

# **Прикладной статистический анализ данных**

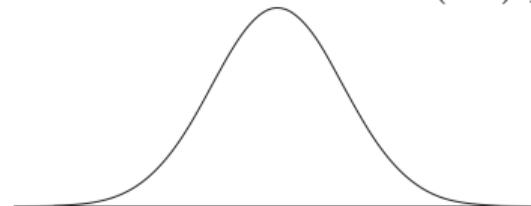
## **Проверка параметрических гипотез**

Олег Бахтеев  
psad@phystech.edu

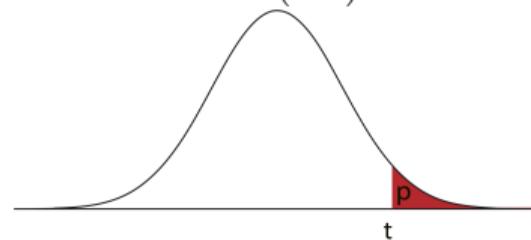
2022

## Проверка гипотез

выборка:  $X^n = (X_1, \dots, X_n)$ ,  $X \sim P \in \Omega$   
нулевая гипотеза:  $H_0: P \in \omega$ ,  $\omega \in \Omega$   
альтернатива:  $H_1: P \notin \omega$   
статистика:  $T(X^n)$ ,  $T(X^n) \sim F(x)$  при  $P \in \omega$   
 $T(X^n) \not\sim F(x)$  при  $P \notin \omega$



реализация выборки:  $x^n = (x_1, \dots, x_n)$   
реализация статистики:  $t = T(x^n)$   
достигаемый уровень значимости:  $p(x^n)$  — вероятность при  $H_0$  получить  
 $T(X^n) = t$  или ещё более экстремальное



$$p(x^n) = P(T \geq t | H_0)$$

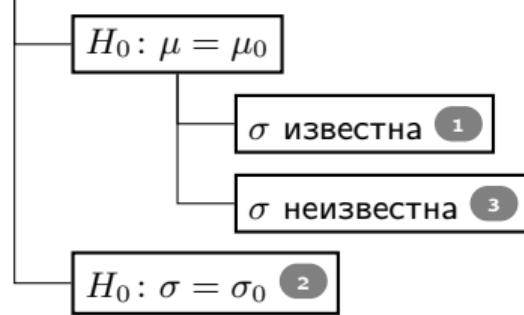
Гипотеза отвергается при  $p(x^n) \leq \alpha$ ,  $\alpha$  — уровень значимости

## Проверка гипотез



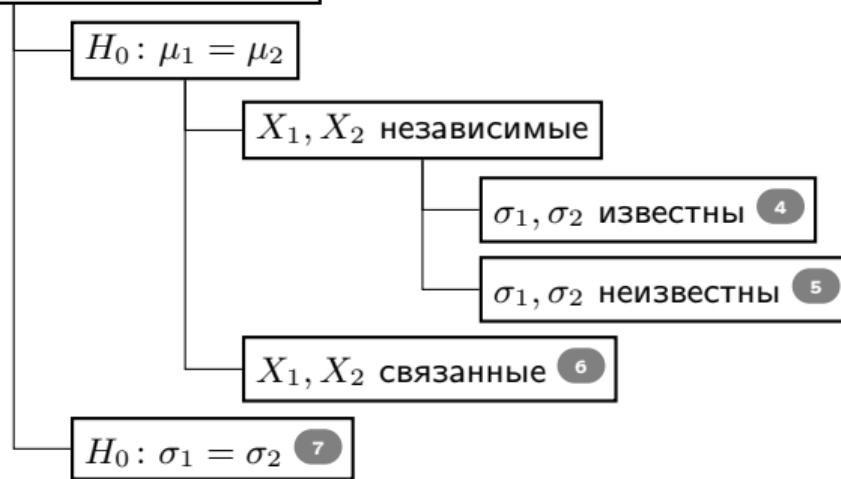
## Виды задач: одновыборочные

$$X^n \sim N(\mu, \sigma^2)$$



## Виды задач: двухвыборочные

$$X_1^{n_1} \sim N(\mu_1, \sigma_1^2), \quad X_2^{n_2} \sim N(\mu_2, \sigma_2^2)$$



