Статистические

инструменты для проведения A/B-теста

Цель модуля

Научиться применять математическую статистику для оценки данных и готовить данные для проведения A/B-тестирования.

Что вы будете уметь
по итогам изучения модуля?

- Познакомитесь с популярными критериями
- Изучите условия применения статистических критериев
- Поймёте, что можно сделать в случае, когда даже распределение не является нормальным

Статистические инструменты для проведения A/B-теста

Параметрические
и непараметрические критерии, или как доказать гипотезу

Цель урока

Изучить основные параметрические и непараметрические критерии и понять принцип их работы.

Задачи урока

- Узнать, что такое критерий и как его интерпретировать
- Узнать, как работает Т-тест (Критерий Стьюдента) и Z-тест и когда какой применяют
- Узнать, как работает Т-критерий Уилкоксона и когда его применяют
- Узнать, как работает U-критерий Манна — Уитни

Статистический критерий

Статистический критерий — это решающее правило, обеспечивающее надёжное поведение, то есть принятие истинной или отклонение ложной гипотезы с высокой вероятностью (Суходольский).

Статистические критерии обозначают также метод расчёта определённого числа и само это число.

Когда мы говорим, что достоверность различий определялась по t-критерию Стьюдента, то имеем в виду, что для расчёта определённого числа t использовали метод t-тест.

Эмпирическое значение критерия

Например, когда мы говорим, что $t = 12{,}6$, то имеем в виду определённое число, рассчитанное по t-критерию Стьюдента. Это число обозначается как **эмпирическое значение критерия**.

По **соотношению эмпирического и критического значений критерия** мы можем судить о том, подтверждается или опровергается нулевая гипотеза. Например, если $|t_{эмп}|>|t_{кр}|$, H_0 отвергается.

Эмпирическое значение критерия

Примечание: в большинстве случаев для того, чтобы мы признали различия значимыми, необходимо, чтобы эмпирическое значение критерия превышало критическое, хотя есть критерии (например, критерий Манна — Уитни), в которых мы должны придерживаться противоположного правила. Эти правила оговариваются в описании каждого из критериев.

