Лекция 4
Статистические гипотезы и параметрические критерии

Жигалов Августин

Data Engineer

РСХБ

План лекции

- Статистические гипотезы
- Р-значение
- Ошибки 1 и 2 рода
- Параметрические критерии

Что такое гипотеза

Гипотеза – любое утверждение, которое возникло в нашей голове и которое мы собираемся проверить по данным.

- В Питере лягушек любят больше, чем в Москве
- Жрец Лофир может предсказывать среднее выпадение лягушек
- Размер лягушек нормально распределен
- Загрязнение воды влияет на популяцию лягушек

Что значит проверить гипотезу

- Собрать данные и посмотреть, не противоречит ли им наше утверждение
- Любая выборка – случайная, просто посчитать описательные статистики недостаточно
- Описательные статистики – случайные величины
- При любом объёме выборки можно допустить ошибку

Что значит проверить гипотезу

- Собрали данные о предсказаниях лягушки
- Данные не противоречат тому, что она провидец
- Это не означает, что лягушка правда пророк. Если мы соберём ещё данных, они могут начать этому противоречить

Грязная или чистая вода

ИТМО

- Ученый обучает лягушек, чтобы они могли отличить грязную воду из водоема
- Правда ли лягушка отличает грязную воду от чистой

Грязная или чистая вода

- Мы хотим проверить, действительно ли лягушка способна отличить воду или просто угадывает. Поэтому даем две одинаковые емкости с водой из разных прудов
- Случайная величина $X_i = 1$ если лягушка отличила мутную воду от чистой

Грязная или чистая вода

- Нулевая гипотеза ($H0$): Лягушка не различает, то есть вероятность правильного выбора p=0.5
- Альтернативная гипотеза ($H1$): Лягушка способна отличить загрязненную воду, то есть p>0.5

Нужно формализовать гипотезу **на языке статистики**

Формализация задачи

Выборка состоит из единиц и нулей:

$X_i=1$, если правильно выбрана вода

$X_i=0$, если не правильно выбрана

Если лягушка не различает воду и выбирает наугад, вероятность успеха p должна быть равна 0.5

$\mathrm{H}_0\colon p = 0.5 \quad \Leftrightarrow$ Лягушка не различает воду

$\mathrm{H}_a\colon p \neq 0.5 \quad \Leftrightarrow$ Лягушка различает воду

Поиск критерия

$H_0\colon\ p = 0.5$

$H_a\colon\ p \neq 0.5$

$X_1, \dots, X_n \sim iid\ Bern(p)$

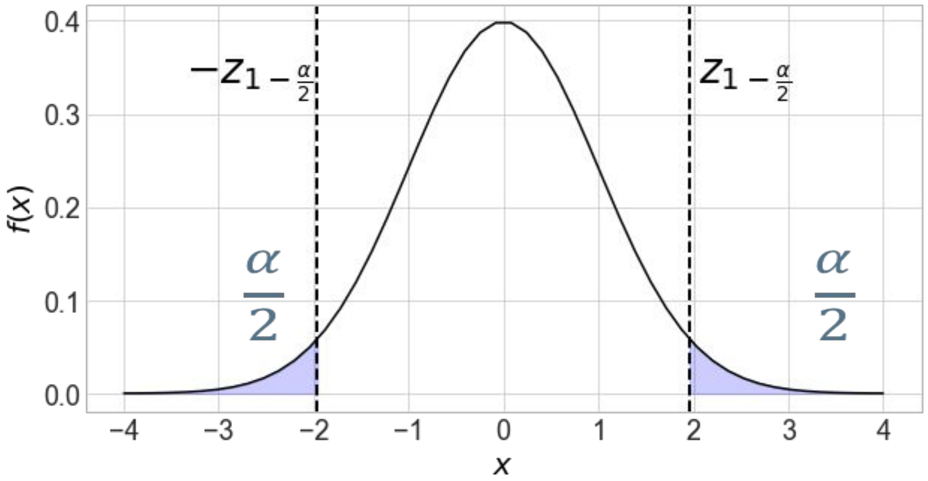
$\hat{p}$ – случайная величина

$\hat{p} - 0.5$ – случайная величина

Если расстояние от $\hat{p}$ до 0.5 достаточно маленькое, данные не противоречат нулевой гипотезе

Проверка гипотезы

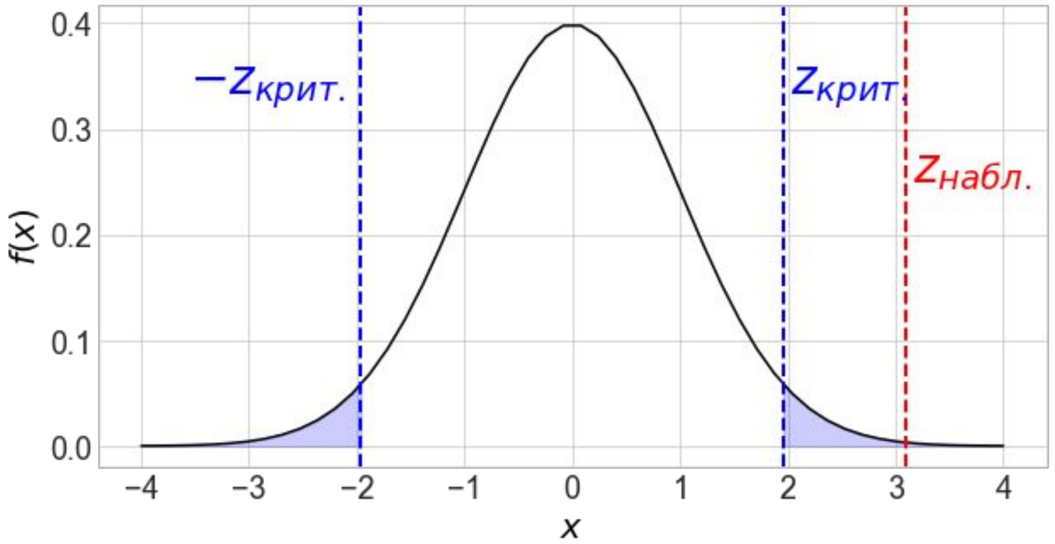
Мы рассуждаем о распределении в терминах нулевой гипотезы. Близкие к центру значения z-статистики показывают, что H_0 не противоречат данным



Мы не решаем, верна ли гипотеза, а проверяем, противоречат ли ей данные

Проверка гипотезы

Если наблюдаемое значение статистики попало в хвост (левый или правый), гипотеза отвергается, расстояние между $\hat{p}$ и 0.5 оказывается слишком большим



Уровень значимости

ITMO

- Если мы отвергаем нулевую гипотезу, когда она верна – мы ошибаемся
- Выбирая порог для отсечения, мы фиксируем вероятность такой ошибки
- Вероятность такой ошибки α – **уровень значимости**, также эту ошибку называют **ошибкой первого рода**
- Если мы 100 раз попытаемся сесть на поезд на уровне значимости 0.05, в среднем мы будем опаздывать 5 раз
- Обычно α выбирают равным 0.05, 0.01 или 0.001

Процедура проверки гипотез

ИТМО

1. Фиксируем уровень значимости: $\alpha = 0.05$
2. Формулируем нулевую гипотезу и альтернативную

$$\mathrm{H}_0\colon\ p = 0.5 \qquad\qquad \mathrm{H}_a\colon\ p \neq 0.5$$

3. Выбираем **союзника** (статистический тест) для проверки гипотезы
4. Находим наблюдаемое значение статистики:

$$z_{obs} = \frac{0.6-0.5}{\sqrt{\frac{0.5\,(1-0.5)}{10}}} = 0.654$$

5. Находим критическое значение с помощью **союзника**:

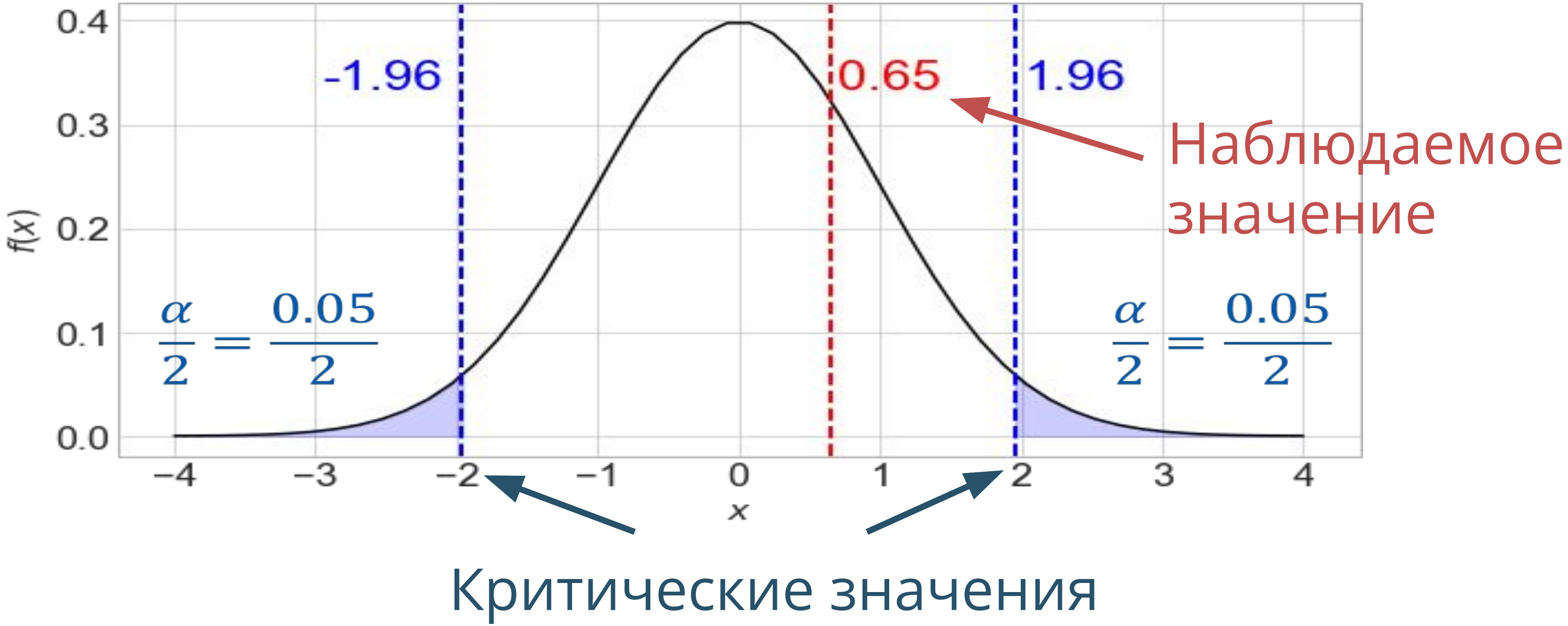
ЦПТ

$$z \underset{\mathrm{H}_0}{\overset{asy}{\sim}} N(0,1)$$

Процедура проверки гипотез

ИТМО

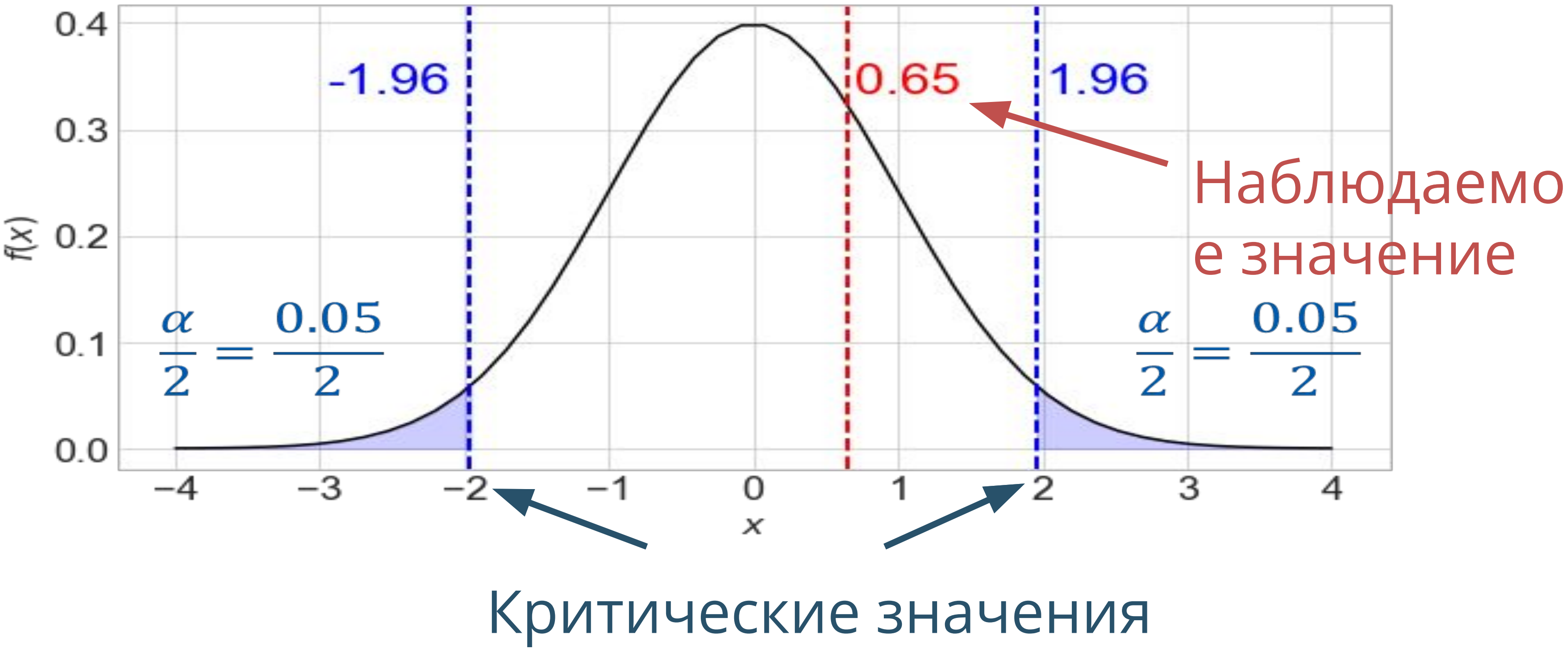
6. Сравниваем наблюдаемое значение с критическим
и делаем выводы



