

# Проверка гипотез



# План

- Что такое гипотезы
- p\_value
- Ошибки 1 и 2 рода
- Какими бывают критерии
- Параметрические критерии для долей, средних и дисперсий



# Что такое гипотезы и как их проверять



# Что такое гипотеза

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

- В Австралии женщин дискриминируют на рынке труда
- В Питере кофе любят больше, чем в Москве
- Осьминог Пауль может предсказывать будущее
- Доходности по акциям Яндекса нормально распределены
- Подгузники и пиво часто покупают вместе



# 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 осьминога



memepedia.com

- Данные не противоречат тому, что он провидец
- Это не означает, что осьминог правда пророк. Если мы соберём ещё данных, они могут начать этому противоречить.



# Что значит проверить гипотезу

- Собрали данные о предсказаниях осьминога



Если данные не противоречат утверждению, это не является доказательством его верности

- Данные не противоречат тому, что он провидец
- Это не означает, что осьминог правда пророк. Если мы соберём ещё данных, они могут начать этому противоречить



# Взболтать, но не смешивать



James Bond Casino Royale (2006)

«Взболтать,  
но не смешивать»

© Бонд, Джеймс  
Бонд

- Правда ли агент 007 отличает взбитанный мартини от смешанного?



# Взболтать, но не смешивать



James Bond Casino Royale (2006)

«Взболтать,  
но не смешивать»

© Бонд, Джеймс  
Бонд

- Завязываем агенту 007 глаза и даём два мартини: смешанный и взболтанный  $n$  раз
- Случайная величина  $X_i = 1$ , если Бонд отличил взболтанный мартини от смешанного



# Взболтать, но не смеивать



James Bond Casino Royale (2006)

«Взболтать,  
но не смеивать»

© Бонд, Джеймс  
Бонд

$H_0$ : агент 007 не различает два вида мартини

$H_a$ : агент 007 различает два вида мартини

Нужно формализовать гипотезу **на языке статистики**



# Формализация задачи

Выборка состоит из единиц и нулей:

- $X_i = 1$ , если правильно назвал вид мартини
- $X_i = 0$ , если не правильно назвал

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

Если Бонд не различает напитки и выбирает наугад, вероятность успеха  $p$  должна быть равна 0.5

$H_0: p = 0.5 \Leftrightarrow$  Бонд не различает напитки

$H_a: p \neq 0.5 \Leftrightarrow$  Бонд различает напитки



# Поиск критерия

$$H_0: p = 0.5$$

$$H_a: p \neq 0.5$$

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

Представим, что мы дали Бонду мартини 10 раз и получили, что  $\hat{p} = 0.6$

~~Выходит, что Бонд  
отвечает за свои  
слова~~

$\hat{p}$  – случайная величина, которая зависит от выборки



# Поиск критерия

$$H_0: p = 0.5$$

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

$$H_a: p \neq 0.5$$

Представим, что мы дали Бонду мартини 10 раз и получили, что  $\hat{p} = 0.6$

**Задача:** узнать при текущем объёме выборки насколько  $\hat{p}$  близка к 0.5



# Поиск критерия

$$H_0: p = 0.5$$

$$H_a: p \neq 0.5$$

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

$\hat{p}$  – случайная величина

$\hat{p} - 0.5$  – случайная величина

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



# Поиск критерия

$$H_0: p = 0.5$$

$$H_a: p \neq 0.5$$

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

По ЦПТ:

$$\hat{p} \stackrel{asy}{\sim} N \left( p, \frac{p(1-p)}{n} \right) \Leftrightarrow$$

$$\hat{p} \underset{H_0}{\stackrel{asy}{\sim}} N \left( 0.5, \frac{0.5(1-0.5)}{n} \right)$$

$$\hat{p} - 0.5 \underset{H_0}{\stackrel{asy}{\sim}} N \left( 0, \frac{0.5(1-0.5)}{n} \right)$$

при верности  
нулевой  
гипотезы

$$\frac{\hat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} \underset{H_0}{\stackrel{asy}{\sim}} N(0, 1)$$



# Поиск критерия

$$H_0: p = 0.5$$

$$H_a: p \neq 0.5$$

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

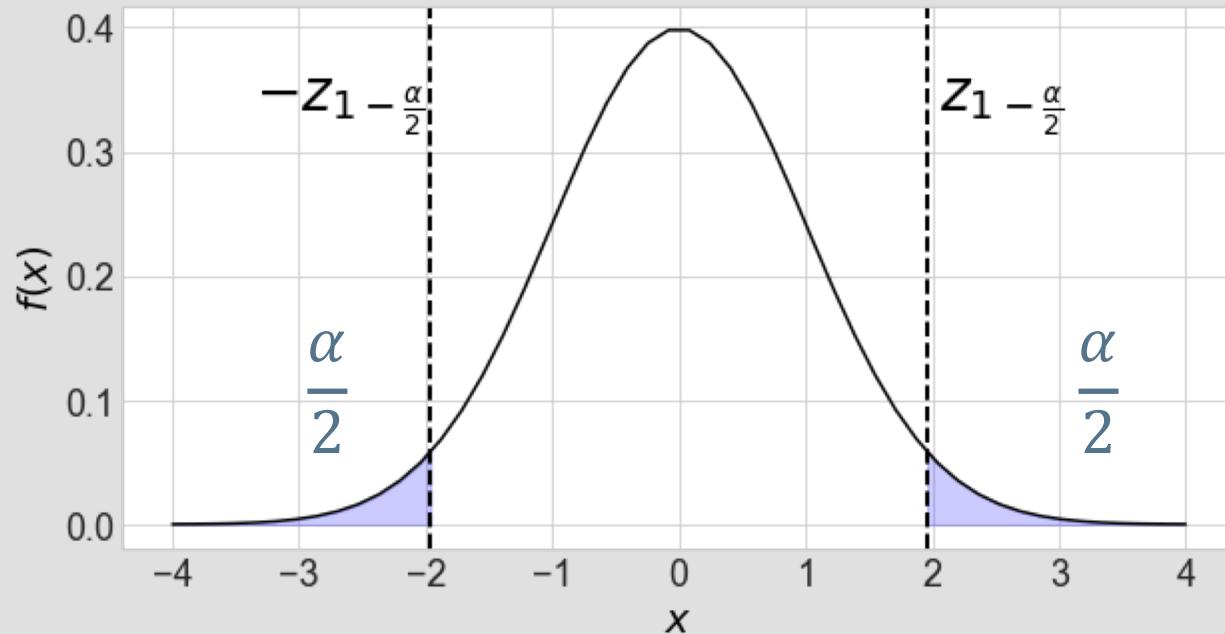
$$z = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} \stackrel{asy}{\underset{H_0}{\sim}} N(0, 1)$$

- Мы выяснили, как распределено расстояние до 0.5 при верности нулевой гипотезы
- Осталось выбрать порог, при котором мы будем считать, что нулевая гипотеза  $H_0$  не отвергается
- Против нашей гипотезы говорят очень большие либо очень маленькие значения статистики



# Проверка гипотезы

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

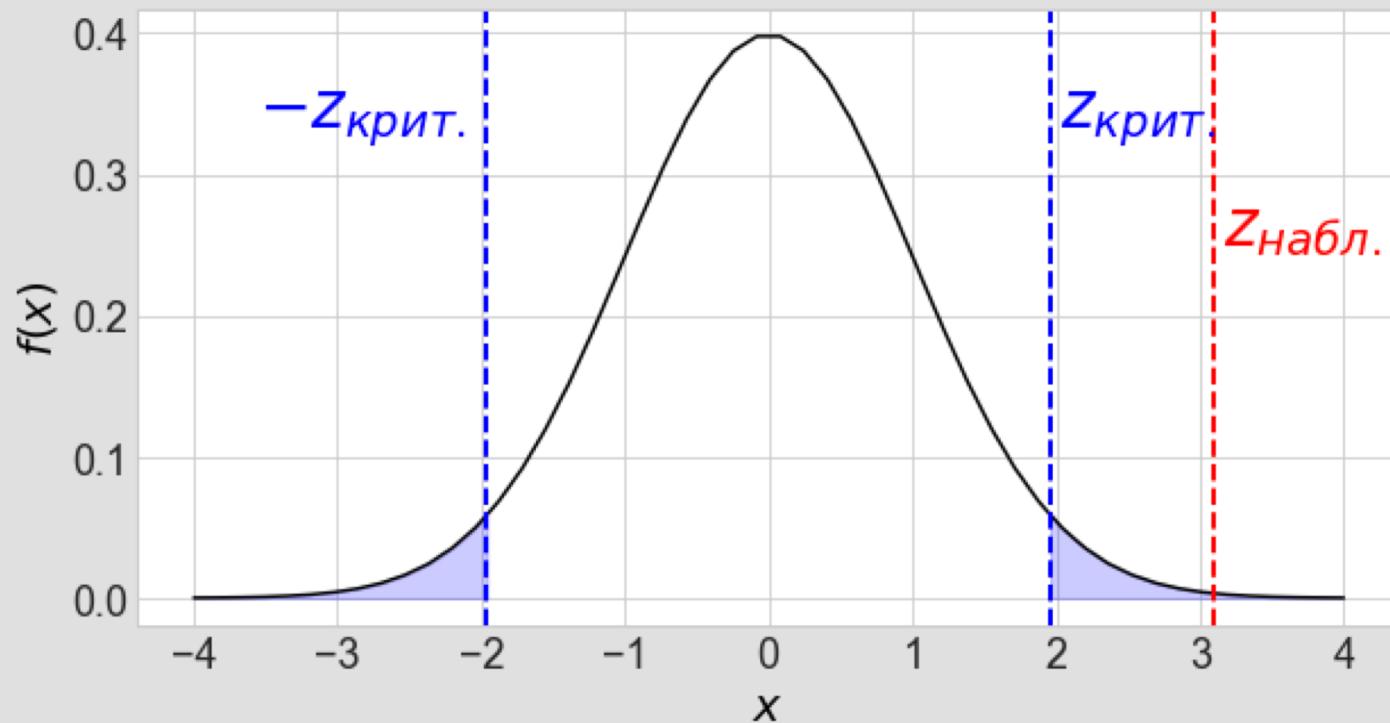


Мы не решаем, верна ли гипотеза,  
а проверяем, противоречат ли ей данные



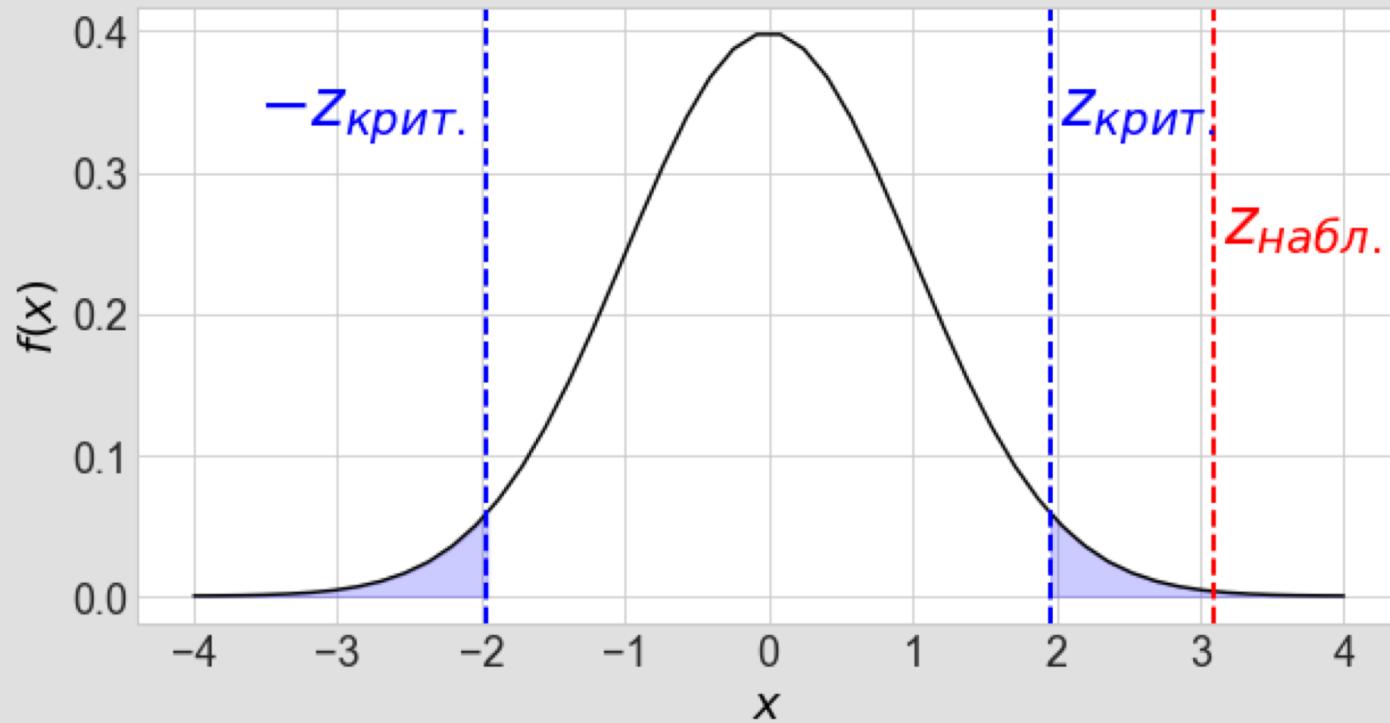
# Проверка гипотезы

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



# Проверка гипотезы

Если наблюдаемое значение попало между критическими, данные не противоречат гипотезе