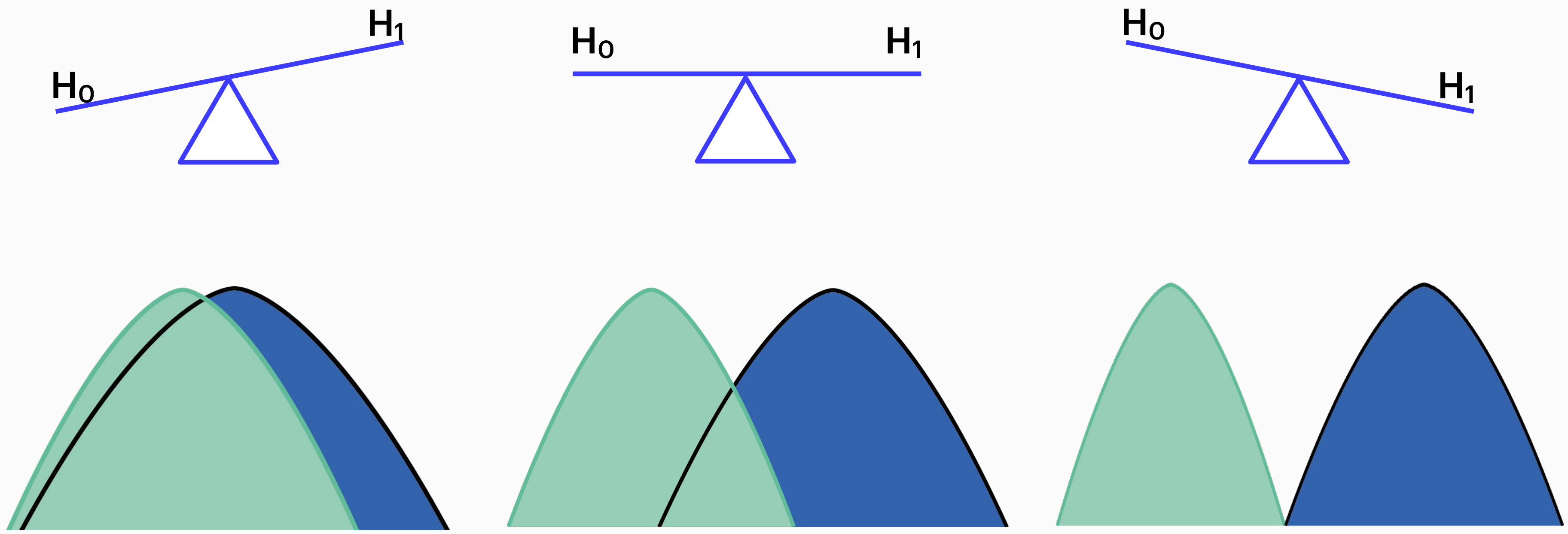
Основная цель критерия

Основная цель критерия — дать вам оценку возможной ошибки.

То есть дать вам **p-value**.

Значение **p-value** вы уже легко можете интерпретировать.