Процедура проверки гипотез

ИТМО

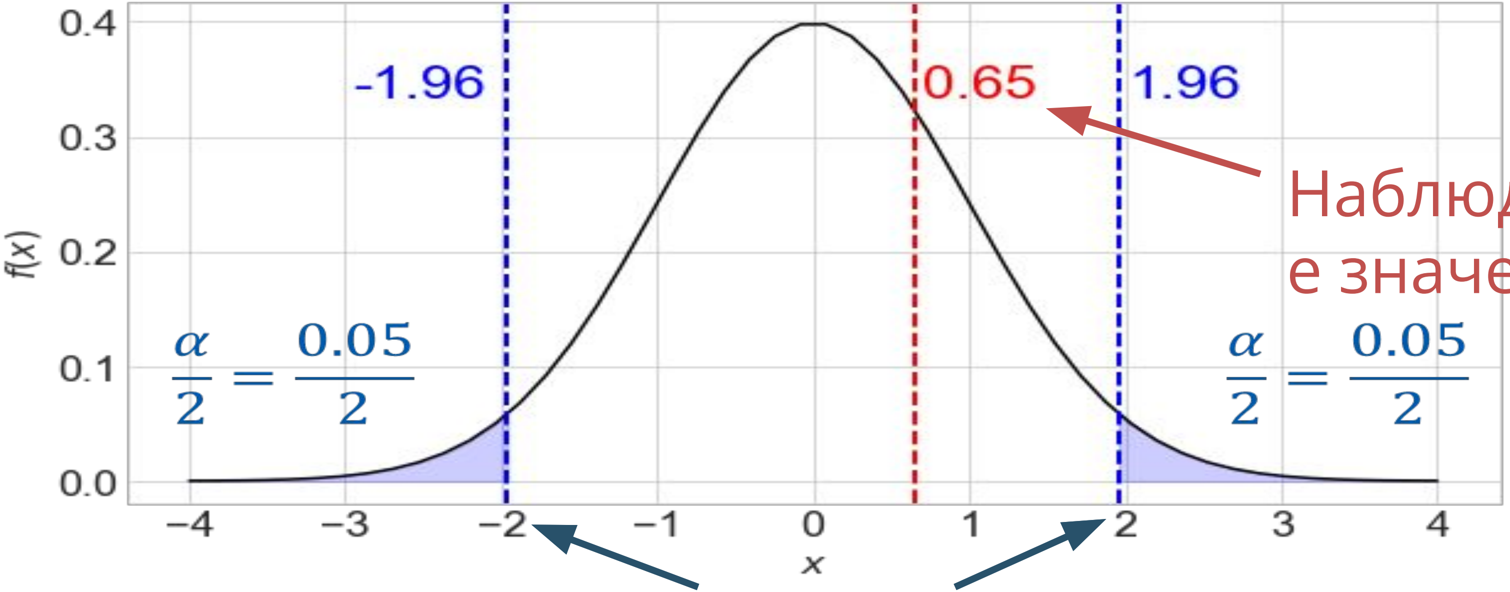
- Наблюдаемое значение попало в область между критическими ⇒ гипотеза не отвергается
- Голубая площадь под хвостами – уровень значимости



Процедура проверки гипотез

ИТМО

Гипотеза, что лягушка не различает воду, не отвергается на уровне значимости 5%



Наблюдаемое значение

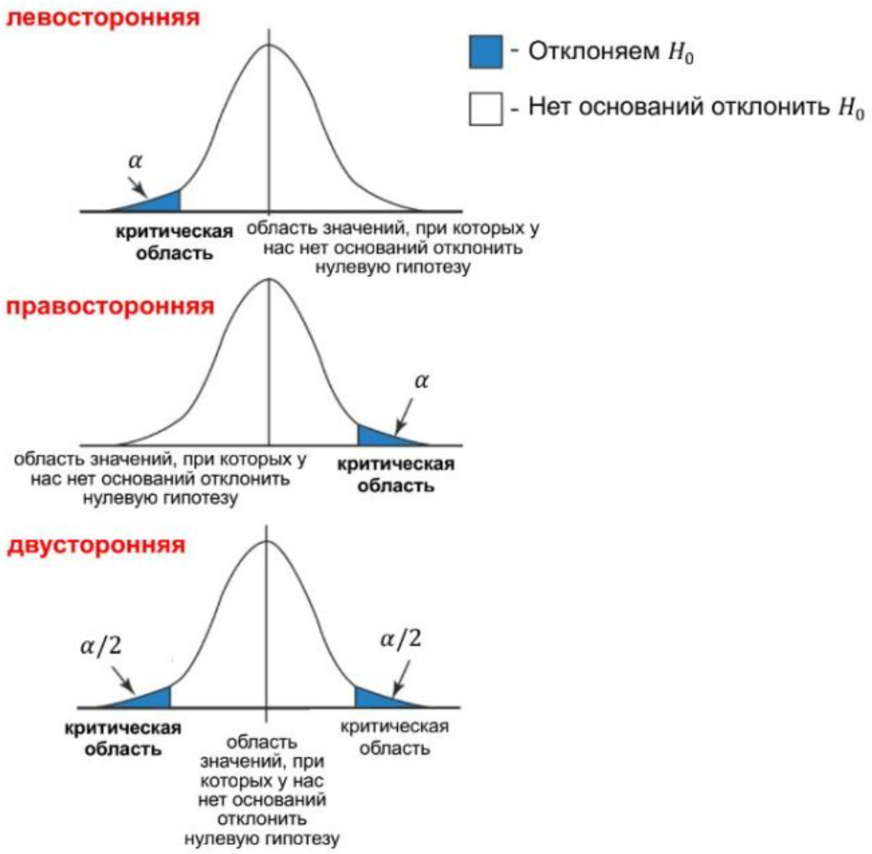
Критические значения

Нельзя принять нулевую гипотезу ИТМО

- Если при проверке нулевая гипотеза не отвергается, нельзя считать её доказанной
- Говорят, что данные не противоречат нулевой гипотезе
- Новые данные могут показать, что гипотеза неверна

Альтернативы

- Иногда рассматривают односторонние альтернативы
- Обычно это делается в ситуациях, когда мы уверены в направлении ожидаемых различий
- При таких альтернативах ошибку первого рода полностью переносят на один из двух хвостов

