Общий семинар, неделя 11

"Исследование данных с Python. Описательные статистики. Статистика вывода."

Жарова Мария

Сегодня в программе

- **Описательная статистика**
- **Введение в статистические гипотезы**
  - выборка и генеральная совокупность
  - ошибки I и II рода
  - p-value
  - статистические критерии

МФТИ

Описательная статистика

Пример из жизни

В новостях пишут: “Средняя зарплата в Краснодарском крае составляет 83 000 руб. Это число даёт нам представление о среднестатистическом жителе этого края.”

Что такое средняя ЗП и среднестатистический житель?

МФТИ

Пример из жизни

В новостях пишут: “Средняя зарплата в Краснодарском крае составляет 83 000 руб. Это число даёт нам представление о среднестатистическом жителе этого края.”

Что такое средняя ЗП и среднестатистический житель?

Мера центральной тенденции – это число, которое описывает так называемое «среднее» признака.
Она может рассчитываться по-разному в зависимости от типа признака или от его распределения.

МФТИ

Среднее арифметическое

– сумма всех элементов, поделённая на количество элементов в числовом ряду или признаке (Series).

$$\text{среднее арифметическое} = \frac{\text{сумма всех чисел}}{\text{количество слагаемых}}$$

Пример: оценки Миши по математике за 3-ю четверть 5, 2, 5, 5, 5, 4, 5, 5. Какую оценку получит Миша за четверть, если округление в спорных ситуациях происходит в пользу ученика?

МФТИ

Среднее арифметическое

– сумма всех элементов, поделённая на количество элементов в числовом ряду или признаке (Series).

$$\text{среднее арифметическое} = \frac{\text{сумма всех чисел}}{\text{количество слагаемых}}$$

Пример: оценки Миши по математике за 3-ю четверть 5, 2, 5, 5, 5, 4, 5, 5. Какую оценку получит Миша за четверть, если округление в спорных ситуациях происходит в пользу ученика?

Ответ: (5+2+5+5+5+4+5+5) / 8 = 4.5 ⇒ получит 5 за четверть.

Плохой пример среднего арифметического

10 друзей собрались на вечеринку и заказали пиццу.
Количество съеденных кусков пиццы каждым соответственно:
1, 2, 2, 2, 3, 3, 4, 4, 5, 20.

Среднее арифметическое:
1 + 2 + 2 + 2 + 3 + 3 + 4 + 4 + 5 + 20) / 10 = 4.6 куска/чел.

Справедливо и показательно ли это?

Плохой пример среднего арифметического

10 друзей собрались на вечеринку и заказали пиццу.
Количество съеденных кусков пиццы каждым соответственно:
1, 2, 2, 2, 3, 3, 4, 4, 5, 20.

Среднее арифметическое:
1 + 2 + 2 + 2 + 3 + 3 + 4 + 4 + 5 + 20) / 10 = 4.6 куска/чел.

Справедливо и показательно ли это?

Вывод: среднее арифметическое хорошо работает для признаков с небольшим разбросом в значениях. Но не подходит для признаков с **большим разбросом в значениях**.

Медиана

Медиана – это средний элемент распределения в предварительно отсортированном числовом ряду.

Алгоритм расчёта медианы:

Возьмите ваши наблюдения:	80, 87, 95, 83, 92
Расположите их в возрастающем порядке:	80, 83, 87, 92, 95
Среднее значение и есть медиана	80, 83, **87**, 92, 95
Если значений чётное кол-во, то медианой будет среднее арифметическое двух средних значений	$80, 83, \overbrace{\mathbf{87}, \mathbf{92}}^{\mathbf{89.5}}, 95, 98$

Медиана

Пример 1. В классе 7 детей, они выпили такое количество сока (в мл):

150, 200, 250, 300, 350, 400, 450.

Найдите медиану.

Медиана

Пример 1. В классе 7 детей, они выпили такое количество сока (в мл):

150, 200, 250, 300, 350, 400, 450.

Найдите медиану.

Ответ: 300 мл.

Медиана

Пример 2. Решим старую задачу про пиццу: 10 друзей собрались на вечеринку и заказали пиццу. Количество съеденных кусков пиццы каждым соответственно:
1, 2, 2, 2, 3, 3, 4, 4, 5, 20.
Найдите медиану.

Медиана

Пример 2. Решим старую задачу про пиццу: 10 друзей собрались на вечеринку и заказали пиццу. Количество съеденных кусков пиццы каждым соответственно:
1, 2, 2, 2, 3, 3, 4, 4, 5, 20.
Найдите медиану.

Ответ: 3 шт.

Медиана

Вывод: в отличие от среднего арифметического медиана **хорошо справляется с разбросом в значениях**, поэтому её чаще используют. Также медиана разбивает данные на две группы, состоящие из одинакового количества элементов.

Средние значения, например уровень дохода или цена на недвижимость, часто вычисляются именно по медиане, потому что в этом случае важен средний уровень доходов **большей части** населения.

МФТИ

Мода

Мода – это самое часто встречающееся значение в числовом или нечисловом ряду данных.

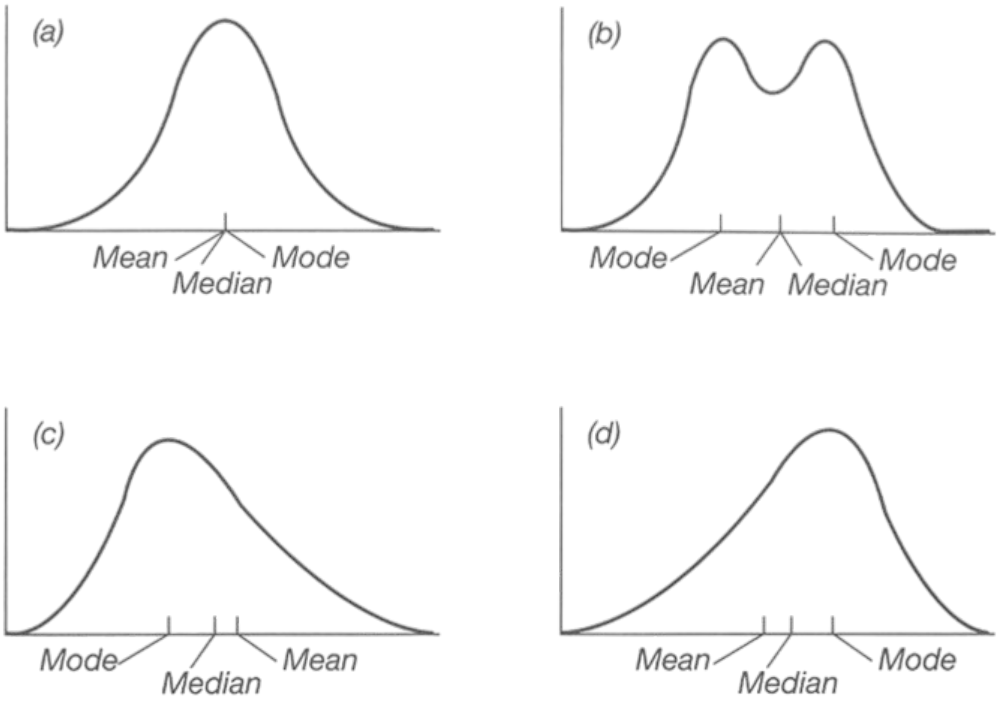
Пример: допустим, Вы решили полететь в отпуск со своими друзьями и решили выбрать месяц. Вы провели опрос, кому в какой месяц было бы удобнее:

- Дима и Серёжа свободны в феврале
- Толя может взять отпуск в любой месяц года
- Ксюша может только в июне
- Какой месяц наилучший для общего отпуска?