# 1 Z-критерий

выборка:  $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$

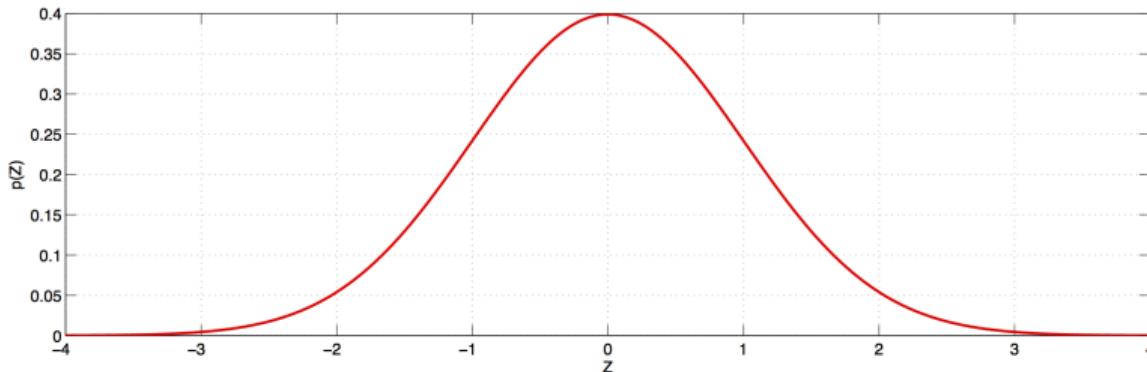
$\sigma$  известна

нулевая гипотеза:  $H_0: \mu = \mu_0$

альтернатива:  $H_1: \mu < \neq > \mu_0$

статистика:  $Z(X^n) = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

нулевое распределение:  $N(0, 1)$



достигаемый уровень значимости:

$$p(Z) = \begin{cases} 1 - F_{N(0,1)}(Z), & H_1: \theta > \theta_0, \\ F_{N(0,1)}(Z), & H_1: \theta < \theta_0, \\ 2(1 - F_{N(0,1)}(|Z|)), & H_1: \theta \neq \theta_0. \end{cases}$$

# 1 Z-критерий

выборка:  $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$

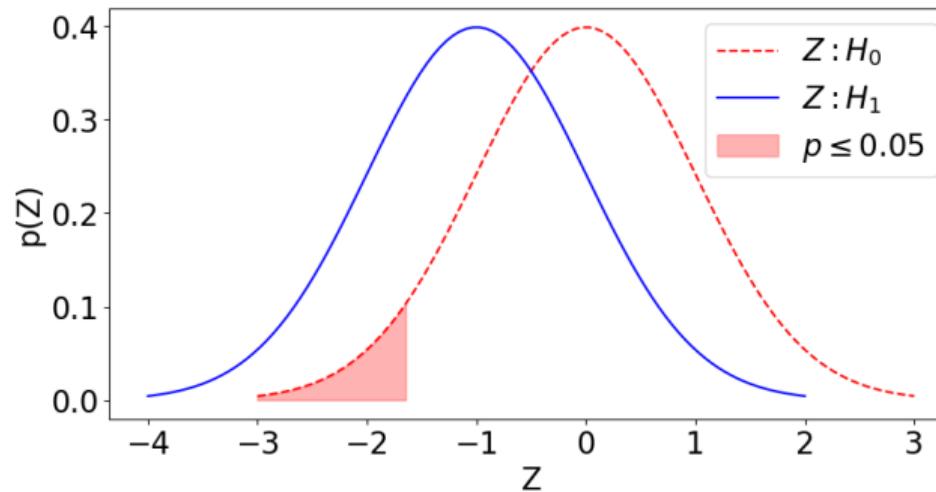
$\sigma$  известна

нулевая гипотеза:  $H_0: \mu = \mu_0$

альтернатива:  $H_1: \mu < \mu_0$

статистика:  $Z(X^n) = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

нулевое распределение:  $N(0, 1)$



Достигаемый уровень значимости:

$$p(Z) = F_{N(0,1)}(Z).$$

# 1 Z-критерий

выборка:  $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$

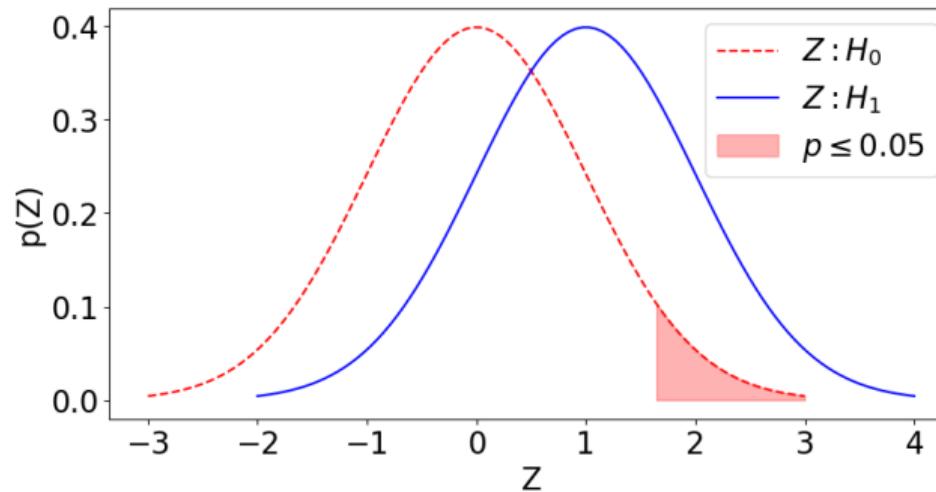
$\sigma$  известна

нулевая гипотеза:  $H_0: \mu = \mu_0$

альтернатива:  $H_1: \mu > \mu_0$

статистика:  $Z(X^n) = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

нулевое распределение:  $N(0, 1)$



Достигаемый уровень значимости:

$$p(Z) = 1 - F_{N(0,1)}(Z).$$

# 1 Z-критерий

## Пример, Kanji, критерий 1

Линия по производству пудры должна обеспечивать средний вес пудры в упаковке 4 грамма, заявленное стандартное отклонение — 1 грамм.

В ходе инспекции выбрано 9 упаковок, средний вес продукта в них составляет 4.6 грамма.

$H_0$ : средний вес пудры в упаковке соответствует норме.

$H_1$ : средний вес пудры в упаковке не соответствует норме  $\Rightarrow p = 0.0719$ , 95% доверительный интервал для среднего веса — [3.95, 5.25] г.

$H_1$ : средний вес пудры в упаковке превышает норму  $\Rightarrow p = 0.0359$ , нижний 95% доверительный предел для среднего веса — 4.05 г.

**Одностороннюю альтернативу можно использовать, если знак изменения среднего известен заранее.**

## 2 Критерий хи-квадрат

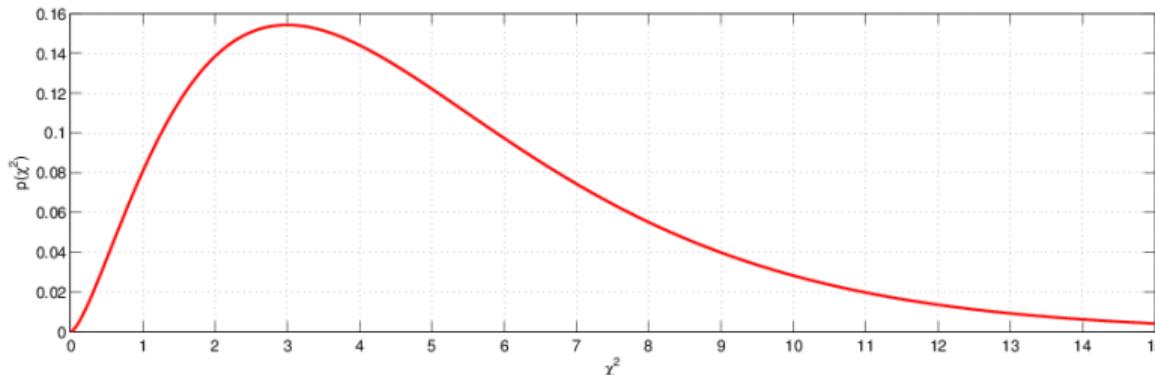
выборка:  $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$

нулевая гипотеза:  $H_0: \sigma = \sigma_0$

альтернатива:  $H_1: \sigma < \neq > \sigma_0$

статистика:  $\chi^2(X^n) = \frac{(n-1)S^2}{\sigma_0^2}$

нулевое распределение:  $\chi_{n-1}^2$



достигаемый уровень значимости:

$$p(\chi^2) = \begin{cases} 1 - F_{\chi_{n-1}^2}(\chi^2), & H_1: \sigma > \sigma_0, \\ F_{\chi_{n-1}^2}(\chi^2), & H_1: \sigma < \sigma_0, \\ 2 \min \left( 1 - F_{\chi_{n-1}^2}(\chi^2), F_{\chi_{n-1}^2}(\chi^2) \right), & H_1: \sigma \neq \sigma_0. \end{cases}$$

## 2 Критерий хи-квадрат

### Пример, Kanji, критерий 15

При производстве микрогидравлической системы делается инъекция жидкости. Дисперсия объёма жидкости — критически важный параметр, установленный стандартом на уровне 9 кв. мл. В выборке из 25 микрогидравлических систем выборочная дисперсия объёма жидкости составляет 12 кв. мл.

$H_0$ : дисперсия объёма жидкости соответствует стандарту.