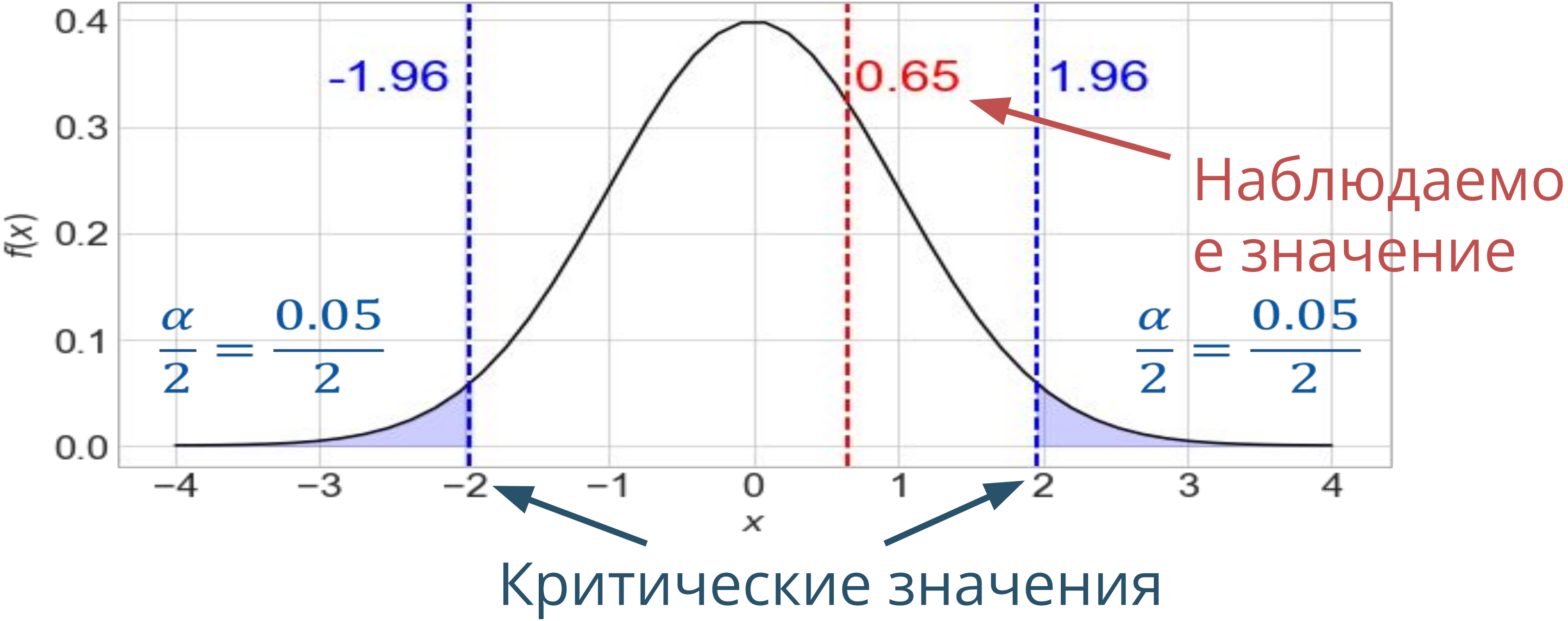


P-значение

P-значение

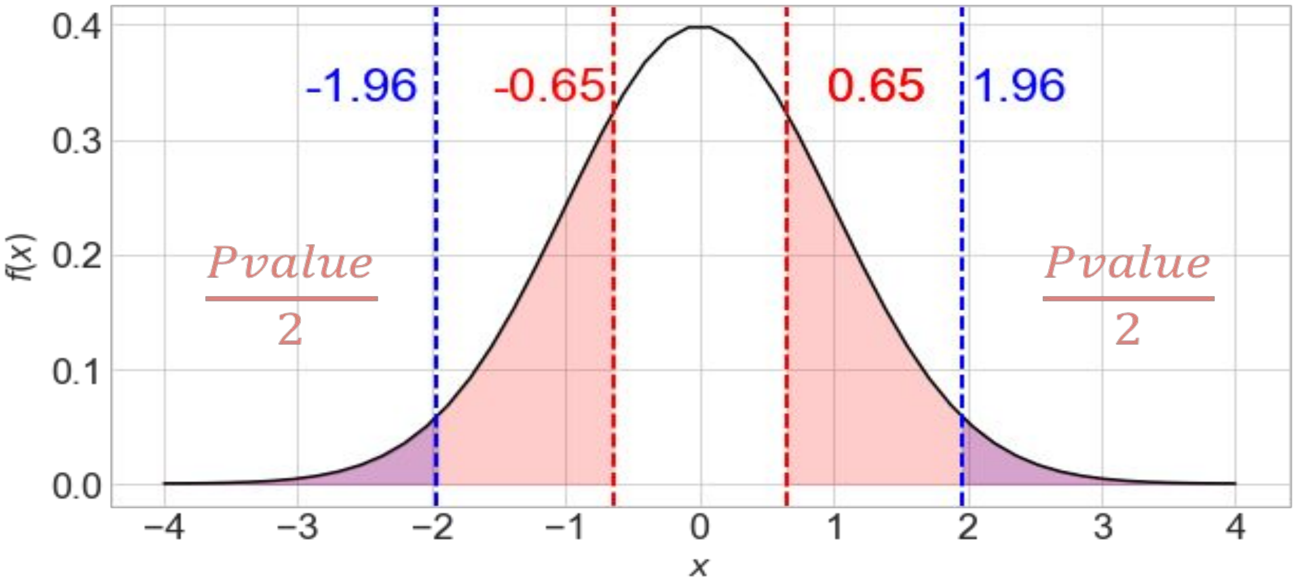
ИТМО

- Наблюдаемое значение попало в область между критическими ⇒ гипотеза не отвергается
- Голубая площадь под хвостами – уровень значимости



P-значение

- Красная площадь под хвостами – **p-значение** (достигаемый уровень значимости)



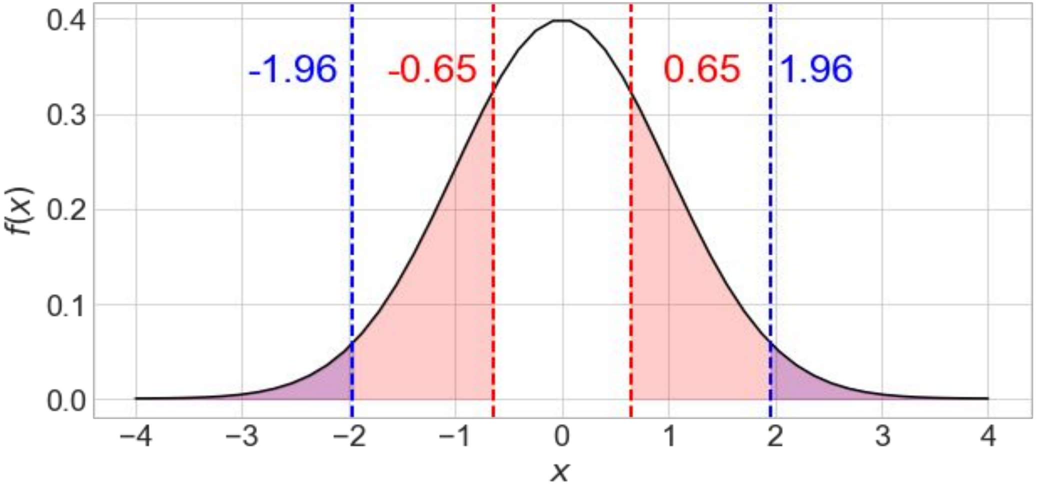
P-значение упрощает проверку гипотез

ИТМО

Если красная площадь оказалась больше синей

$$p\_value > \alpha,$$

$\Rightarrow$ наблюдаемое значение попало в область между критическими, гипотеза не отвергается.



$p\_value > \alpha \Rightarrow$ не отвергается

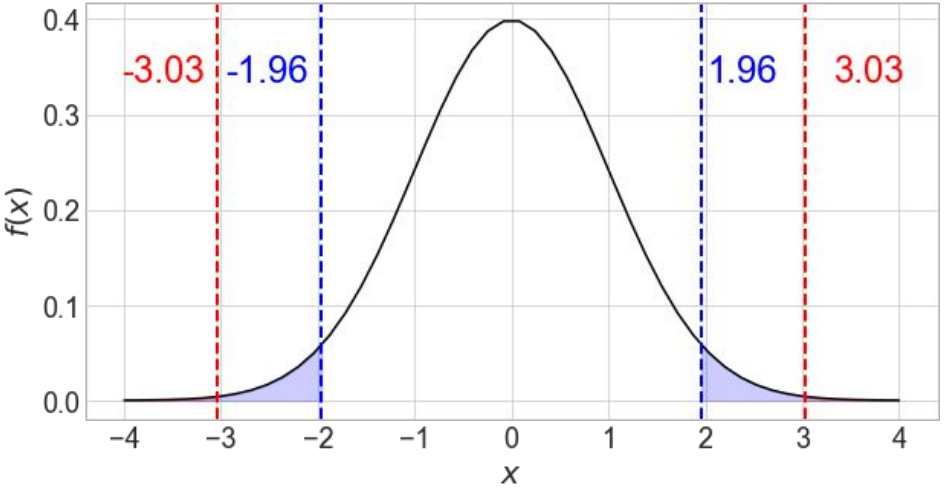
P-значение упрощает проверку гипотез

ИТМО

Если красная площадь оказалась больше синей

$$p\_value > \alpha,$$

$\Rightarrow$ наблюдаемое значение попало в область между критическими, гипотеза не отвергается.