Мода

Вывод: чаще всего мода используется в нечисловых рядах. Например самая популярная порода собак, фильм года, лучшие рестораны вычисляются именно модой.

Симметричные и несимметричные распределения



На графиках a и b – симметричные распределения, c и d – несимметричные.

b – также мультимодальное распределение!

Чем интересна мультимодальность?

Мультимодальные распределения неудобны по нескольким причинам:

- Среднее и медиана вводят в заблуждение.
Между пиками может быть "яма", но именно туда попадёт среднее значение. Оно не отражает типичного случая.

- Они сигнализируют о скрытых группах.
Возможно, данные на самом деле состоят из разных подгрупп (например, мужчины и женщины, дети и взрослые), и анализ без учёта этих подгрупп может быть некорректным.

- Плохо подходят для ML-моделей, предполагающих "одну гору".
Например, линейные модели или распределения, близкие к нормальному, не справятся.

Другие меры центральной тенденции

Применяются реже, но также могут быть полезны следующие меры центральной тенденции:

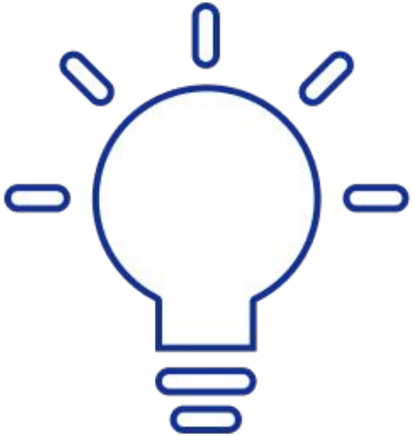
среднее геометрическое $\bar{x}_{\text{геом.}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$

среднее гармоническое $\bar{x}_{\text{гарм.}} = \dfrac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$

средневзвешенное и другие $\bar{x}_{\text{взвеш.}} = \dfrac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n}$

МФТИ

Введение в статистические гипотезы



Совокупность VS выборка

Генеральная совокупность – это всё множество интересующих нас объектов.

Выборка — случайное подмножество из генеральной совокупности.

Часто множество объектов генеральной совокупности настолько большое, что мы не можем посчитать все объекты ⇒ приходится делать расчёты не на ген. совокупности, а на выборке.

Совокупность VS выборка

В таком случае важно, чтобы выборка была **репрезентативна** генеральной совокупности.

Репрезентативность – это насколько хорошо выборка отражает всю генеральную совокупность (всю популяцию, которую мы хотим изучить).

Плохой пример: “Большинство детей имеют математический склад ума”, но исследование проведено среди учеников физ-мат лицея.

Репрезентативность

На какие параметры обратить внимание?

Выборка (то, что измеряем)	Генеральная совокупность (то, что хотим узнать)
Выборочное среднее $\bar{x}$	Истинное среднее μ (мат. ожидание)
Выборочное стандартное отклонение s	Истинное стандартное отклонение σ
Выборочная пропорция $\hat{p}$	Истинная пропорция p

Достаточный объём: чем больше, тем лучше (точно ≥ 30, но зависит от задачи).
Случайность: каждый элемент должен иметь равный шанс попасть в выборку.
Разнообразие: важные группы (пол, возраст, регион и т.п.) должны быть представлены.
Отсутствие смещения: или систематических искажений.

Статистические гипотезы

Статистическая гипотеза — это некоторое утверждение о параметрах генеральной совокупности, которое мы будем проверять на основе данных выборки.

Например, утверждение «новый дизайн сайта влияет на средний чек покупки (на всей аудитории)», которое мы будем проверять на выборке – это статистическая гипотеза.

Алгоритм проверки гипотезы

Состоит из 6-ти этапов:

- Построение нулевой и альтернативной гипотез.
- Выбор уровня значимости.
- Сбор данных для проверки гипотезы.
- Выбор статистического теста.
- Проведение статистического теста, вычисление p-value.
- Сравнение p-value с уровнем значимости и вывод, **отклонить или не отклонить нулевую гипотезу**.

Алгоритм проверки гипотезы

Состоит из 6-ти этапов:

- **Построение нулевой и альтернативной гипотез.**
- Выбор уровня значимости.
- Сбор данных для проверки гипотезы.
- Выбор статистического теста.
- Проведение статистического теста, вычисление p-value.
- Сравнение p-value с уровнем значимости и вывод, **отклонить или не отклонить нулевую гипотезу**.

Нулевая и альтернативная гипотезы

Это конкурирующие и противоположные друг другу утверждения.

В качестве нулевой гипотезы (H_0) обычно берётся утверждение об **отсутствии эффекта**, т.е. тех изменений, которых мы ожидали, нет.
Пример нулевой гипотезы: новый дизайн не влияет на средний чек покупки.

В качестве альтернативной гипотезы (H_1) берётся утверждение **о наличии эффекта** (обратное нулевой гипотезе), т.е. те изменения, которые мы ожидали, есть.
Пример альтернативной гипотезы: новый дизайн влияет на средний чек покупки.

Одно/двусторонние гипотезы

То, что было на предыдущем слайде – гипотезы для **двувыборочных тестов**:

- Альтернативная гипотеза вида H_1: $\mu_1 \neq \mu_2$ называется **двусторонней**.
- Альтернативные гипотезы H_1: $\mu_1 < \mu_2$ и H_1: $\mu_1 > \mu_2$ — **односторонними**.
- При этом гипотеза H_1: $\mu_1 < \mu_2$ называется **левосторонней**, а гипотеза H_1: $\mu_1 > \mu_2$ — **правосторонней**.

МФТИ

Одновыборочные тесты

Могут быть **одновыборочные** тесты, где сравниваем средние не для двух групп, а для одной (с конкретным значением).

- H_0: $\mu \leq 300$
- H_1: $\mu > 300$

Одновыборочные тесты: пример

Аналитик в кулинарной компании знает, что содержание натрия должно быть не более 300 мг на 100 г соуса, и хочет проверить, не превышено ли оно.

С точки зрения статистики это утверждение будет значить, что на ген. совокупности среднее (мат. ожидание) содержание натрия (μ), должно быть меньше или равно 300 мг на 100 г.

Тогда нулевая и альтернативная гипотезы будут следующими:

- H_0: $\mu \leq 300$
- H_1: $\mu > 300$

Алгоритм проверки гипотезы

Состоит из 6-ти этапов:

- ~~Построение нулевой и альтернативной гипотез.~~
- **Выбор уровня значимости.**
- Сбор данных для проверки гипотезы.
- Выбор статистического теста.
- **Проведение статистического теста, вычисление p-value.**
- **Сравнение p-value с уровнем значимости и вывод, отклонить или не отклонить нулевую гипотезу.**

МФТИ

Ошибки при проверке гипотез

Ошибка I (первого) рода: отклонение нулевой гипотезы, когда она на самом деле верна. Эта ошибка также называется α-ошибкой, или ложноположительным результатом.

Ошибка II (второго) рода: неотклонение нулевой гипотезы, когда она на самом деле ложна. Эта ошибка также называется β-ошибкой, или ложноотрицательным результатом.

H_0	ЛОЖНАЯ	ИСТИННАЯ
Отклоняется	**Ошибки нет**	**Ошибка I рода**
Не отклоняется	**Ошибка II рода**	**Ошибки нет**

МФТИ

Ошибки при проверке гипотез

Вспомним пример про новый дизайн сайта…

- Какая нулевая гипотеза?
- Какая альтернативная?

МФТИ

Ошибки при проверке гипотез

Вспомним пример про новый дизайн сайта...

- Какая нулевая гипотеза? **Новый дизайн не повлиял на ср. чек.**
- Какая альтернативная? **Новый дизайн повлиял на ср. чек.**

Заполним табличку:

	Описание	**Последствия**
Ошибка I рода	???	???
Ошибка II рода	???	???