# Уровень значимости

- Если мы отвергаем нулевую гипотезу, когда она верна – мы ошибаемся
- Выбирая порог для отсечения, мы фиксируем вероятность такой ошибки
- Вероятность такой ошибки  $\alpha$  – **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  $\alpha$  выбирают равным 0.05, 0.01 или 0.001



# Процедура проверки гипотез



# Процедура проверки гипотез

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

$$H_0: p = 0.5$$

$$H_a: p \neq 0.5$$

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:  $z \underset{H_0}{\sim} N(0, 1)$



# Процедура проверки гипотез

- Сравниваем наблюдаемое значение с критическим и делаем выводы



# Процедура проверки гипотез

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



# Процедура проверки гипотез

- ! Гипотеза, что Джеймс Бонд не различает напитки, не отвергается на уровне значимости 5%



# Нельзя принять нулевую гипотезу

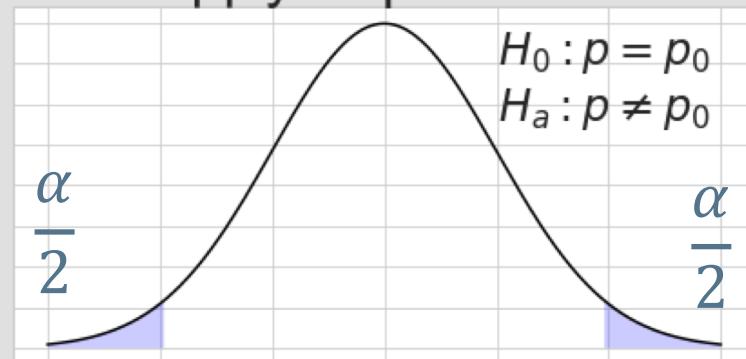
- Если при проверке нулевая гипотеза не отвергается, нельзя считать её доказанной
- Говорят, что данные не противоречат нулевой гипотезе
- Новые данные могут показать, что гипотеза неверна



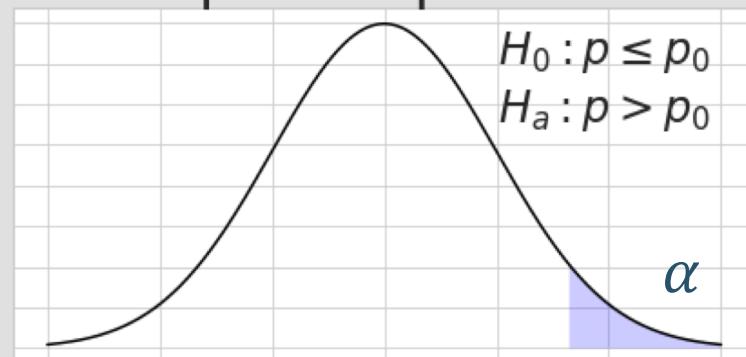
# Альтернативы

- 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

## Двусторонняя



## Правосторонняя



## Левосторонняя



# P-значение



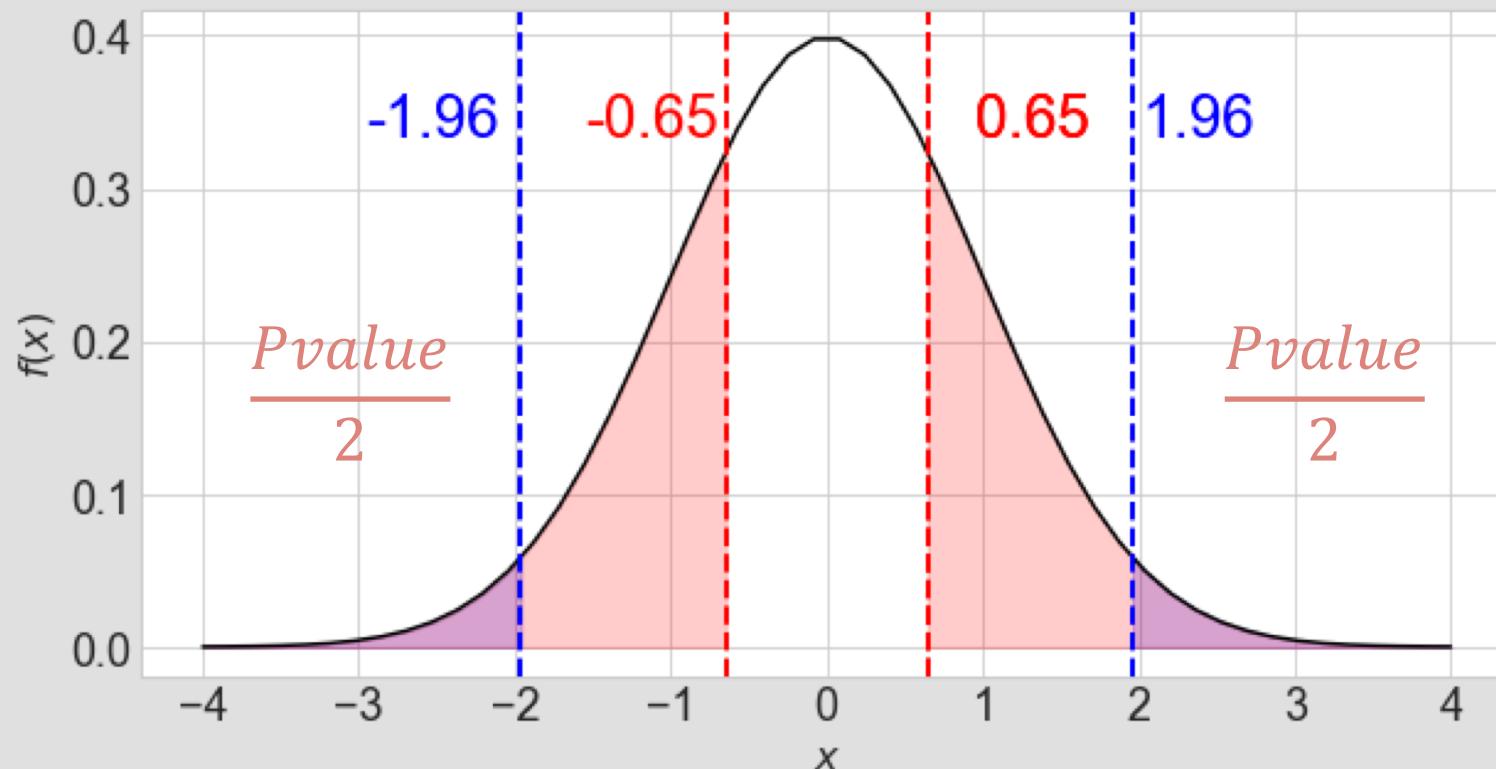
# P-значение

- Наблюданное значение попало в область между критическими  $\Rightarrow$  гипотеза не отвергается
- Голубая площадь под хвостами – уровень значимости



# P-значение

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

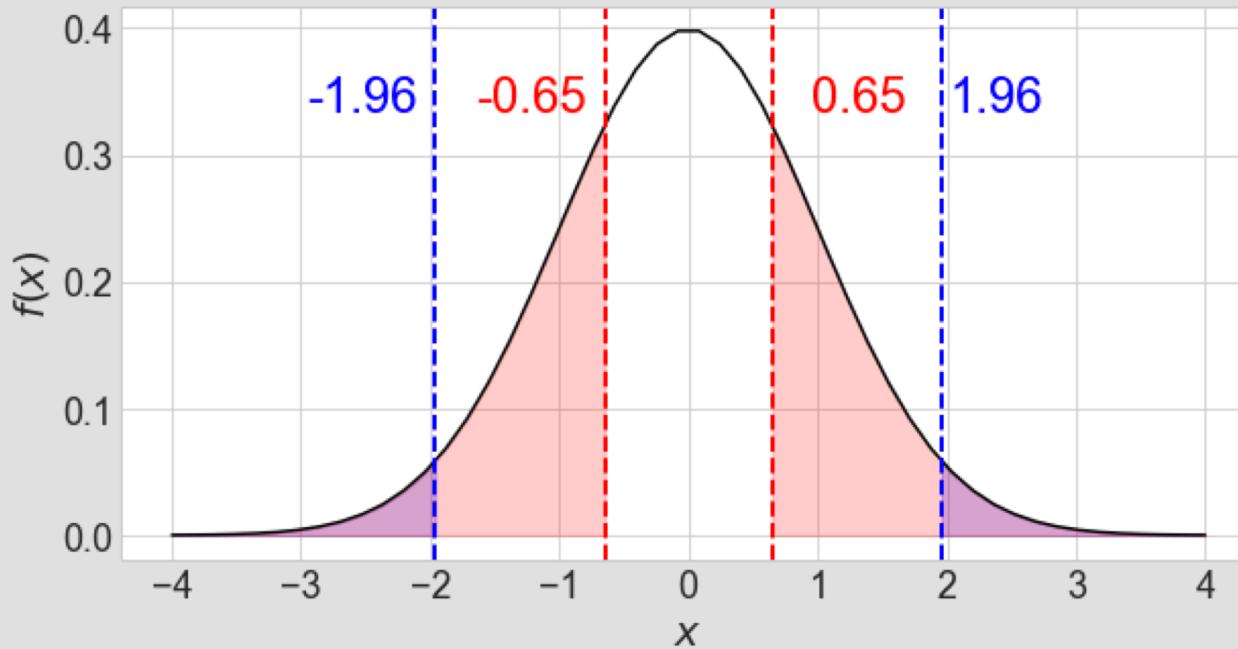


# P-значение упрощает проверку гипотез

Если красная площадь оказалась больше синей

$$p\_value > \alpha,$$

⇒ наблюдаемое значение попало в область между критическими, гипотеза не отвергается.



