УНИВЕРСИТЕТ ИННОПОЛИС

Основы статистики. p-value, t-value. Проверка гипотез

Воробьёва Мария

- maria.vorobyova.ser@gmail.com
- @SparrowMaria

Статистика. Факты

Математи́ческая стати́стика — наука, разрабатывающая математические методы систематизации и использования статистических данных для научных и практических выводов.

Во многих своих разделах математическая статистика опирается на теорию вероятностей, дающую возможность оценить надежность и точность выводов, делаемых на основании ограниченного статистического материала

Статистика

Генеральная совокупность — это полный набор всех элементов, которые исследуют в рамках задачи

Выборка — это отдельный набор элементов, отобранных из генеральной совокупности некоторым случайным процессом

Выборка должна обладать следующими свойствами:
- Репрезентативность
- Достаточный размер выборки
- Случайный отбор

Статистика. Факты

- анализируя выборку мы делаем выводы для генеральной совокупности
- выборка - случайная
- для одной генеральной совокупности **можно** создать много разных выборок, но обычно доступна только 1!

Статистика

Меры центральной тенденции:

- выборочное среднее
- выборочная медиана
- выборочная мода
- выборочные квантили

Статистика

Что такое квантиль? Квантиль задает значение, ниже которого находится определенная доля данных в распределении.

1) отсортируем ряд
2) разделим ряд на две части. Точка, которой мы делим - это и есть медиана. То есть ниже медианы 50% данных
3) Если мы ряд разделим на 3 равные части - то получим квартили. Ниже 1 квартиля 25% всех данных, ниже 2 квартиля - 50%, ниже 3 квартиля 75%

Получается, Квантиль задает значение, ниже которого находится определенная доля данных в распределении.
Медиана = 0.5 квантиль = 50% персентиль = 2 квартиль - это значение,

Статистика

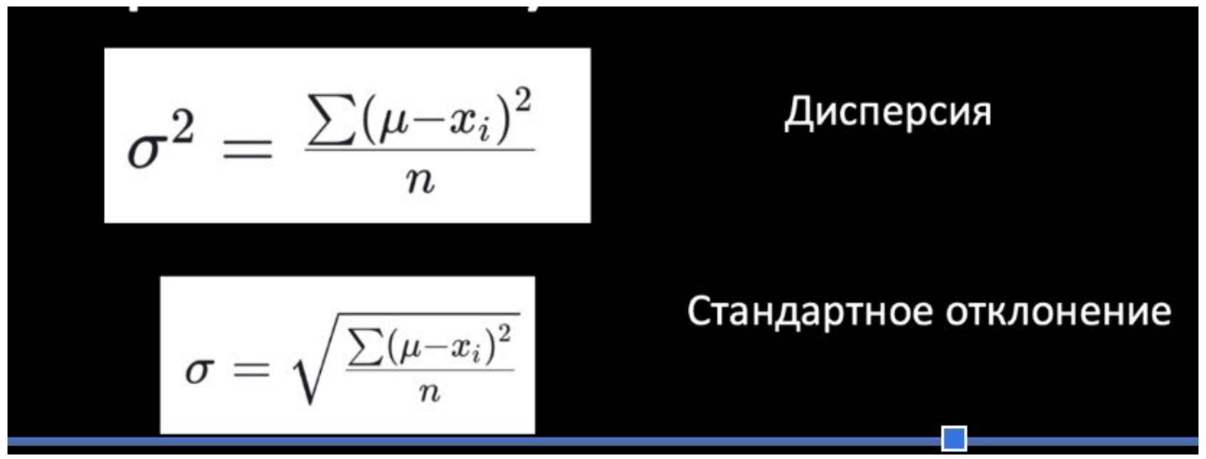
Меры изменчивости:

