

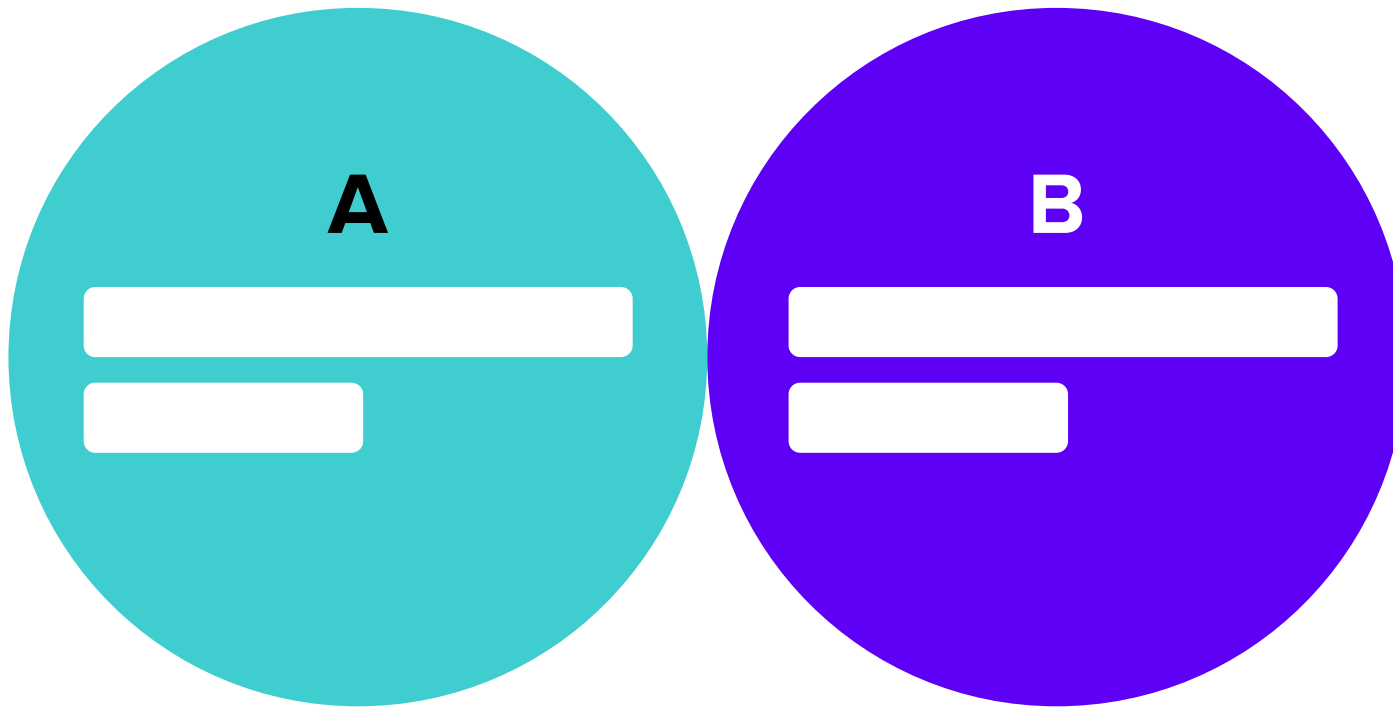
The background of the slide is an aerial photograph of a dense city skyline, likely New York City, with numerous skyscrapers. The image is overlaid with a semi-transparent blue layer. A network of thin, light blue lines connects various points across the blue area, creating a web-like pattern. The text «А/В-тестирование» is centered in white, with a thin white horizontal line underneath it.

«А/В-тестирование»

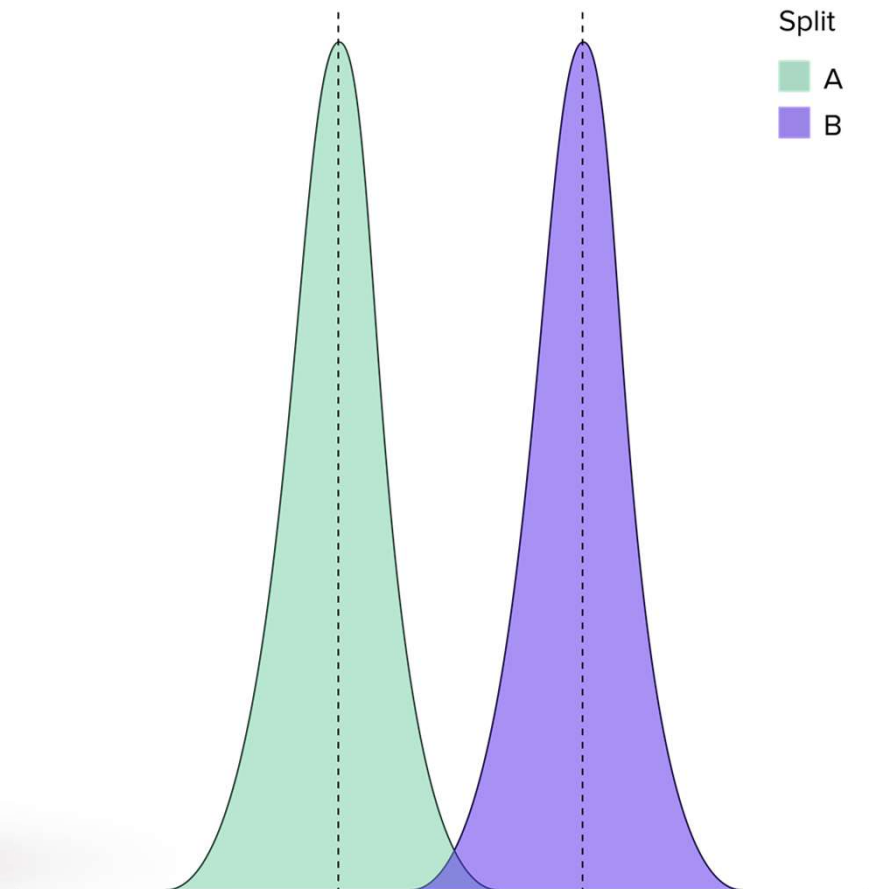
Вопросы

- 1 A/B тестирование
- 2 Статистическая проверка гипотез для связанных выборок
- 3 Дисперсионный анализ
- 4 Проблема множественного сравнения

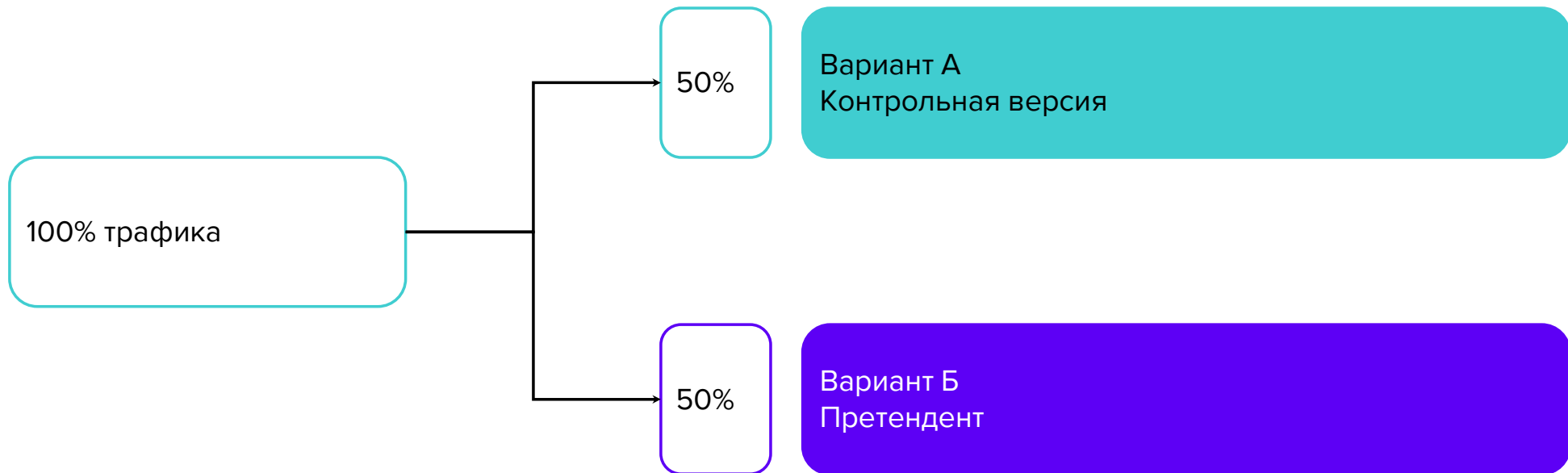
**Допустим у нас есть предположение,
что, если поменять цвет сайта,
конверсия увеличится.
Как проверить?**



Другие примеры



А/В тестирование

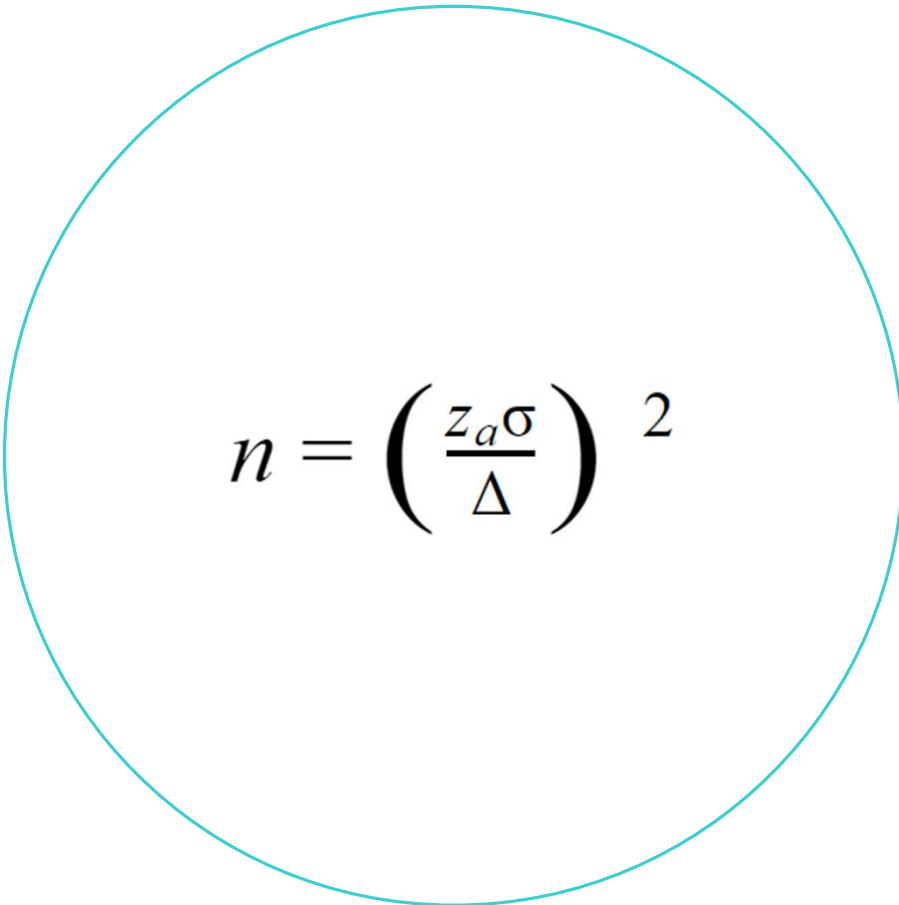


Требования к тестам

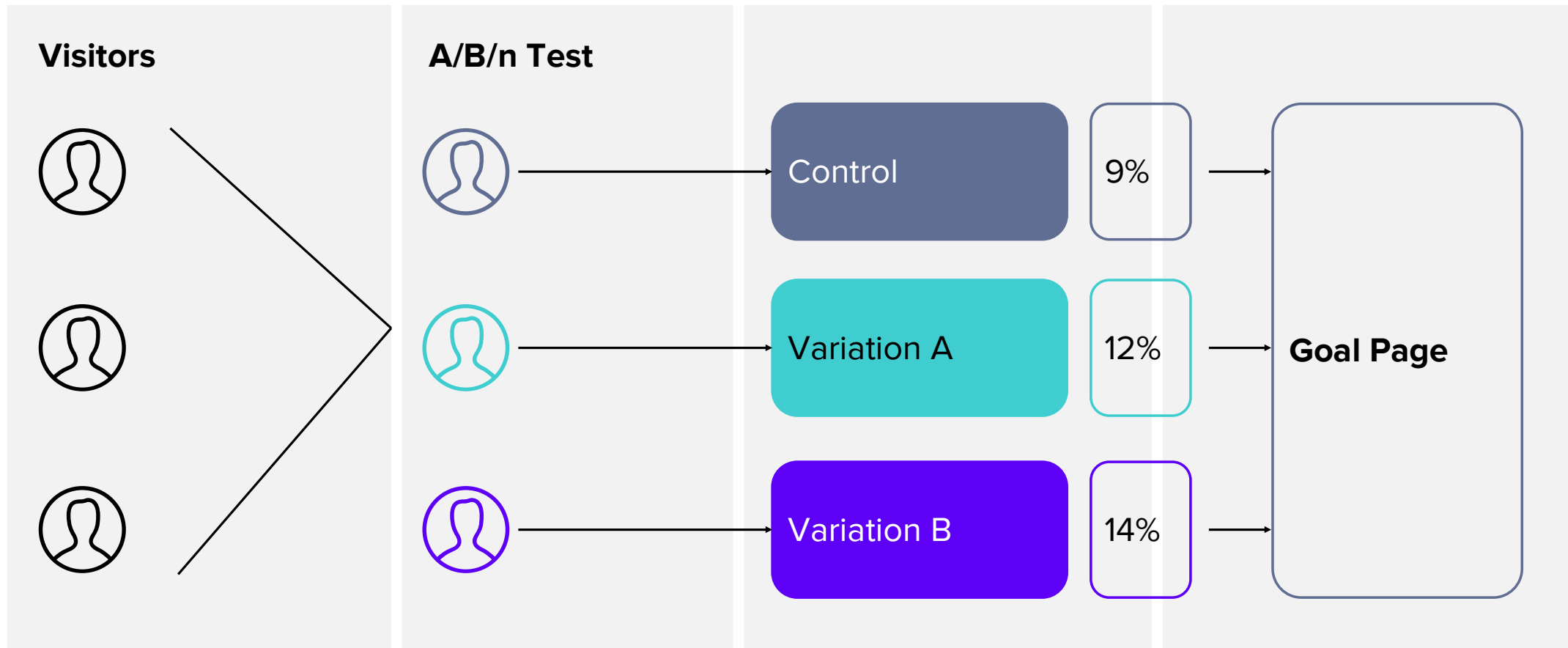
- Понимание метрики/цели
- Одновременность
- Случайность
- Достаточный объём выборки
- Независимость