Ошибки при проверке гипотез

Вспомним пример про новый дизайн сайта...

- Какая нулевая гипотеза?
- Какая альтернативная?

Заполним табличку:

	Описание	**Последствия**
Ошибка I рода	Сделан вывод, что новый дизайн влияет на средний размер чека, когда на самом деле она не влияет.	Потрачены ресурсы на запуск нового дизайна, но прибыль не получена.
Ошибка II рода	Сделан вывод, что новый дизайн не влияет на средний размер чека, когда на самом деле она влияет.	Новый дизайн не запущен, прибыль от его использования не получена.

Ошибки при проверке гипотез

А на этом рисунке какого рода случились ошибки на обеих картинках?:)

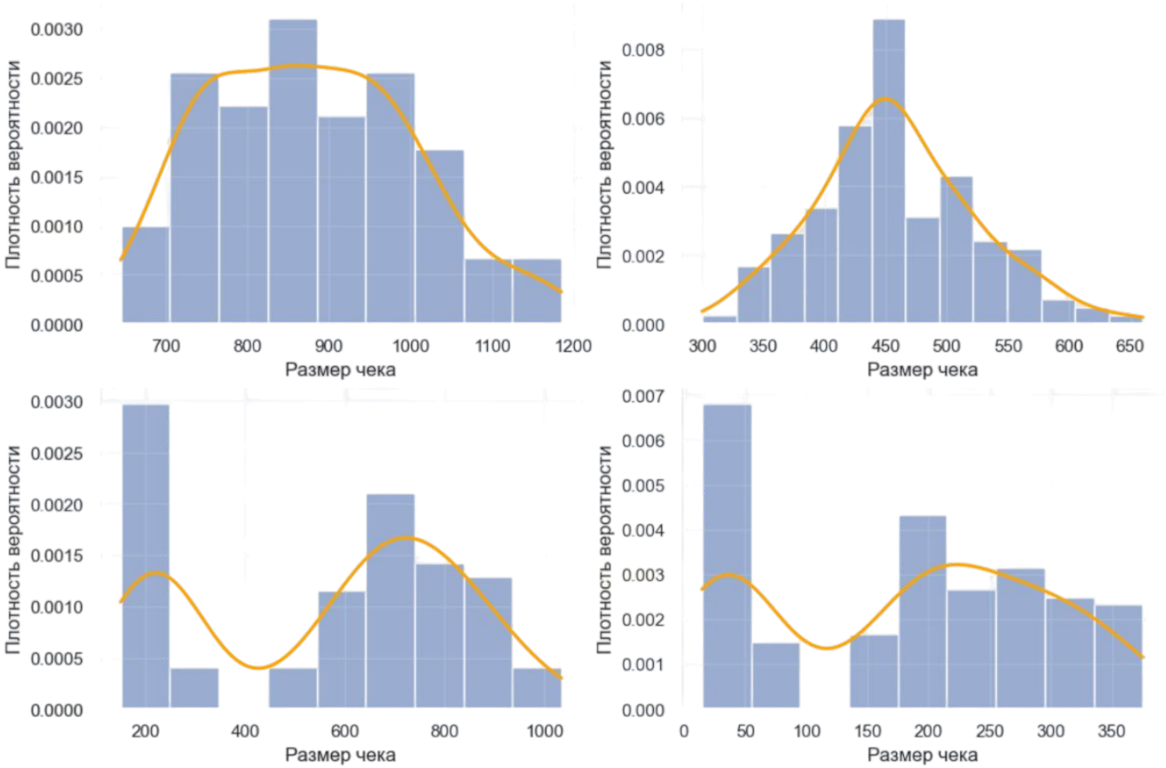
Подсказка:

H_0 всегда берём как “отсутствие эффекта” – т.е. в данном случае H_0 == “беременности нет”.

Функции плотности распределения

График ФПР построен так, что **вся площадь под кривой равна 1** — это вероятность всех возможных значений размера чека.

Это следует из **основного правила теории вероятностей**: сумма вероятностей всех исходов (всех возможных значений чека) всегда равна 1.



МФТИ

p-value

Уровень статистической значимости (p-value) – это вероятность получить наблюдаемые или более экстремальные значения при условии, что нулевая гипотеза верна.

Пример: «Средний чек покупки в онлайн-магазине больше 500 рублей?».

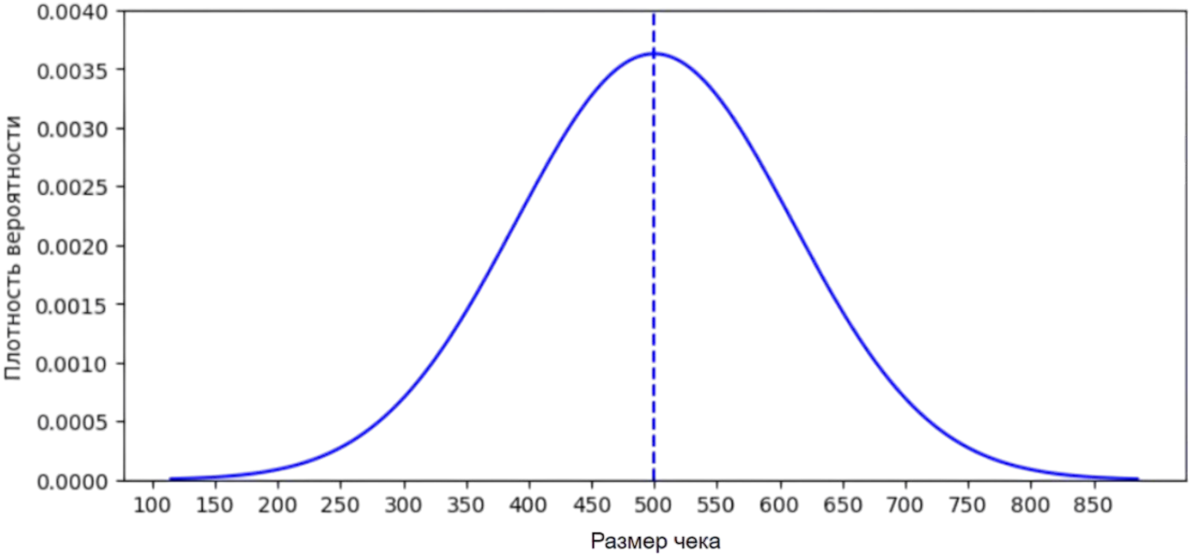
Гипотезы будут следующими:

НУЛЕВАЯ ГИПОТЕЗА	**АЛЬТЕРНАТИВНАЯ ГИПОТЕЗА**
H_0: $\mu_1 \leq 500$	H_1: $\mu_1 > 500$
Средний чек покупки меньше или равен 500 рублей.	Средний чек покупки больше 500 рублей.

p-value

Неизвестно, какой будет **истинная форма плотности вероятности чека**, которая получается на основе **генеральной совокупности**.