$H_1$ : дисперсия объёма жидкости не соответствует стандарту  $\Rightarrow p = 0.254$ , 95% доверительный интервал для дисперсии — [7.3, 23.2] кв. мл.

$H_1$ : дисперсия объёма жидкости превышает допустимое значение  $\Rightarrow p = 0.127$ , односторонний нижний 95% доверительный предел — 7.9 кв. мл.

### 3 t-критерий Стьюдента

выборка:  $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$

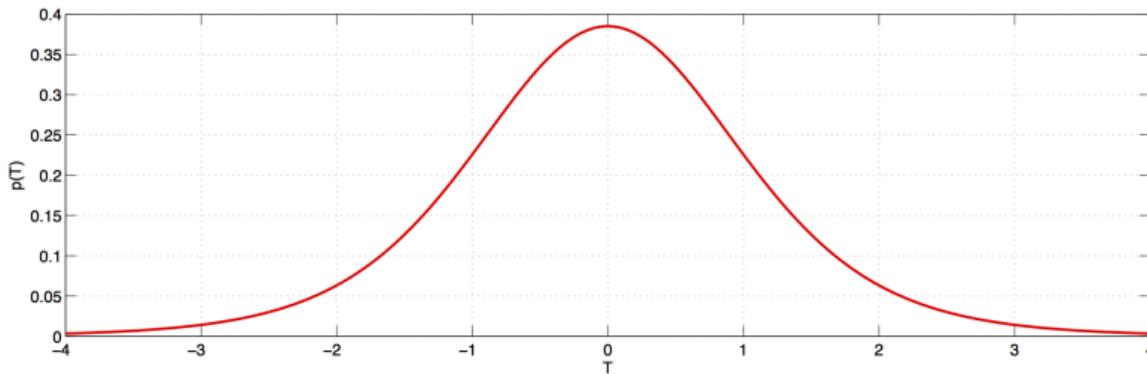
$\sigma$  неизвестна

нулевая гипотеза:  $H_0: \mu = \mu_0$

альтернатива:  $H_1: \mu < \neq > \mu_0$

статистика:  $T(X^n) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

нулевое распределение:  $St(n - 1)$



С ростом объема выборки разница между t- и z-критериями уменьшается.

### 3 t-критерий Стьюдента

#### Пример

Средний вес детей при рождении составляет 3300 г. В то же время, если мать ребёнка живёт за чертой бедности, то средний вес таких детей — 2800 г.

С целью увеличить вес тех детей, чьи матери живут за чертой бедности, разработана экспериментальная программа ведения беременности. Чтобы проверить ее эффективность, проводится эксперимент. В нём принимают участие 25 женщин, живущих за чертой бедности. У всех них рождаются дети, и их средний вес составляет 3075 г, выборочное стандартное отклонение — 500 г.

Эффективна ли программа?

$H_0$ : программа не влияет на вес детей,  $\mu = 2800$

$H_1$ : программа как-то влияет на вес детей,  $\mu \neq 2800 \Rightarrow p = 0.0111$ , 95% доверительный интервал для изменения веса — [68.6, 481.4] г.

$H_1$ : программа увеличивает вес детей,  $\mu > 2800 \Rightarrow p = 0.0056$ , 95% нижний доверительный предел для увеличения веса — 103.9 г.

## Выбор альтернативы

Одностороннюю альтернативу можно использовать, если знак изменения среднего известен заранее.

**Альтернатива должна выбираться до получения данных!**

## Z-критерий

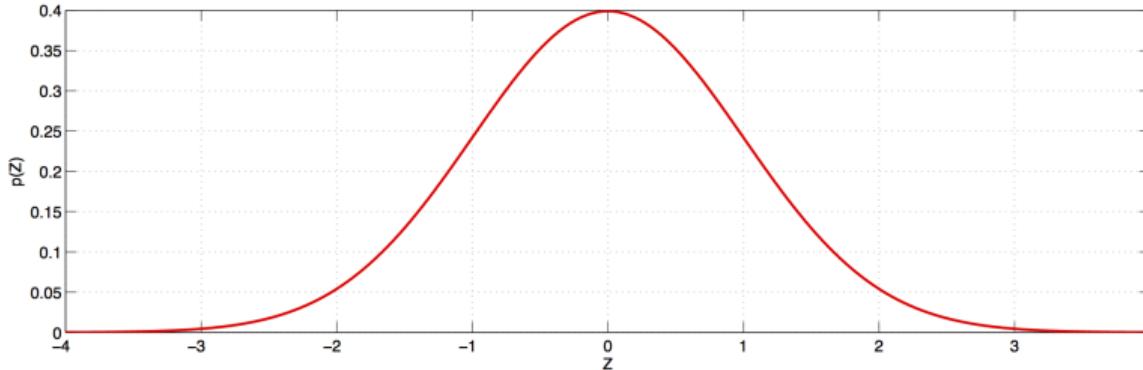
выборки:  $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2)$   
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2)$   
 $\sigma_1, \sigma_2$  известны