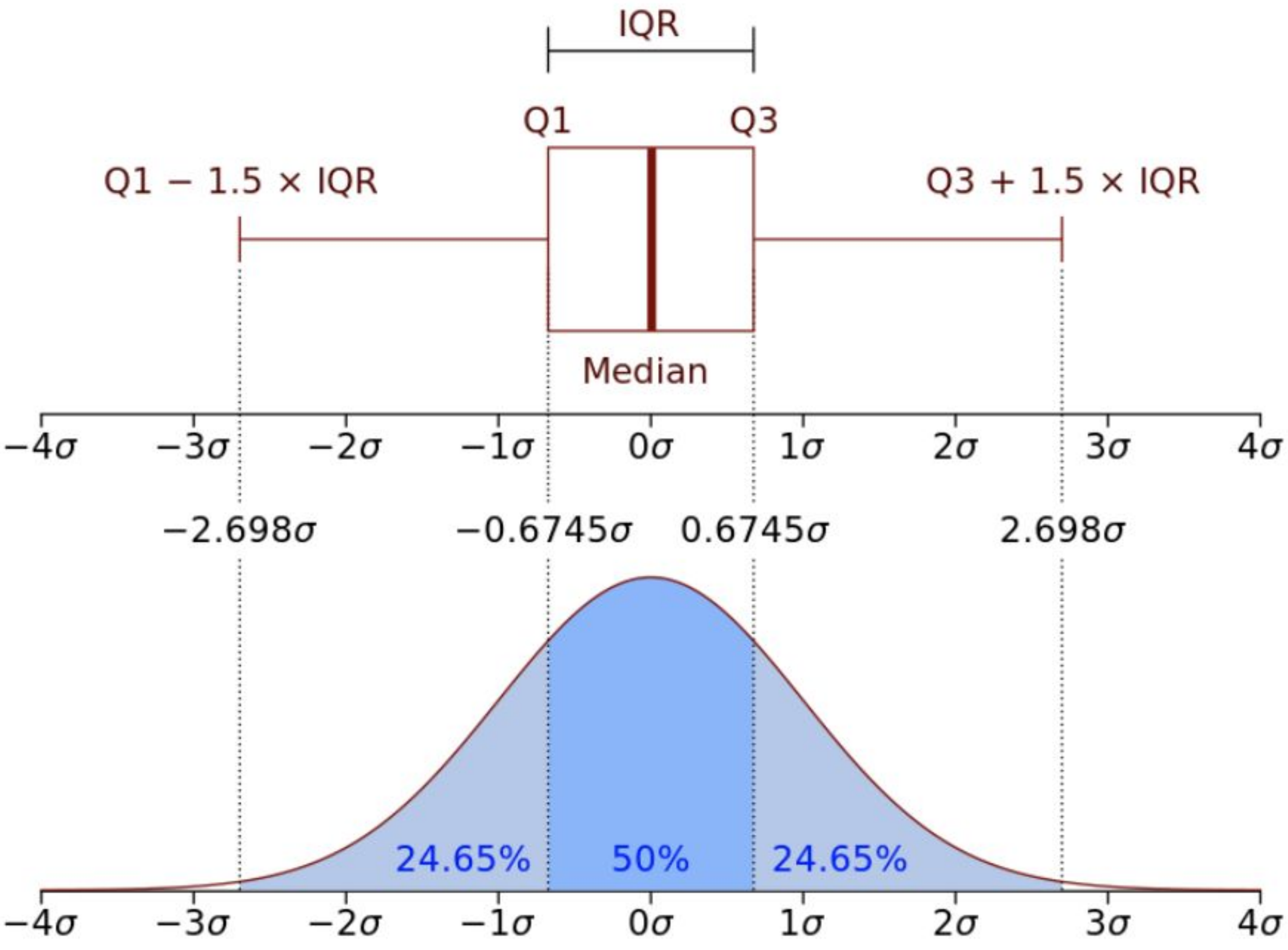
- выборочный размах
- выборочный межквартильный размах
- выборочная дисперсия
- выборочное стандартное отклонение

Статистика

Меры изменчивости:

- выборочный размах
- выборочный межквартильный размах
- выборочная дисперсия
- выборочное стандартное отклонение





Статистика. Доверительный интервал

Мы рассмотрели методы оценивания, позволяющие находить **точечные оценки** параметров.