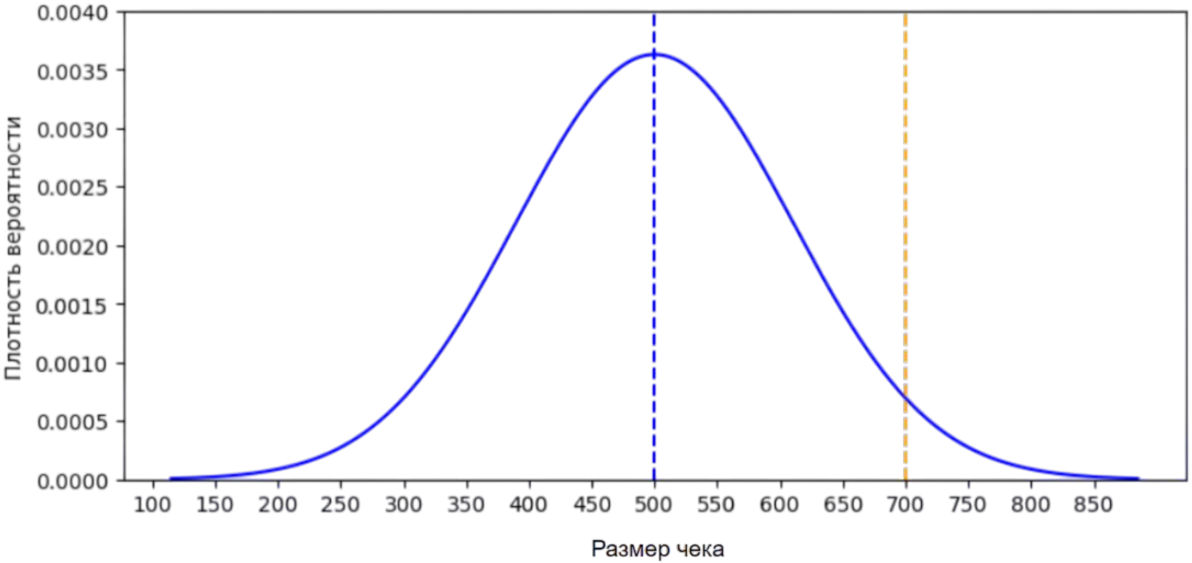
Обычно предполагают какую-то форму из стандартных законов: например, нормальное распределение.



Синяя пунктирная линия соответствует среднему чеку в 500 рублей, вокруг которого мы строим нулевую гипотезу.

p-value

Пусть на выборке мы получили средний размер чека 700 рублей. Отметим это значение на графике **оранжевой линией**.

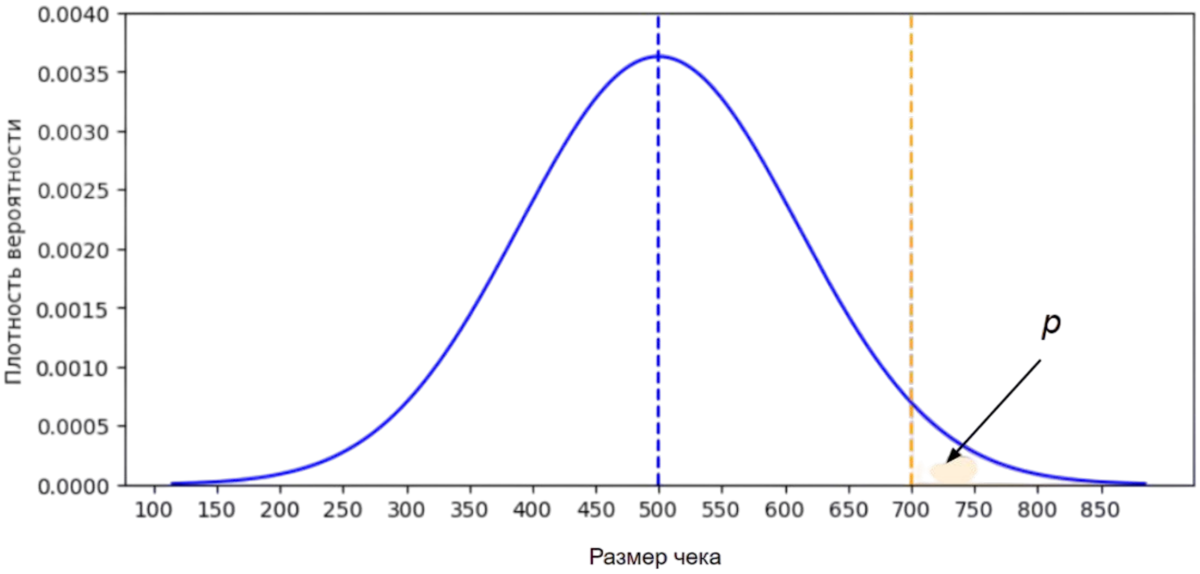


МФТИ

p-value

Какова вероятность получить такой же (700 рублей) или ещё больший средний чек при условии, что средний чек равен 500 рублям?

Этой вероятностью будет площадь под кривой от 700 к большим значениям (p-value = p).



МФТИ

p-value

Чем больше было бы число, которое мы получили на выборке:

- тем меньше была бы площадь под кривой, т. е. меньше полученное значение p-value;
- тем меньше полученное наблюдение соответствовало бы нулевой гипотезе, согласно которой среднее значение чека – 500 рублей или меньше.

Важно! p-value – это именно вероятность получить наблюдаемые или более экстремальные значения, если верна нулевая гипотеза, а не вероятность верности нулевой гипотезы.

МФТИ

Уровень значимости

Уровень значимости – это пороговая вероятность допустить ошибку I рода (вероятность отклонить нулевую гипотезу, когда она верна). Обычно это число обозначается за α.

- Если полученное p-value ниже уровня значимости ($p < \alpha$), результаты статистически значимы и согласуются с альтернативной гипотезой.
- Если полученное p-value выше уровня значимости ($p > \alpha$), результаты считаются статистически незначимыми, нельзя отвергнуть нулевую гипотезу.

```
if p_value < alpha:
    print('Принимаем альтернативную гипотезу')
else:
    print('Принимаем нулевую гипотезу')
```

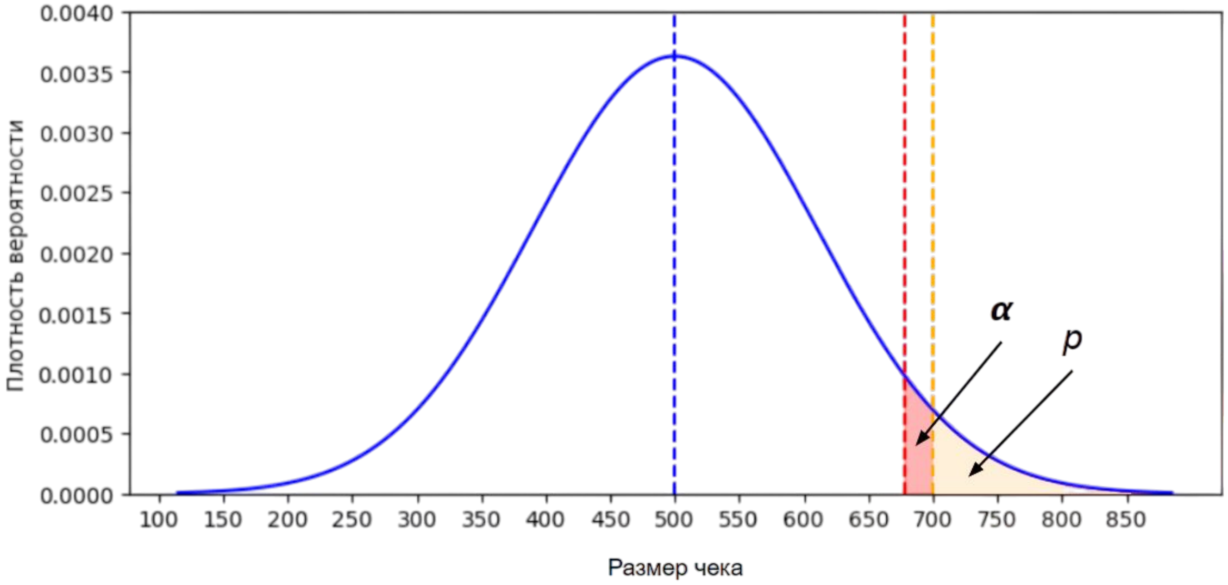
Уровень значимости

Иллюстрация принятия решения для проверки гипотезы:

МФТИ

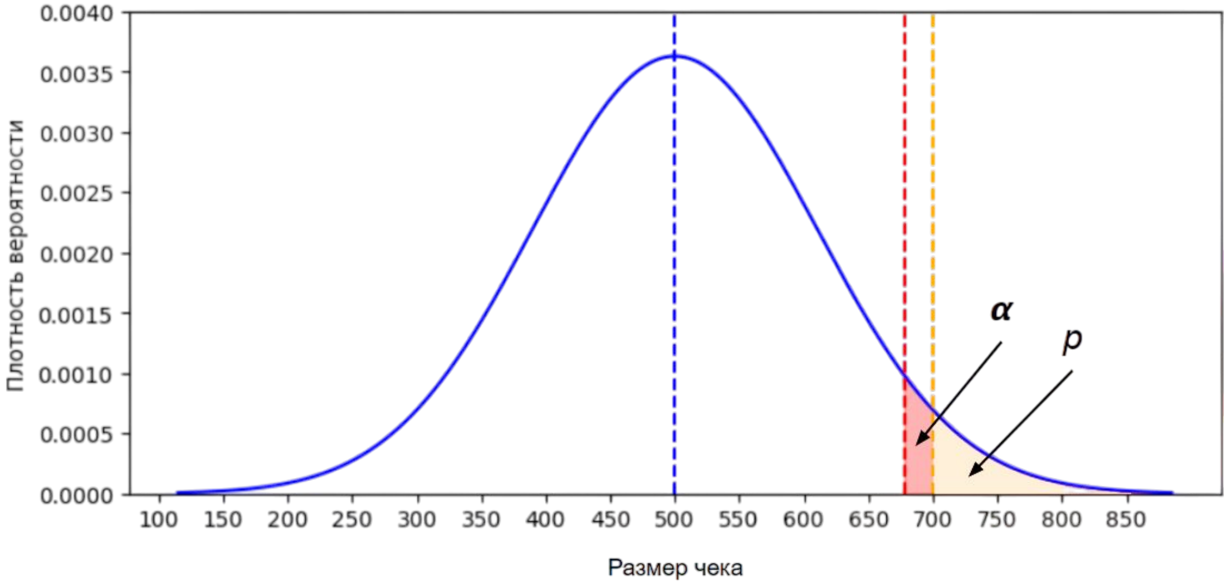
Тренировка

В таком случае мы отвергаем или принимаем нулевую гипотезу?



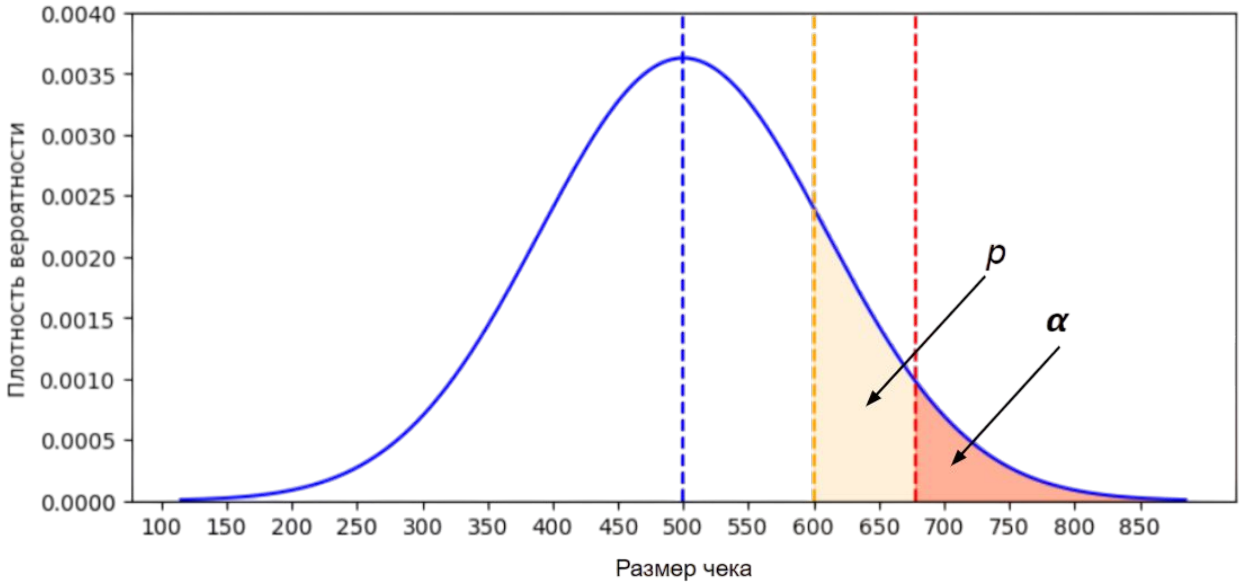
Тренировка

В таком случае мы отвергаем или принимаем нулевую гипотезу?
Отвергаем



Тренировка

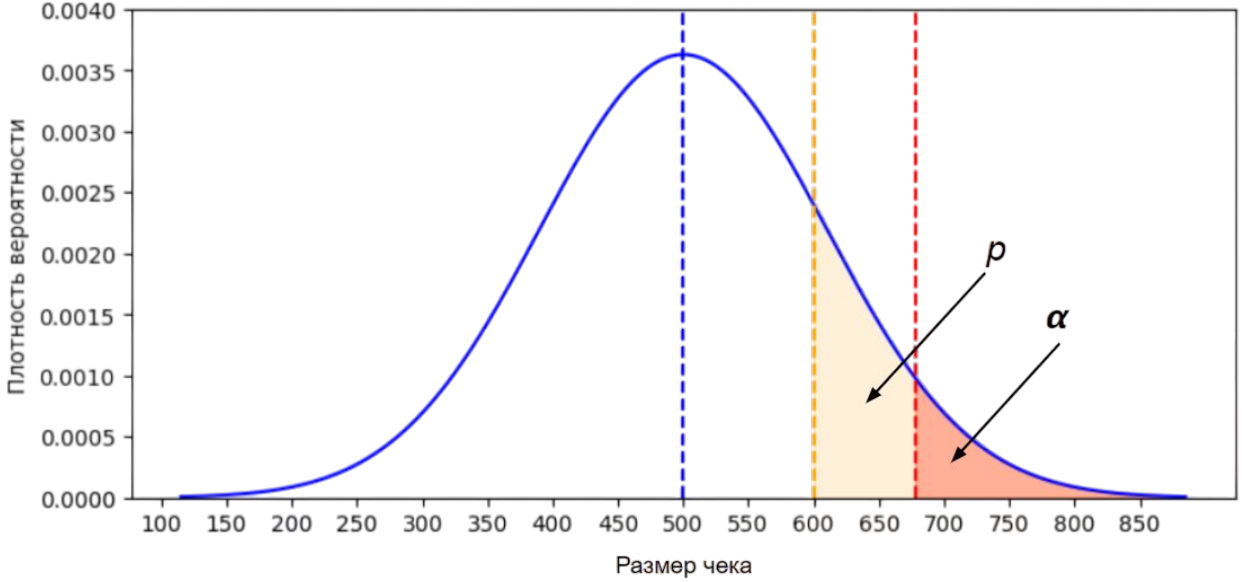
А в таком?



МФТИ

Тренировка

А в таком? Принимаем нулевую гипотезу



Левосторонние и правосторонние гипотезы

На прошлых рисунках были правосторонние гипотезы. С тем же успехом можно проверять и левосторонние:

Правосторонняя:
Интернет-магазин внедрил новую систему рекомендаций и хочет доказать, что средний чек вырос.