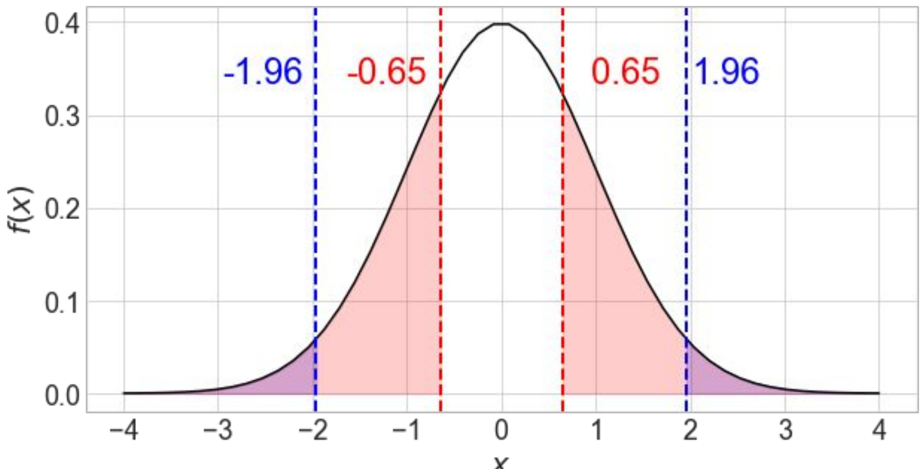
$p\_value < \alpha \Rightarrow$ отвергается

P-значение

Вопрос: какой уровень значимости надо выбрать, чтобы гипотеза впервые отверглась?

Ответ: равный P-значению

Из-за этого P-значение также называют **достигаемым уровнем значимости**

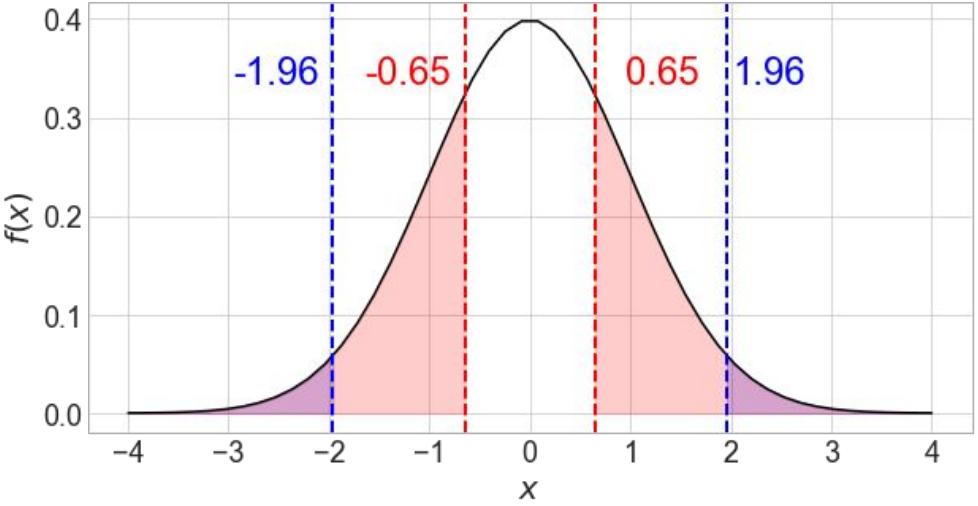


Р-значение

Достигаемый уровень значимости (**Р-значение**) – это вероятность при справедливости нулевой гипотезы получить такое же наблюдаемое значение статистики,

$$p\_value = \mathbb{P}(|z| > z_{\text{набл.}} \mid H_0)$$

как в эксперименте либо ещё более экстремальное



P-значение

Достигаемый уровень значимости (**P-значение**) – это вероятность при справедливости нулевой гипотезы получить такое же наблюдаемое значение статистики, $\text{p\_}value = \mathbb{P}(|z| > z_{\text{набл.}} \mid H_0)$
как в эксперименте либо ещё более экстремальное

В нашей ситуации $p\_value = 0.518$, то есть вероятность получить наше или ещё более экстремальное значение статистики, при верности H_0 высока, что говорит в пользу гипотезы

Ошибки 1 и 2 рода

Ошибки

Перенесемся в город Фрогус, где обитают лягушки, туда затесался подозрительный субъект, который хочет доказать, что он лягушка.
Послушный гражданин Прыгус что-то заподозрил, поэтому хочет попытаться помочь городу узнать истину.

Ошибки

	H_0: не лягушка	H_a: лягушка
	  Вы не лягушка	  Вы не лягушка
	  Вы лягушка	  Вы лягушка

Ошибки

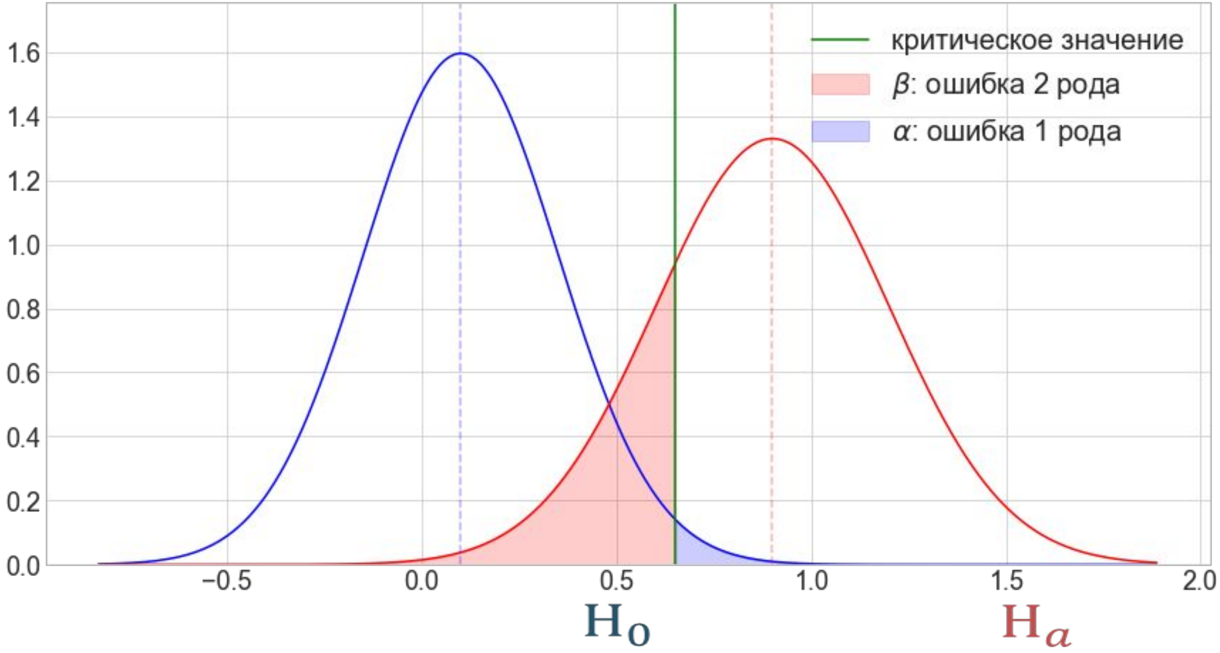
	H_0 верна	H_0 неверна
H_0 принимается	H_0 верно принята	Ошибка II рода
H_0 отвергается	Ошибка I рода	H_0 верно отвергнута