Помним, что для одной генеральной совокупности можно построить несколько выборок

Допустим, мы взяли выборку и получили выборочное среднее, равное 501. Это наша оценка среднего всей генеральной совокупности, но как понять, насколько она точна?

Может быть, случайно попалась выборка со средним, очень далёким от истинного среднего? Очень маловероятно, но возможно.

Статистика. Доверительный интервал

γ**-доверительный интервал** - это интервал, внутри которого с заданной вероятностью γ находится истинное значение параметра.

Доверительный интервал (англ. confidence interval) — отрезок числовой оси, в который с заданной вероятностью попадает нужный нам параметр генеральной совокупности. Параметр неизвестен, но его можно оценить по выборке.

Если величина с вероятностью 99% попадает в интервал от 300 до 500, то 99%-й доверительный интервал для неё — это (300, 500).

Статистика. Доверительный интервал

Доверительный интервал — это не просто диапазон значений случайной величины.

Величина, которую мы оцениваем, изначально неслучайная. Например, средний чек — это просто одно число. Это параметр поведения пользователей, у него нет разброса значений и дисперсии.

Вероятность возникает потому, что число неизвестно по генеральной совокупности и оценивается по выборке. Случайность выборки вносит случайность и в оценку. Доверительный интервал оценивает уверенность в этой оценке.

Статистика. Предельная теорема теории вероятностей

Центральная предельная теорема

С увеличением числа одинаково распределенных случайных величин с математическим ожиданием μ и дисперсией σ^2: распределение их среднего арифметического стремится к **нормальному распределению** с математическим ожиданием μ и дисперсией $\frac{\sigma^2}{n}$.

$$\bar{X} \sim \mathbf{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Стандартное отклонение этого распределения называется стандартной ошибкой (англ. standard error of mean, или SEM):