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

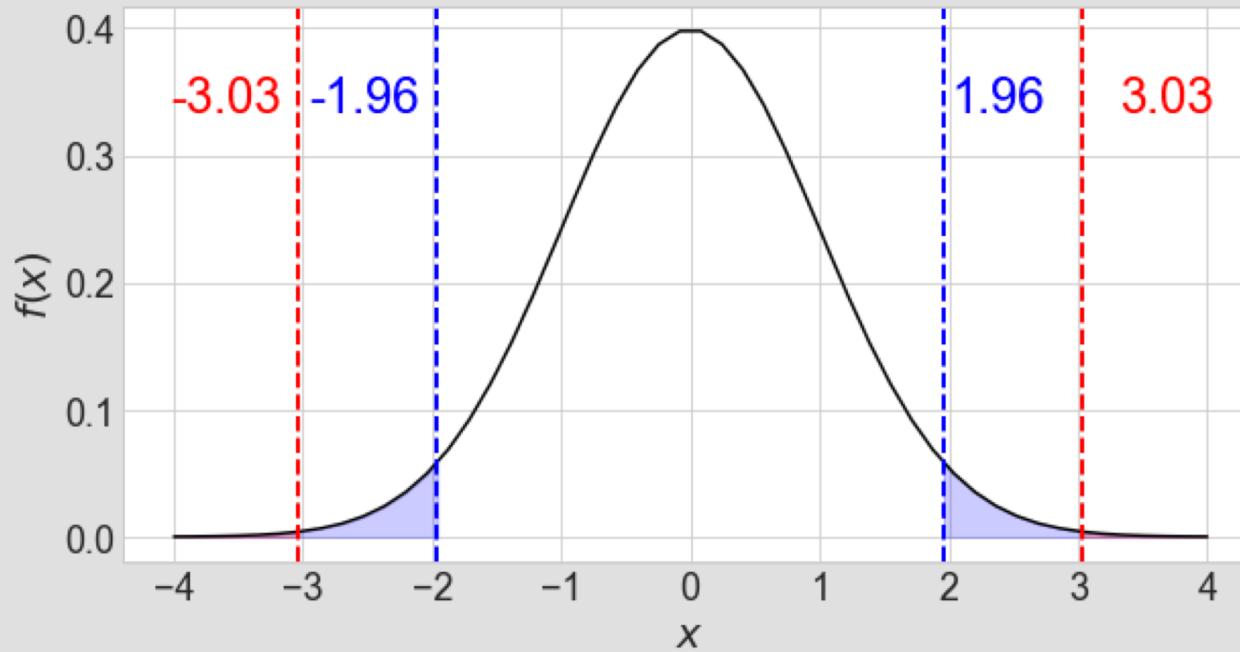


# P-значение упрощает проверку гипотез

Если красная площадь оказалась больше синей

$$p\_value > \alpha,$$

⇒ 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

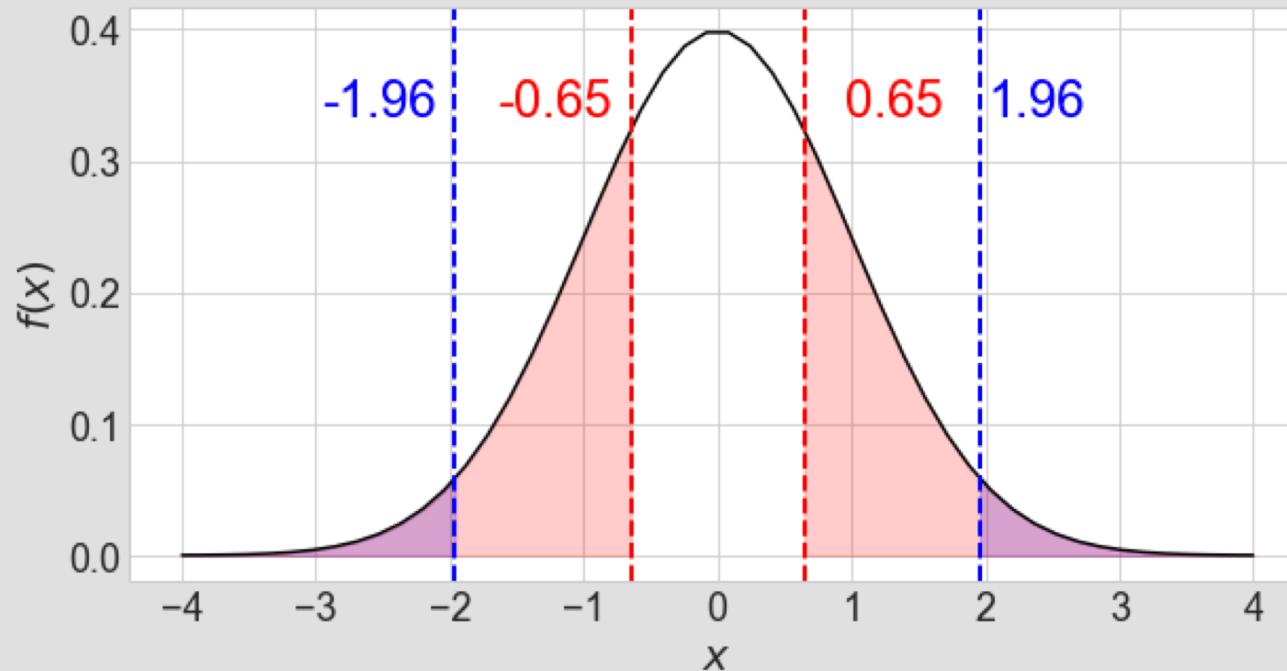


# P-значение

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

**Ответ:** равный P-значению

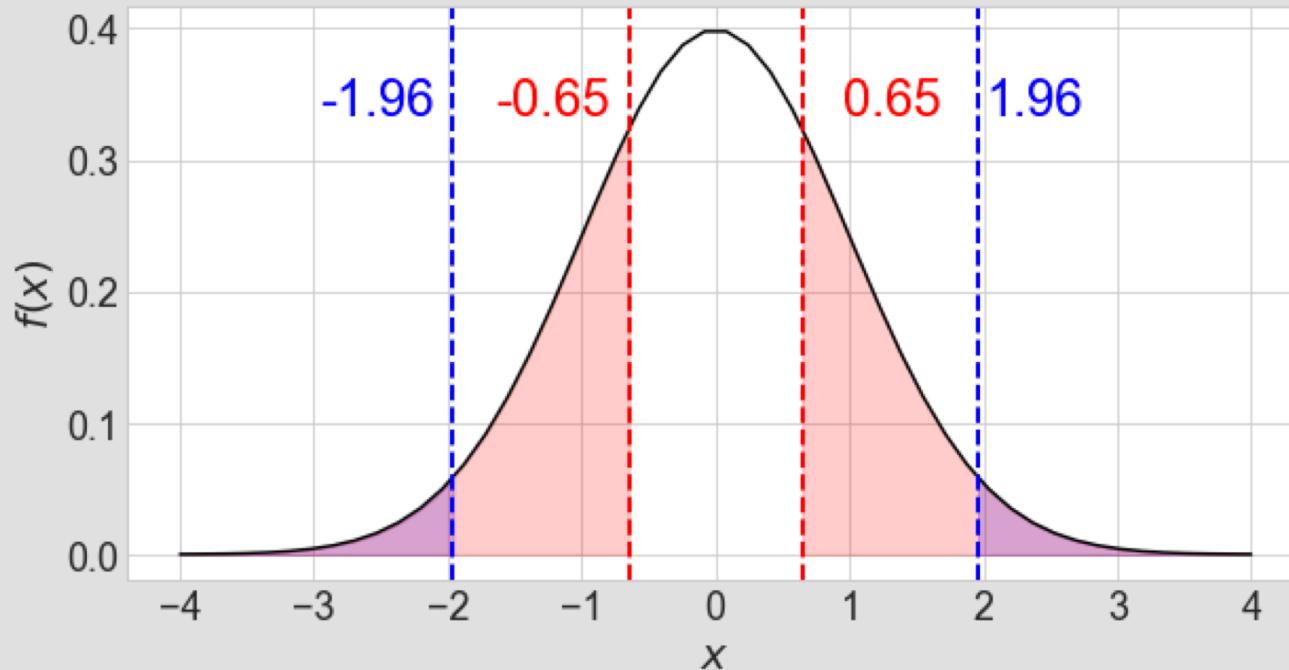
Из-за этого P-значение также называют **достигаемым уровнем значимости**



# P-значение

Достигаемый уровень значимости (**P-значение**) – это вероятность при справедливости нулевой гипотезы получить такое же наблюдаемое значение статистики, как в эксперименте либо ещё более экстремальное

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



# P-значение

Достигаемый уровень значимости (**P-значение**) – это вероятность при справедливости нулевой гипотезы получить такое же наблюдаемое значение статистики, как в эксперименте либо ещё более экстремальное

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

- ! 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

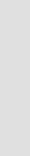


# Резюме

- С помощью Р-значения удобно проверять гипотезы
- Оно помогает не акцентировать внимание на том, как именно вычисляется критическое значение, и упрощает понимание статистических протоколов

Типичный протокол из статистического пакета:

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-2.5988	1.918	-1.355	0.175	-6.358	1.161
y	0.4170	0.297	1.404	0.160	-0.165	0.999



*p\_value*



# Ошибки 1 и 2 рода



# Ошибки, что мы совершаем

	$H_0$ верна	$H_0$ неверна	
$H_0$ не отвергается	<i>ok</i>	$\beta$	ошибка 2 рода
$H_0$ отвергается	$\alpha$	<i>ok</i>	ошибка 1 рода