нулевая гипотеза:  $H_0: \mu_1 = \mu_2$

альтернатива:  $H_1: \mu_1 < \neq > \mu_2$

статистика:  $Z(X_1^{n_1}, X_2^{n_2}) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

нулевое распределение:  $N(0, 1)$



### Пример, Kanji, критерий З

известно, что одна из линий по расфасовке чипсов даёт упаковки с более вариабельным весом продукта, чем вторая. Дисперсии равны  $0.000576 \text{ г}^2$  и  $0.001089 \text{ г}^2$  соответственно, средние значения веса в выборках из 13 и 8 элементов — 80.02 г и 79.98 г.

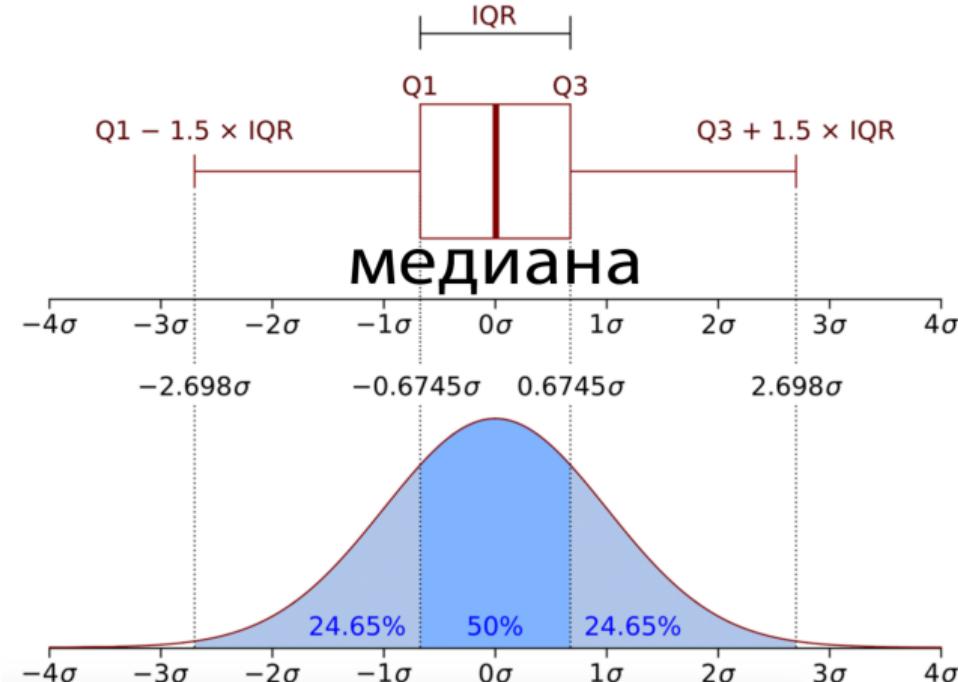
$H_0$ : средний вес продукта в упаковках, произведённых на двух линиях, совпадает.

$H_1$ : средние веса продукта в упаковках, произведённых на двух линиях, различаются

$\Rightarrow p = 0.001$ , 95% доверительный интервал для разности — [0.039, 0.041].

## Боксплот

Ящик с усами — способ визуализации основных характеристик распределения:



Длина усов отличается в разных реализациях.

## 5 t-критерий Стьюдента / Аспина-Уэлша (проблема Беренца-Фишера)

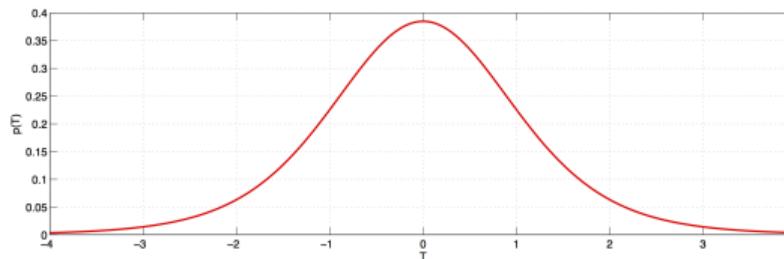
выборки:  $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2)$   
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2)$   
 $\sigma_1, \sigma_2$  неизвестны