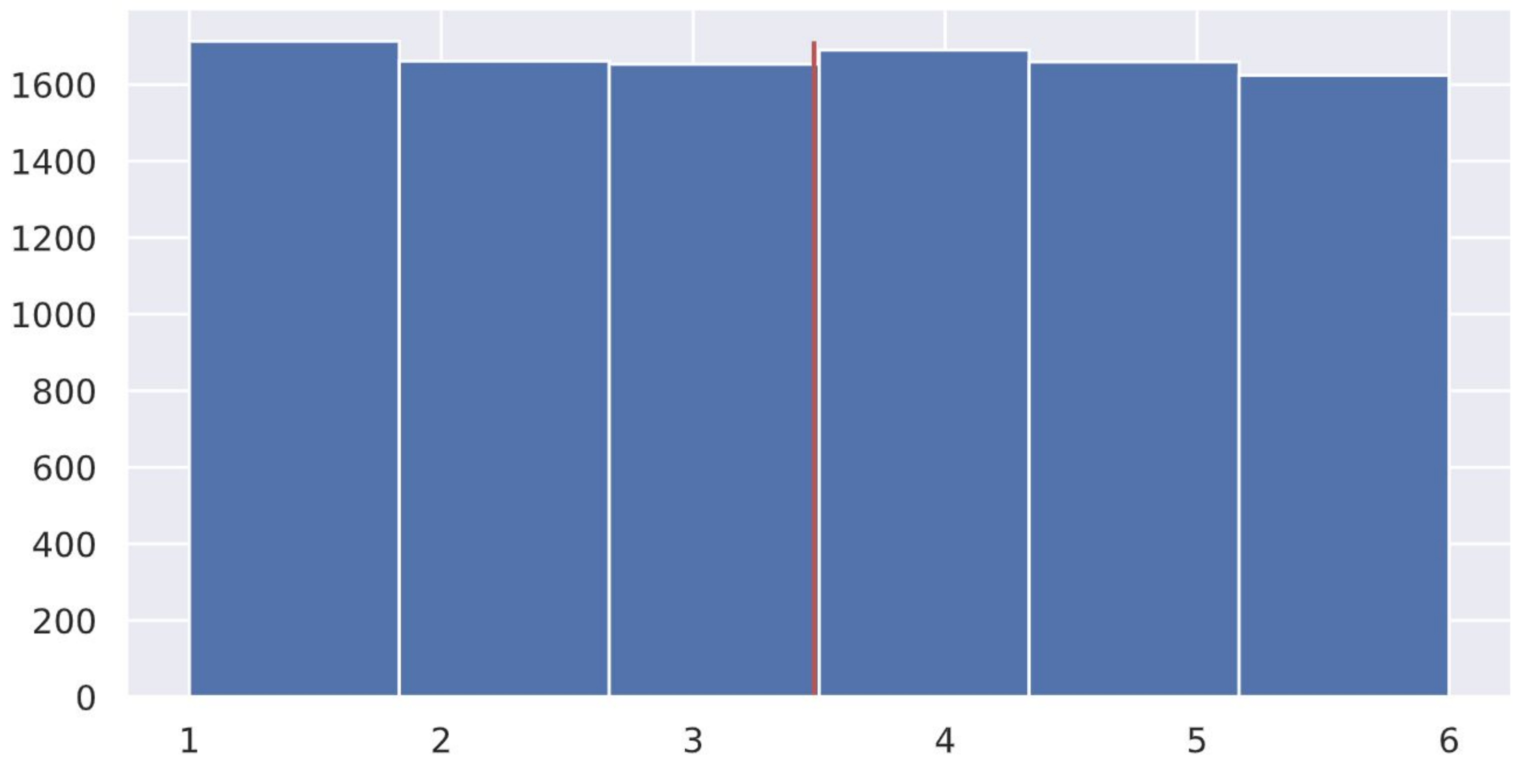
$$\mathrm{SEM}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Статистика. Предельная теорема теории вероятностей

Центральная предельная теорема

С увеличением числа одинаково распределенных случайных величин с математическим ожиданием μ и дисперсией σ^2: распределение их среднего арифметического стремится к **нормальному распределению** с математическим ожиданием μ и дисперсией $\frac{\sigma^2}{n}$.