Ошибки

$H_0: p = p_0$

$H_a: p = p_a$

При уменьшении ошибки первого рода всегда возрастает ошибка второго рода



Презумпция нулевой гипотезы

Презумпция невиновности: человек не считается лягушкой, пока его случай не будет доказан

Презумпция нулевой гипотезы: мы верим в нулевую гипотезу, пока данные не опровергнут её

Вывод: “данные противоречат гипотезе” всегда весомее и категоричнее, чем “данные **не** противоречат гипотезе”

Наиболее мощный критерий

ИТМО

- **Статистический критерий** – способ посчитать расстояние между наблюдаемым значением и предполагаемым
- Подобные расстояния можно считать разными способами
- Хочется выбрать такой способ, который при фиксированном размере выборки и фиксированной ошибке первого рода будет давать наименьшую ошибку второго рода
- Такой критерий называется **наиболее мощным**

Сводка

- **Ошибки первого и второго рода неравнозначны**
- **Имеется презумпция нулевой гипотезы**
- **Обычно нулевую гипотезу формулируют так, что нет значимого эффекта**

Параметрические критерии

Какими бывают критерии

Критерии

Параметрические

Включают в себя расчёт параметров конкретного распределения

Непараметрические

Не завязаны на конкретное распределение

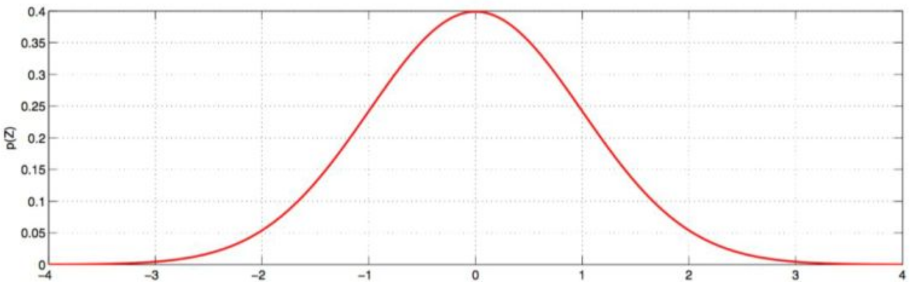
Согласия

Проверяется гипотеза о виде неизвестного закона распределения

Z-критерий для доли

$$H_0: p = p_0$$

$$H_a: p \neq p_0$$



ЦПТ:

$$\hat{p} \underset{H_0}{\overset{asy}{\sim}} N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$$

Критерий для проверки:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0\,(1-p_0)}{n}}} \underset{H_0}{\overset{asy}{\sim}} N(0,1)$$

Условия применения:

- **Выполняется ЦПТ для средних (распределения асимптотически нормальны)**
- **Большой размер выборки**
- **Подходит для случайных величин с распределением Бернулли**

Пример (отбор лягушек):

- В одном заповеднике обитают различные виды лягушек, и ученые проводят исследование для изучения распределения их видов.
- Из 300 лягушек, отобранных для изучения, оказалось, что 90 принадлежат к виду зелёных лягушек, а остальные к желтым
- Экологи обеспокоены тем, что этот вид может быть недооценен при отборе.

Пример (отбор лягушек):

ИТМО

H_0: отбор правильный $\Leftrightarrow$ $\text{H}_0: p = 0.5$

H_a: лягушек неверно взяли $\quad \text{H}_a: p < 0.5$

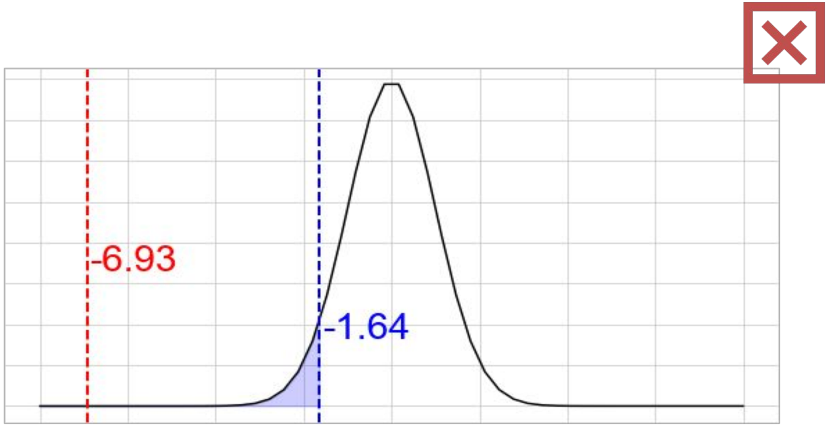
$X_1, \dots, X_n \sim iid\ Bern(p)$

$X_i = 1$, если лягушка - зеленая

$$\hat{p} = \frac{90}{300}$$

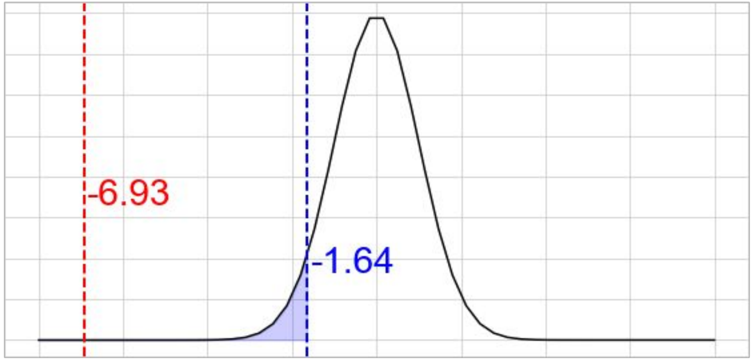
$$z_{obs} = \frac{0.3 - 0.5}{\sqrt{\frac{0.5\,(1 - 0.5)}{300}}}$$

$$z_{0.05} = -1.64$$



Гипотеза о правильном отборе **отвергается**

Z-критерий для разности независимых долей



Выборки независимые

$$H_0\colon p_x = p_y \qquad H_a\colon p_x \neq p_y$$

Критерий для проверки:

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{P(1-P)\cdot\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}} \overset{asy}{\underset{H_0}{\sim}} N(0,1)$$

$$P = \frac{m_x + m_y}{n_x + n_y}$$

Условия применения:
- Выполняется ЦПТ для средних (распределения асимптотически нормальны)
- Большой размер выборки
- Подходит для случайных величин с распределением Бернулли
- Группы (подвыборки) независимы

m_x - число в выборке

Пример (о любви к лягушкам):

- В Москве и Питере 100 человек спрашивают о том, любят ли они лягушек
- В Питере сказали да 50 человек, в Москве 60 человек
- Правда ли, что в Москве лягушек любят сильнее?
- Проверяем гипотезу на уровне значимости 5%