$$\alpha = \mathbb{P}(H_0 \text{ отвергнута} \mid H_0 \text{ верна})$$

$$\beta = \mathbb{P}(H_0 \text{ не отвергнута} \mid H_0 \text{ не верна})$$

Величину  $1 - \beta$  называют **мощностью** критерия



# Ошибки, что мы совершаем

$H_0$ : нет беременности

$H_0$  верна

$H_a$ : есть беременность

$H_0$  неверна

$H_0$  не  
отвергается



Вы не беременны



Вы не беременны

$H_0$   
отвергается



Вы беременны



Вы беременны



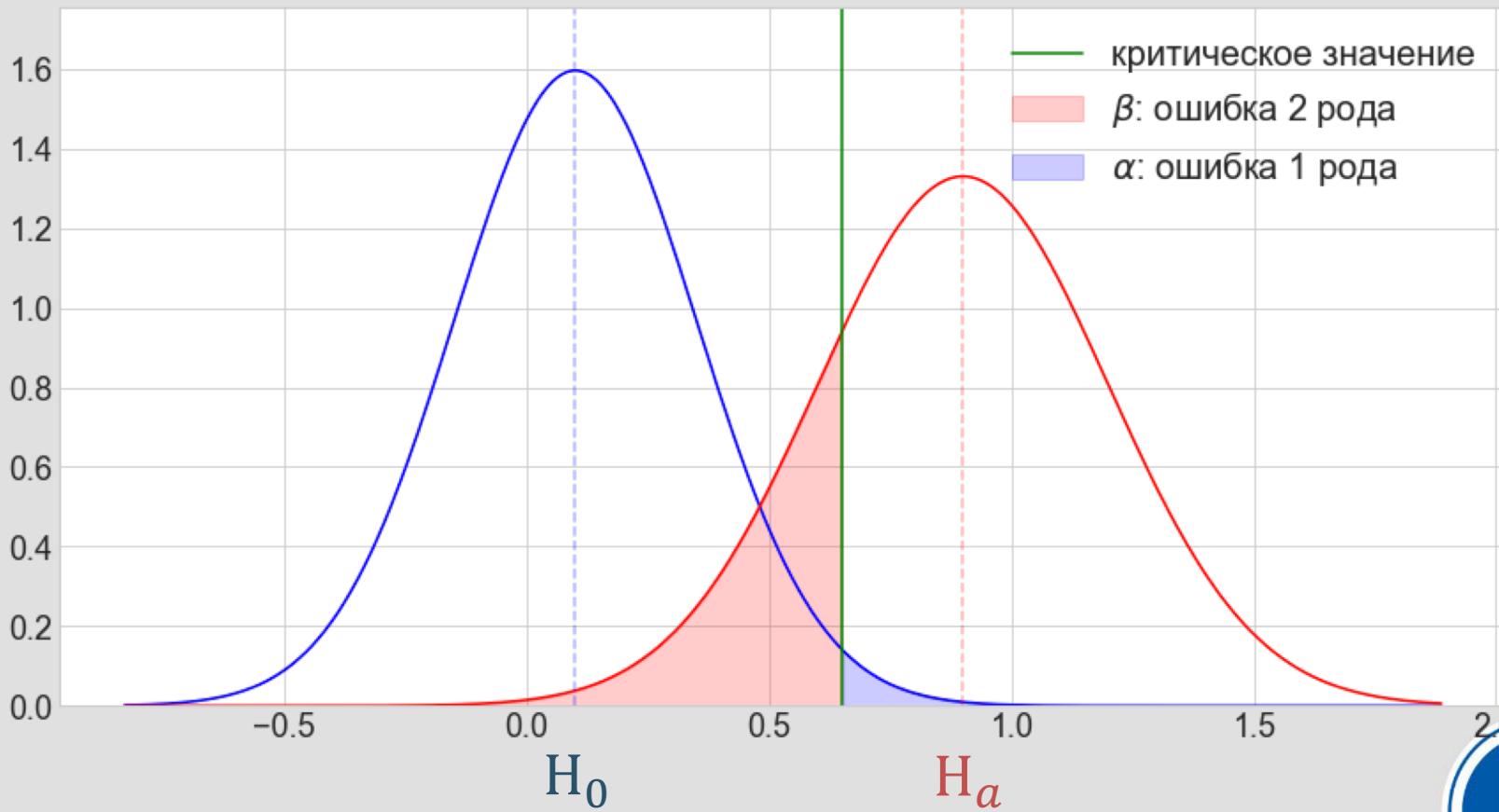
# Ошибки, что мы совершаем

$H_0: p = p_0$

$H_a: p = p_a$



При уменьшении ошибки первого рода  
всегда возрастает ошибка второго рода



# Презумпция нулевой гипотезы

**Презумпция невиновности:** человек считается невиновным, пока его вина не будет доказана в суде

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

**Ошибки первого и второго рода неравнозначны:**

мы перед экспериментом фиксируем  $\alpha$ , а  $\beta$  минимизируется по остаточному принципу

- ❗ Вывод “данные противоречат гипотезе” всегда весомее и категоричнее, чем “данные не противоречат гипотезе”



# Аналогия с классификацией

	$y = 1$	$y = 0$	
$\hat{y} = 1$	$TP$	$FP$	ошибка 2 рода
$\hat{y} = 0$	$FN$	$TN$	ошибка 1 рода

**Пример:** хотим, чтобы классификатор удалял спам и задел минимум хороших документов

**Подбор порога:** зафиксировать  $FPR = \frac{FP}{FP+TN} \leq 0.05$

(доля зря удалённых), а дальше максимизировать полноту

$$Recall = TPR = \frac{TP}{TP+FN} = 1 - \frac{FN}{TP+FN}$$



# Наиболее мощный критерий

- 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**



# Резюме

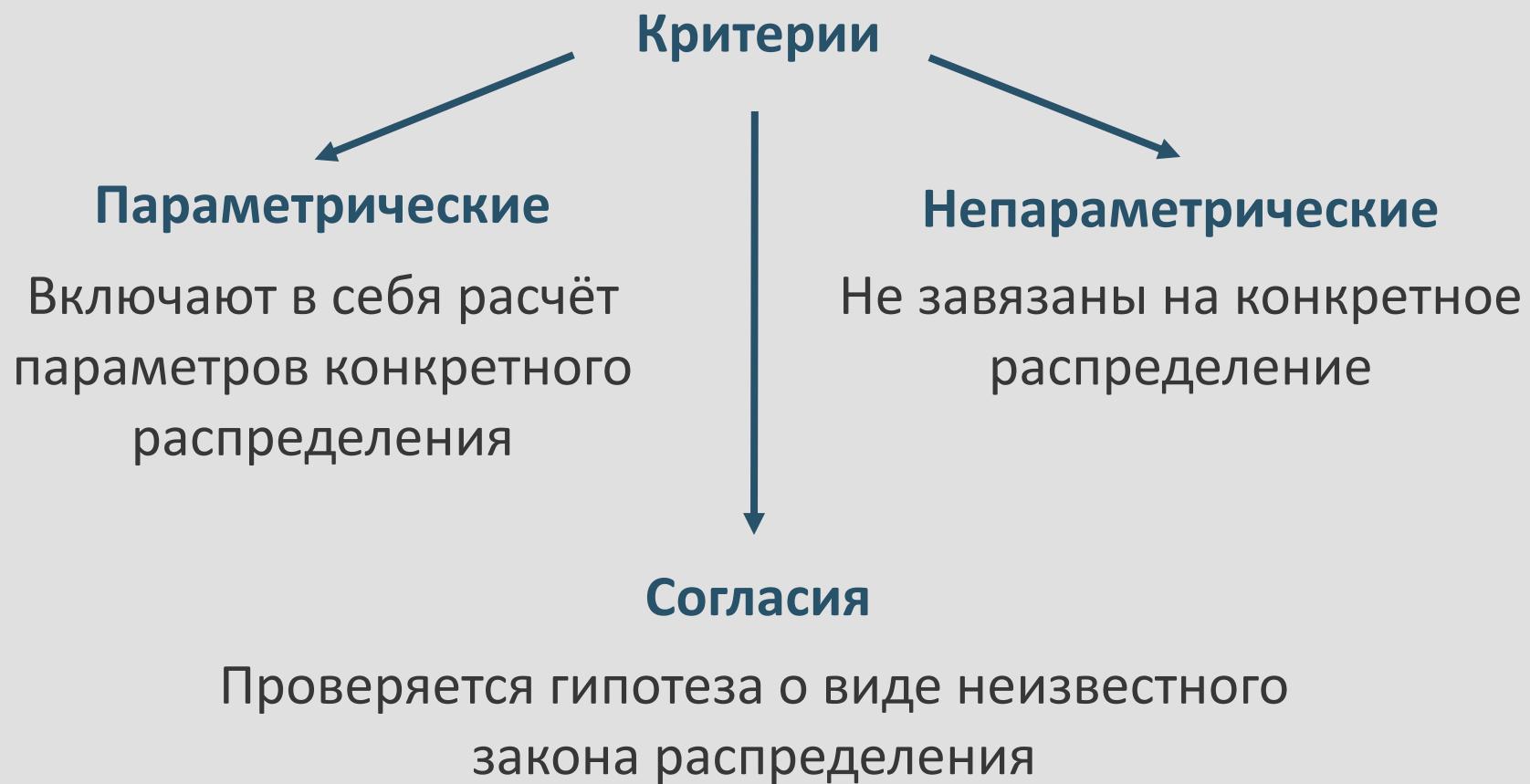
- Ошибки первого и второго рода неравнозначны
- Имеется презумпция нулевой гипотезы
- Обычно нулевую гипотезу формулируют так, что нет значимого эффекта



# Параметрические критерии



# Какими бывают 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



# 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

**Есть и другие группы критериев.** Любое математически формализованное правило, по которому проверяется гипотеза, можно назвать критерием.



# Схема математической статистики

Выборка:  $x_1, \dots, x_n$  Параметр:  $\theta$

$\hat{\theta}$



$f_{\hat{\theta}}(t)$

Как оценить

- Метод моментов
- Метод максимального правдоподобия

Хорошие свойства

- Несмещенная
- Состоятельная
- Эффективная

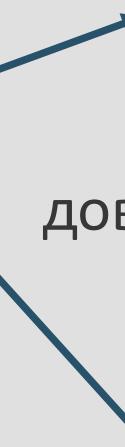
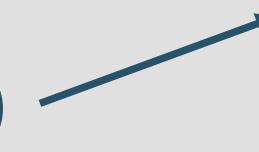
Союзники

Асимптотические  
(при большом  $n$ )

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi_n^2, t_n, F_{n,k}$
- Ещё союзники!



Точность  
оценки,  
прогнозов

доверительные  
интервалы

Ответы на  
вопросы  
проверка  
гипотез



# Схема математической статистики

Выборка:  $x_1, \dots, x_n$  Параметр:  $\theta$

$\hat{\theta}$

$f_{\hat{\theta}}(t)$

Асимптотические критерии

Точные критерии

## Союзники

Асимптотические (при большом  $n$ )

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi_n^2, t_n, F_{n,k}$
- Ещё союзники!

Точность оценки, прогнозов

доверительные интервалы

Ответы на вопросы проверка гипотез



# Гипотезы оолях

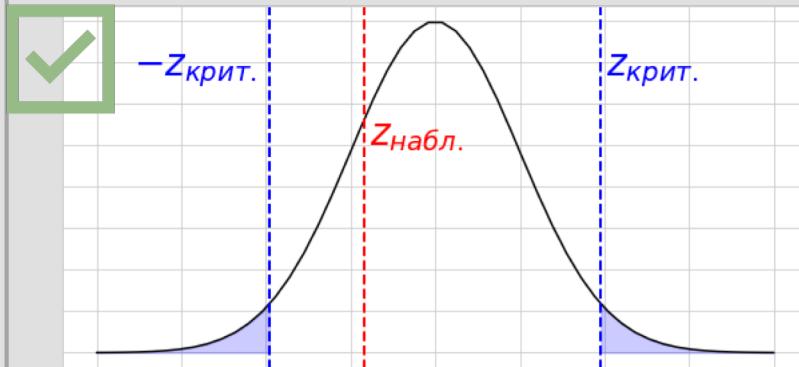
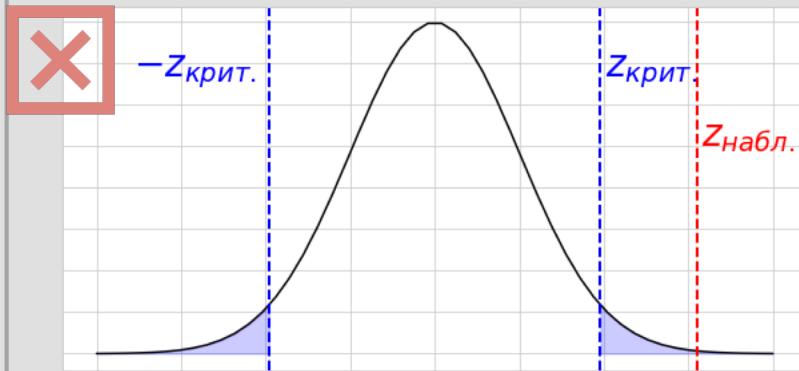


# Z-критерий для доли

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

$H_0: p = p_0$

$H_a: p \neq p_0$



ЦПТ:

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

Критерий для проверки:

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



Критерий  
асимптотический,  
т.к. использует ЦПТ



## Пример (дискриминация):

- В 70-х за участие в антивоенной демонстрации арестован педиатр, автор книг о воспитании детей, Бенджамин Спок
- Суд присяжных назначается сложной многоступенчатой процедурой отбора
- Отобрано 300 человек, из них 90 женщин
- Адвокаты протестуют на предвзятость выбора
- Проверяем гипотезу о предвзятости на уровне значимости 5%



# Пример (дискриминация):

$H_0$ : отбор беспристрастный



$H_0: p = 0.5$

$H_a$ : женщин дискриминируют

$H_a: p < 0.5$



Левосторонняя альтернатива, так  
как нас интересует именно  
дискриминация женщин