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

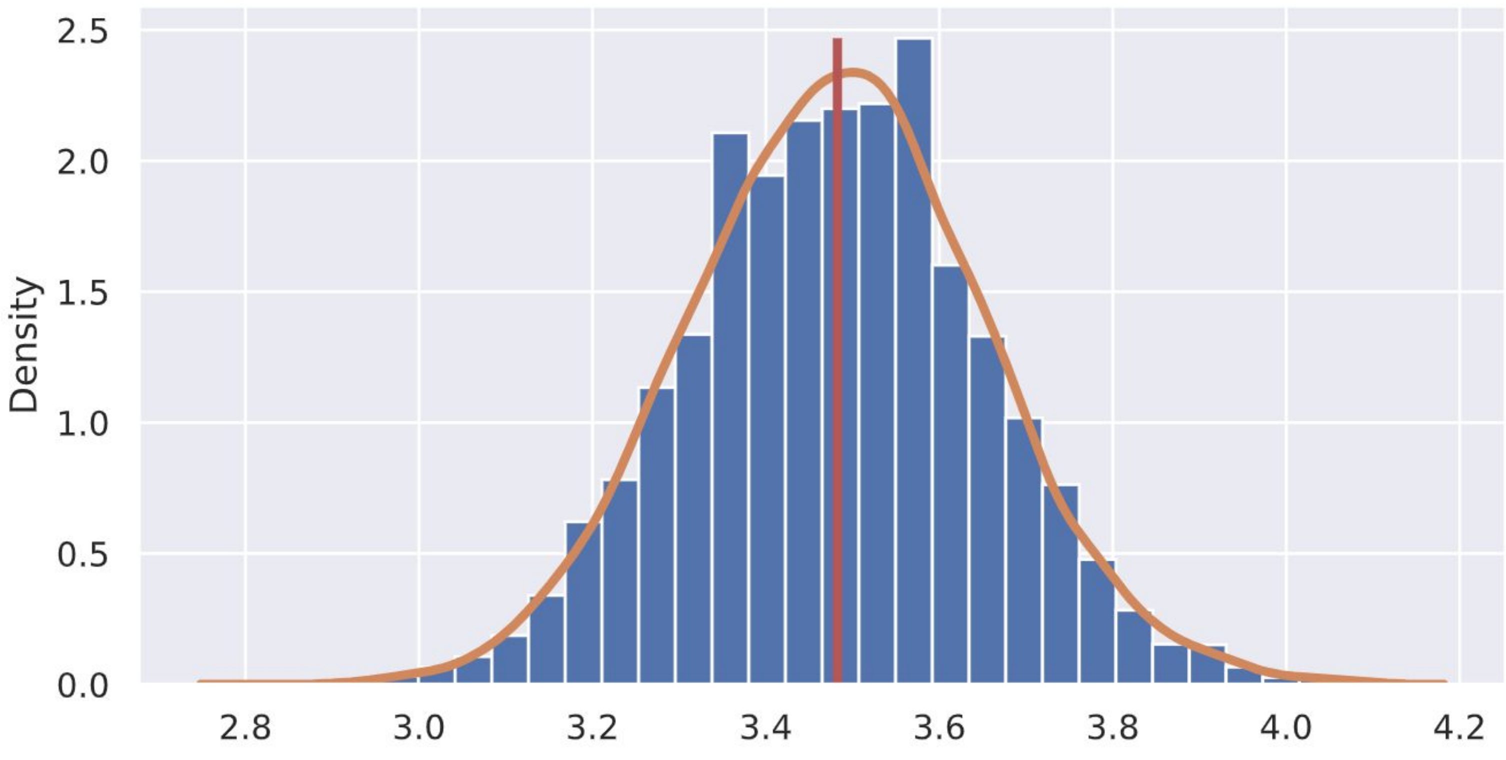


3.4793

Выбираем выборку и считаем среднее

повторяем предыдущий шаг много раз

получаем что среднее будет распределено по нормальному закону




```
np.mean(res_sum)
```

3.482507

Статистика. Доверительный интервал

У стандартного нормального распределения возьмём 5%-й квантиль F(0.05) и 95%-й квантиль F(0.95) для 90%-го доверительного интервала:

$$P\left(F(0.05) < \frac{\mu-\bar{X}}{\mathrm{SEM}(\bar{X})} < F(0.95)\right) = 90\%$$

Преобразуем и получим 90% интервал :

$$P\left(\bar{X} + F(0.05) \cdot \mathrm{SEM}(\bar{X}) < \mu < \bar{X} + F(0.95) \cdot \mathrm{SEM}(\bar{X})\right) = 90\%$$

$$\bar{X} + F(0.05) \cdot \mathrm{SEM}(\bar{X}) < \mu < \bar{X} + F(0.95) \cdot \mathrm{SEM}(\bar{X})$$

Статистика. Доверительный интервал

$$\bar{X} + F(0.05) \cdot \mathrm{SEM}(\bar{X}) < \mu < \bar{X} + F(0.95) \cdot \mathrm{SEM}(\bar{X})$$

$$\mathrm{SEM}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Осталась только одна проблема: при вычислении стандартной ошибки берут дисперсию генеральной совокупности, но она нам неизвестна! Оценивать её нужно по выборке.