Принятие гипотез и p-value



Группы не отличаются
Эффект случайный

p-value > 0,05

Мы ничего не можем доказать

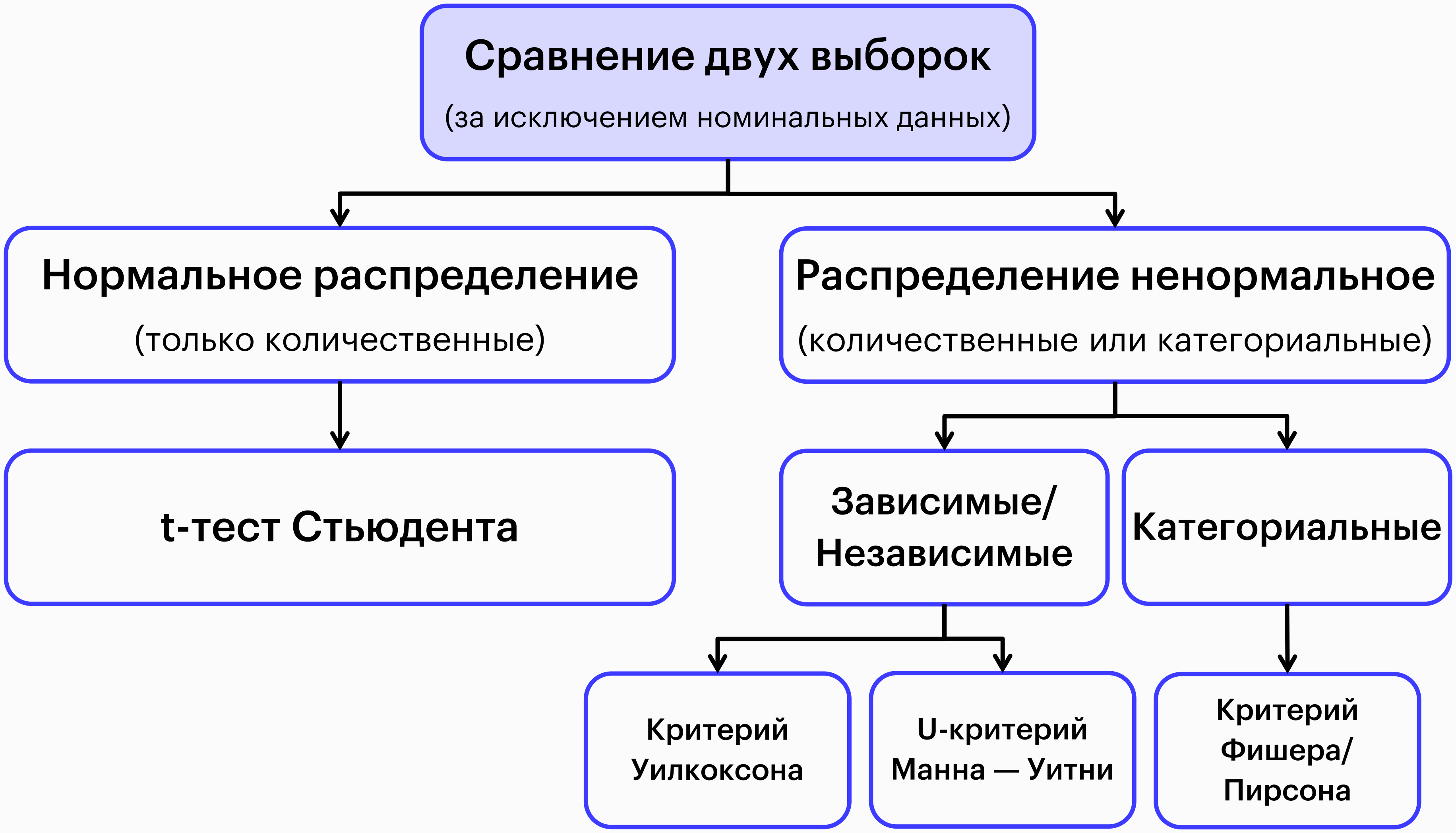
Неоднозначно

p-value = 0,05

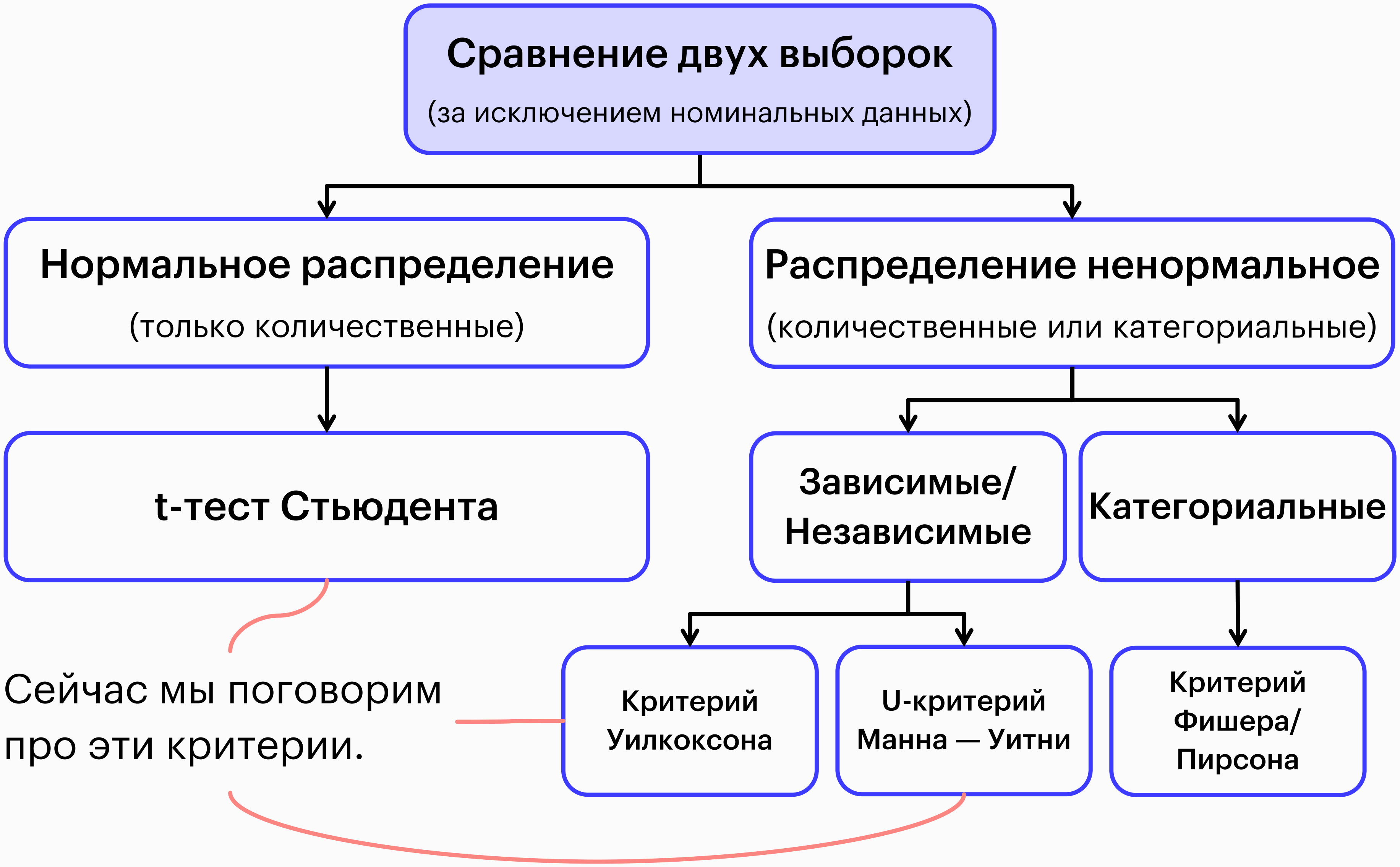
Группы отличаются
Эффект не случайный

p-value < 0,05

Карта статистических тестов гипотез



Карта статистических тестов гипотез



Критерии
для проверки гипотез: t-тест

С помощью t-теста (также называемого t-критерием Стьюдента) можно сравнить два средних значения. Другими словами, он позволяет узнать, могли ли эти различия возникнуть случайно.

