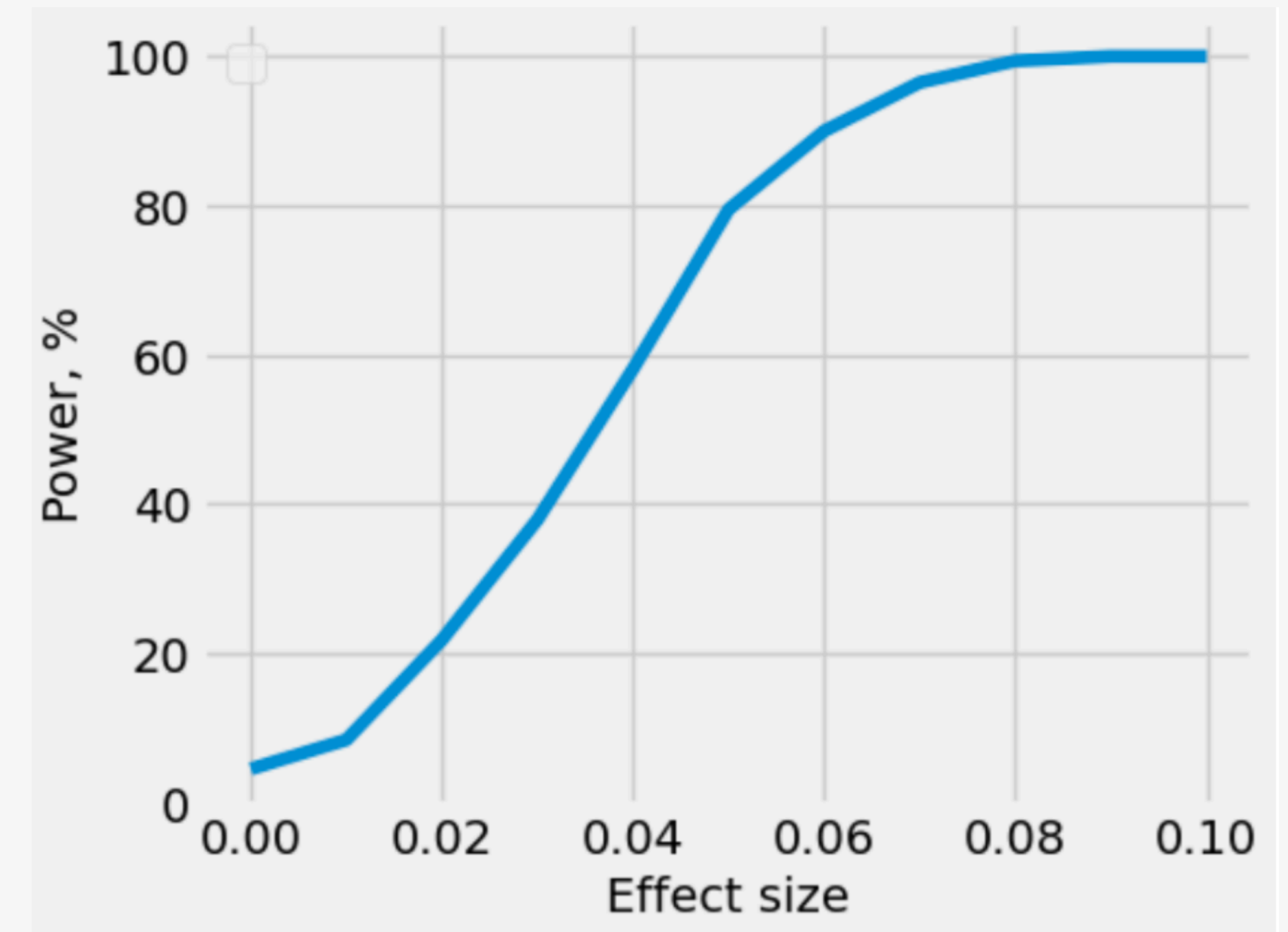
Вопрос аудитории

Reminder: Синтетический A/B тест

Способ смоделировать реальный эксперимент

Алгоритм:

1. Берем две группы из сплит системы
2. Генерируем шум, чтобы добавить одной из групп
3. Считаем метрики, направляем тест
4. Повторяем много-много раз с разными значениями среднего у шума (это будет размер эффекта)
5. Получаем картинку как справа