Пример (о любви к лягушкам): ИТМО

H_0: лягушек любят одинаково
H_a: в Москве любят сильнее

$\Leftrightarrow$

$$H_0: p_П = p_М$$
$$H_a: p_П < p_М$$

$$X_1, \dots, X_{100} \sim iid\ Bern(p_М)$$

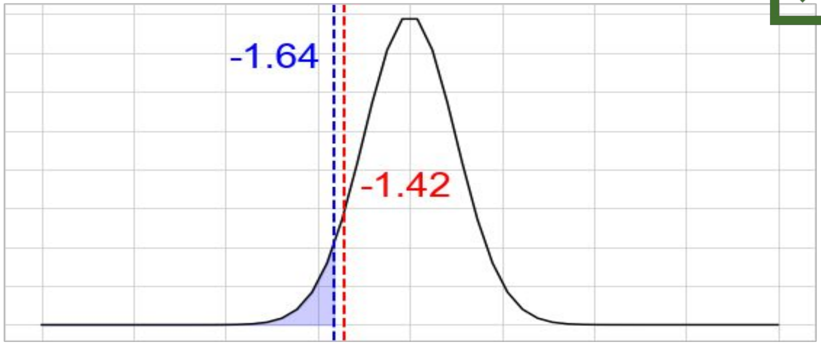
$$Y_1, \dots, Y_{100} \sim iid\ Bern(p_П)$$

$$\hat{p}_M = 0.6, \qquad \hat{p}_П = 0.5$$

$$P = \frac{50+60}{200} = 0.55$$

$$z_{obs} = \frac{0.5 - 0.6}{\sqrt{0.55 \cdot 0.45 \cdot \frac{2}{100}}}$$

$$z_{0.05} = -1.64$$



Гипотеза об одинаковой любви **не отвергается**

Z-критерий для среднего

ИТМО

$$X_1, \dots, X_n \sim iid\ (\mu, \sigma^2)$$

$$\mathrm{H}_0 : \mu = \mu_0$$

$$\mathrm{H}_a : \mu \neq \mu_0$$

ЦПТ:

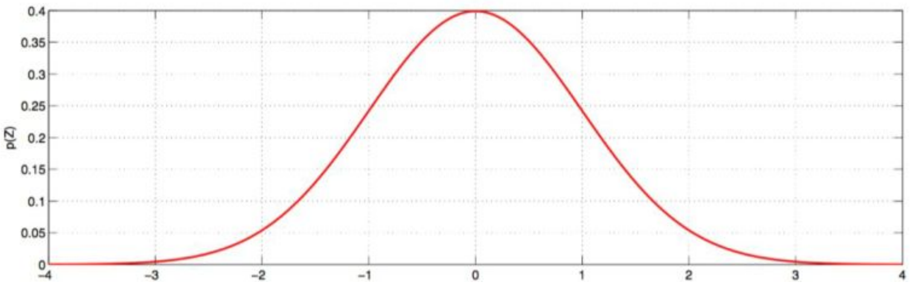
$$\bar{x} \underset{\mathrm{H}_0}{\overset{asy}{\sim}} N\left(\mu_0, \frac{\hat{\sigma}^2}{\sqrt{n}}\right)$$

Критерий для проверки:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \underset{\mathrm{H}_0}{\overset{asy}{\sim}} N(0,1)$$

Условия применения:

- Выполняется ЦПТ для средних (распределения асимптотически нормальны)
- Большой размер выборки
- Известно стандартное отклонение генеральной совокупности



t-критерий для среднего

ИТМО

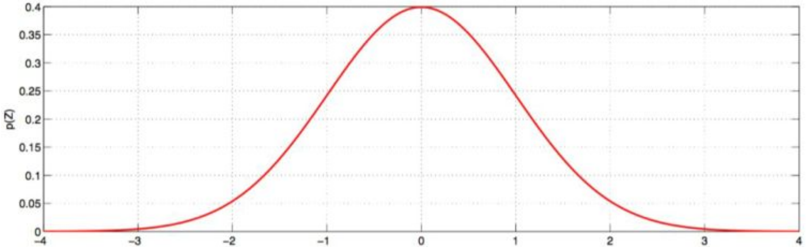
$X_1, \ldots, X_n \sim iid\ N(\mu, \sigma^2)$

$\mathrm{H}_0 : \mu = \mu_0$

$\mathrm{H}_a : \mu \neq \mu_0$ σ^2 – НЕизвестна

Критерий для проверки:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} \underset{\mathrm{H}_0}{\sim} t(n-1)$$



Условия применения:
- Предположение о нормальности выборки
- Может быть малый размер выборки
- Неизвестно стандартное отклонение генеральной совокупности

Пример (обучение прыжкам):

- Каждый год племя лягушек проходит ежегодное испытание по прыжкам через водоемы. В обычной ситуации средняя длина прыжка у лягушек составляет 15 метров.
- Вождь племени Квакис открыл свою школу обучения прыжкам, и его группа из 100 лягушек показала средний результат в 17 метров, а стандартное отклонение составило 4 метра.
- Правда ли, что школа Квакиса помогает лягушкам прыгать дальше? Проверим гипотезу на уровне значимости 5%.

Пример (обучение прыжкам):

Нулевая гипотеза H0: средняя длина прыжка лягушек, прошедших обучение у Квакиса, такая же, как и у всех лягушек племени, μ=15 метров.

Альтернативная гипотеза H1: средняя длина прыжка лягушек, прошедших обучение у Квакиса, больше 15 метров, μ>15 метров (односторонний тест).

критическое значение по степеням свободы 100-1 и уровню значимости 0.05 t≈1.66

$$t = \frac{17 - 15}{\frac{4}{\sqrt{100}}} = \frac{2}{\frac{4}{10}} = \frac{2}{0.4} = 5$$

Так как рассчитанное значение t = 5 значительно больше критического значения 1.66, мы **отвергаем нулевую гипотезу**. Это означает, что школа Квакиса действительно помогает лягушкам прыгать дальше, чем обычно.

t-критерий для разности средних

ИТМО

$$X_1, \dots, X_n \sim iid\,(\mu, \sigma^2)$$

$$\mathrm{H}_0:\ \mu_1 = \mu_2$$

$$\mathrm{H}_a:\ \mu_1 \neq \mu_2$$

ЦПТ:

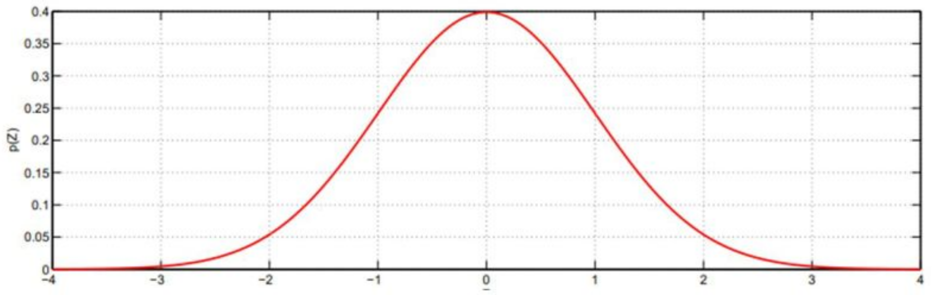
$$\bar{X}_1 - \bar{X}_2 \ \underset{\mathrm{H}_0}{\overset{asy}{\sim}}\ N\left(0, \frac{s_1^2}{n} + \frac{s_2^2}{m}\right)$$

Критерий для проверки:

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}} \ \underset{\mathrm{H}_0}{\overset{asy}{\sim}}\ N(0,1)$$

$$\bar{X}_1 - \bar{X}_2 \ \sim\ N\left(0, \frac{s_1^2}{n} + \frac{s_2^2}{m}\right)$$

Выборки независимые



Точные критерии для разности средних

ITMO

Критерий для проверки:

дисперсии известны

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \underset{H_0}{\sim} N(0,1)$$