Нулевая гипотеза: средние двух выборок равные.

Альтернативная гипотеза: средние двух выборок не равны.

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

$\overline{X}$ — это среднее выборки

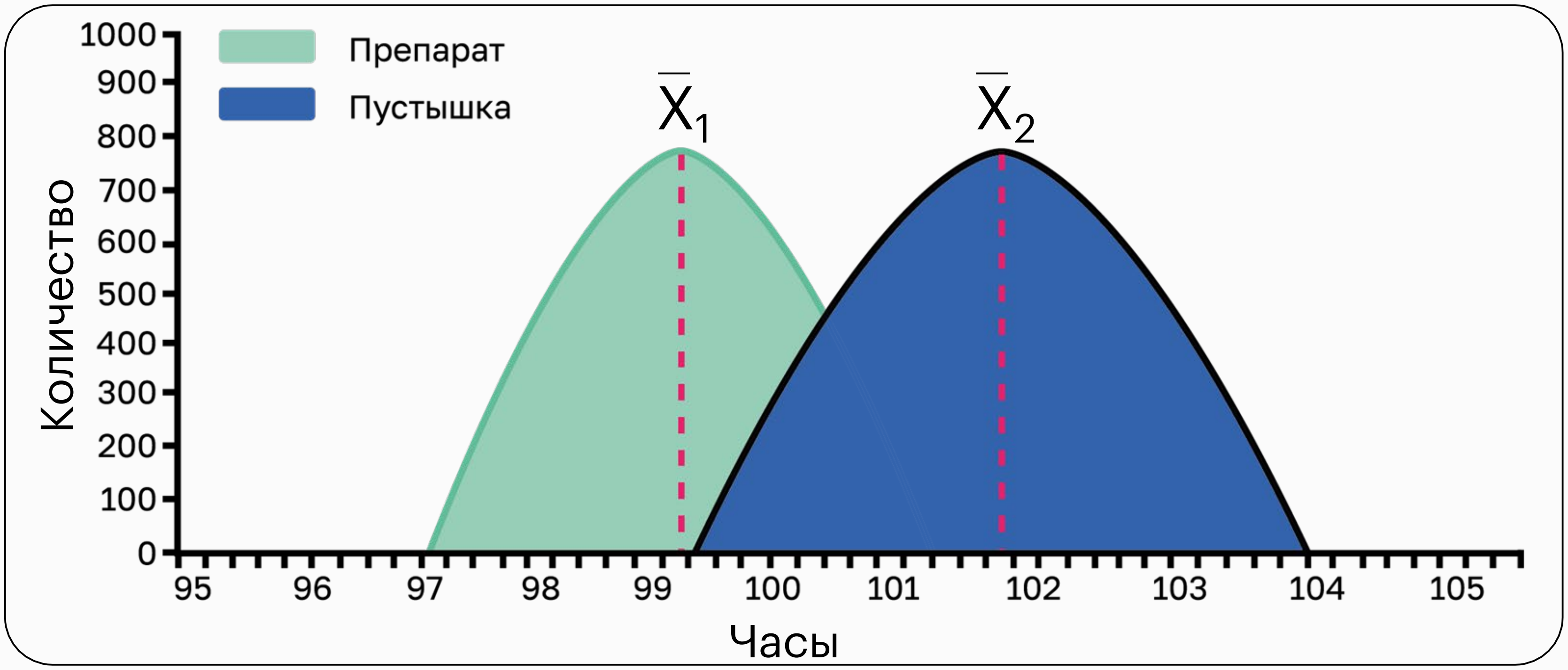
s — это выборочная дисперсия выборки

n — это объём (размер) выборки

Величина t количественно отражает аргументы в наборе данных против нулевой гипотезы