# Пример (дискриминация):

$H_0$ : отбор беспристрастный



$H_0: p = 0.5$

$H_a$ : женщин дискриминируют

$H_a: p < 0.5$

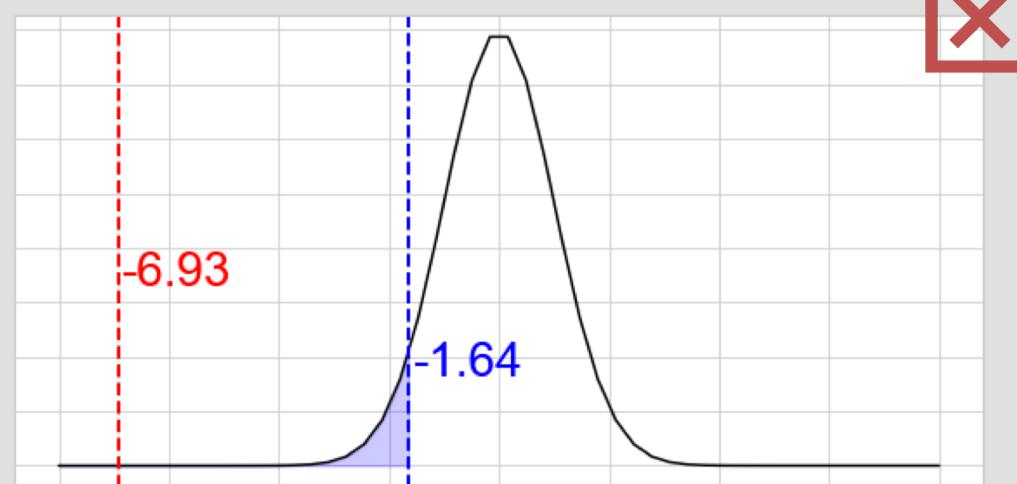
$X_1, \dots, X_n \sim iid Bern(p)$

$X_i = 1$ , если присяжный – женщина

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

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

$$z_{0.05} = -1.64$$



Гипотеза о беспристрастном  
отборе **отвергается**



# Z-критерий для разности независимых долей

$X_1, \dots, X_{n_x} \sim iid Bern(p_x)$

$Y_1, \dots, Y_{n_y} \sim iid Bern(p_y)$

**Выборки независимые**

$H_0: p_x = p_y$        $H_a: p_x \neq p_y$

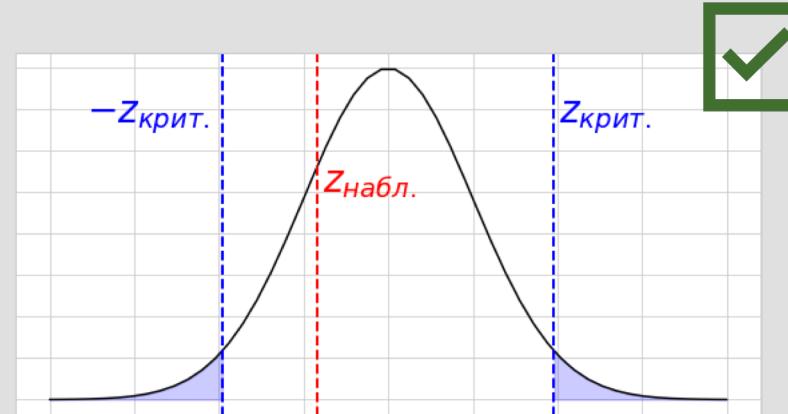
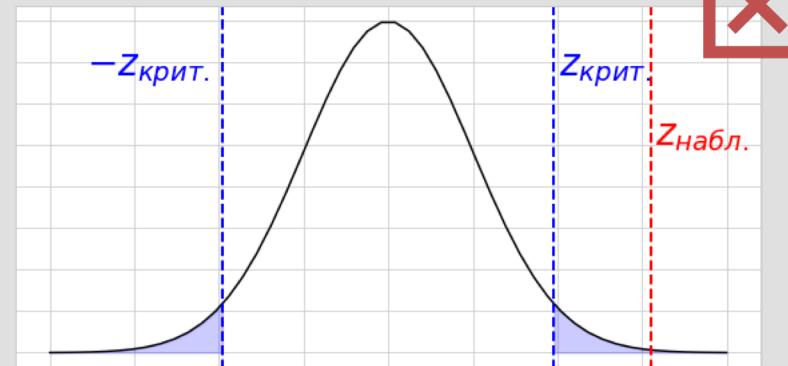
❗ Критерий  
асимптотический,  
т.к. использует ЦПТ

**Критерий для проверки:**

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{P(1-P) \cdot \left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$

$$\underset{H_0}{\overset{asy}{\sim}} N(0, 1)$$

$m_i$  – число 1 в выборке



## Пример (кофе):

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



# Пример (кофе):

$H_0$ : кофе любят одинаково



$H_0: p_{\Pi} = p_M$

$H_a$ : в Москве любят сильнее

$H_a: p_{\Pi} < p_M$

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

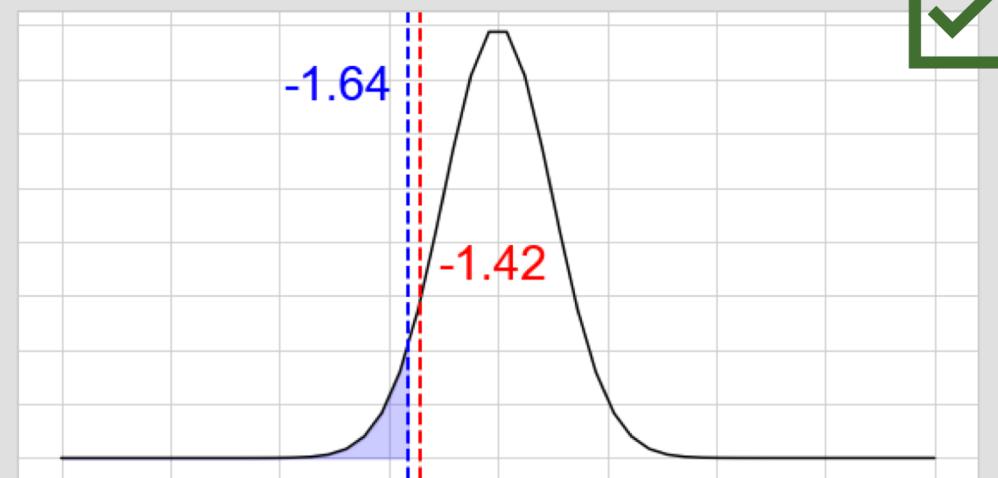
$Y_1, \dots, Y_{100} \sim iid Bern(p_{\Pi})$

$$\hat{p}_M = 0.6, \quad \hat{p}_{\Pi} = 0.5$$

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

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

$$z_{0.05} = -1.64$$



Гипотеза об одинаковой любви **не отвергается**



# Z-критерий для разности зависимых долей

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

$Y_1, \dots, Y_n \sim iid Bern(p_y)$

Выборки зависимые

$H_0: p_x = p_y$

$H_a: p_x \neq p_y$

Критерий для проверки:

$$z = \frac{c - b}{\sqrt{c + b - \frac{(c - b)^2}{n}}} \underset{H_0}{\sim} N(0, 1)$$

		$X$
		0
		1
		$b$
$Y$		$a$
		$d$
		$c$

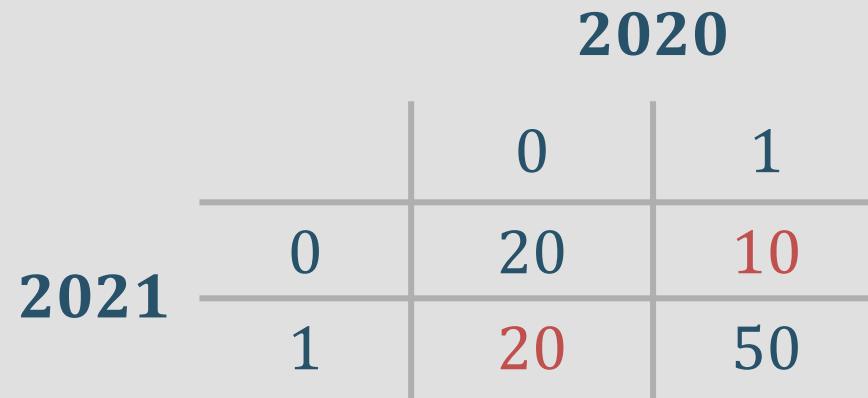
Люди одни и те же, нас интересуют те, кто поменял мнение

! Критерий асимптотический



## Пример (кофе):

- В 2020 году в Москве сотню человек спросили, любят ли они кофе
- Через год у этих же ста человек снова спросили, любят ли они кофе по-прежнему
- Правда ли, что число любителей кофе изменилось?



# Пример (кофе):

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

$Y_1, \dots, Y_n \sim iid Bern(p_y)$

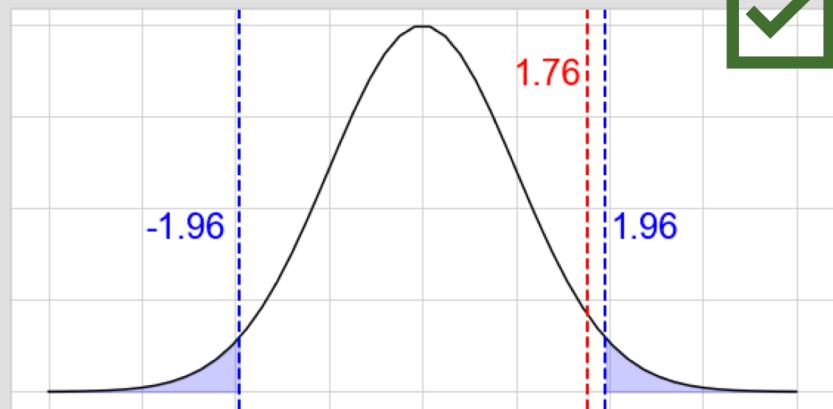
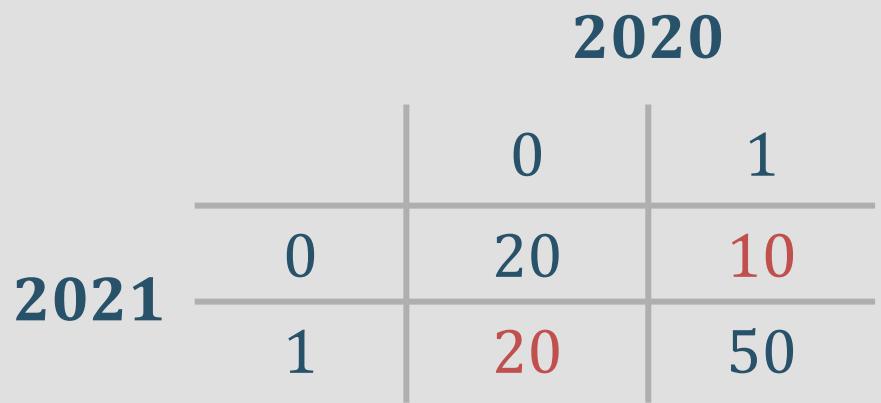
Выборки зависимые

$H_0: p_x = p_y$

$H_a: p_x \neq p_y$

$$z_{obs} = \frac{20 - 10}{\sqrt{20 + 10 - \frac{(20 - 10)^2}{100}}} = 1.76$$

$$z_{0.975} = 1.96$$



Гипотеза о том, что люди не  
поменяли вкусовых  
предпочтений **не отвергается**



# Резюме

- Для проверки гипотезы о долях используется z-тест, основанный на ЦПТ
- В случае независимых и зависимых выборок статистика считается немного по-разному, для зависимых мы акцентируем внимание только на изменениях
- Обе имеют асимптотически нормальное распределение



С помощью ровно этих же статистик мы до этого строили для долей асимптотические доверительные интервалы