Левосторонняя:
Производитель смартфонов хочет проверить, ухудшилась ли автономность батареи после обновления прошивки.

МФТИ

Двусторонние гипотезы

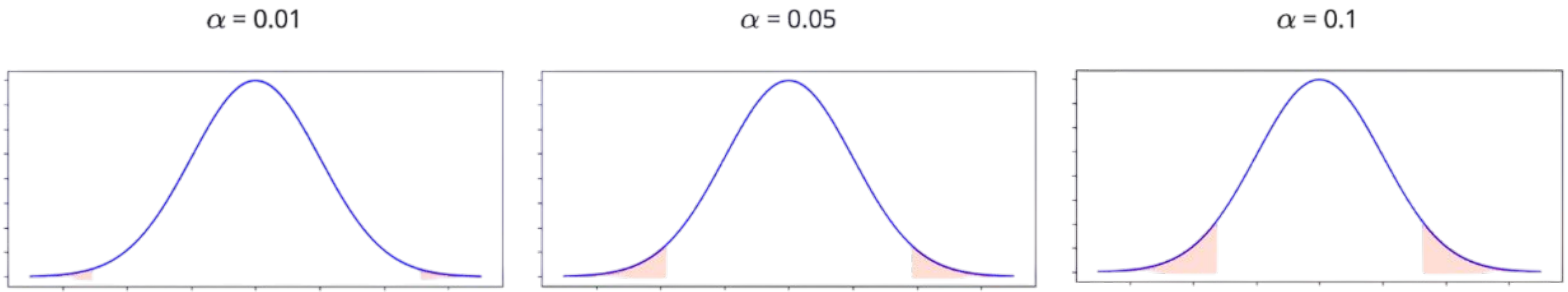
Левосторонняя + правосторонняя = двусторонняя гипотеза!

Двусторонняя:
Компания выпускает новую партию товара и хочет убедиться, что средний вес упаковки не отличается от нормы (например, 1 кг). И увеличение, и уменьшение веса нежелательны.

МФТИ

Настройка уровня значимости

В зависимости от важности вопроса можно настраивать уровень значимости (допустимую ошибку I рода):



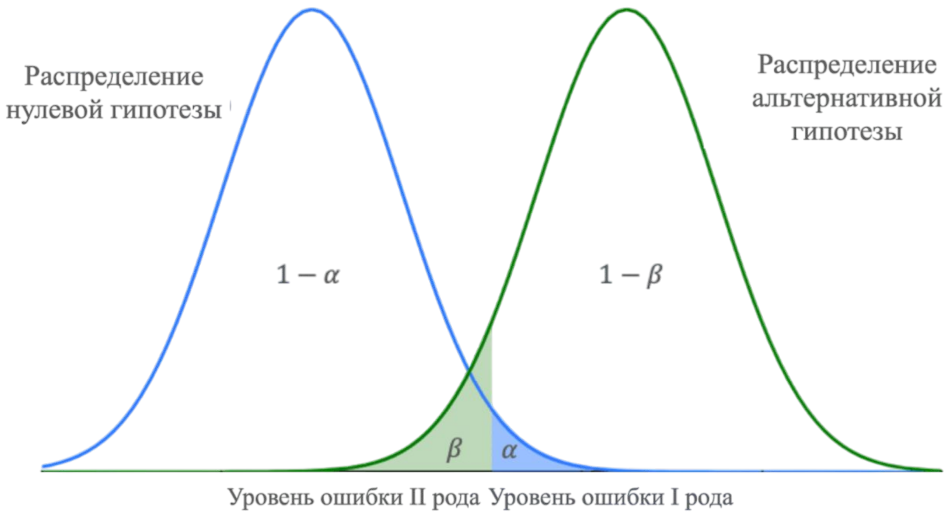
Визуализация баланса ошибок

Построим распределения для H_0 и H_1:

- Под H_0 предполагается, что *средняя выручка в группе 1 (тестовой) == средней в группе 2 (контрольной)*.
  Тогда разница средних $\bar{X}_{test} - \bar{X}_{control}$ будет иметь нормальное распределение с центром в нуле (если выборка большая, по ЦПТ).

- Под H_1 предполагается, что *средняя выручка в тесте != средней в контроле*.
- Этому соответствует нормальное распределение, но со сдвинутым центром.

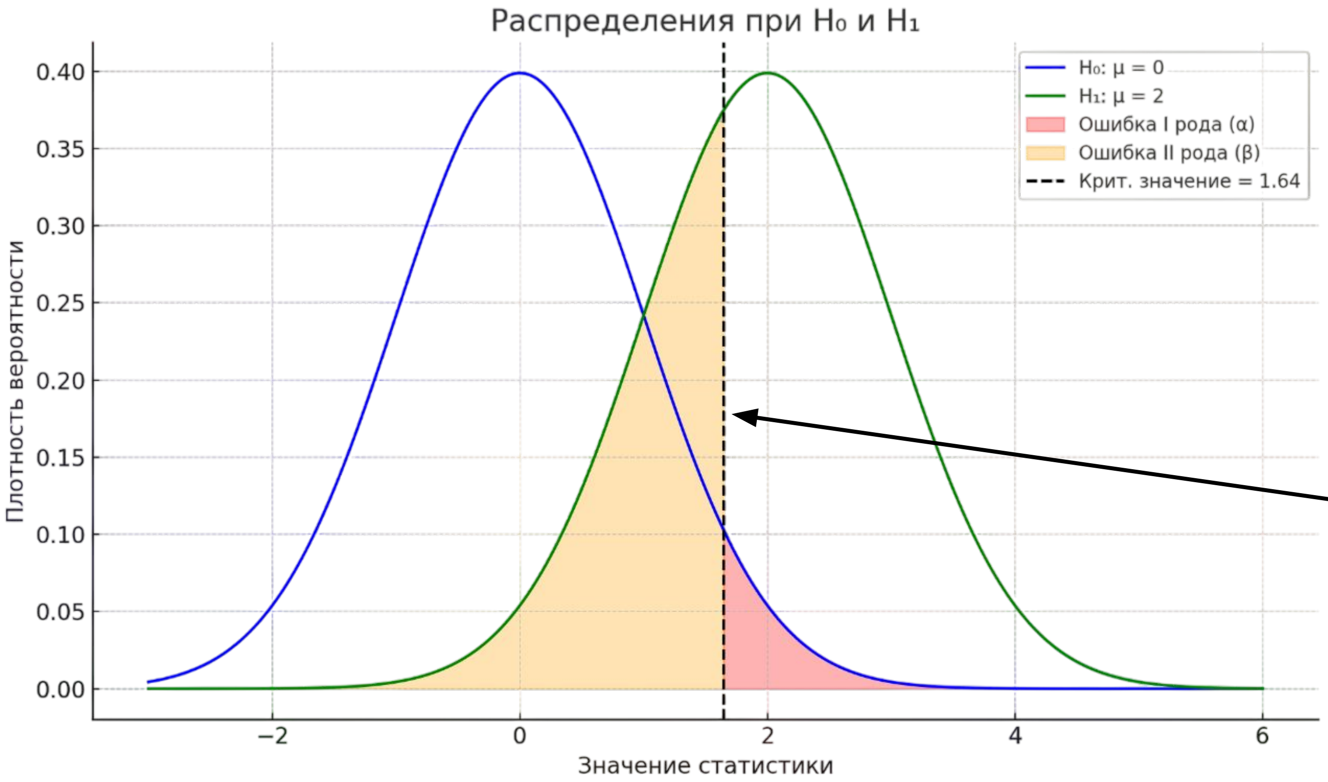
Баланс ошибок I и II рода



Возникает вполне логичный вопрос: почему бы не принять α = 0, тем самым исключив возможность ошибки?