Критерии
для проверки гипотез: t-тест

- Требует нормального распределения
- Легко интерпретируемый: смотрит на **разницу средних** двух выборок



Критерии
для проверки гипотез: t-тест

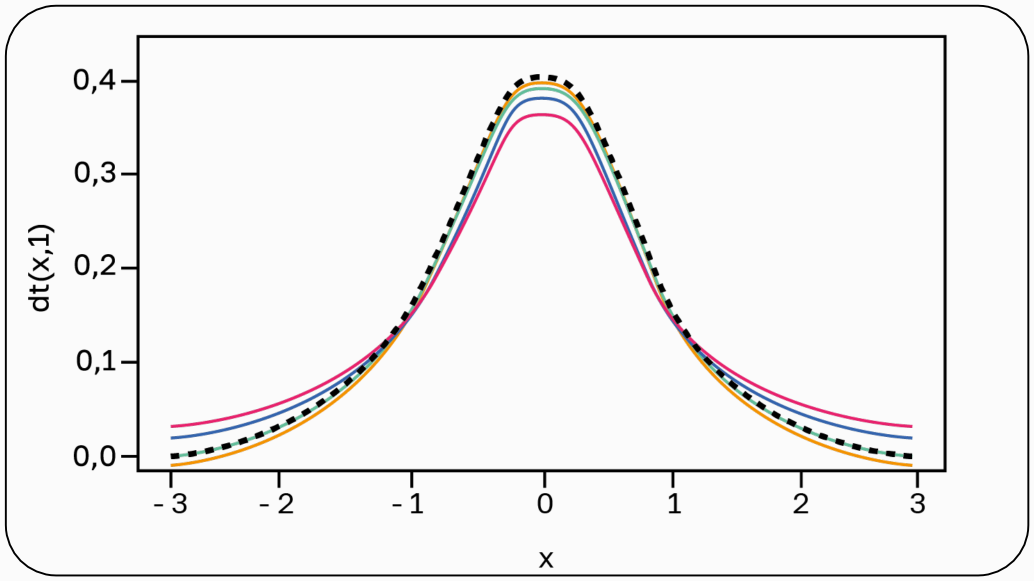
Получаемая величина **t** имеет распределение Стьюдента.

Знаная распределение, мы можем сопоставить значению **t** его **p-value**.

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

t ~ распределение Стьюдента

Распределение t-критерия Стьюдента
(очень похоже на нормальное)



Критерии
для проверки гипотез: t-тест

Т-критерий Стьюдента уже реализован в большинстве языков программирования, в том числе на языке Python. Поэтому алгоритм использования такой:

- берём исследуемые данные
- применяем t-критерий Стьюдента
- анализируем полученный **p-value**

Критерии
для проверки гипотез: t-тест

Позже воспользуемся его реализацией на Python в библиотеке **scipy**.

```
from scipy import stats  #Подключаем библиотеку

t, p  = stats.ttest_ind(A, B)  # Получаем t-value и p-value
```

Обратите внимание, на выходе у нас, помимо **t**-значения, подсчитывается ещё и **p-value**.

Критерии
для проверки гипотез: Z-тест

С помощью Z-теста (также называемого Z-критерием) можно сравнить два средних значения. Иными словами, он позволяет узнать, могли ли эти различия возникнуть случайно.

Нулевая гипотеза: средние двух выборок равные.

Альтернативная гипотеза: средние двух выборок не равны.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