# Гипотезы о средних



# Z-критерий для среднего

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

$$H_0: \mu = \mu_0$$

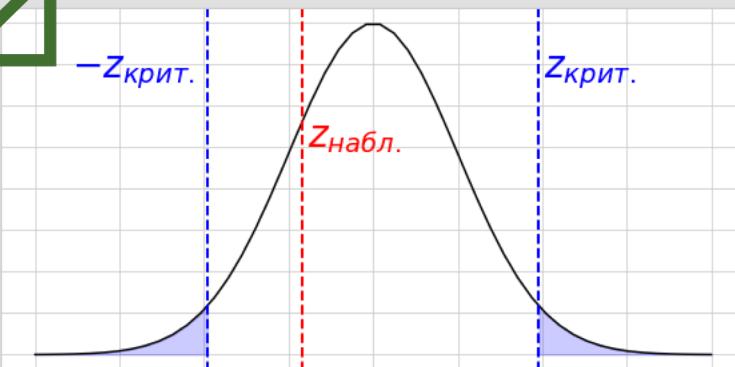
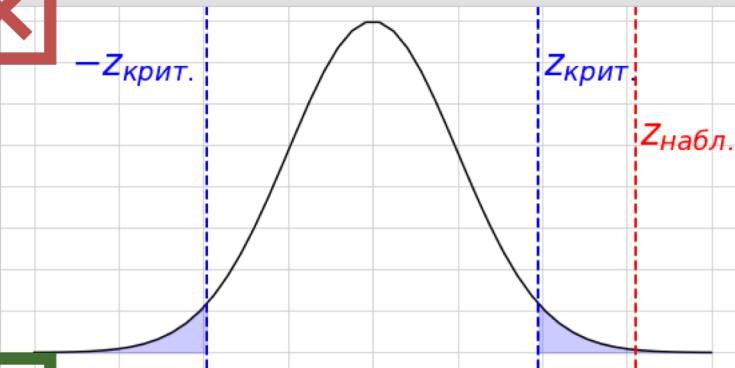
$$H_a: \mu \neq \mu_0$$

ЦПТ:

$$\bar{x} \underset{H_0}{\overset{asy}{\sim}} N\left(\mu_0, \frac{\hat{\sigma}^2}{\sqrt{n}}\right)$$

Критерий для проверки:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \underset{H_0}{\overset{asy}{\sim}} N(0, 1)$$



Критерий асимптотический, т.к. использует ЦПТ, применим для любых средних

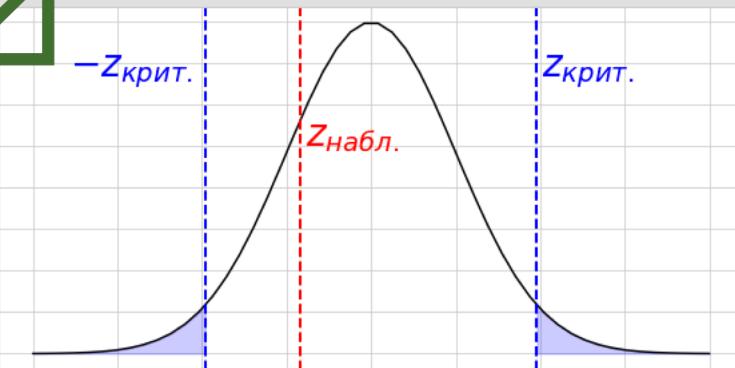
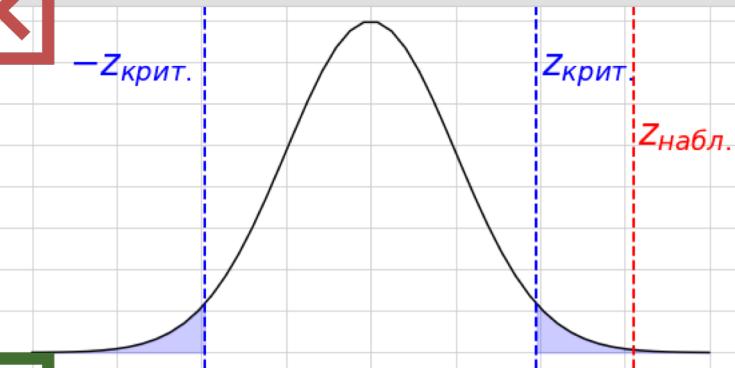


# t-критерий для среднего

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

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0 \quad \sigma^2 - \text{известна}$$



ЦПТ:

$$\bar{x} \underset{H_0}{\overset{\alpha \vee}{\sim}} N\left(\mu_0, \frac{\hat{\sigma}^2}{\sqrt{n}}\right)$$

Критерий для проверки:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \underset{H_0}{\overset{\alpha \vee}{\sim}} N(0, 1)$$



Критерий точный,  
используется  
предположение  
о нормальности выборки  
и известности  $\sigma^2$

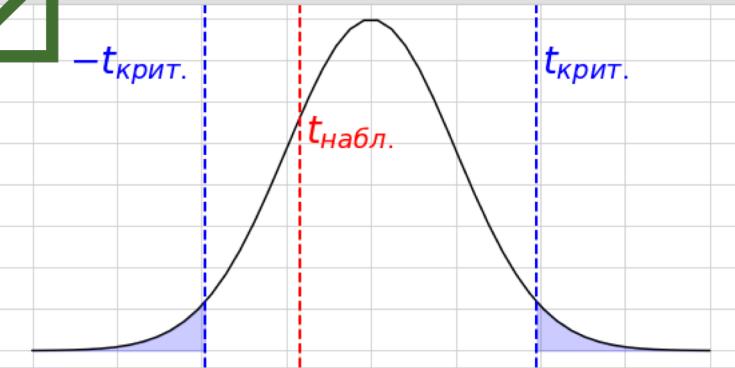
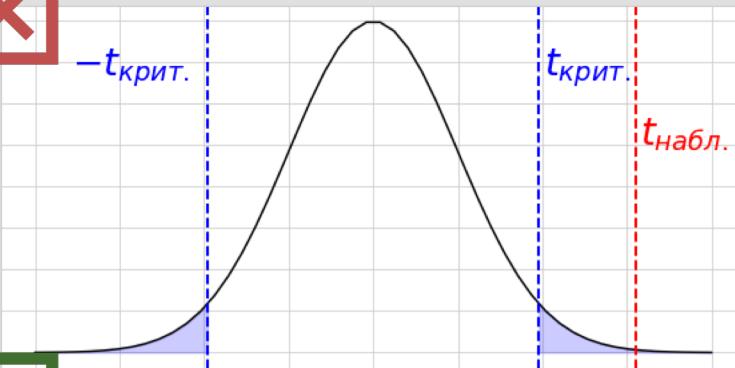


# t-критерий для среднего

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

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0 \quad \sigma^2 - \text{НЕизвестна}$$



ЦПТ:

$$\bar{x} \underset{H_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{\sqrt{n}}\right)$$

Критерий для проверки:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} \underset{H_0}{\sim} t(n-1)$$



Критерий точный,  
используется  
предположение  
о нормальности выборки



## Пример (курсы подготовки):

- Каждый год люди сдают ЕГЭ по математике. В обычной ситуации средний результат составляет 65 баллов.
- Игорь открыл свои курсы подготовки к ЕГЭ, его группа из 100 человек получила в среднем 70 баллов, стандартное отклонение составило 20 баллов.
- Правда ли, что курсы Игоря помогают получить более высокий балл?
- Проверяем гипотезу на уровне значимости 5%



# Пример (курсы подготовки):

$H_0$ : курсы Игоря неэффективны



$H_0: \mu = 60$

$H_a$ : курсы помогают

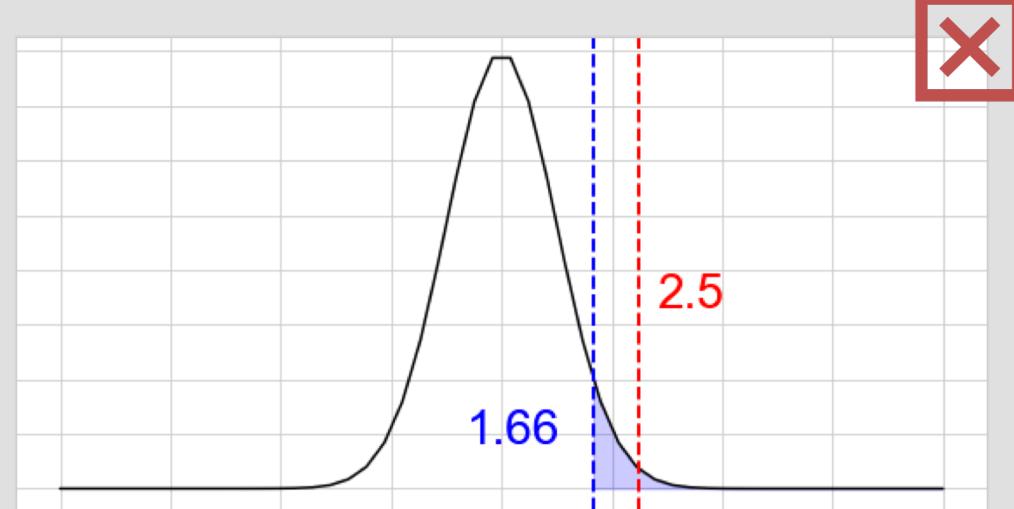
$H_a: \mu > 60$

**Предположение:**

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

$$t_{obs} = \frac{70 - 65}{\frac{20}{\sqrt{100}}}$$

$$t_{0.95} = 1.66$$



Гипотеза о неэффективности курсов  
Игоря **отвергается**,  
они помогают получить более  
высокий балл



# Z-критерий для разности средних

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

$$Y_1, \dots, Y_m \sim iid (\mu, \sigma^2)$$



Критерий асимптотический,  
т.к. использует ЦПТ

Выборки независимые

$$H_0: \mu_1 = \mu_2$$

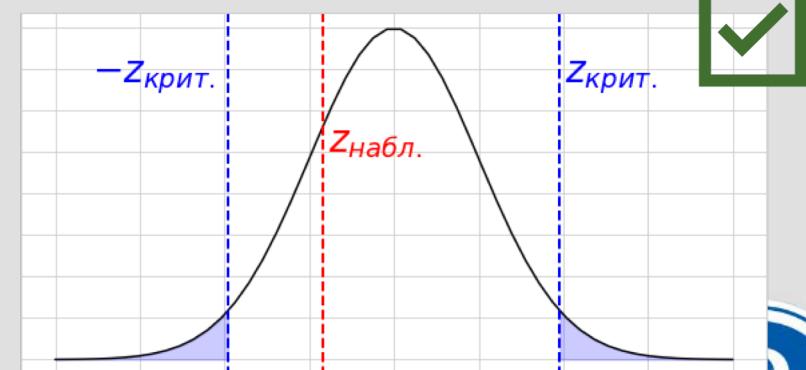
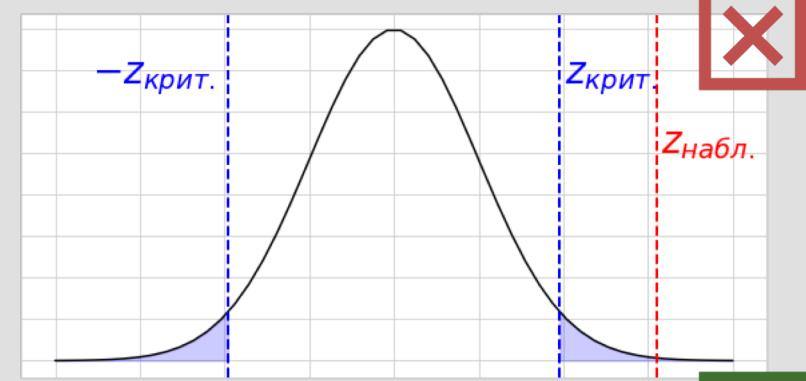
$$H_a: \mu_1 \neq \mu_2$$

ЦПТ:

$$\bar{X}_1 - \bar{X}_2 \underset{H_0}{\overset{asy}{\sim}} N\left(0, \frac{s_1^2}{n} + \frac{s_2^2}{m}\right)$$

Критерий для проверки:

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}} \underset{H_0}{\overset{asy}{\sim}} N(0, 1)$$



# Точные критерии для разности средних

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

$$H_0: \mu_1 = \mu_2$$

$$Y_1, \dots, Y_m \sim iid N(\mu, \sigma^2)$$

$$H_a: \mu_1 \neq \mu_2$$

Выборки независимые

Критерий для проверки:

$$\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

дисперсии  
известны

дисперсии  
неизвестны,  
но равны

дисперсии  
неизвестны,  
различаются



# Точные критерии для разности средних

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

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \underset{H_0}{\sim} t(\nu)$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$$