Это влияет и на распределение выборочных средних: если дисперсия неизвестна, его нужно описывать уже не нормальным распределением, а распределением Стьюдента.

Статистика. Доверительный интервал

$$\bar{X} + F(0.05) \cdot \mathrm{SEM}(\bar{X}) < \mu < \bar{X} + F(0.95) \cdot \mathrm{SEM}(\bar{X})$$

$$\mathrm{SEM}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Подставив в формулу 5%-й квантиль t(0.05) и 95%-й квантиль t(0.95), получаем:

$$\bar{X} + t(0.05) \cdot \mathrm{SEM}(\bar{X}) < \mu < \bar{X} + t(0.95) \cdot \mathrm{SEM}(\bar{X})$$

Статистические гипотезы и их проверки

- Конверсия в покупку в двух версиях дизайна сайта
- Какой промокод дает больше продаж
- Сравнение цен на товар в двух магазинах
- Сколько килограмм сбросили участники двух групп
- Какая группа курьеров доставляет быстрее
- Когда лучше бегать – утром или вечером

Статистические гипотезы и их проверки

Мы везде использовали выборки.

А вдруг мы получили результат случайно?

Статистические гипотезы и их проверки

В интернет-магазине в среднем клиент тратил 18р на покупку.

Интернет-магазин решил ввести систему лояльности на 100 пользователях. Были получены следующие результаты: средний чек увеличился до 19 при стандартном отклонении равном 5.

Какой вывод можно сделать, основываясь на этих данных?

Гипотезы и их проверка

- Нулевая гипотеза – никакой разницы/эффекта нет
- Альтернативная гипотеза – разница/эффект есть

α – значимость

	Нулевая гипотеза (H_0) верна	Нулевая гипотеза (H_0) ложна
Нулевая гипотеза **отвергнута**	ложноположительное (ошибка первого рода)	H_0 верно принята
Нулевая гипотеза **не отвергнута**	H_0 верно отвергнута	ложноотрицательное (ошибка второго рода)

β

$(1-\beta)$ – чувствительность, мощность

Гипотезы и их проверка

- Нулевая гипотеза – никакой разницы/эффекта нет
- Альтернативная гипотеза – разница/эффект есть