$\overline{X}$ — это среднее выборки

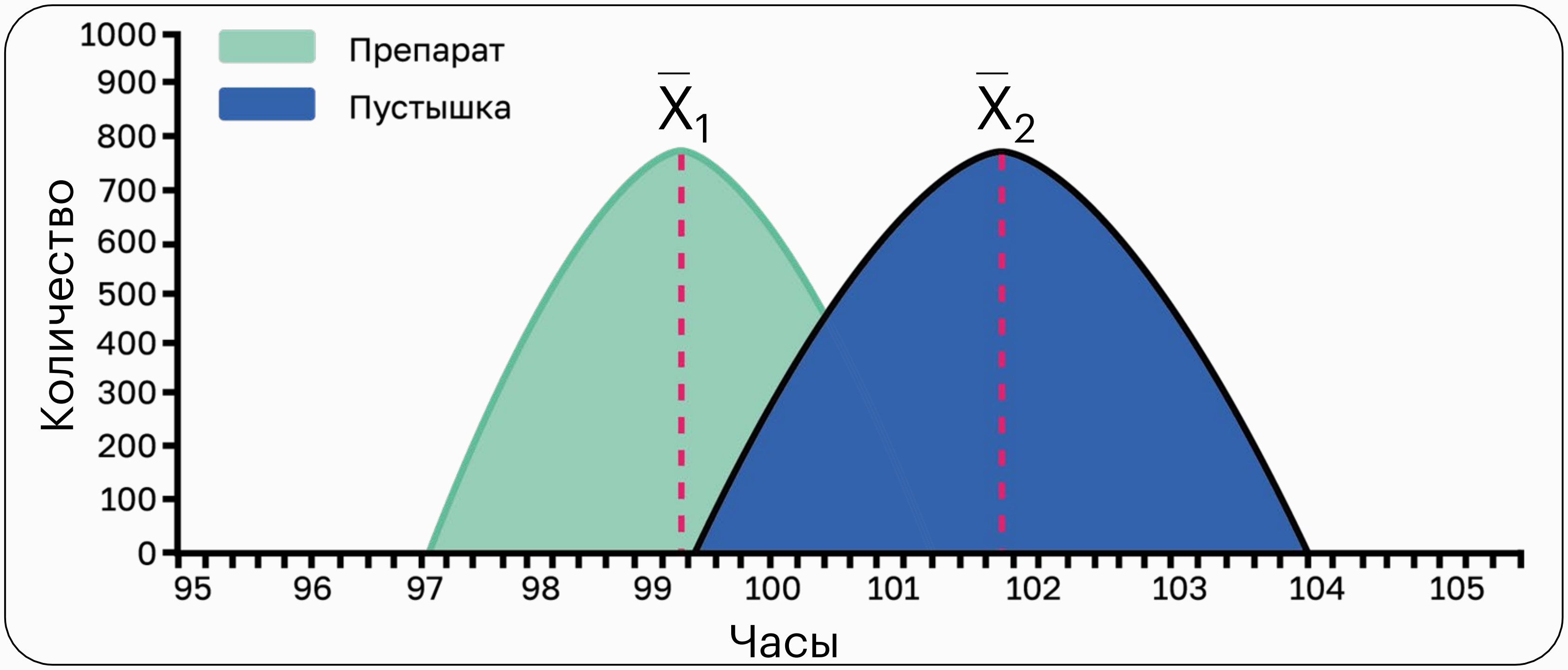
σ — это дисперсия выборки

n — это объём (размер) выборки

Отличие от t-теста в обычной дисперсии

Критерии
для проверки гипотез: Z-тест

- Требует нормального распределения
- Легко интерпретируемый: смотрит на **разницу средних** двух выборок



Критерии
для проверки гипотез: Z-тест

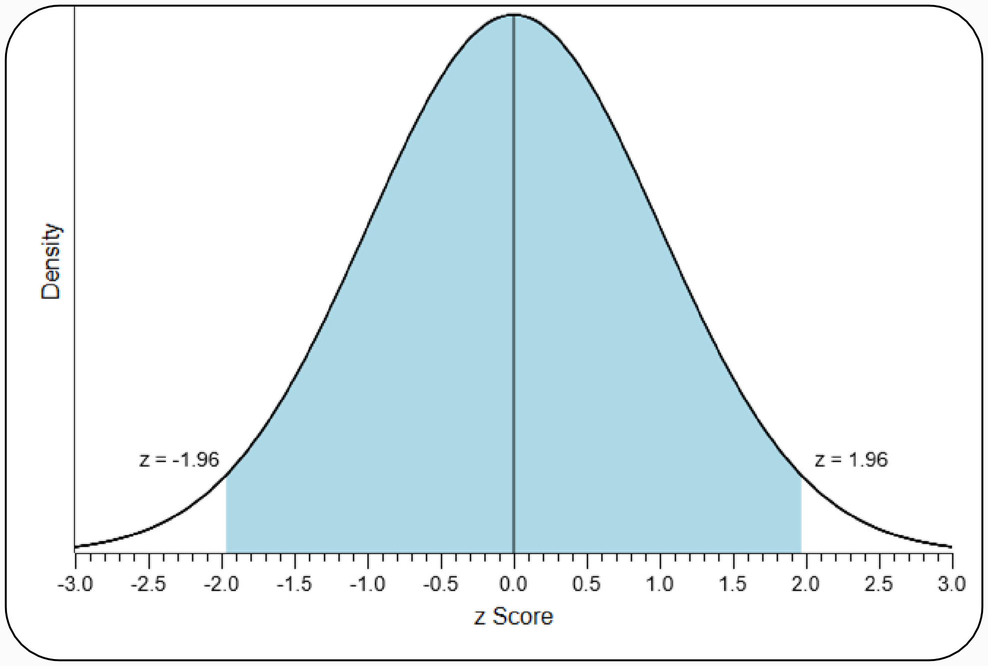
Получаемая величина **z** имеет нормальное распределение.