нулевая гипотеза:  $H_0: \mu_1 = \mu_2$

альтернатива:  $H_1: \mu_1 < \neq > \mu_2$

статистика:  $T(X_1^{n_1}, X_2^{n_2}) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$   
 $\nu = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}$

нулевое распределение:  $\approx St(\nu)$

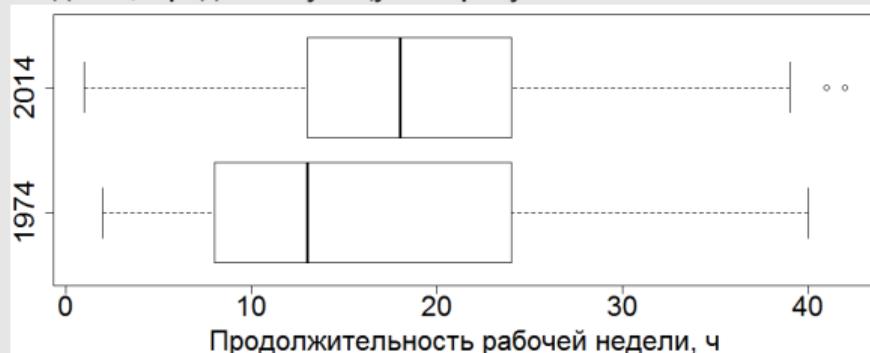


Приближение достаточно точно при  $n_1 = n_2$  или  $[n_1 > n_2] = [\sigma_1 > \sigma_2]$ .

## 5 t-критерий Стьюдента / Аспина-Уэлша (проблема Беренца-Фишера)

### Пример

В 1974 году 108 респондентов GSS работали неполный день, в 2014 — 196. Для каждого из них известно количество рабочих часов за неделю, предшествующую опросу.



Изменилось ли среднее время работы у работающих неполный день?

$H_0$ : среднее время работы не изменилось,  $\mu_1 = \mu_2$ .

$H_0$ : среднее время работы изменилось,  $\mu_1 \neq \mu_2$ .

t-критерий:  $p = 0.02707$ , средняя продолжительность рабочей недели увеличилась на 2.57 часов (95% доверительный интервал — [0.29, 4.85] ч).

## t-критерий Стьюдента для связанных выборок

выборки:  $X_1^n = (X_{11}, \dots, X_{1n}), X_1 \sim N(\mu_1, \sigma_1^2)$   
 $X_2^n = (X_{21}, \dots, X_{2n}), X_2 \sim N(\mu_2, \sigma_2^2)$

выборки связанные

нулевая гипотеза:  $H_0: \mu_1 = \mu_2$

альтернатива:  $H_1: \mu_1 < \neq > \mu_2$

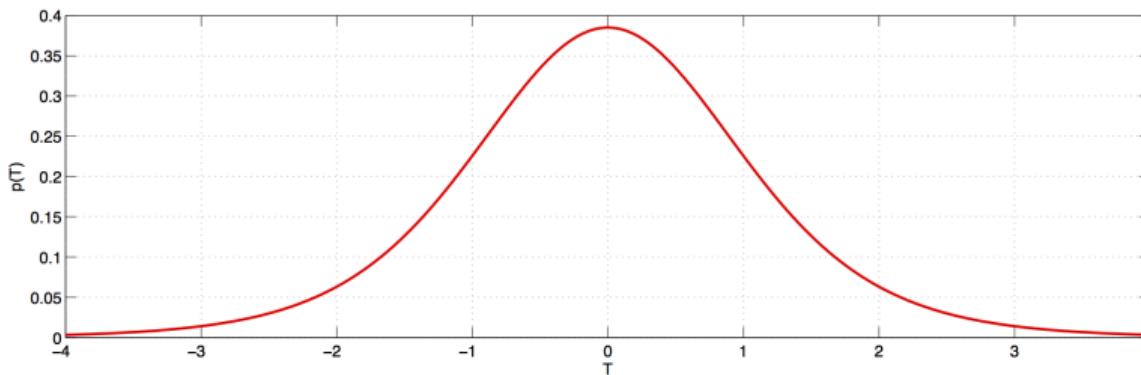
статистика:  $T(X_1^n, X_2^n) = \frac{\bar{X}_1 - \bar{X}_2}{S/\sqrt{n}}$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2}$$

$$D_i = X_{1i} - X_{2i}, \bar{D} = \frac{1}{n} \sum_i D_i$$

$$St(n-1)$$

нулевое распределение:



## t-критерий Стьюдента для связанных выборок

### Пример, Kanji, критерий 10

На 10 испытуемых сравниваются два лекарства против респираторного заболевания. Каждый из испытуемых вдыхает первое лекарство с помощью ингалятора, после чего проходит упражнение беговой дорожке. Измеряется время достижения максимальной нагрузки. Затем после периода восстановления эксперимент повторяется со вторым лекарством.