Увы, всё не так просто. Дело в том, что, **помимо α, мы должны учитывать и β – вероятность ошибки II рода** (риск ложноотрицательного результата). Две этих величины противоречат друг другу.

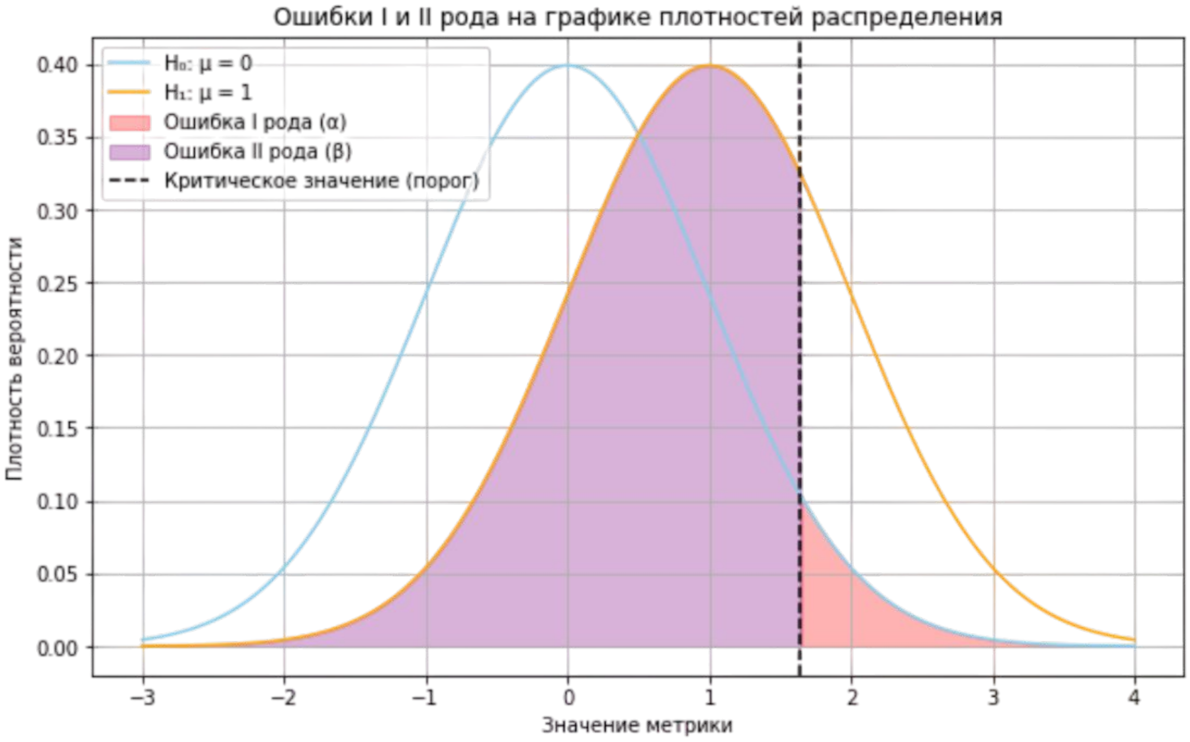
Визуализация баланса ошибок



порог для
ошибки I рода

МФТИ

Визуализация баланса ошибок



здесь ситуация похуже:)

Баланс ошибок I и II рода

Пример 1: Вы супер-осторожны, чтобы не внедрить бесполезную модель. Для этого ставите α = 0.01 вместо 0.05.

Что это значит: отвергнуть H_0 теперь труднее → чаще оставляете H_0.

Итог: снижается α, но растёт β (можно упустить что-то действительно полезное).

Пример 2: Вы не хотите упустить реально хорошую модель. Делаете тест более чувствительным, чтобы быстрее заметить эффект (например, α = 0.1 вместо α = 0.05).

Что это значит: принять H_0 теперь труднее → чаще оставляете H_1.

Итог: снижается β, но растёт α (больше риск внедрить ложноположительный результат).

Баланс ошибок I и II рода

Итого:

- Уменьшим α → увеличим β.
- НО увеличим выборку → уменьшим β, сохранив α!

Почему так?

- больше данных ⇒ меньшая дисперсия оценок ⇒ легче выявить настоящий эффект!

Пример:

- С 10 зернами вес скачет: ±1 г.
- С 1000 зернами вес точнее: ±0.1 г.

Если зерно изменится на 0.5 г, в первом случае мы этого не заметим, во втором – заметим.
Здесь увеличение выборки == более точные весы.

Алгоритм проверки гипотезы

Состоит из 6-ти этапов:

- ~~Построение нулевой и альтернативной гипотез.~~
- ~~Выбор уровня значимости.~~
- **Сбор данных для проверки гипотезы.**
- Выбор статистического теста.
- ~~Проведение статистического теста, вычисление p-value.~~
- ~~Сравнение p-value с уровнем значимости и вывод,~~ ~~**отклонить или не отклонить нулевую гипотезу**~~.

МФТИ

Подготовка данных

Главные правила:

- все выборки должны быть **репрезентативны** генеральной совокупности
- между собой выборки должны быть **однородны** по ключевым признакам

Идеи для проверок однородности:

- исходные характеристики клиентов (регион, пол, возраст)
- показатели целевых метрик на исторических данных (CR, CTR, GMV, AOV...)
- показатели метрик, которые косвенно могут повлиять на целевые (наличие каких-либо продуктов, эффект каннибализации...)

Подготовка данных

Плохой пример:

Проверяем новый дизайн сайта, сравнивая две выборки – одна его не видела, другая видела.

При этом в одной выборке только женщины, в другой – только мужчины; среди всех пользователей сайта ж:м = 6:4.

Или так: в одной выборке только постоянные посетители; в другой – неактивные, которые заходили на сайт более месяца назад.

Алгоритм проверки гипотезы

Состоит из 6-ти этапов:

- ~~Построение нулевой и альтернативной гипотез.~~
- ~~Выбор уровня значимости.~~
- ~~Сбор данных для проверки гипотезы.~~
- **Выбор статистического теста.**
- ~~Проведение статистического теста, вычисление p-value.~~
- ~~Сравнение p-value с уровнем значимости и вывод, **отклонить или не отклонить нулевую гипотезу**.~~

МФТИ

Что такое статистический критерий?

Статистический критерий — это правило или метод, с помощью которого мы принимаем решение о том, насколько данные соответствуют некоторому предположению (гипотезе).

Или инструмент, который помогает ответить на вопрос:
«Могли ли полученные данные возникнуть случайно, если бы гипотеза была верна?»

Что такое статистический критерий?

Расчёт везде аналогичный:

1. Сначала по особой формуле вычисляется значение **статистики**. Например, такое:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

2. Далее ему **сопоставляется p-value** (по распределению, делается по таблице).
3. Теперь только остается только **сравнить p-value с α!**

Хорошая новость:

Python делает все описанные выше шаги самостоятельно, поэтому за нами только выбор подходящего стат. критерия и сравнение полученного p-value с заданным уровнем значимости (подробнее математический расчёт изучите на статистике).

Алгоритм выбора критерия

Шаг 1: Какой тип переменной сравниваем?

1. Количественная (числовая):
   - *Примеры: выручка, CTR, средний чек.*
2. Категориальная (дискретная):
   - *Примеры: конверсия (0/1), победил ли вариант A или B, категория товара.*

Шаг 2: Сколько групп сравнивается?

- 1 группа
- 2 группы.
- >2 групп.

Алгоритм выбора критерия

Шаг 3: Группы зависимы или независимы?