Требования к тестам

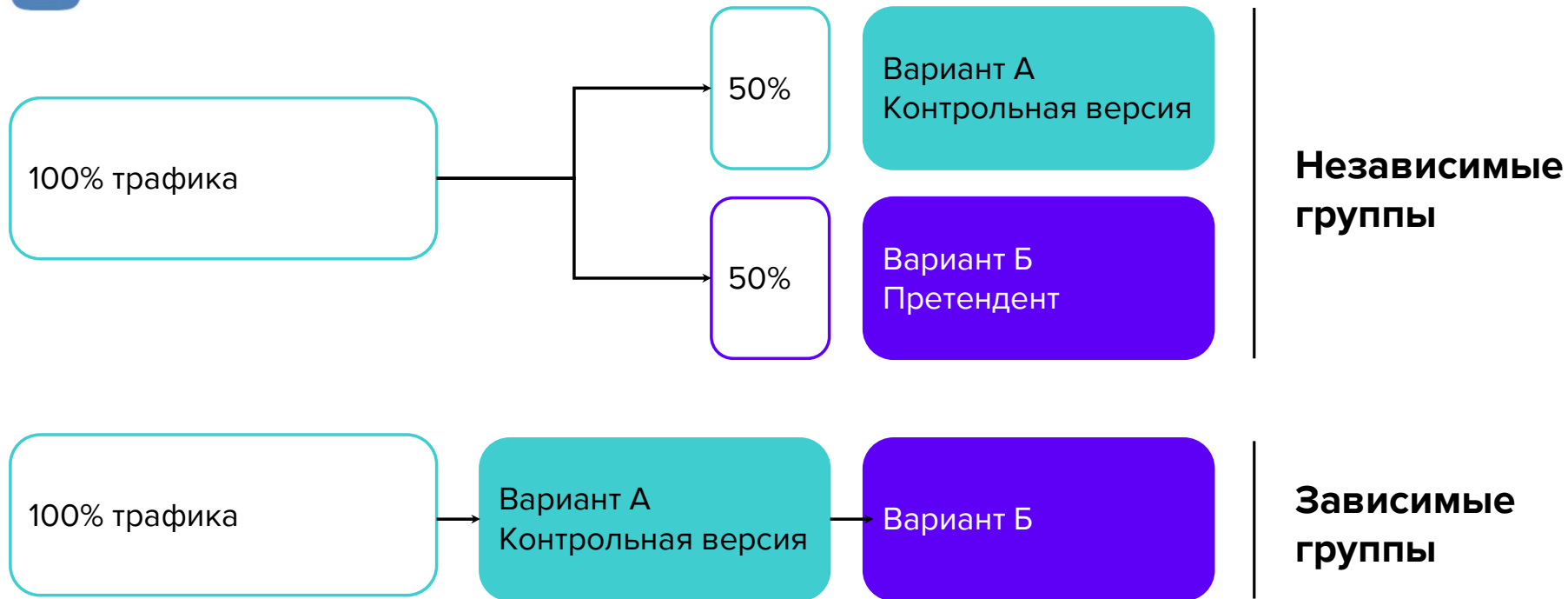
- Понимание метрики/цели
- Одновременность
- Случайность
- Достаточный объём выборки
- Независимость
- и тд (Похожесть/Стратификация)