Нормальное распределение

дисперсии неизвестны, но равны

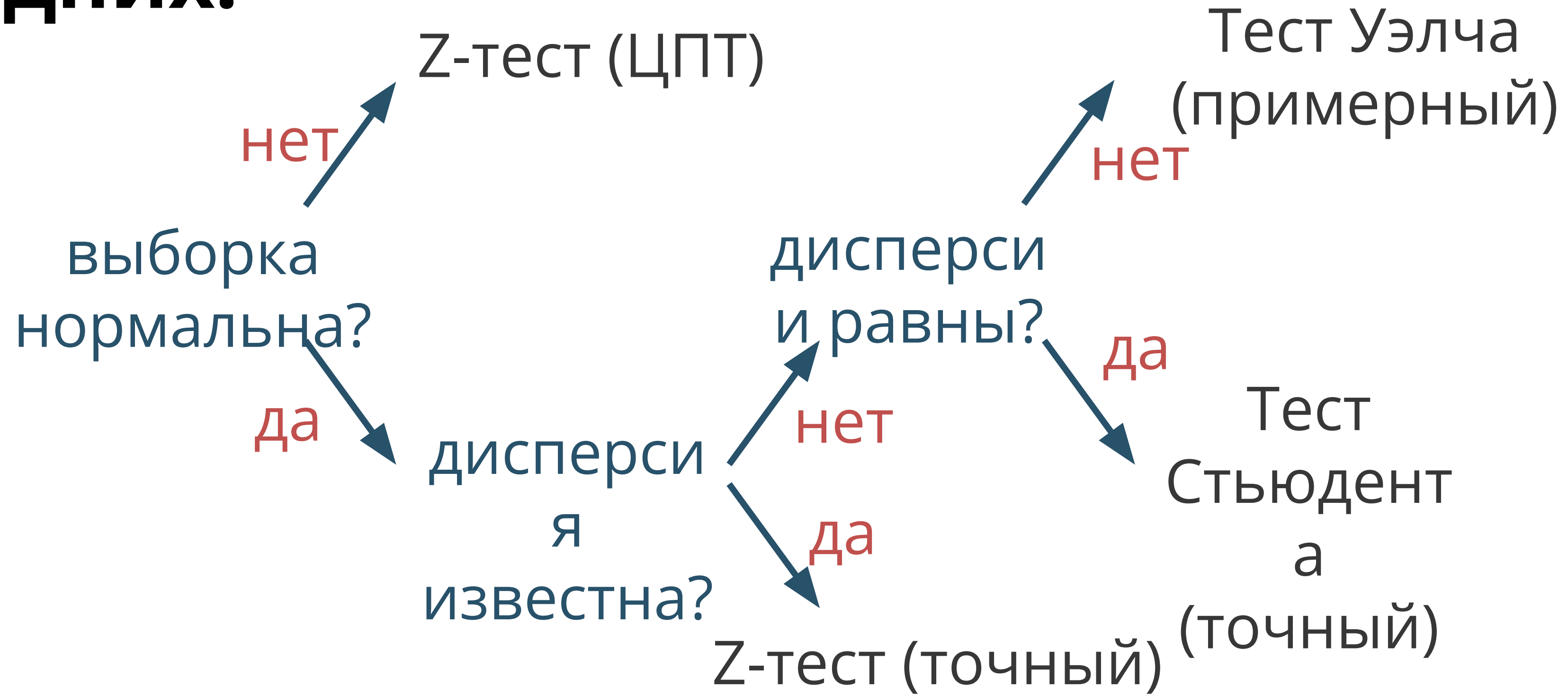
$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s^2}{n} + \frac{s^2}{m}}} \underset{H_0}{\sim} t(n+m-2)$$

Распределение Стьюдента

дисперсии неизвестны

Распределение Уэлча

Как выбрать критерий для средних:



Пример результаты прыжков

- До засухи средняя длина прыжка лягушек из пруда составляла 105 сантиметров при стандартном отклонении 40 сантиметров.
- После засухи средняя длина прыжка лягушек составила 90 сантиметров при стандартном отклонении 50 сантиметров.

Пример результаты прыжков

Нулевая гипотеза H_0: Средняя длина прыжков лягушек не изменилась или не сократилась

Альтернативная гипотеза H_1: Средняя длина прыжков лягушек уменьшилась

Уровень значимости $\alpha=0.05$.

Поскольку у нас есть две независимые группы (прыжки до и после засухи) и различающиеся дисперсии, используем **двухвыборочный t-тест для независимых выборок с разными дисперсиями**.

Для одностороннего t-теста при уровне значимости α и приблизительных степенях свободы около 40, критическое значение для левого хвоста равно **-1.684**.

Значение t = -1.17 **больше** критического значения -1.684. Это означает, что мы **не отвергаем нулевую гипотезу**. У нас **недостаточно доказательств**, чтобы утверждать, что длина прыжков лягушек сократилась после засухи.

$$t = \frac{90 - 105}{\sqrt{\frac{40^2}{20} + \frac{50^2}{30}}} \approx -1.17$$

Разность средних (зависимые выборки)

ITMO

Выборки зависят друг от друга:

$X_1, \dots, X_n \sim iid\, N(\mu_1, \sigma_1^2)$ $Y_1, \dots, Y_n \sim iid\, N(\mu_2, \sigma_2^2)$

- Измерения делаются на одних и тех же объектах
- Можем посмотреть разницу на отдельных объектах

$$d_i = X_i - Y_i \qquad \bar{x} - \bar{y} = \overline{x - y}$$

- Используем распределение Стьюдента, дисперсию считаем по формуле:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

Пример различие двух трав

В племени лягушек шаманы проводят испытания двух разных трав для лечения дыхательных проблем. Каждая лягушка сначала дышит через листья первой травы и прыгает, после чего измеряют, сколько времени она может прыгать, пока не устанет. Затем лягушка пробует вторую траву, и процедура повторяется.

Проверим гипотезу о том, что две травы **не отличаются** по своей эффективности. Предполагается, что выборки независимы и распределены нормально.

Пример различие двух трав

Нулевая гипотеза : Среднее время прыжков после использования двух трав одинаково, **μ1=μ2**.

Альтернативная гипотеза H1: Среднее время прыжков после использования двух трав различается, Уровень значимости α=0.05

Поскольку мы сравниваем зависимые выборки (одни и те же лягушки испытывают обе травы), применяем **парный t-тест**. Основное отличие — мы будем работать с разностями между значениями для каждой лягушки, а не с абсолютными значениями времени прыжков.

Лягушка	Трава 1 (время, сек)	Трава 2 (время, сек)	Разность $d = \text{Трава } 1 - \text{Трава } 2$
1	120	130	-10
2	135	140	-5
3	150	155	-5
4	160	165	-5
5	145	150	-5

Пример различие двух трав

Найдём среднюю разность

$$\bar{d} = \frac{-10 + (-5) + (-5) + (-5) + (-5)}{5} = \frac{-30}{5} = -6$$

Найдём стандартное отклонение разностей

$$s_d = \sqrt{\frac{16 + 1 + 1 + 1 + 1}{4}} = \sqrt{\frac{20}{4}} = \sqrt{5} \approx 2.24$$

Теперь подставим значения в формулу для t-статистики

$$t = \frac{-6}{2.24/\sqrt{5}} = \frac{-6}{1.00} = -6$$

критическое значение будет равно ±2.776.

Значение t = -6 **меньше** критического значения -2.776. Это означает, что мы **отвергаем нулевую гипотезу**. Есть статистически значимая разница между двумя травами.

Сколько было лягушек на слайдах?