Зная распределение, мы можем сопоставить значение **z** и его **p-value**.

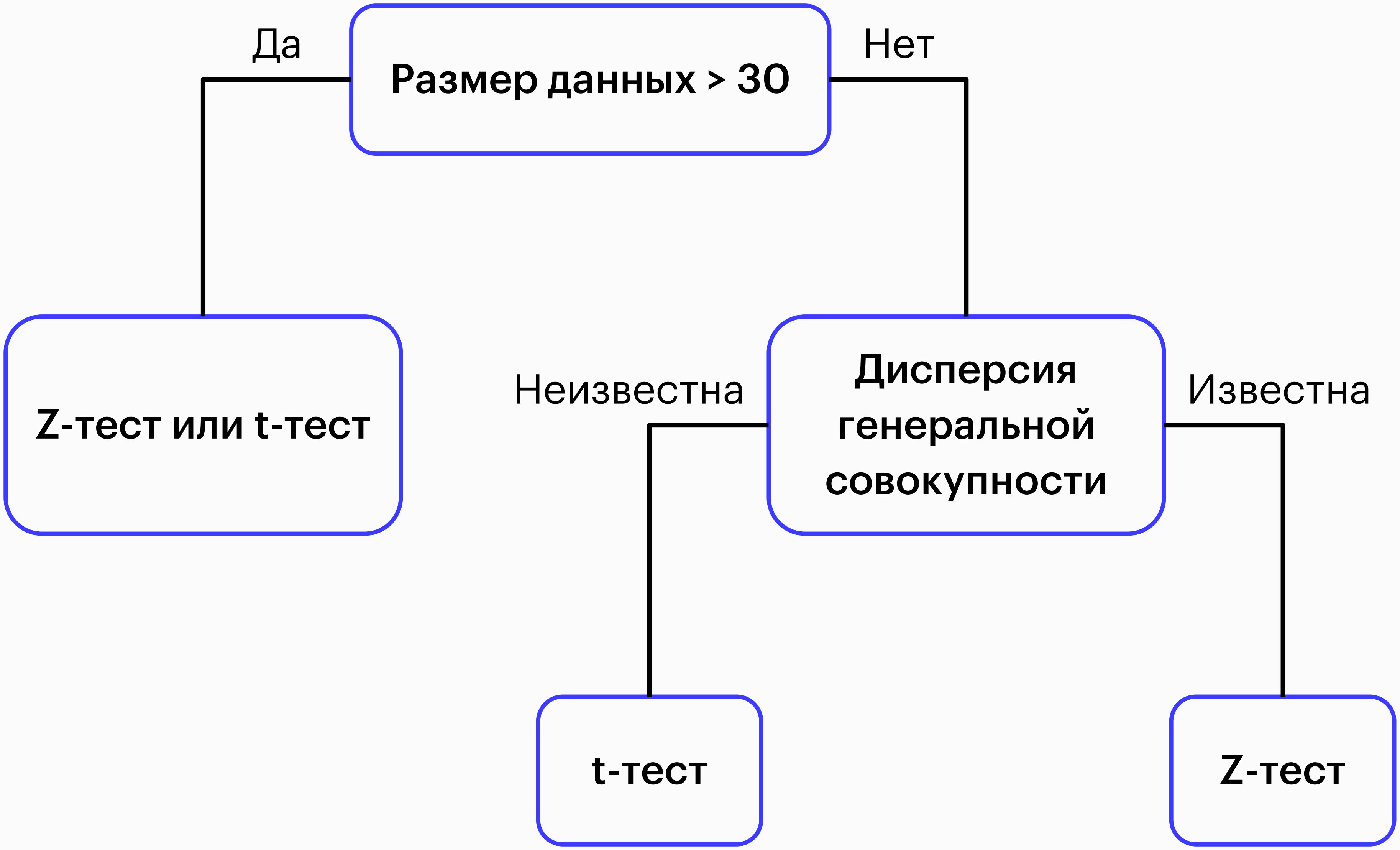
$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}}$$

z ~ N(0,1)