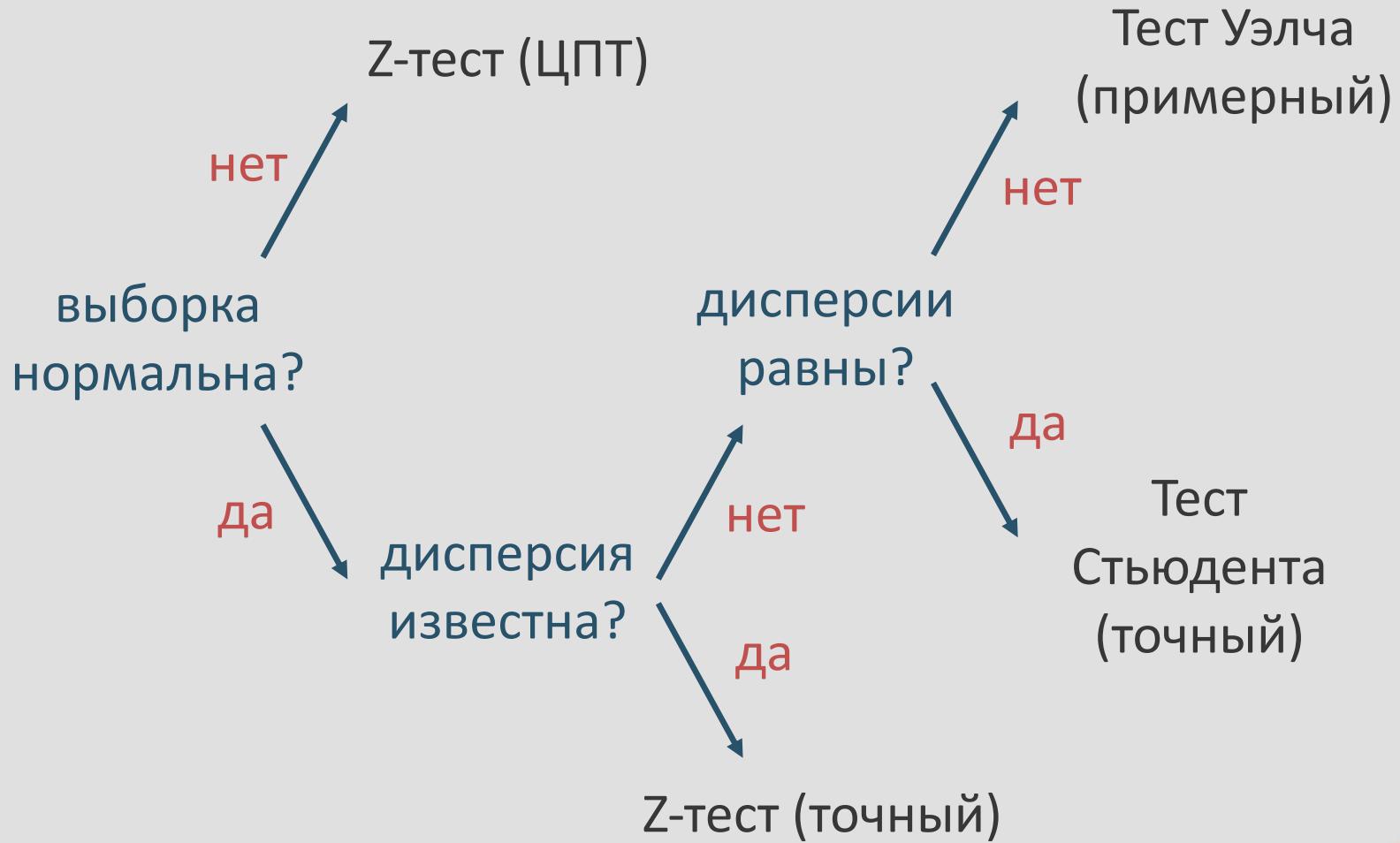


## Пример (стоимость недвижимости):

- До коронакризиса один квадратный метр в Москве стоил 105 тыс. рублей при стандартном отклонении в 40 тыс.
- После коронакризиса один квадратный метр стоит 90 тыс. рублей при стандартном отклонении в 50 тыс.
- Упала ли стоимость, если  $n_x = 20$ ,  $n_y = 30$ ?



# Пример (стоимость недвижимости):



# Пример (стоимость недвижимости):

$H_0$ : стоимость не поменялась



$H_0: \mu_{2021} = \mu_{2020}$

$H_a$ : стоимость упала

$H_a: \mu_{2021} < \mu_{2020}$

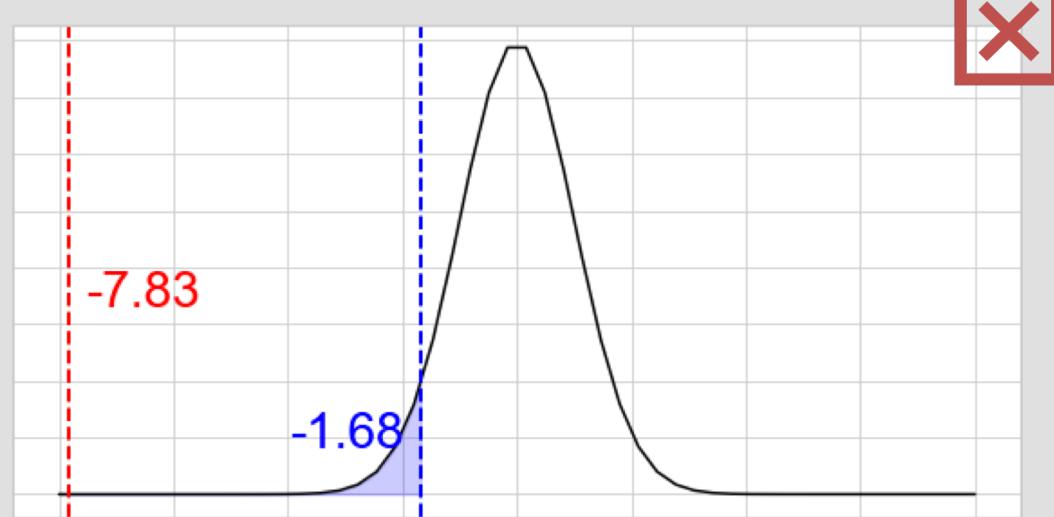
## Предположение:

Выборки независимы,  
нормальны

$$t_{obs} = \frac{90 - 105}{\sqrt{\frac{40}{20} + \frac{50}{30}}}$$

$$\nu = 43.89$$

$$t_{0.95} = -1.67$$



Гипотеза о неизменности  
цен **отвергается**



# Разность средних (зависимые выборки)

Выборки зависят друг от друга:

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

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

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

- Используем распределение Стьюдента, дисперсию считаем по формуле:

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



## Пример (польза лекарства):

- На 5 испытуемых сравнивают два лекарства против респираторных заболеваний
- Каждый вдыхает лекарство из ингалятора, а после принимает участие в упражнении на беговой дорожке
- Измеряется время достижения максимальной нагрузки
- Затем то же самое проделывается со вторым лекарством, правда ли что лекарства не отличаются?



# Пример (польза лекарства):

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Предположение:

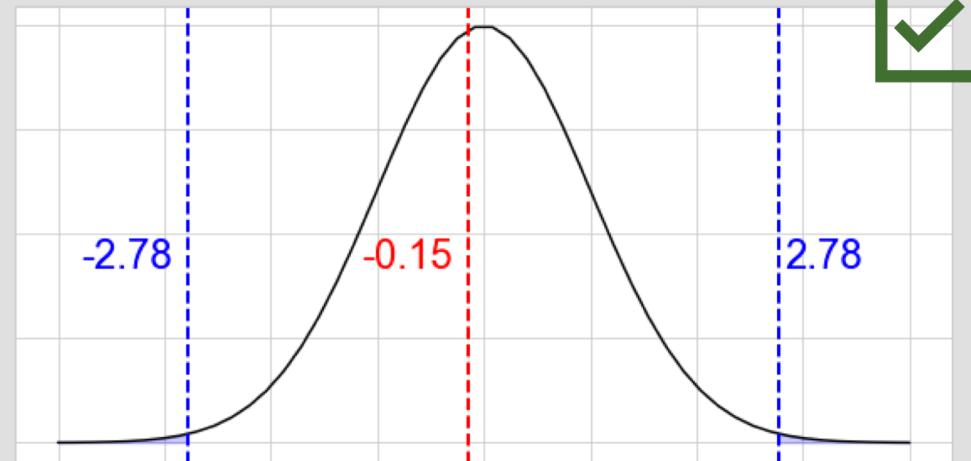
Выборки независимы,  
нормальны

$$\bar{d} = \frac{1}{5} \sum_{i=1}^5 d_i = -6$$

$$s^2 = \frac{1}{5-1} \sum_{i=1}^5 (d_i - \bar{d})^2 = 92.5$$

$$t_{obs} = \frac{-6 - 0}{92.5/\sqrt{5}} \quad t_{0.95} = 2.78$$

	Лекарство 1	50	40	45	45	35
Лекарство 2	60	30	30	35	30	
	$d_i$	10	-10	-15	-10	-5



Гипотеза, что время  
достижения максимальной  
нагрузки не отличается  
не отвергается



# Резюме

- Благодаря удобству ЦПТ, критерии основанные на ней широко распространены
- Для маленьких выборок бывает лучше использовать точные критерии
- Если эти критерии основаны на нормальном распределении, надо проверить выборку на нормальность
- Если предполагается равенство дисперсий, это тоже не помешает проверить



С помощью ровно этих же статистик мы до этого строили для средних доверительные интервалы.



# Гипотезы о дисперсиях



# $\chi^2$ – критерий для дисперсии

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

Критерий для проверки:

$\mu$  – известно

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_a: \sigma^2 > \sigma_0^2$$

$$\frac{n \cdot s^2}{\sigma_0^2} = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} \stackrel{H_0}{\sim} \chi_n^2$$

- Высокая дисперсия связана с риском и нестабильностью
- Мы хотим знать, принимает ли дисперсия своё значение ниже  $\sigma_0^2$
- Из-за этого в качестве альтернативы обычно используют правостороннюю



# $\chi^2$ – критерий для дисперсии

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

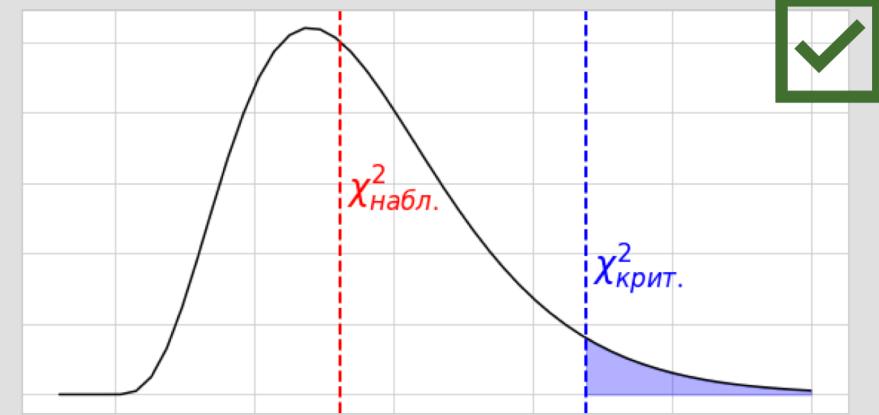
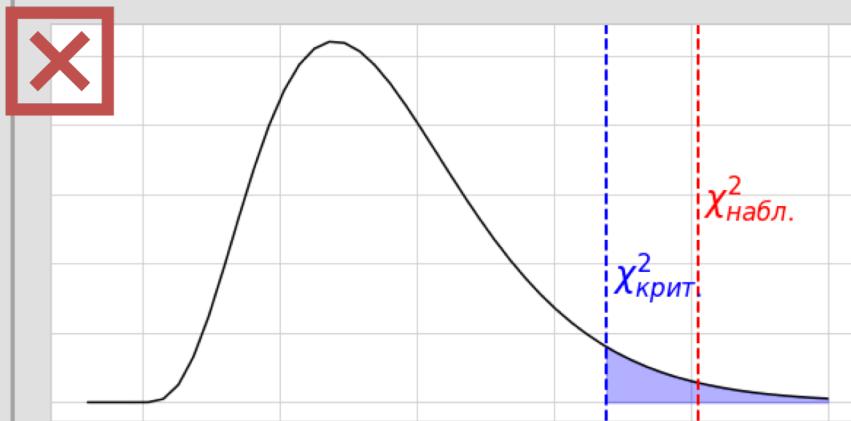
$\mu$  – известно