$$n = \left(\frac{z_{\alpha} \sigma}{\Delta} \right)^2$$

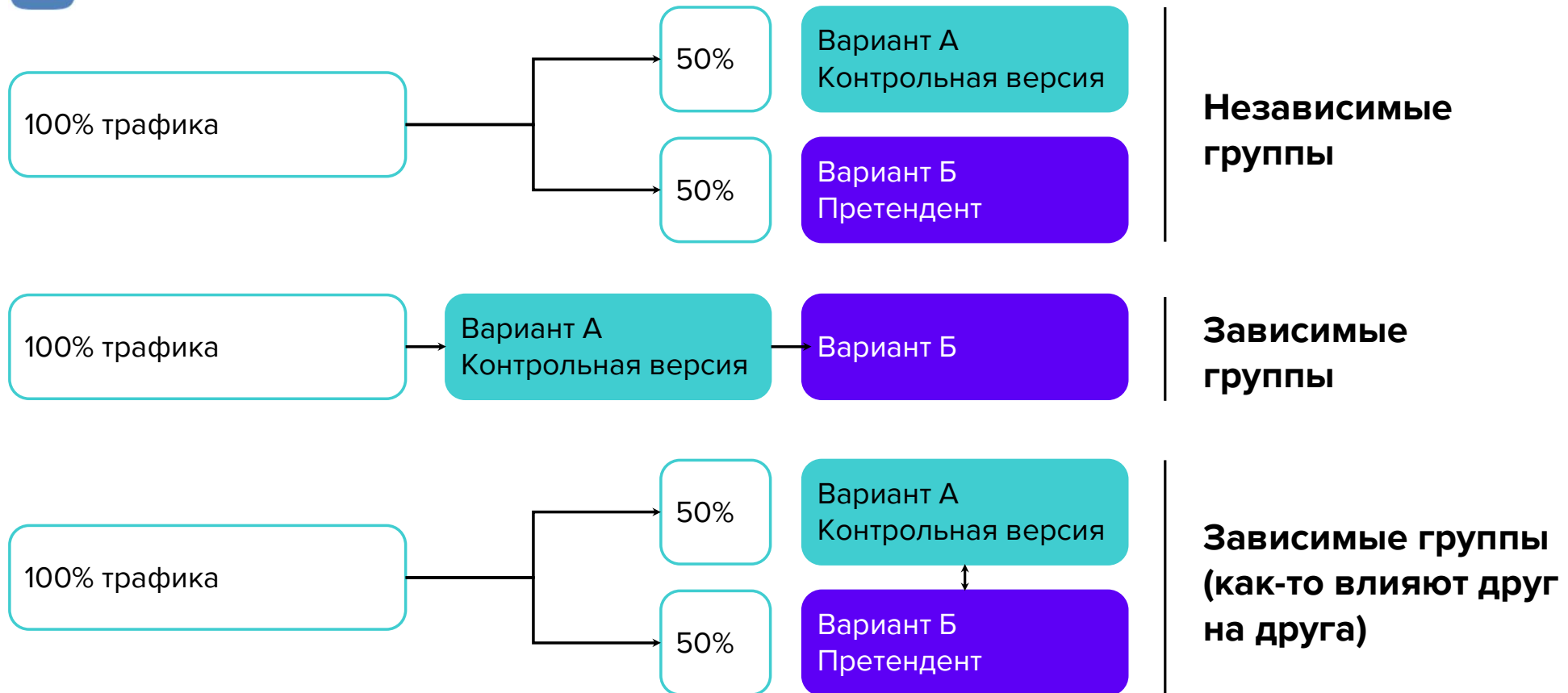
Multi A/B тесты и как их проводить



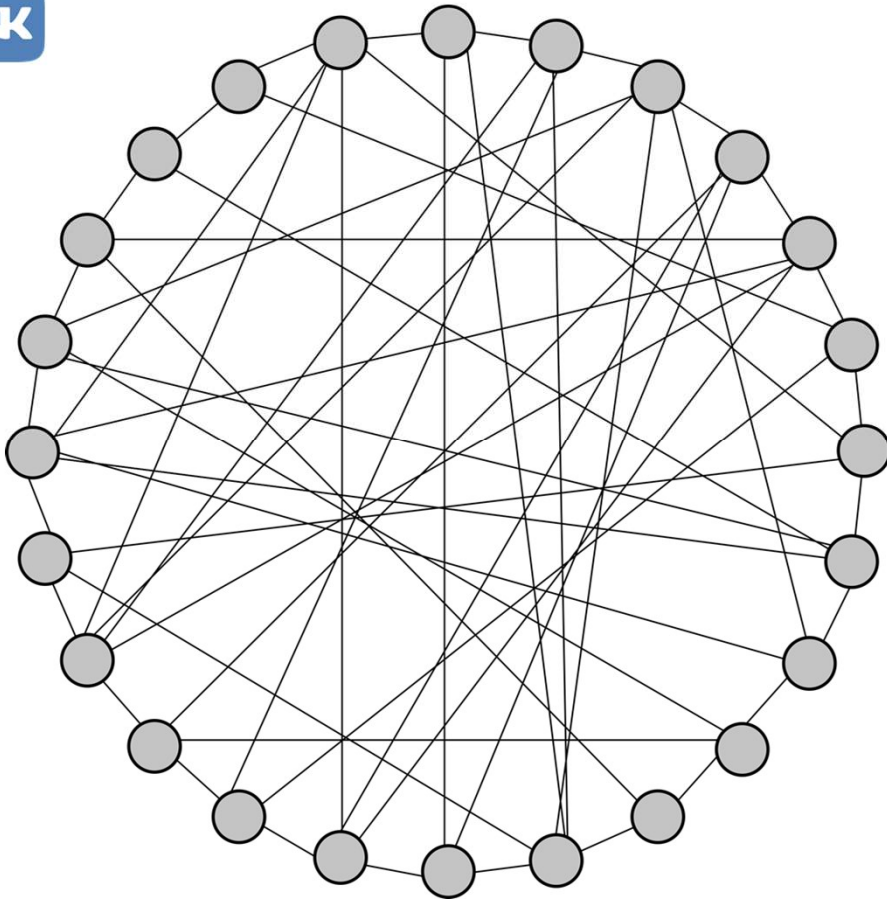
А/В тесты при зависимых группах



А/В тесты при зависимых группах

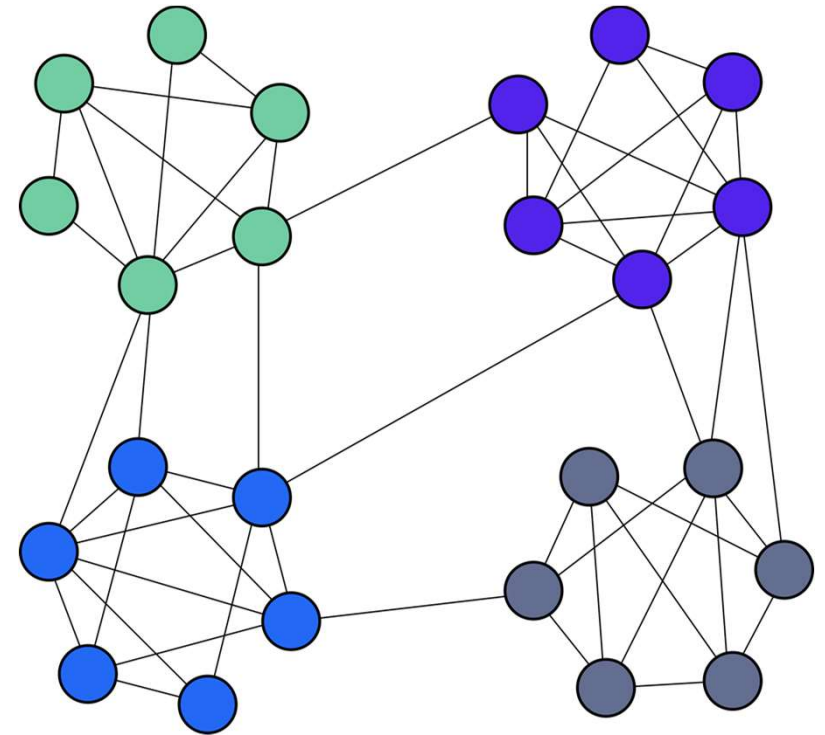


Multi A/B тесты при зависимых группах



=

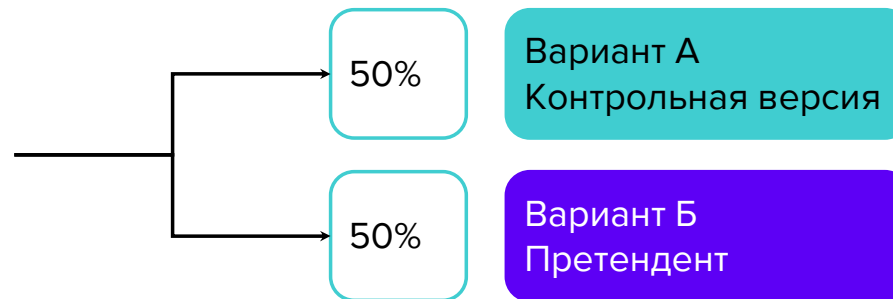
Группы людей



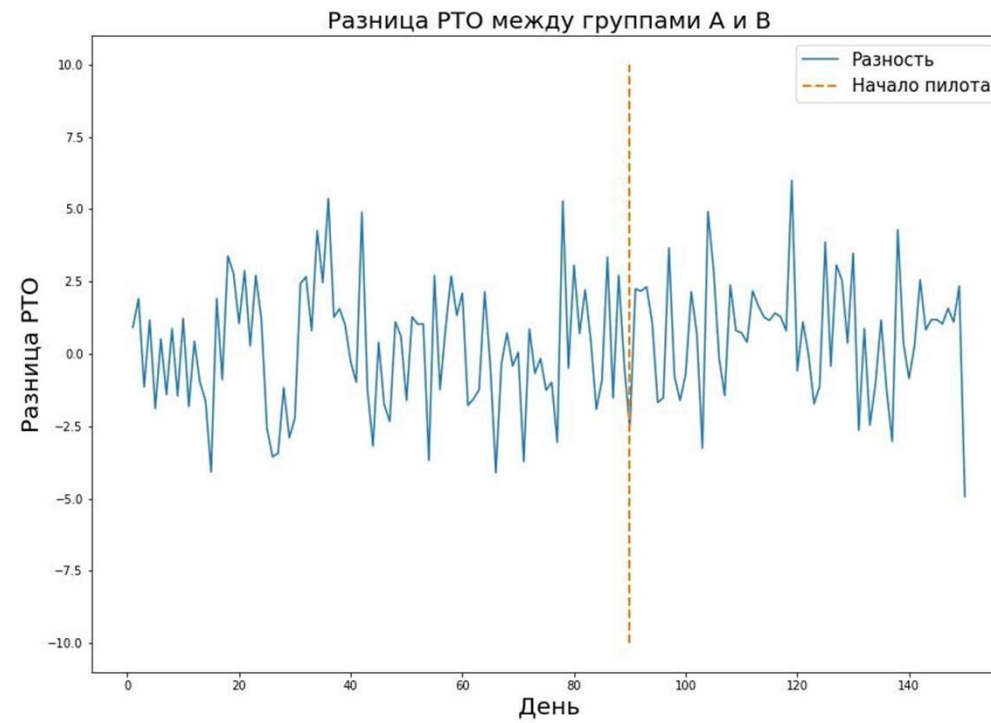
Multi A/B тесты и похожие группы



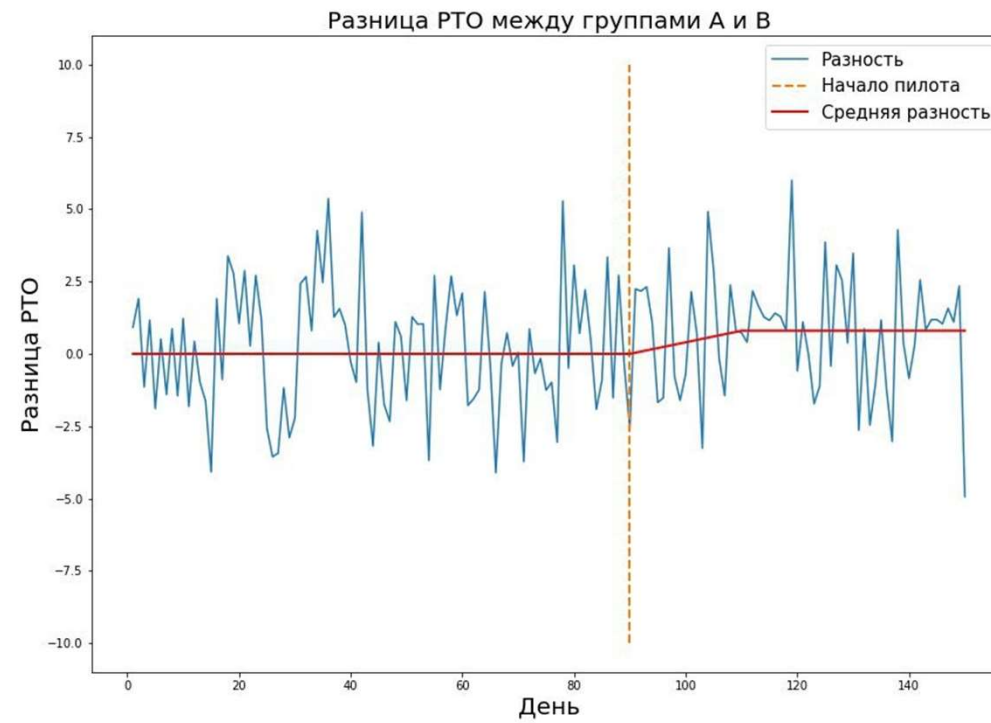
Как делить разные магазины на группы
чтоб группы были похожи ?!?!?



Multi A/B тесты и похожие группы



Multi A/B тесты и похожие группы



С А/В тестами ты один в поле воин!

- В 2012 году сотрудник Microsoft, работавший над поисковой системой Bing, провёл эксперимент по тестированию различных способов отображения рекламных заголовков.

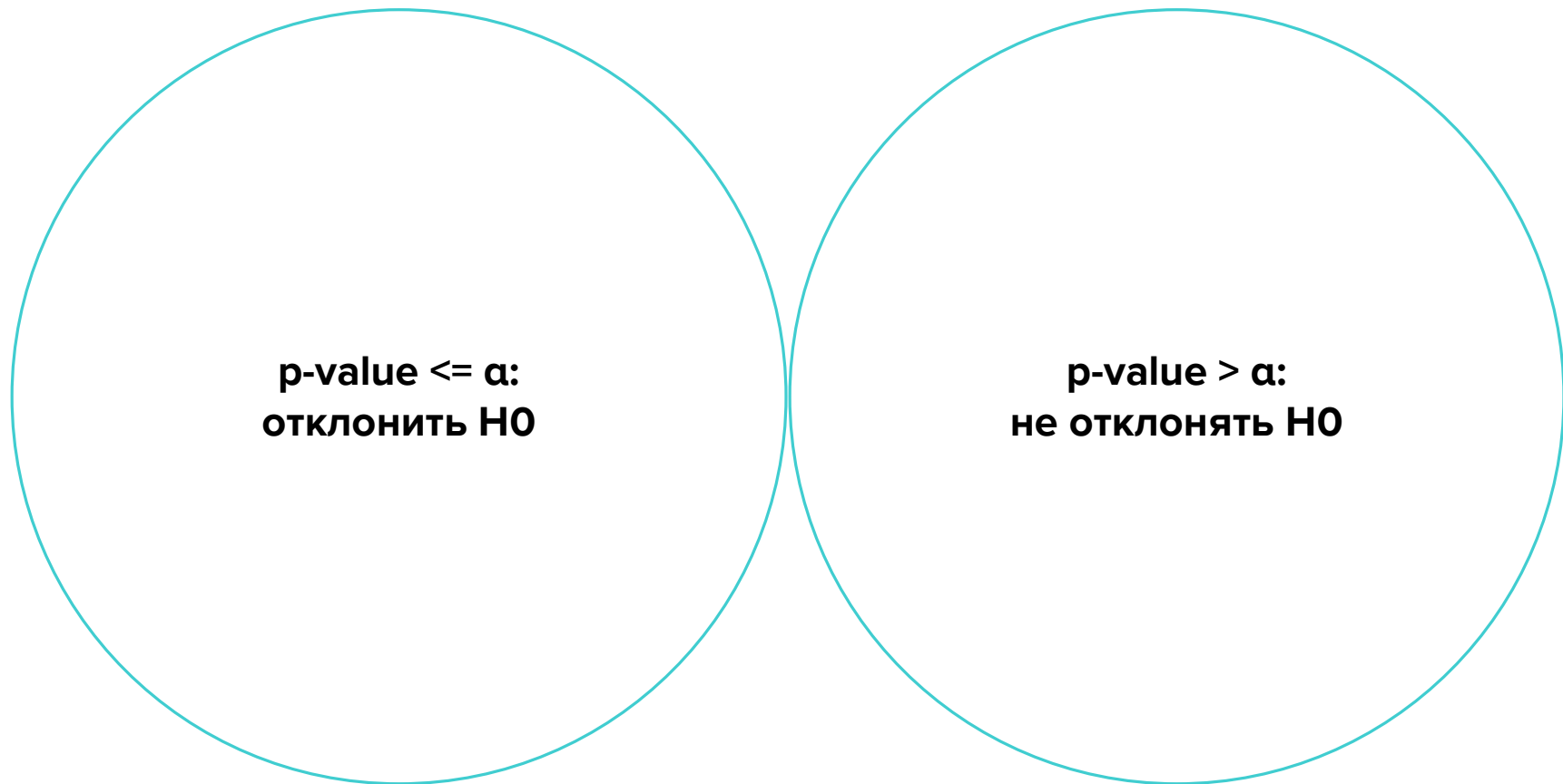
В течение нескольких часов альтернативный формат привёл к увеличению доходов на 12% без влияния на показатели взаимодействия с пользователем.

- По итогам 2012 года выручка Microsoft составила \$74 млрд, чистая прибыль — около \$17 млрд.



Прирост чистой
прибыли \$2,04 млрд —
круто?

Статистическая проверка гипотез



Статистическая проверка гипотез

Что такое гипотеза и что такое p-value?

Понятие p-value

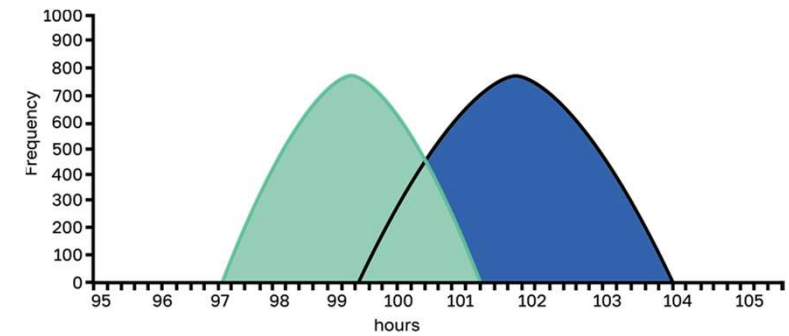
Значение статистики критерия, полученное из выборки, связывают с уже известным распределением, которому она подчиняется, чтобы получить значение p , площадь обоих «хвостов» (или одного «хвоста» в случае односторонней гипотезы) распределения вероятности.

Большинство компьютерных пакетов обеспечивают автоматическое вычисление двустороннего значения p (Python).

Определение:

Значение p — это вероятность получения нашего вычисленного значения или его ещё большего значения, если нулевая гипотеза верна.

Иными словами, p — это вероятность отвергнуть нулевую гипотезу при условии, что она верна.



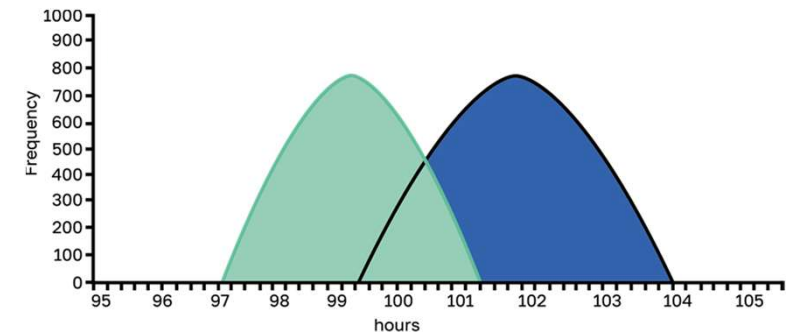
Применение значения P-value

Как по p-value определить, есть ли основания отвергнуть нулевую гипотезу? Тут важно сначала зафиксировать уровень значимости α , а потом уже делать выводы.

Уровень значимости α — это вероятность отвергнуть нулевую гипотезу при условии, что она верна.

p-value — это минимальный уровень значимости, на котором нулевая гипотеза может быть отвергнута.