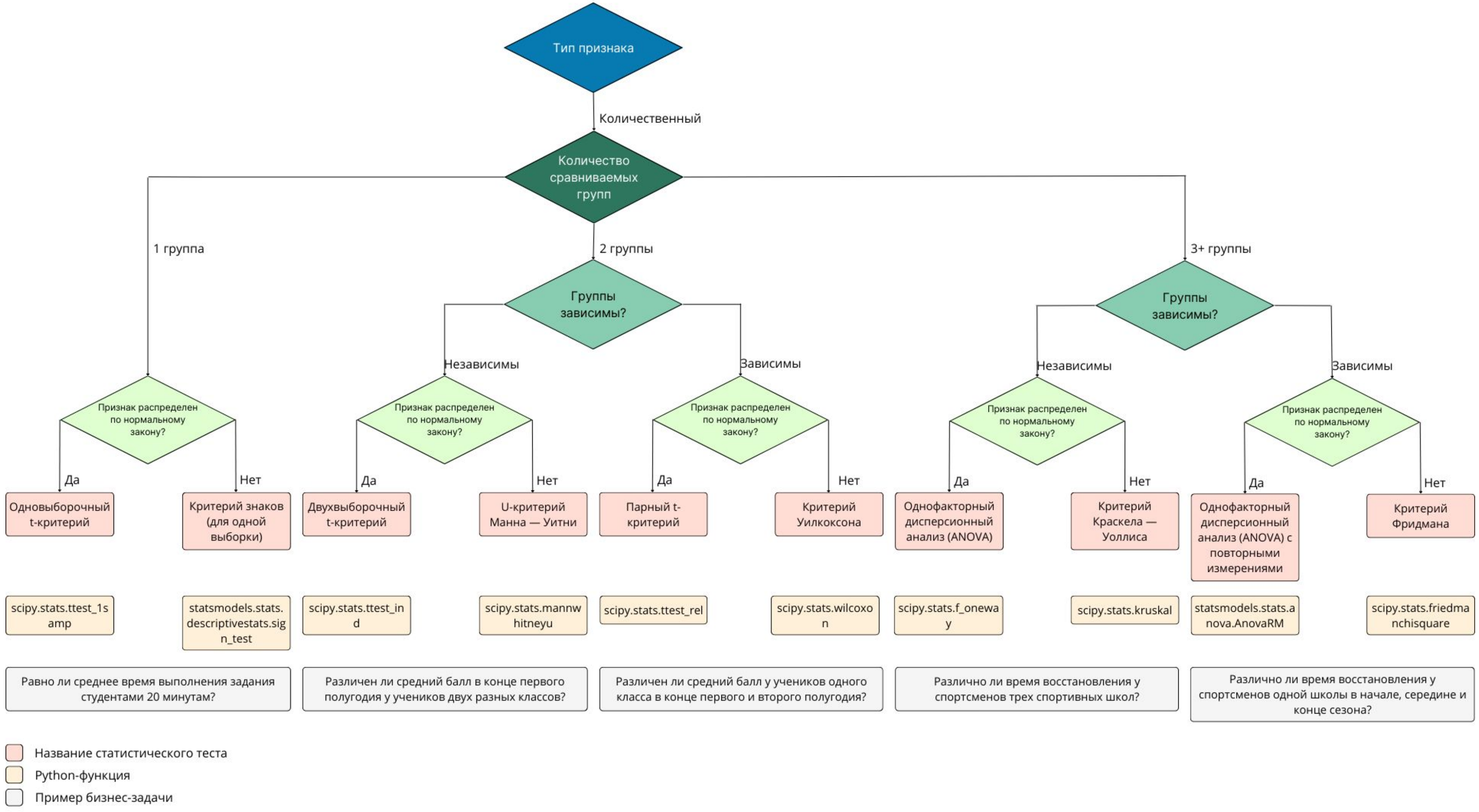
- Независимые группы: *сравниваются разные пользователи в одно и то же время.*
- Зависимые группы (парные): *один и тот же пользователь “до” и “после”, или matched-пары.*

Шаг 4: Есть ли нормальность распределения?

Для количественных переменных проверяем нормальность *(например, с помощью теста Шапиро-Уилка или визуально).*

Для количественных переменных

Кол-во групп	Зависимость	Нормальность	Критерий	Что сравнивается?
2	Независимые	Да	**t-критерий Стьюдента**	Разница средних значений между двумя группами
2	Независимые	Нет	**Манна-Уитни (U-критерий)**	Ранги (порядок значений) между группами
2	Зависимые	Да	**Парный t-критерий**	Разница средних в парных наблюдениях
2	Зависимые	Нет	**Знаковый тест / Вилкоксона**	Ранги разностей между парами
>2	Независимые	Да	**ANOVA (дисперсионный анализ)**	Сравнение дисперсий (между группами / внутри групп)
>2	Независимые	Нет	**Критерий Краскела-Уоллиса**	Ранги между группами
>2	Зависимые	Да	**Repeated Measures ANOVA**	Дисперсии между условиями у одних и тех же объектов
>2	Зависимые	Нет	**Критерий Фридмана**	Ранги условий внутри каждого объекта


Тип признака
Количественный
Количество сравниваемых групп
1 группа
2 группы
3+ группы
Группы зависимы?
Независимы
Зависимы
Признак распределен по нормальному закону?
Да
Нет
Одновыборочный t-критерий
Критерий знаков (для одной выборки)
Двухвыборочный t-критерий
U-критерий Манна — Уитни
Парный t-критерий
Критерий Уилкоксона
Однофакторный дисперсионный анализ (ANOVA)
Критерий Краскела — Уоллиса
Однофакторный дисперсионный анализ (ANOVA) с повторными измерениями
Критерий Фридмана
scipy.stats.ttest_1samp
statsmodels.stats.descriptivestats.sign_test
scipy.stats.ttest_ind
scipy.stats.mannwhitneyu
scipy.stats.ttest_rel
scipy.stats.wilcoxon
scipy.stats.f_oneway
scipy.stats.kruskal
statsmodels.stats.anova.AnovaRM
scipy.stats.friedmanchisquare
Равно ли среднее время выполнения задания студентами 20 минутам?
Различен ли средний балл в конце первого полугодия у учеников двух разных классов?
Различен ли средний балл у учеников одного класса в конце первого и второго полугодия?
Различно ли время восстановления у спортсменов трех спортивных школ?
Различно ли время восстановления у спортсменов одной школы в начале, середине и конце сезона?
Название статистического теста
Python-функция
Пример бизнес-задачи

Для категориальных переменных

Кол-во групп	Зависимость	Тип данных	Критерий	Что сравнивается?
2	Независимые	2x2 таблица	**χ^2-критерий (хи-квадрат)**	Частоты фактов vs ожиданий (от H_0)
2	Независимые	Малые выборки	**Точный критерий Фишера**	Вероятность распределения как или более экстремального
2	Зависимые	Двоичный исход	**Критерий МакНемара**	Разница между количеством переходов A→B и B→A
>2	Независимые	>2 категорий	**χ^2-критерий**	Частоты категорий vs ожидания
>2	Зависимые	>2 условий	**Критерий Кохрана / Q-тест**	Пропорции категорий по условиям

Алгоритм проверки гипотезы

Состоит из 6-ти этапов:

- ~~Построение нулевой и альтернативной гипотез.~~
- ~~Выбор уровня значимости.~~
- ~~Сбор данных для проверки гипотезы.~~
- ~~Выбор статистического теста.~~
- ~~Проведение статистического теста, вычисление p-value.~~
- ~~Сравнение p-value с уровнем значимости и вывод,~~ ~~**отклонить или не отклонить нулевую гипотезу**~~.

В следующий раз – практика на Python!

МФТИ

Спасибо за внимание!

МФТИ