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_a: \sigma^2 > \sigma_0^2$$

Критерий для проверки:

$$\frac{n \cdot s^2}{\sigma_0^2} = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} \stackrel{H_0}{\sim} \chi_n^2$$



- ❗ Критерий точный, используется предположение о нормальности выборки и известности  $\mu$



# $\chi^2$ – критерий для дисперсии

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

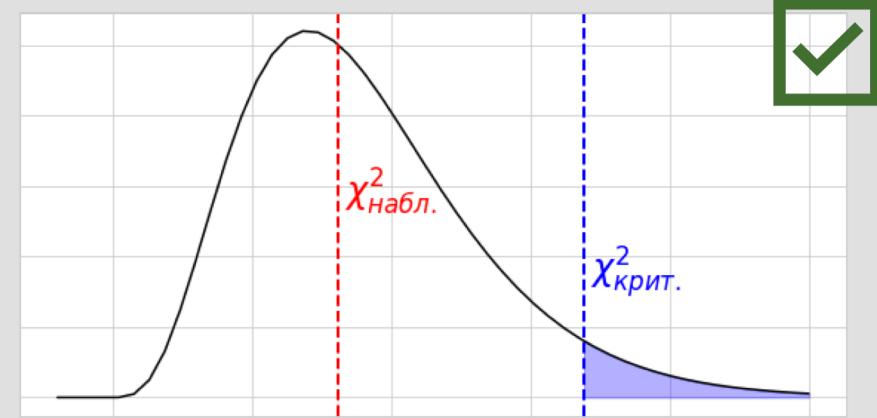
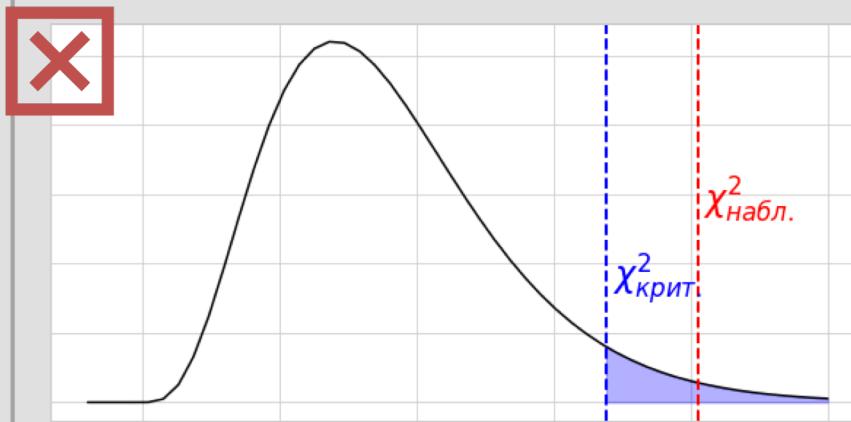
$\mu$  – НЕизвестно

$H_0: \sigma^2 = \sigma_0^2$

$H_a: \sigma^2 > \sigma_0^2$

Критерий для проверки:

$$\frac{(n - 1) \cdot s^2}{\sigma_0^2} \underset{H_0}{\sim} \chi_{n-1}^2$$



- ❗ Критерий точный, используется предположение о нормальности выборки



## Пример (ценная бумага):

- Мистер Белфорд считает, что вкладываться в активы с дисперсией доходности выше **0.04** очень рискованно.
- За последний год дисперсия Chesapeake Energy составила **0.09**. Следует ли вкладываться в эту бумагу? Решение принимается на уровне значимости **5%**.



# Пример (ценная бумага):

$H_0$ : стоит вкладываться

$H_a$ : лучше не надо

↔

$H_0: \sigma^2 \leq 0.04$

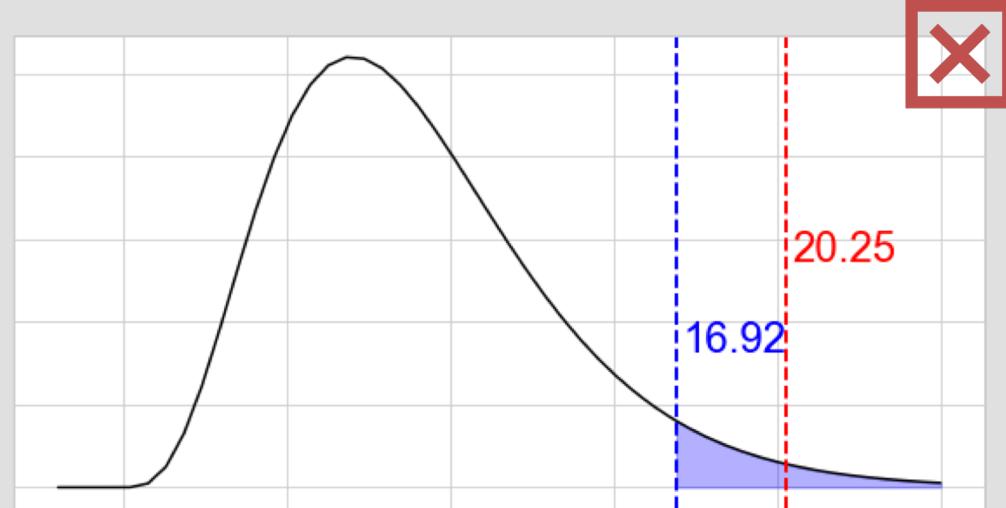
$H_a: \sigma^2 > 0.04$

Предположение:

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

$$\chi^2_{obs} = \frac{(10 - 1) \cdot 0.09}{0.04}$$

$$\chi^2_9(0.95) = 16.92$$



Гипотеза о том, что риск  
бумаги устроит мистера  
Белфорда **отвергается**



# Тест Фишера для отношения дисперсий

$X_1, \dots, X_n \sim iid N(\mu_x, \sigma_x^2)$

$Y_1, \dots, Y_m \sim iid N(\mu_y, \sigma_y^2)$

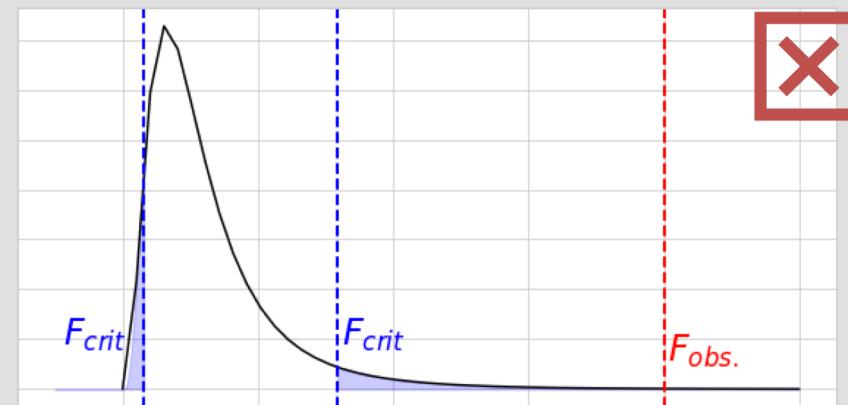
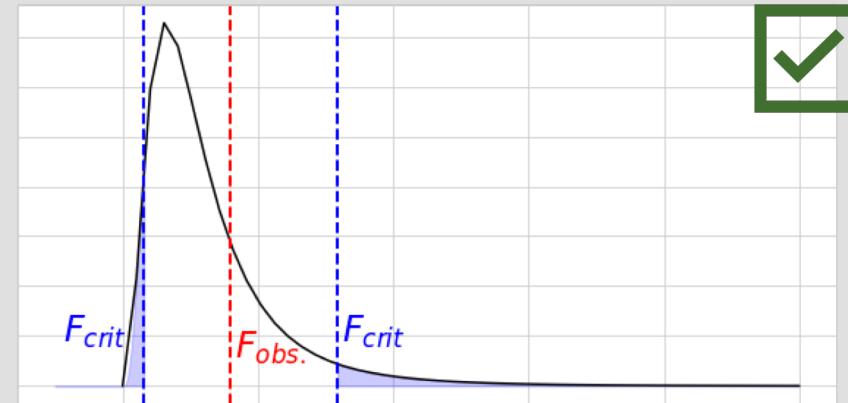
Выборки независимые

$H_0: \sigma_x = \sigma_y$

$H_a: \sigma_x \neq \sigma_y$

Критерий для проверки:

$$\frac{s_x^2 \cdot \sigma_y^2}{s_y^2 \cdot \sigma_x^2} = \frac{s_x^2}{s_y^2} \underset{H_0}{\sim} F_{n-1, m-1}$$



Критерий точный, используется  
предположение о нормальности выборки



## Пример (ценная бумага):

- Мистер Белфорд наблюдает за двумя бумагами. За последние десять лет выборочная дисперсия доходности для первой бумаги составила 0.05. Для второй бумаги 0.08.
- Есть ли основания полагать, что на уровне значимости 5% вложение во вторую бумагу более рискованно.



# Пример (ценная бумага):

$H_0$ : риск одинаковый

$H_a$ : вторая бумага рискованнее



$H_0: \sigma_1^2 = \sigma_2^2$

$H_a: \sigma_1^2 < \sigma_2^2$

**Предположение:**

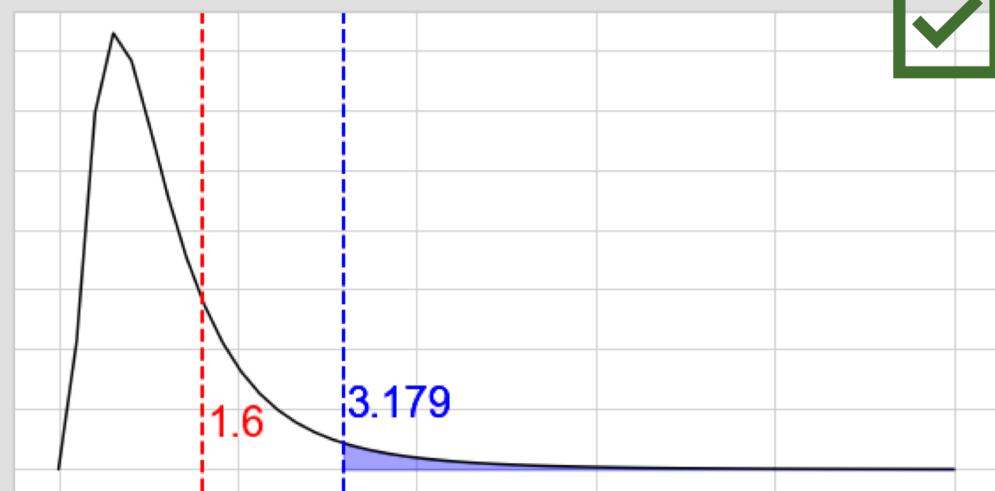
$$X_1, \dots, X_n \sim iid N(\mu_x, \sigma_x^2)$$

$$Y_1, \dots, Y_m \sim iid N(\mu_y, \sigma_y^2)$$

Выборки независимые

$$F_{obs} = \frac{0.08}{0.05}$$

$$F_{9,9}(0.95) = 3.18$$



Гипотеза о том, что бумаги  
обладают одинаковой  
рискованностью  
**не отвергается**



# Резюме

- Для нормальных выборок для проверки гипотез о дисперсиях можно использовать критерий Хи-квадрат
- Для случая двух нормальных выборок подойдёт тест Фишера
- Есть и другие способы проверять гипотезы о дисперсиях



С помощью ровно этих же статистик мы до этого строили для дисперсий доверительные интервалы