Соответственно, если **p-value** меньше нашего фиксированного **уровня значимости α** , на котором мы проверяем гипотезу, то **нулевую гипотезу** следует отвергнуть, если больше — оснований отвергать нулевую гипотезу нет.



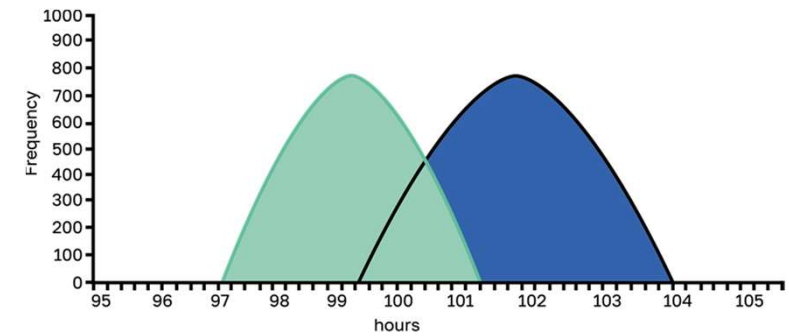
Применение значения P-value - это просто!

Как по p-value определить, есть ли основания отвергнуть нулевую гипотезу? Тут важно сначала зафиксировать уровень значимости α , а потом уже делать выводы.

Уровень значимости α — это вероятность отвергнуть нулевую гипотезу при условии, что она верна.

p-value — это минимальный уровень значимости, на котором нулевая гипотеза может быть отвергнута.

Соответственно, если **p-value** меньше нашего фиксированного **уровня значимости α** , на котором мы проверяем гипотезу, то **нулевую гипотезу** следует отвергнуть, если больше — отвергать нулевую гипотезу нельзя (**и принимать тоже**).



НЕПРАВИЛЬНЫЕ, но **понятные**
определения **p-value**

p-value — это вероятность, получить различия случайно

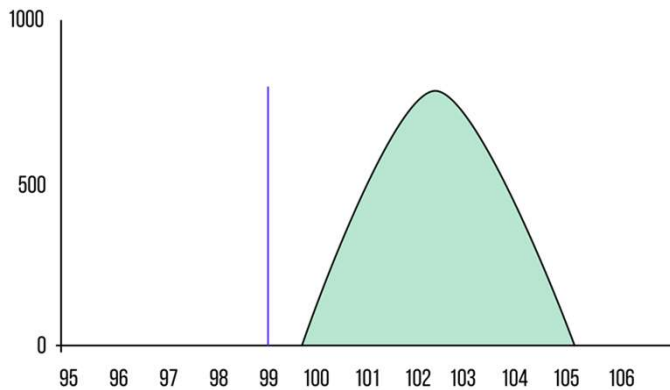
p-value — это вероятность ошибки

НЕ говорите никому, что я вас такому учил =)

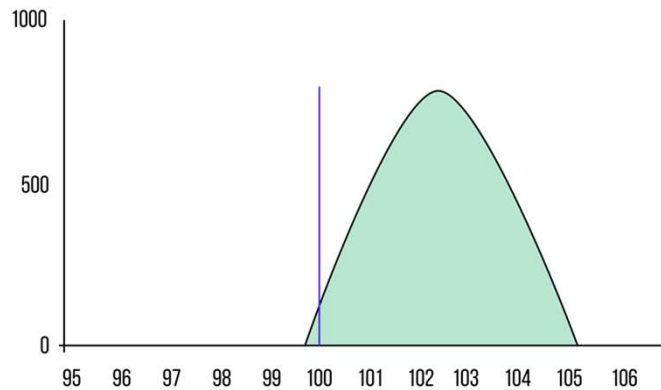
Статистическая проверка гипотез

- $p\text{-value} \leq \alpha$: отклонить H_0
- $p\text{-value} > \alpha$: не отклонять H_0

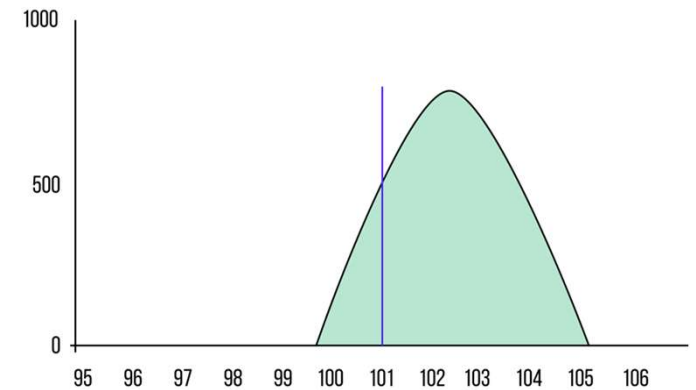
$p\text{-value} < 5\%$



$p\text{-value} = 5\%$



$p\text{-value} > 5\%$

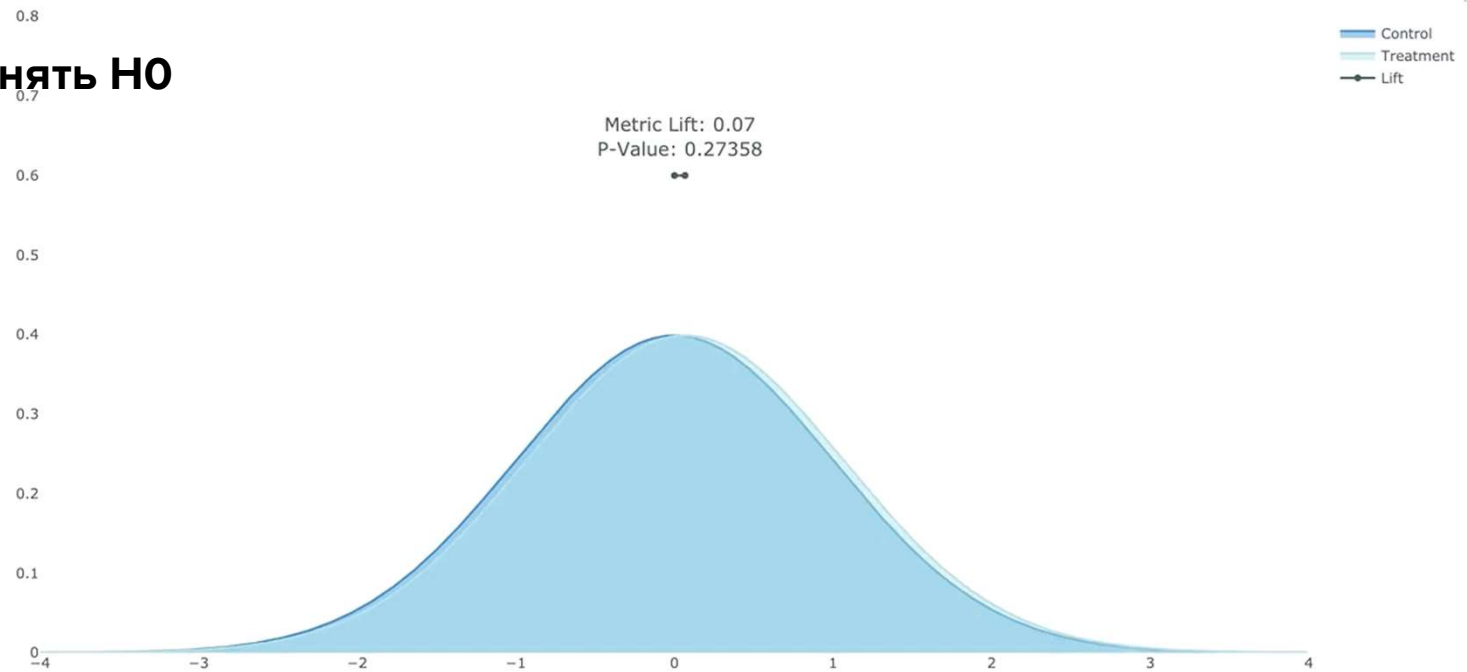


■ Препарат ■ Пустышка

Статистическая проверка гипотез

- $p\text{-value} \leq \alpha$: отклонить H_0

- $p\text{-value} > \alpha$: не отклонять H_0



■ Препарат ■ Пустышка

Статистическая проверка гипотез

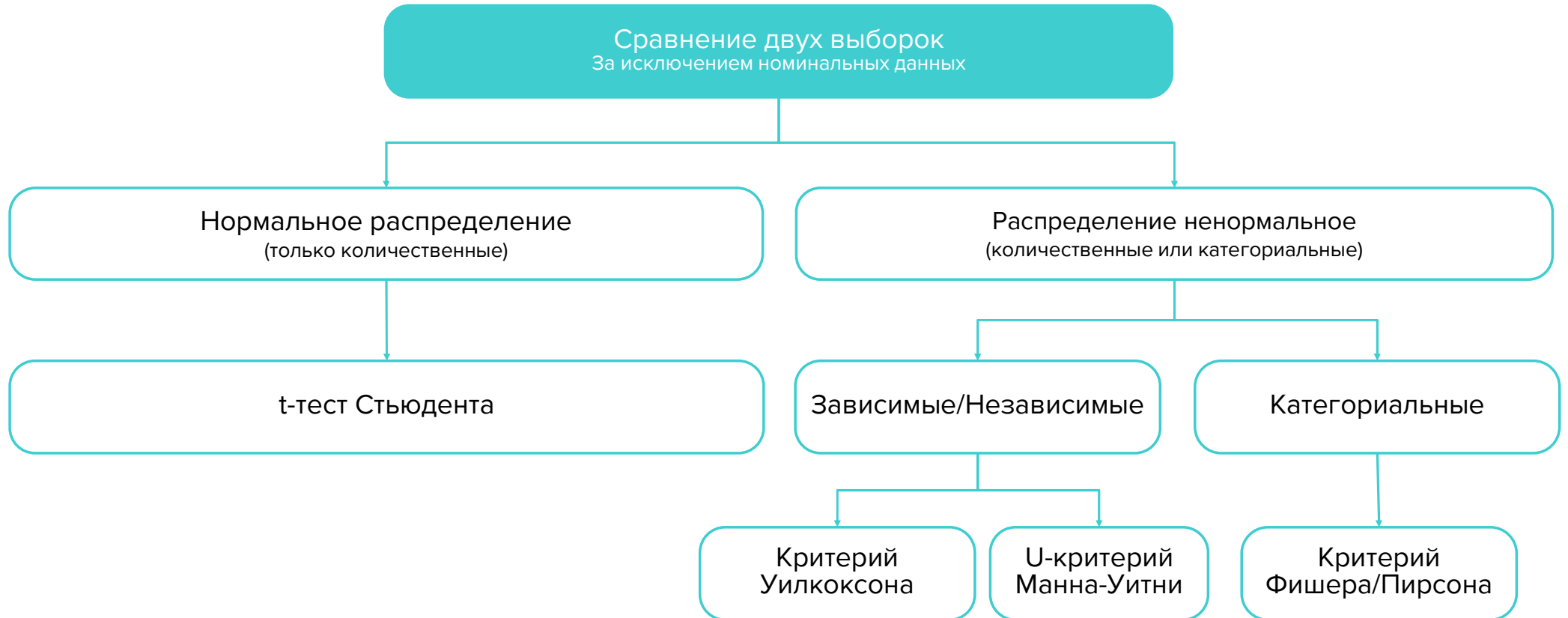
- $p\text{-value} \leq \alpha$: отклонить H_0
- $p\text{-value} > \alpha$: не отклонять H_0



Как получить $p\text{-value}$?