$H_0$ : время достижения максимальной нагрузки не отличается для исследуемых лекарств.

$H_1$ : время достижения максимальной нагрузки для исследуемых лекарств отличается  
 $\Rightarrow p = 0.916$ ; 95% доверительный интервал для разницы —  $[-2.1, 0.9]$ .

## Пример

Пусть имеются следующие связные выборки:

$$X_1^n, \quad X_1 \sim N(0, 1),$$

$$X_2^n, \quad X_2 = X_1 + \varepsilon, \quad \varepsilon \sim N(0.1, 0.25) \Rightarrow X_2 \sim N(0.1, 1.25);$$

требуется оценить разность  $\Delta = \mathbb{E}X_1 - \mathbb{E}X_2$ .

- Если **попарные соответствия элементов известны**, лучшая оценка  $\hat{\Delta}_p = \frac{1}{n} \sum_{i=1}^n (X_{1i} - X_{2i})$  имеет дисперсию

$$\mathbb{D}\hat{\Delta}_p = \frac{1}{n^2} \sum_{i=1}^n \mathbb{D}(X_{1i} - X_{2i}) = \frac{1}{n} \mathbb{D}\varepsilon = \frac{1}{2n};$$

мощность 0.8 достигается при  $n \approx 200$ .

- Если же **попарные соответствия неизвестны**, лучшая оценка —  $\hat{\Delta}_i = \bar{X}_1 - \bar{X}_2$ ; её дисперсия:

$$\mathbb{D}\hat{\Delta}_i = \frac{1}{n^2} \sum_{i=1}^n \mathbb{D}X_{1i} + \frac{1}{n^2} \sum_{i=1}^n \mathbb{D}X_{2i} = \frac{1}{n} + \frac{5}{4n} = \frac{9}{4n}$$

— в 4.5 раза больше; мощность 0.8 достигается при  $n \approx 1900$ .

## F-критерий Фишера

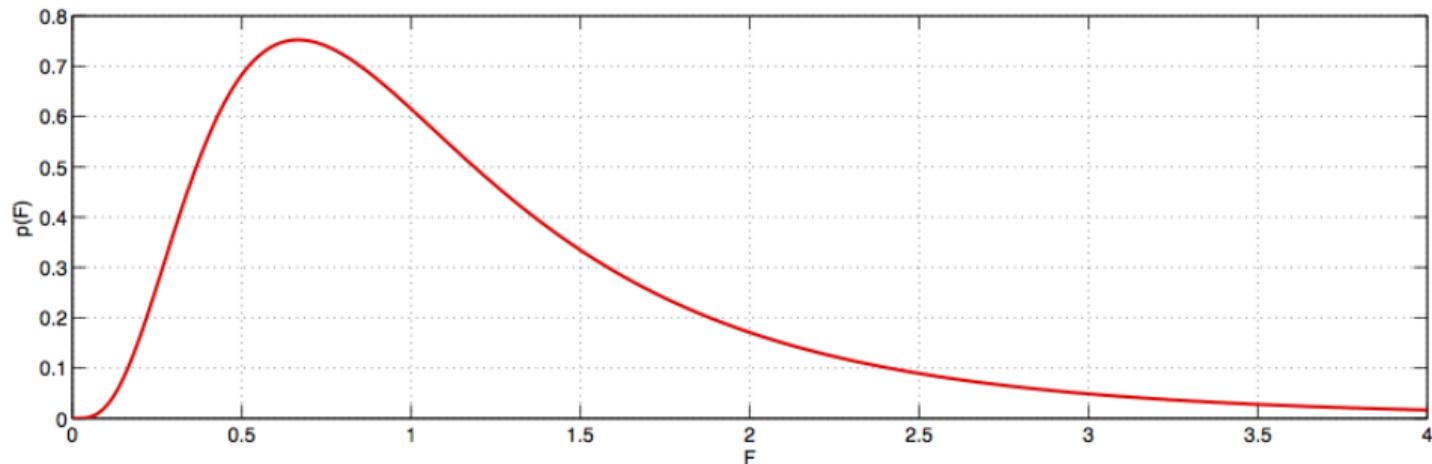
выборки:  $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2)$   
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2)$