H0 = {система лояльности не изменила средний чек пользователя}

HA = H1 = {система лояльности изменила средний чек пользователя}

Гипотезы и их проверка

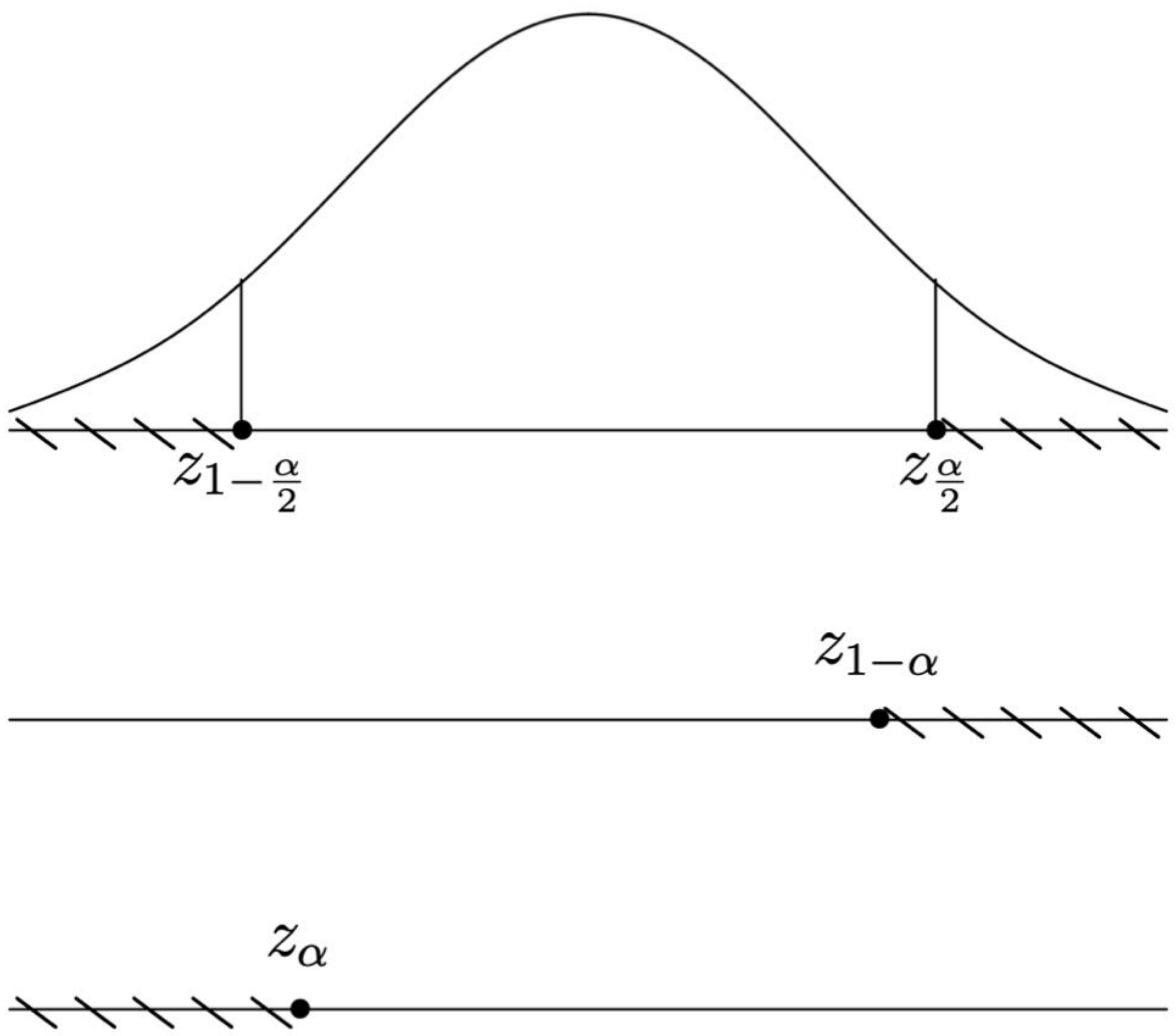
Уровень значимости α – максимально допускаемая исследователем вероятность ошибочно отклонить нулевую гипотезу (совершить ошибку первого рода)

Выберем уровень значимости = 0.05

Гипотезы и их проверка

Критическая область – область значения статистики критерия, при которых отвергается $H0$

Если значение статистики не попадает в интервал между критическими значениями — нулевая гипотеза отвергается. Если попадает, то не отвергается.