Карта тестов гипотез



t-критерий Стьюдента

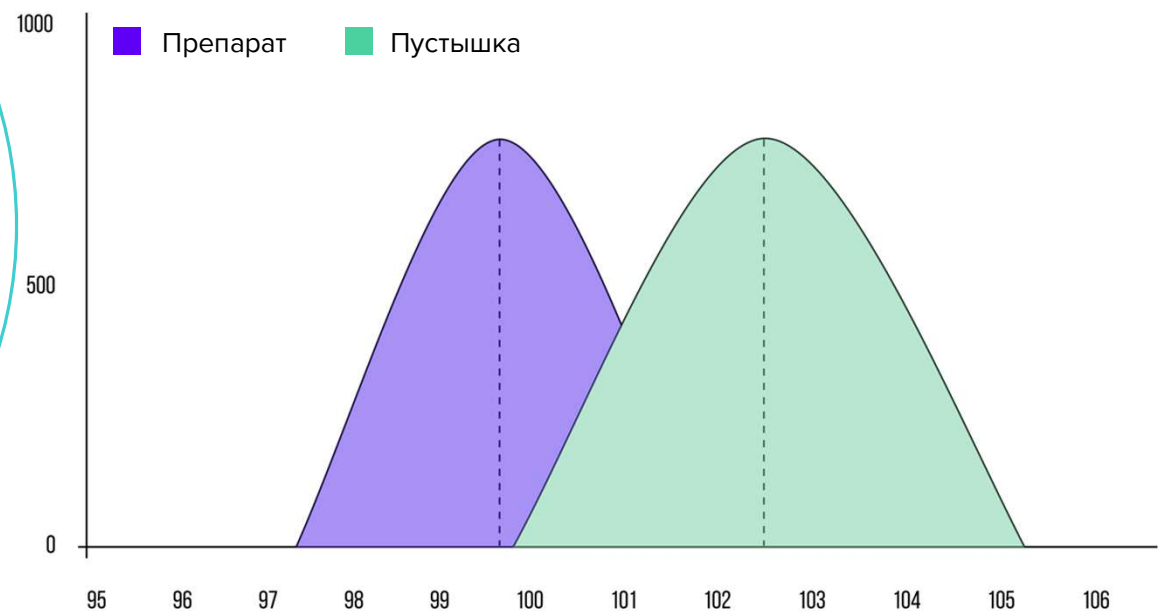
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Рассмотренный ранее
t-критерий Стьюдента предназначен
для сравнения **двух** совокупностей

Требует нормального
распределения данных

Невозможно отклонить $H_0: \mu_1 = \mu_2$

Отклонить H_0 : распределения выборок не равны



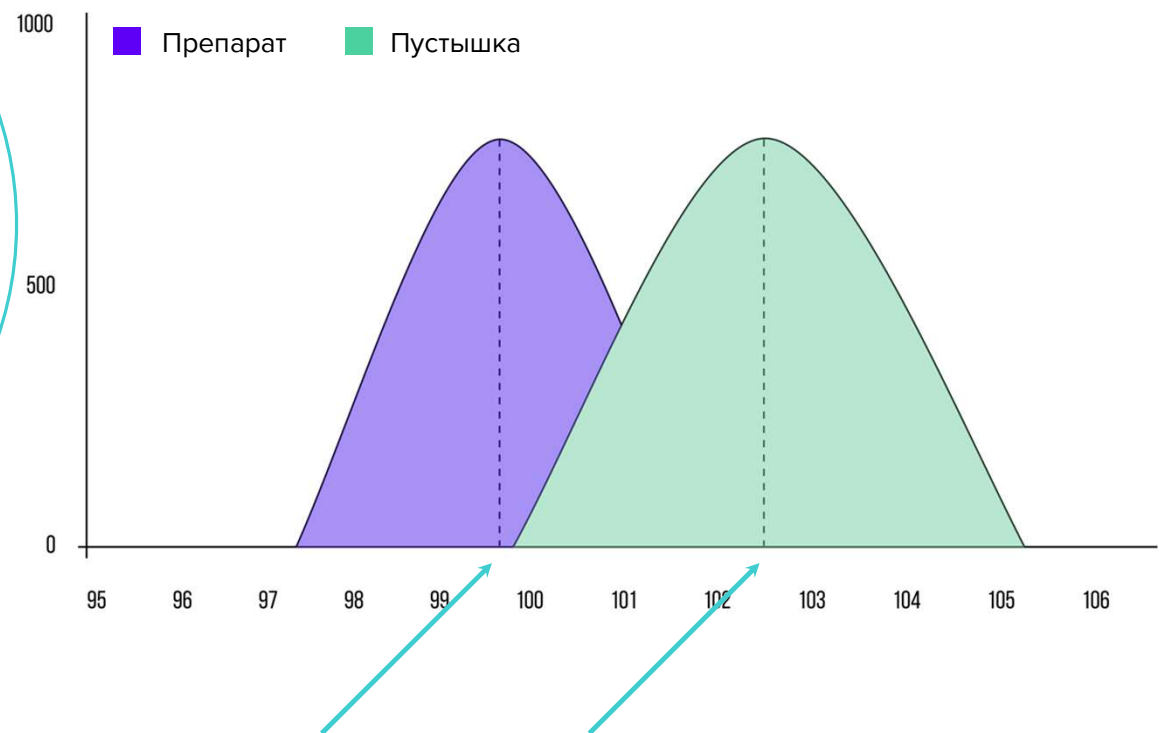
t-критерий Стьюдента

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Легко интерпретируемый
критерий.
Мы просто смотрим на различие
средних

Невозможно отклонить $H_0: \mu_1 = \mu_2$

Отклонить H_0 : распределения выборок не равны



Критерий Вилкоксона и Манна-Уитни

- Используемый для проверки различий между двумя выборками зависимых или независимых по количественному признаку (непрерывного или в порядкового шкале)
- Невозможно отклонить H_0 : распределения выборок равны
Отклонить H_0 : распределения выборок не равны
- Тест Вилкоксона для **независимых** выборок называется критерием Манна-Уитни

$$T = \sum R_r$$

Где R_r — ранговые значения сдвигов с более редким знаком

№	Уровень тревожности (до тренинга)	Уровень тревожности (после тренинга)	Шаг 2: Разность (после-до)	Шаг 3: Значение разности по модулю	Шаг 4: Ранг разности
1	15	14	-1	1	3
2	14	11	-3	3	8
3	16	17	1	1	3

Критерий Вилкоксона и Манна-Уитни

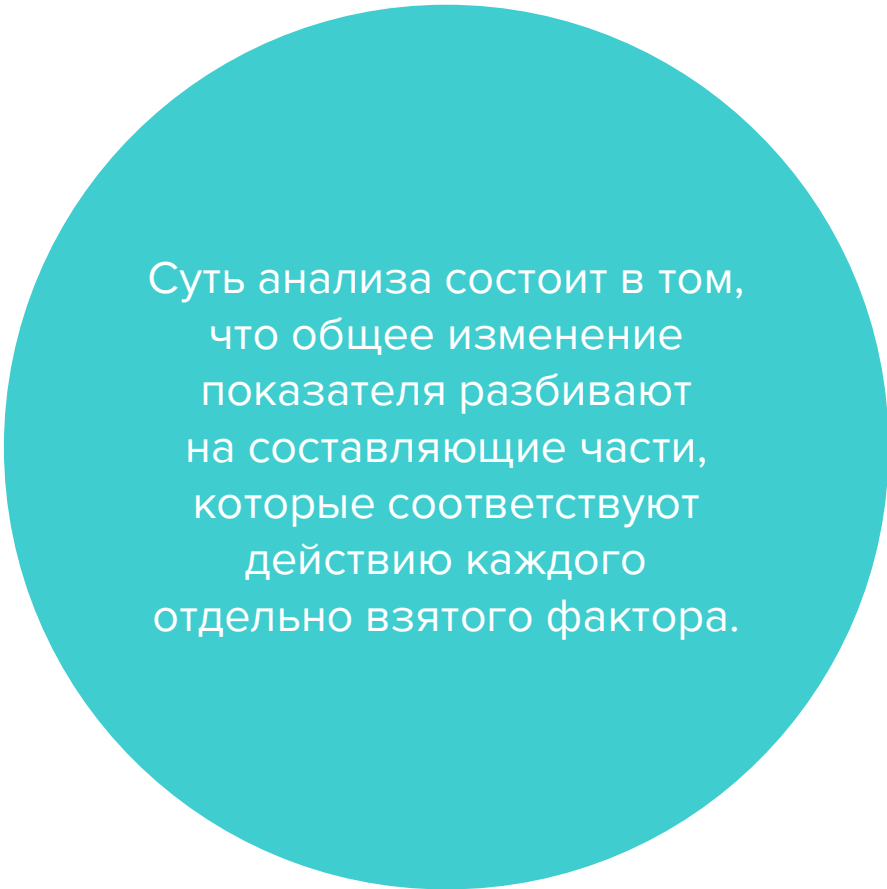
- Используемый для проверки различий между двумя выборками зависимых или независимых по количественному признаку (непрерывного или в порядкового шкале)
- Невозможно отклонить H_0 :
распределения выборок равны
Отклонить H_0 : распределения выборок не равны
- Тест Вилкоксона для **независимых** выборок называется критерием Манна-Уитни

Почему не использовать всегда Манна-Уитни?

В отличии от t-Критерия Стьюдента не требует нормального распределения, **НО** трудно интерпретируемый

Дисперсионный анализ

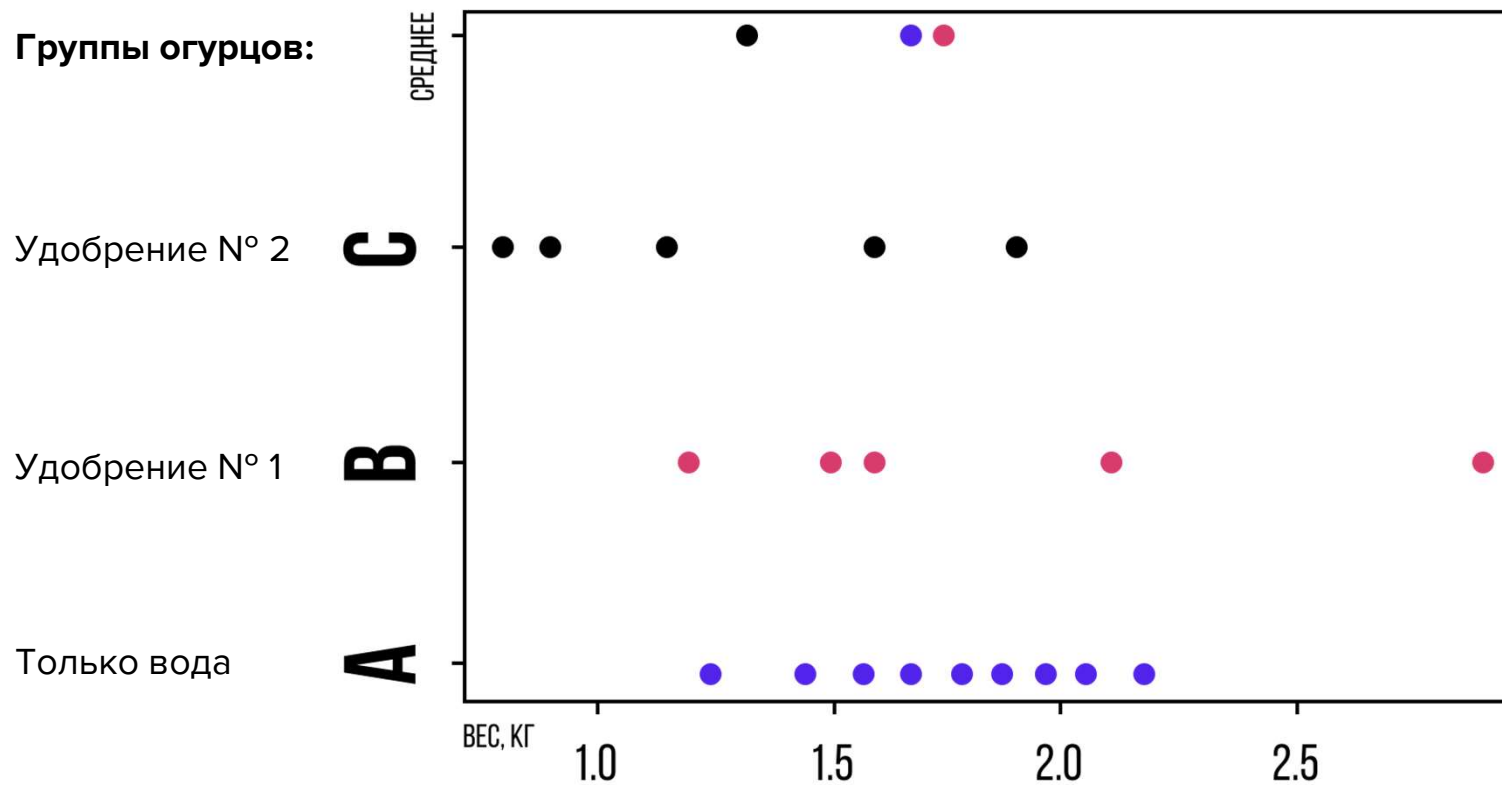
- В случае, когда нужно сравнить две или больше выборок, целесообразно применение дисперсионного анализа.
- Название метода указывает на то, что выводы делают на основе исследования составляющих дисперсии.




Суть анализа состоит в том, что общее изменение показателя разбивают на составляющие части, которые соответствуют действию каждого отдельно взятого фактора.

Дисперсионный анализ

Что можно сказать про эффективность удобрений?