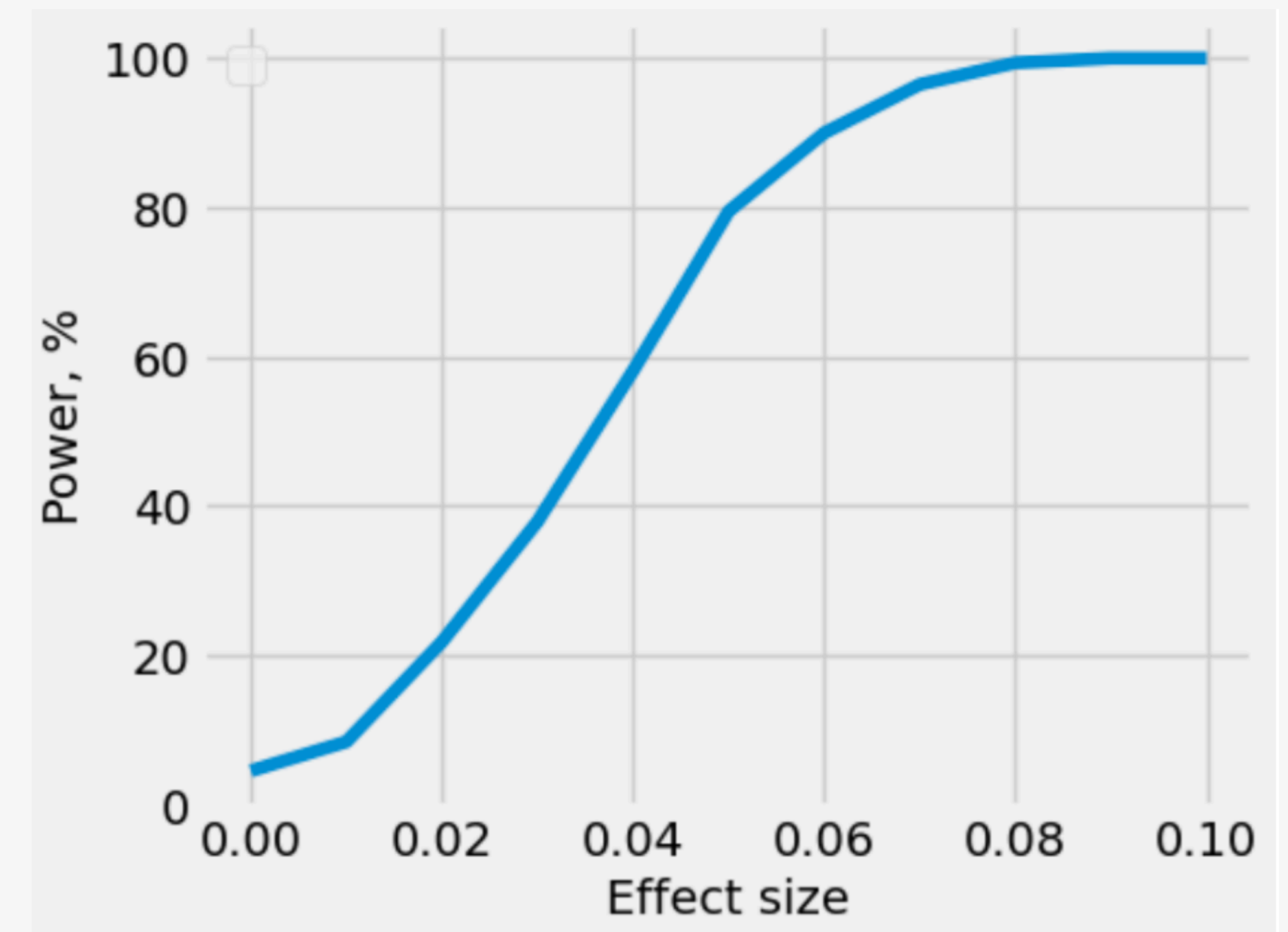


Экспериментальный расчет размера выборки

Способ смоделировать реальный эксперимент

Алгоритм:

1. Берем две группы из сплит системы размера N
2. Генерируем шум, чтобы добавить одной из групп
3. Считаем метрики, направляем тест
4. Повторяем много-много раз с разными значениями среднего у шума (это будет размер эффекта)
5. Получаем картинку как справа для фиксированного N
6. Повторяем для следующего N



Экспериментальный расчет размера выборки

Итоговый результат такого способа

Number of units	MDE	N days
100	1 %	16
100	2 %	4
100	3 %	2
...		
...		
1000	3 %	X

Получаем таблицу, которой можно пользоваться для любого случая жизни и планирования эксперимента. Заметьте, что размер выборки растет квадратично!

Экспериментальный способ работает для всех типа метрик

Средняя

- User Average Metrics (ARPU / ARPPU/etc)

Ratio

- User-level Conversion Metrics (Retention / etc)
- Page-level Conversion Metrics (Global CTR / etc)

Квантиль

- Ну тут просто квантиль (.99 latency / перцентиль чека)

Абсолюты

- Метрики (GMV / Выручка / Просмотры)

Итого

1. Узнали, к чему сводится выбор метрик
2. Посмотрели на разные дизайны экспериментов
3. Разобрались как теоретическим и экспериментальным способом посчитать размер выборки
4. Чуть не умерли от формул (продолжение дальше, когда будем работать с остальными метриками)