Гипотезы и их проверка

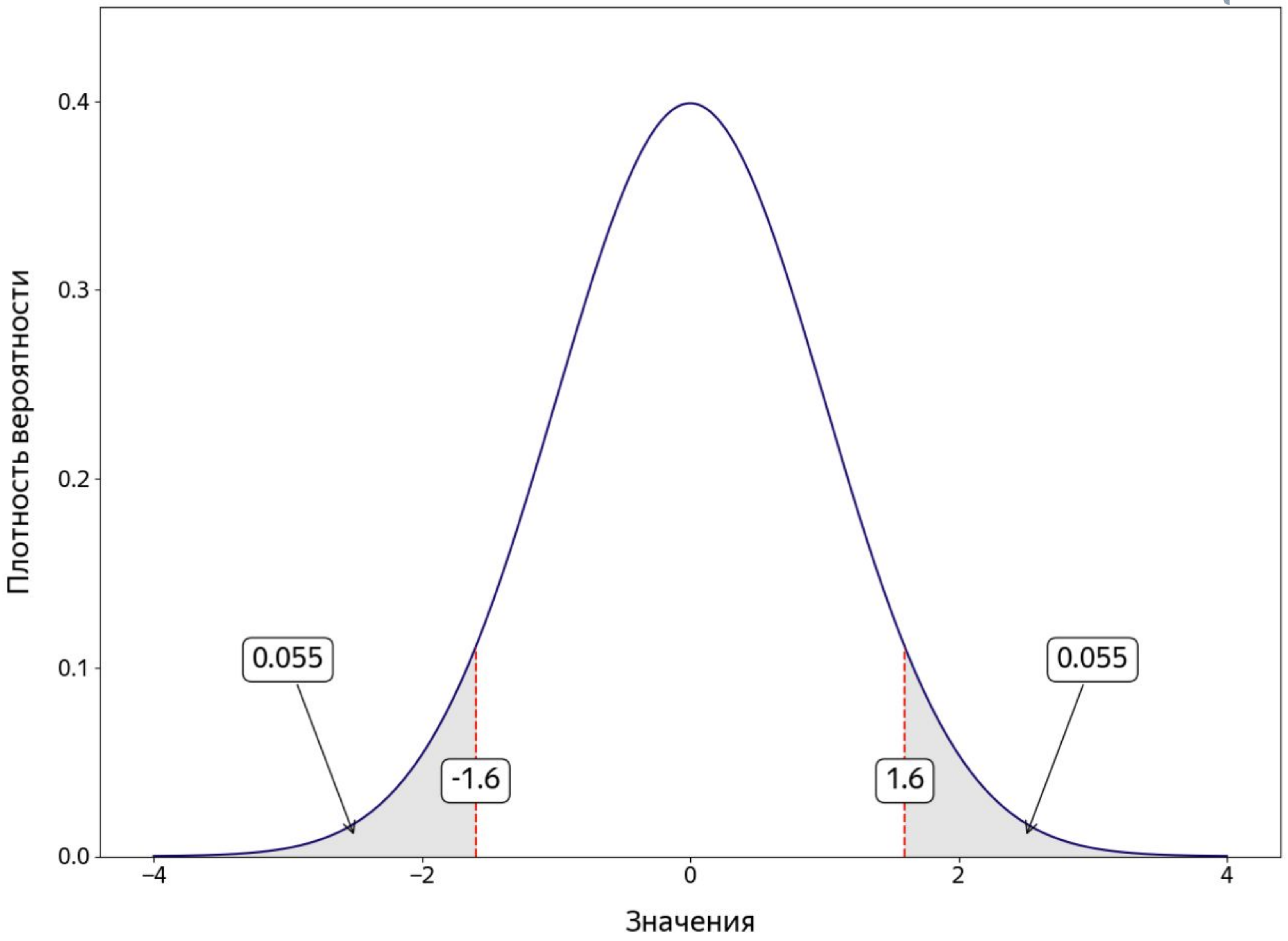
P-value — это вероятность наблюдать такое же или более экстремальное значение статистического критерия, если нулевая гипотеза (H0) верна.

P-value - вероятность получить настолько же или более далекое значение при правдивости $H0$.

Если p-value меньше статистической значимости, то $H0$ отвергается

$p - value < \alpha \rightarrow$ отвергаем H_0

$p - value > \alpha \rightarrow$ недостаточно оснований отвергнуть H_0



Статистические гипотезы и их проверки.

1. Формулируем гипотезы H0 и H1
2. Выбираем уровень значимости а (альфа)
3. Выбираем критерий
4. Считаем статистику критерия
5. Считаем p_value
6. Сравниваем p-value с уровнем значимости:
   - Если p-value меньше уровня значимости, принимаем решение, что надо отклонить нулевую гипотезу.
   - Если p-value больше уровня значимости, то оснований отклонять нулевую гипотезу нет.

УНИВЕРСИТЕТ
ИННОПОЛИС

Спасибо!