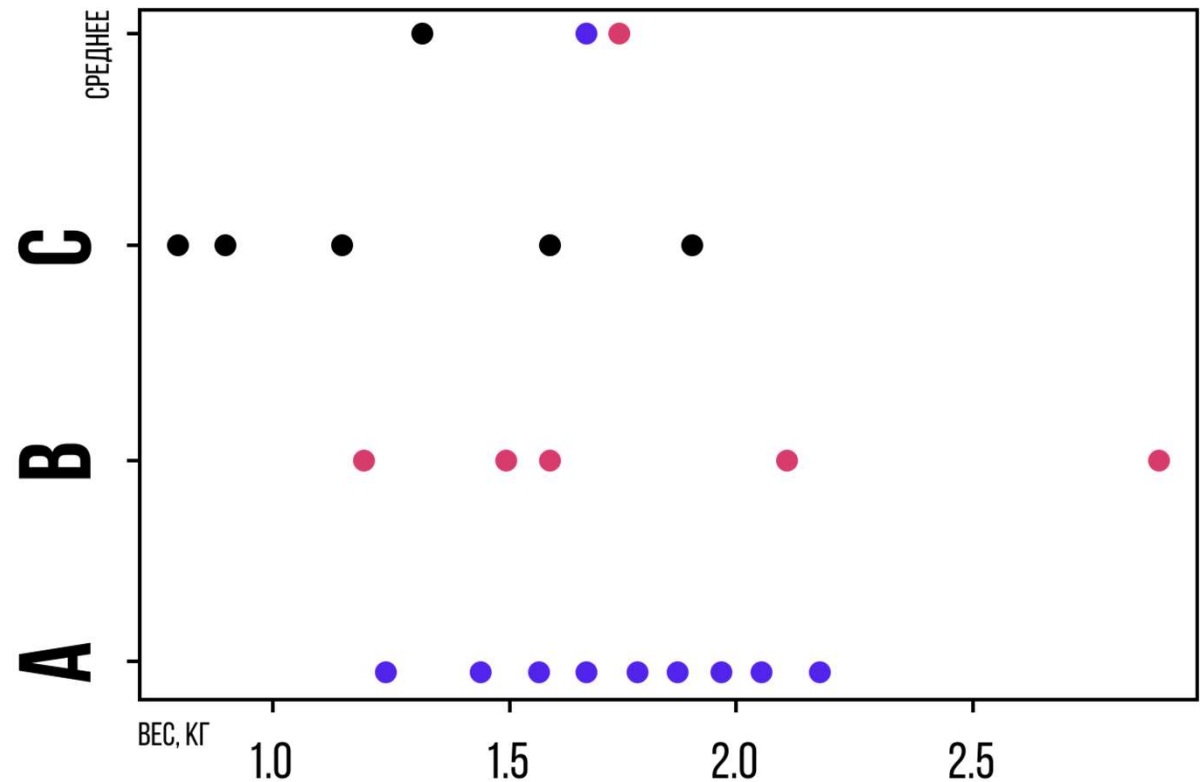
Нам нужен четкий и
понятный критерий
чтобы различить эти
случаи!

Как будем оценивать? t-Тест?

Нам хорошо знаком [t-критерий Стьюдента](#), который мы часто с легкостью рассчитывают для каждой пары сравниваемых групп.

Получив достаточно высокое значение t в каком-либо из этих сравнений, исследователь сообщает, что $p < 0.05$.

Это утверждение означает, что вероятность ошибочного заключения о существовании различий между групповыми средними не превышает 5% .



Но тут проблема множественного сравнения!

К чему приводит тестирование множества гипотез?

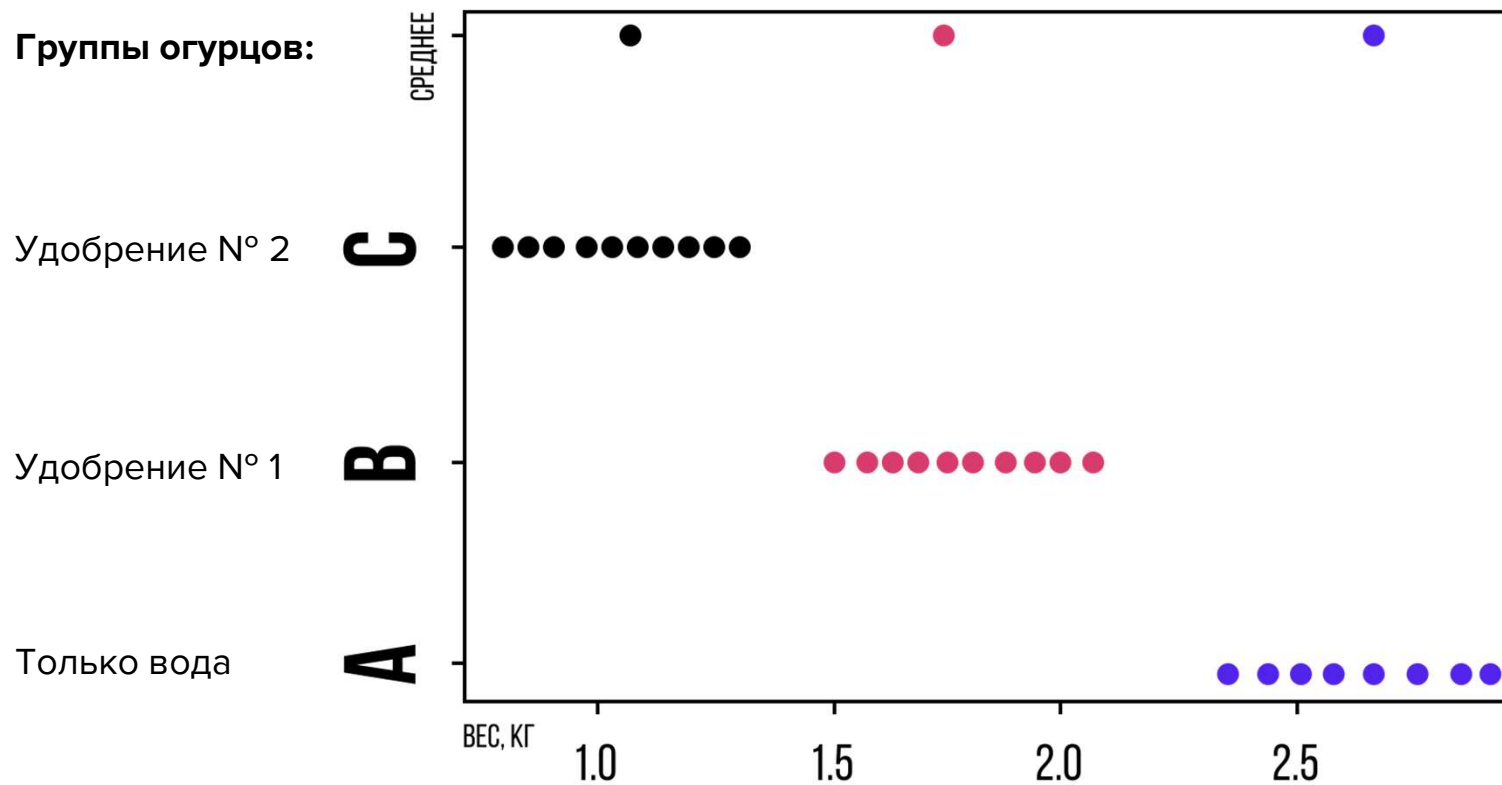
Очевидно, что дальнейшее увеличение числа проверяемых гипотез будет неизбежно сопровождаться возрастанием ошибки первого рода.

t-критерий Стьюдента не подходит для попарного сравнения большего количества групп что вызывает т. н. **эффект множественных сравнений**.

$$P' = 1 - (1 - \alpha)^m = 1 - (1 - 0.05)^3 = 0.143,$$

Дисперсионный анализ

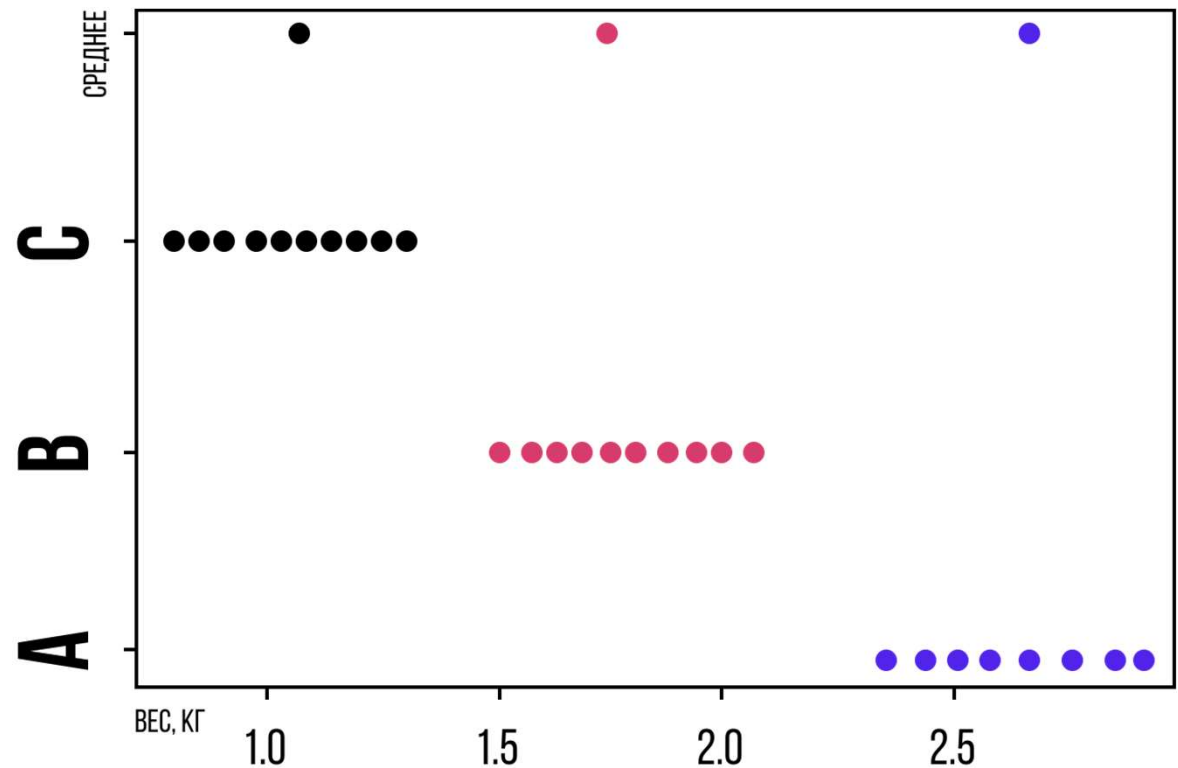
А теперь что можно сказать про эффективность тестов?



Дисперсионный анализ

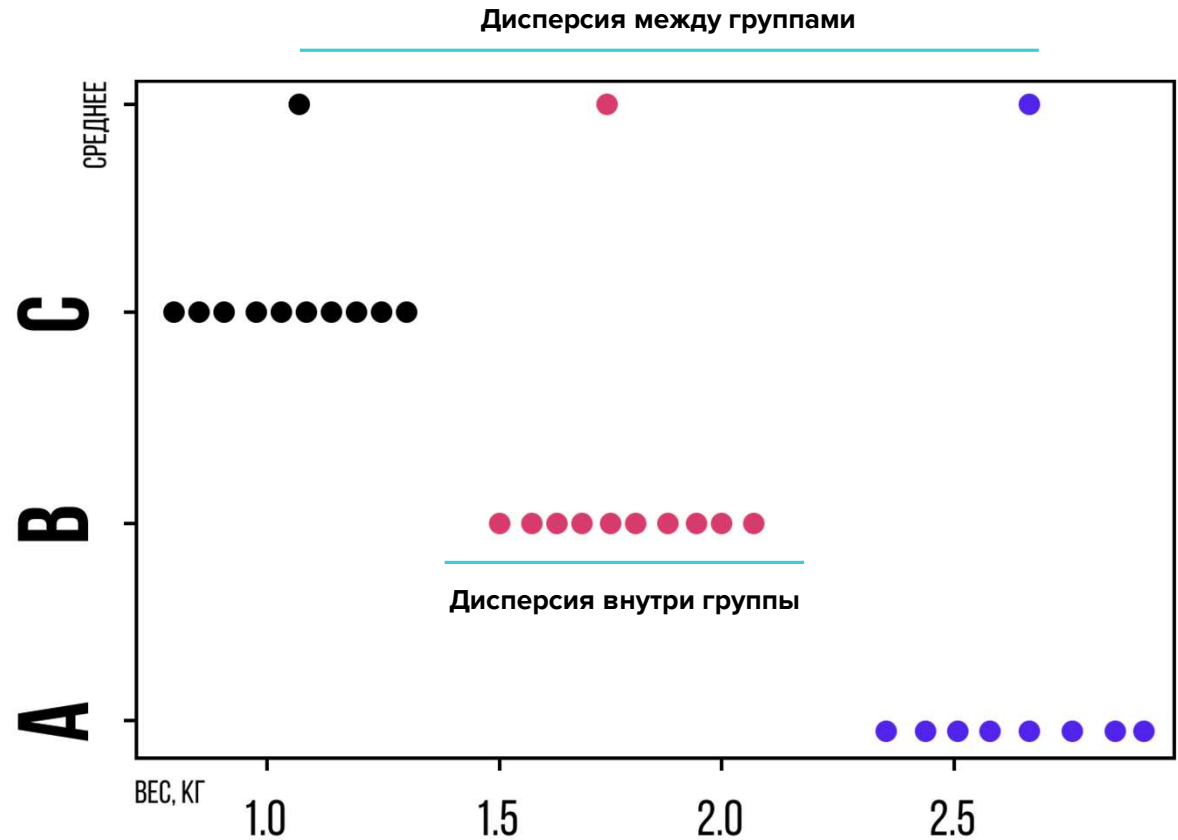
Сравните разброс значений **внутри групп** с разбросом **между** трёх **групповых средних**:

Это ключевая идея дисперсионного анализа!