Нормальное распределение



t-тест vs Z-тест



Далее мы рассмотрим два непараметрических критерия

Непараметрический критерий — это критерий, который применяется без каких-либо предположений относительно параметров изучаемой генеральной совокупности.

Другие словами, непараметрический критерий не требует, чтобы распределение было нормальным.

Ранг

№ испытуемого	Показатели интеллекта	Ранги
1	113	6
2	107	4
3	123	11
4	122	10
5	117	(8) 8,5
6	117	(9) 8,5
7	105	3
8	108	5
9	114	7
10	102	1
11	104	2

Ранг показывает номер значения распределения, если его отсортировать по возрастанию.

Если одинаковые значения идут подряд, то их ранг усредняется.

Ранговые
непараметрические критерии

Допущения: независимые наблюдения между и внутри групп.

Ограничения: только для сравнения средних значений по пользователю.

Если:

- 2 группы — критерий Манна — Уитни (Уилкоксона)
- > 2 групп — критерий Краскела — Уоллиса

Ранговые критерии **устойчивы к выбросам**, потому что абсолютные значения преобразуются в порядковые номера. Они полезны, когда данные даны в номинальной или порядковой шкале.

Критерии проверки гипотез Манна — Уитни

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2$$

n — это объём (размер) выборки

R — ранг значения

$$U = min\{U_1, U_2\}$$

$$p = \frac{|U - {n_1 n_2}/{2}|}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Если U ≤ p, то принимается альтернативная гипотеза — различие в уровне признака в двух выборках существенно.

Не пугайтесь сложных формул.

На практике мы будем использовать готовые реализации на Python.

Сейчас важно понять, что из себя представляет этот критерий.

Python и критерий Манна — Уитни

U-критерий Манна — Уитни уже реализован в большинстве языков программирования, в том числе на языке **Python** в библиотеке **scipy**. Поэтому алгоритм использования такой:

- берём исследуемые данные
- применяем U-критерий Манна — Уитни
- анализируем полученный **p-value**

Критерии проверки гипотез Манна — Уитни

С помощью критериев Манна — Уитни можно осуществить две выборки на значимость разницы рангов. Другими словами, критерий позволяет узнать, могли ли эти различия возникнуть случайно.

Нулевая гипотеза: различие в уровне признака в двух выборках не существенно.

Альтернативная гипотеза: различие в уровне признака в двух выборках существенно.

- Не требует нормального распределения
- Сложно интерпретируемый: опирается на сумму рангов выборок

Если две выборки зависимы, то используем другую форму **критерия — Манна — Уитни**, а именно критерий **Уилкоксона**.

Python и критерий Манна — Уитни

Позже мы воспользуемся реализацией этого критерия на **Python** в библиотеке **scipy**.

```
from scipy import stats  #Подключаем библиотеку

U, p = stats.mannwhitneyu(A, B)   # Получаем U-value и p-value
```

Обратите внимание, на выходе у нас, помимо **U**-значения, подсчитывается ещё и **p-value**, которые вы уже с лёгкостью сможете проинтерпретировать.

Python и Уилкоксон

Позже мы воспользуемся реализацией этого критерия на **Python** в библиотеке **scipy**.

```
import scipy.stats as stats


statistic, p_value = stats.wilcoxon(A, B)
```

Обратите внимание, на выходе у нас, помимо значения статистики подсчитывается ещё и **p-value**, которые вы уже с лёгкостью сможете проинтерпретировать.

Итоги урока

- Мы познакомились с тремя важными критериями: t-критерий Стьюдента, U-критерий Манна — Уитни, критерий Уилкоксона
- Критерии позволяют вам получить p-value (вероятность ошибочно отвергнуть верную гипотезу)
- Критерий Уилкоксона позволяет сделать поправку на зависимость между выборками
- t-критерий Стьюдента легко интерпретируемый — про это мы ещё поговорим дальше