Идея дисперсионного анализа

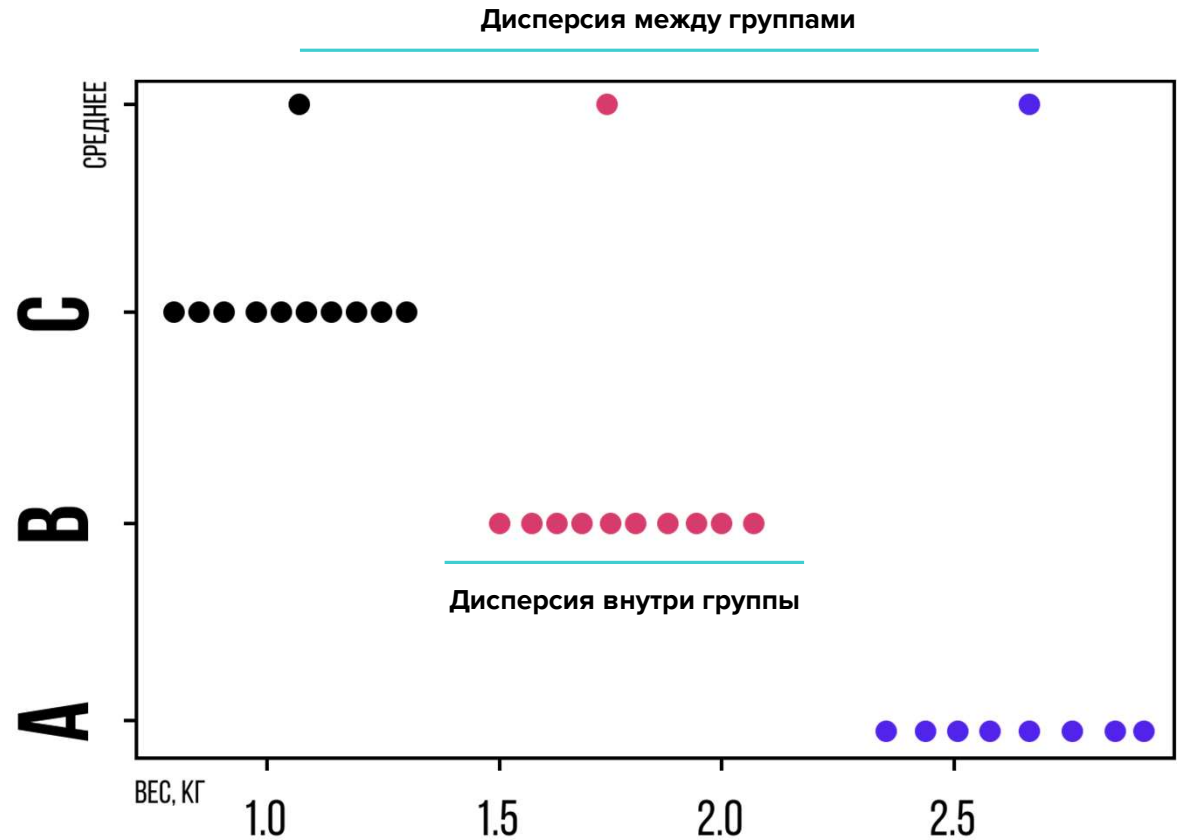
Получается, для дисперсионного анализа важны дисперсия внутри групп и дисперсия между группами!



Идея дисперсионного анализа

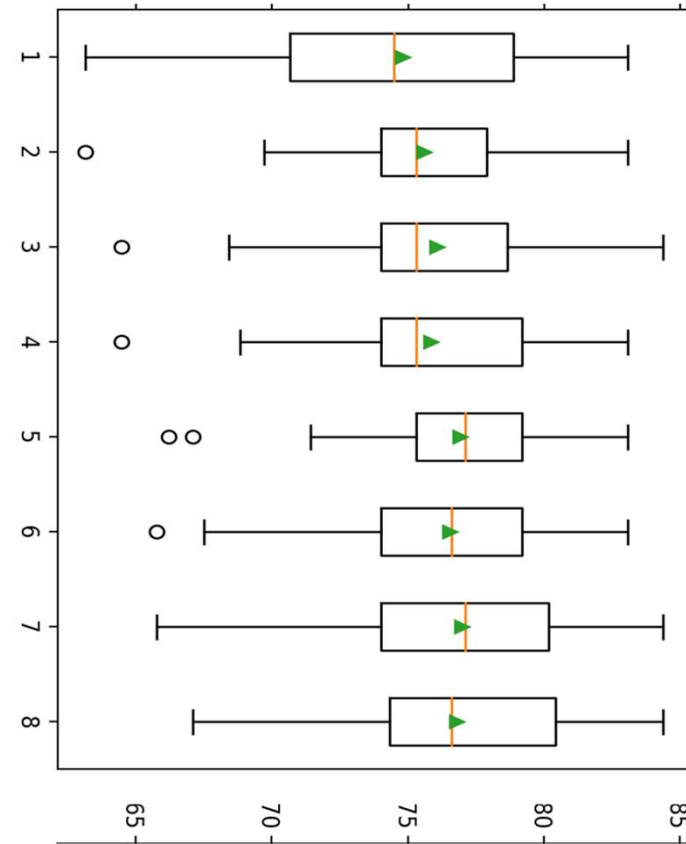
$$F = \frac{\text{Дисперсия между группами}}{\text{Дисперсия внутри группы}}$$

Чем больше F , тем проще различить выборки



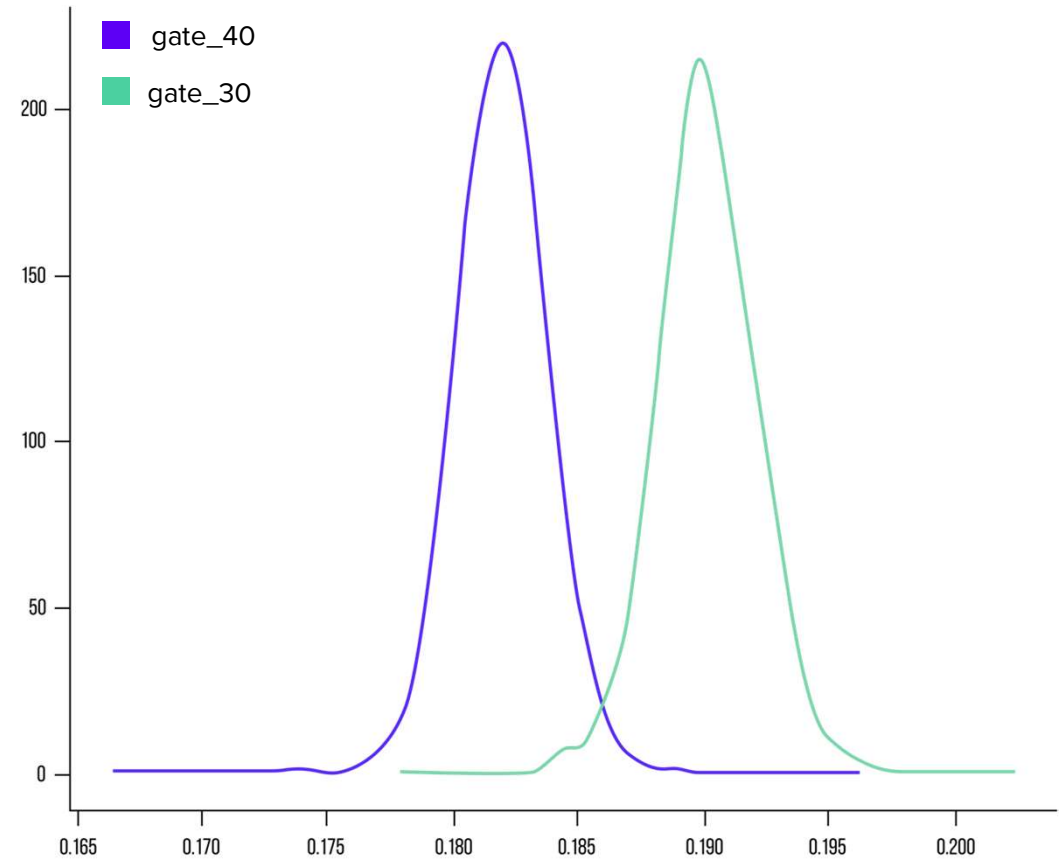
Как будем оценивать?

Принципы
однофакторного
дисперсионного анализа
как раз предназначены
для одновременного
сравнения средних
значений двух и более
групп.



Как еще будем оценивать?

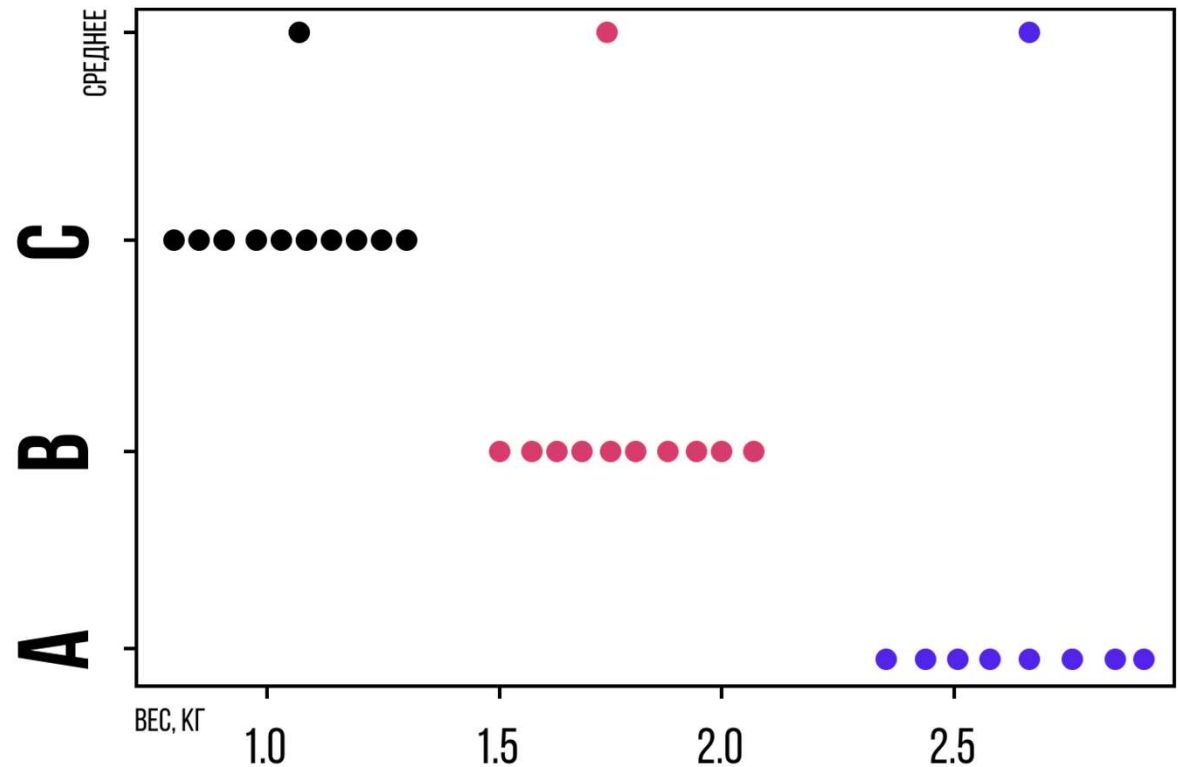
Есть еще бутстрап, но мы
сегодня не об этом



Виды и критерии дисперсионного анализа

Этот метод используется для исследования связи между качественными (номинальными) признаками и количественной (непрерывной) переменной.

В случае двух выборок результаты дисперсионного анализа будут идентичны результатам t-критерия Стьюдента. Однако в отличие от других критериев, это исследование позволяет изучить проблему более детально.



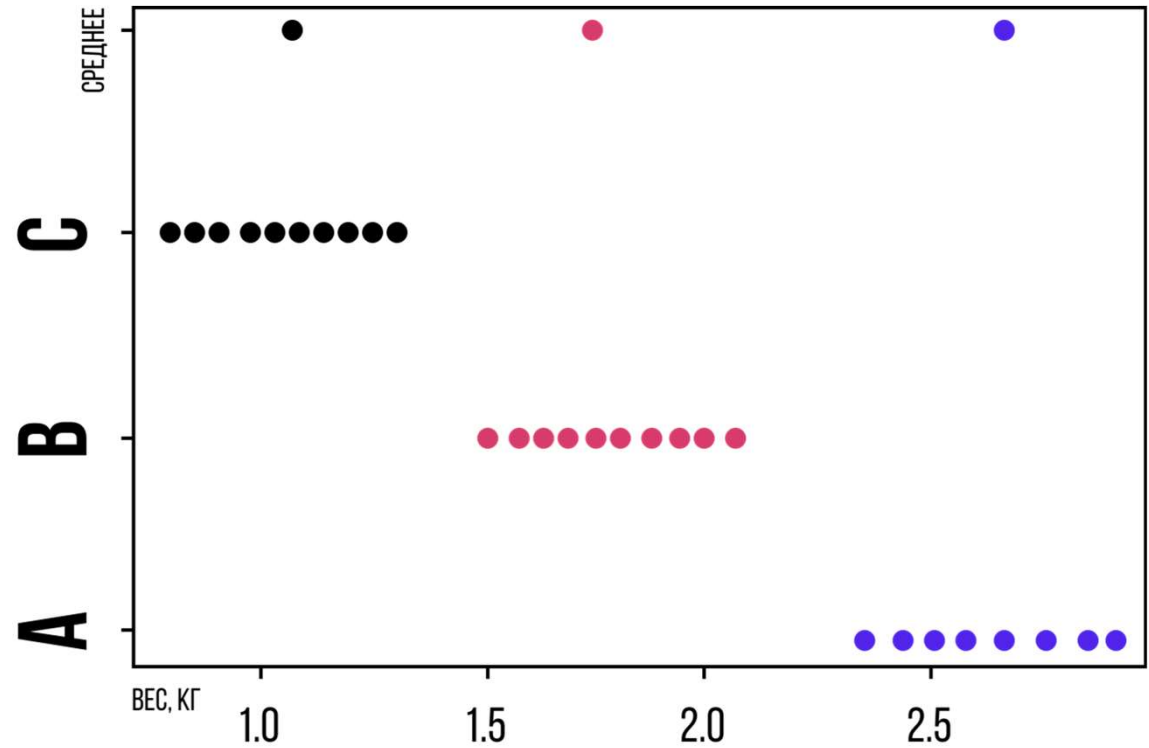
Мотивирующие примеры применения дисперсионного анализа



В троллейбусном депо работают троллейбусы различных типов. Всего на троллейбусных маршрутах оплату собирают 125 контролёров.

Как сравнить экономические показатели работы каждого контролёра (выручку) учитывая различные маршруты и типы троллейбусов?

Evgenii Plusnin on Unsplash



Дисперсионный анализ

$$MS_W = SS_W / (N - k)$$

Внутригрупповая
дисперсия

$$MS_B = SS_B / (k - 1)$$

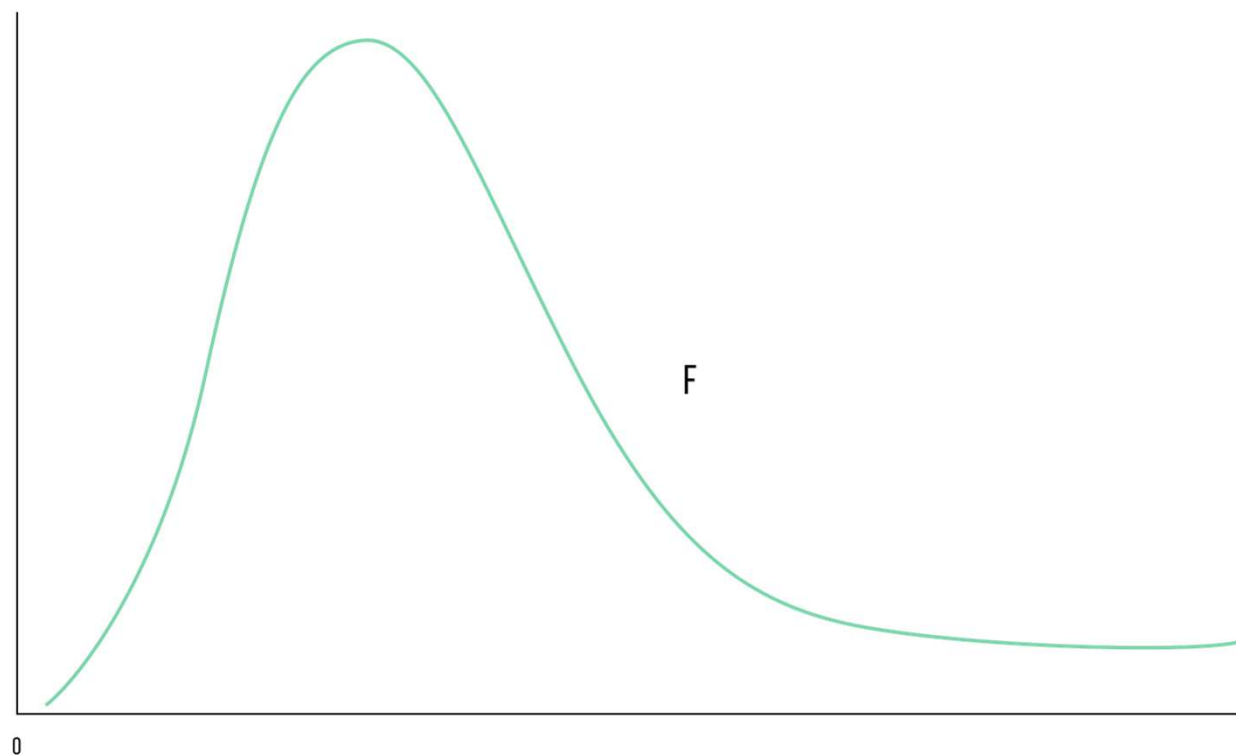
Межгрупповая
дисперсия

$$F = MS_B / MS_W$$

Критерий
Фишера

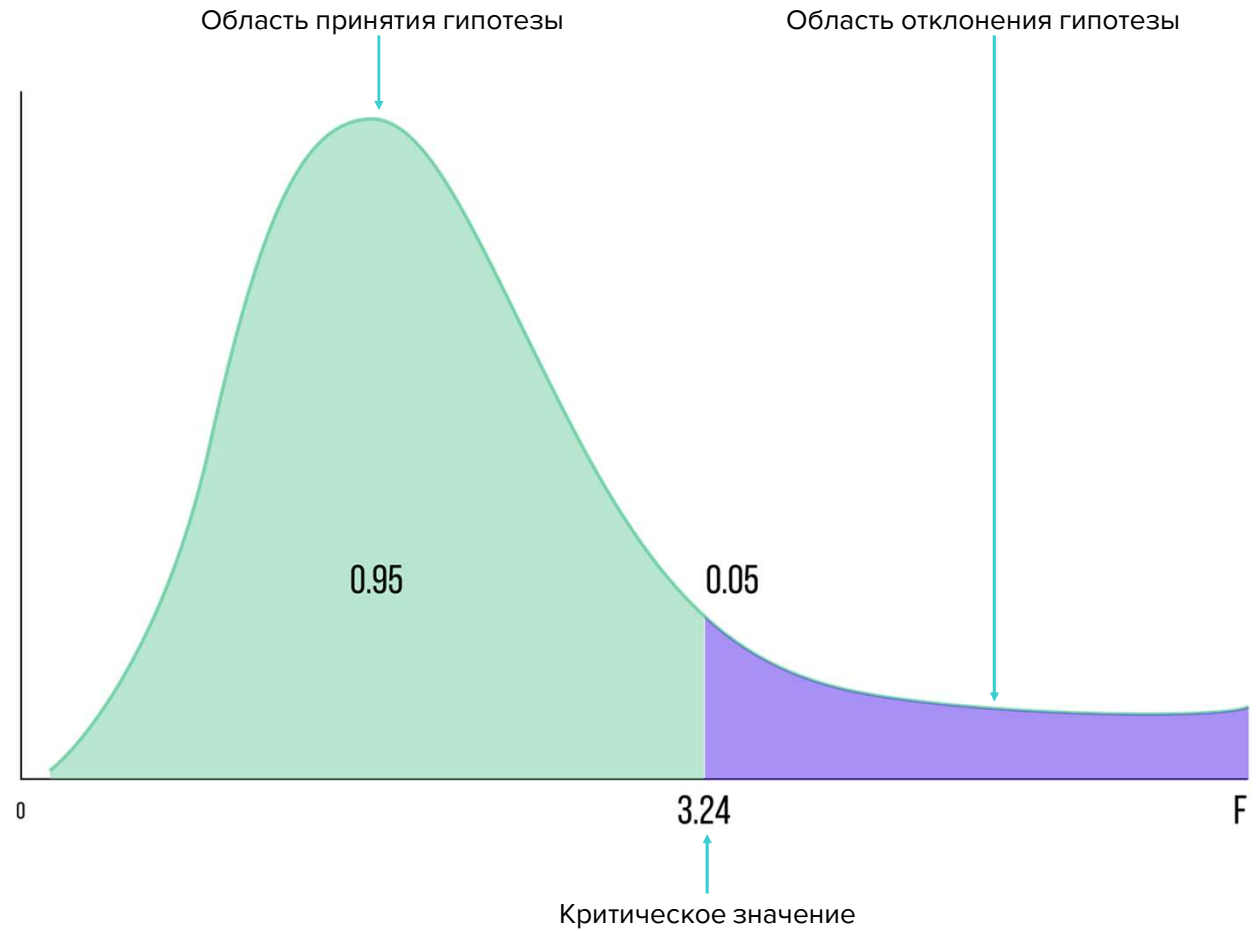
Ну и где тут p-value 5%?

$$F = MS_B / MS_W$$



Ну и где тут p-value 5%?

$$F = MS_B / MS_W$$

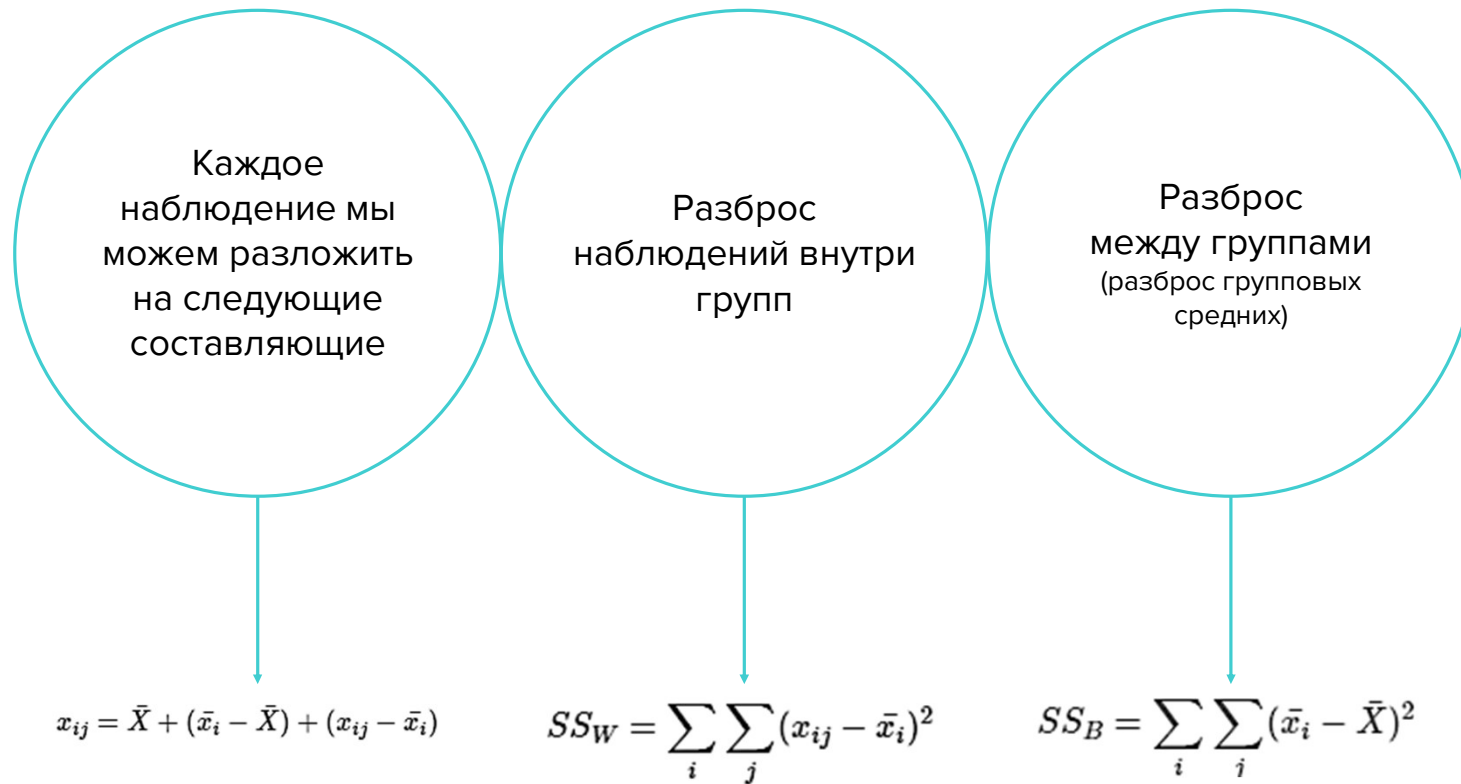


Дисперсионный анализ = ANOVA

Замечание: в специальной литературе его часто называют ANOVA (от англоязычного названия Analysis of Variance).

Впервые этот метод был разработан Р. Фишером в 1925 г.

Дисперсионный анализ



$$(\bar{x}_i - \bar{X})$$

Отклонения групповых средних от общего среднего значения

$$(x_{ij} - \bar{x}_i)$$

Отклонения отдельных наблюдений от среднего значения группы, к которой они принадлежат