нулевая гипотеза:  $H_0: \sigma_1 = \sigma_2$

альтернатива:  $H_1: \sigma_1 < \neq > \sigma_2$

статистика:  $F(X_1^{n_1}, X_2^{n_2}) = \frac{S_1^2}{S_2^2}$

нулевое распределение:  $F(n_1 - 1, n_2 - 1)$

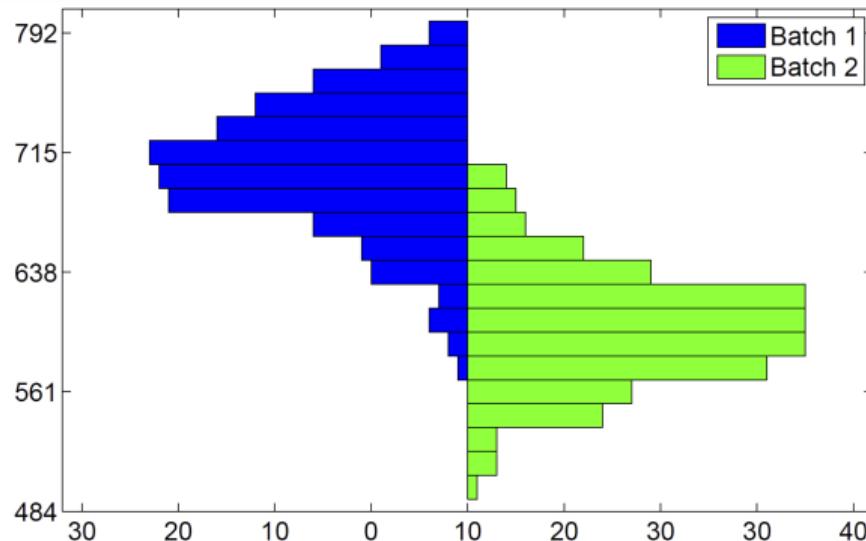


Критерий Фишера неустойчив к отклонениям от нормальности даже асимптотически.

## 7 F-критерий Фишера

Пример, NIST/industry ceramics consortium for strength optimization of ceramic, 1996

Собраны данные о прочности материала 440 керамических изделий из двух партий по 220 в каждой.  
Однакова ли дисперсия прочности в разных партиях?



Критерий Фишера:  $p = 0.1721$ ,  $[C_L, C_U] = [0.9225, 1.5690]$ .

## Коэффициент ассиметрии

- коэффициент ассиметрии (skewness):

$$\gamma_1 = \mathbb{E} \left( \frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}} \right)^3$$



Negative Skew

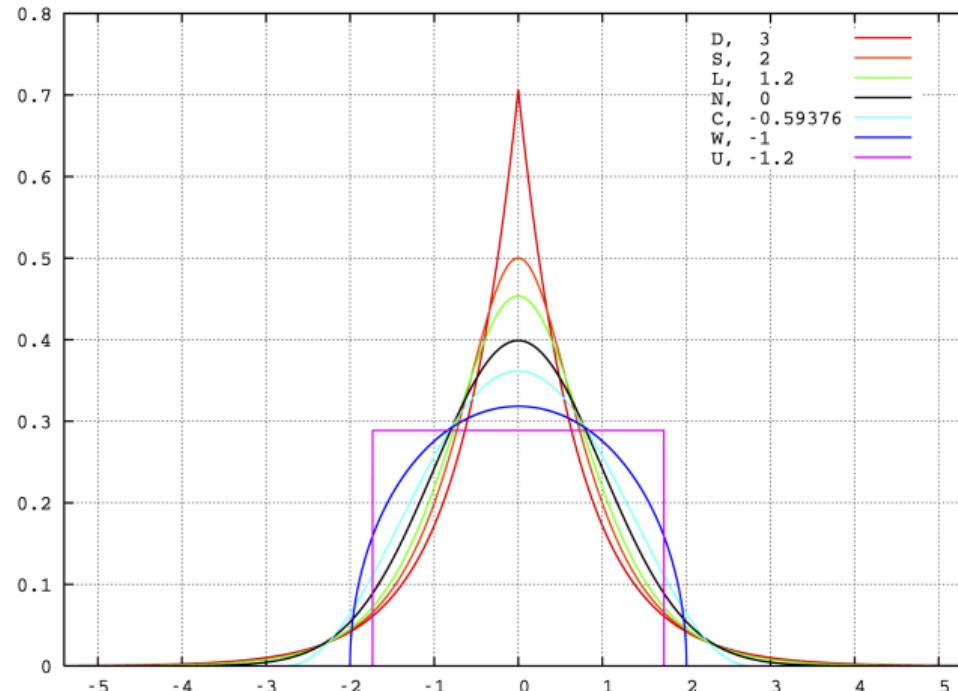


Positive Skew

## Коэффициент эксцесса

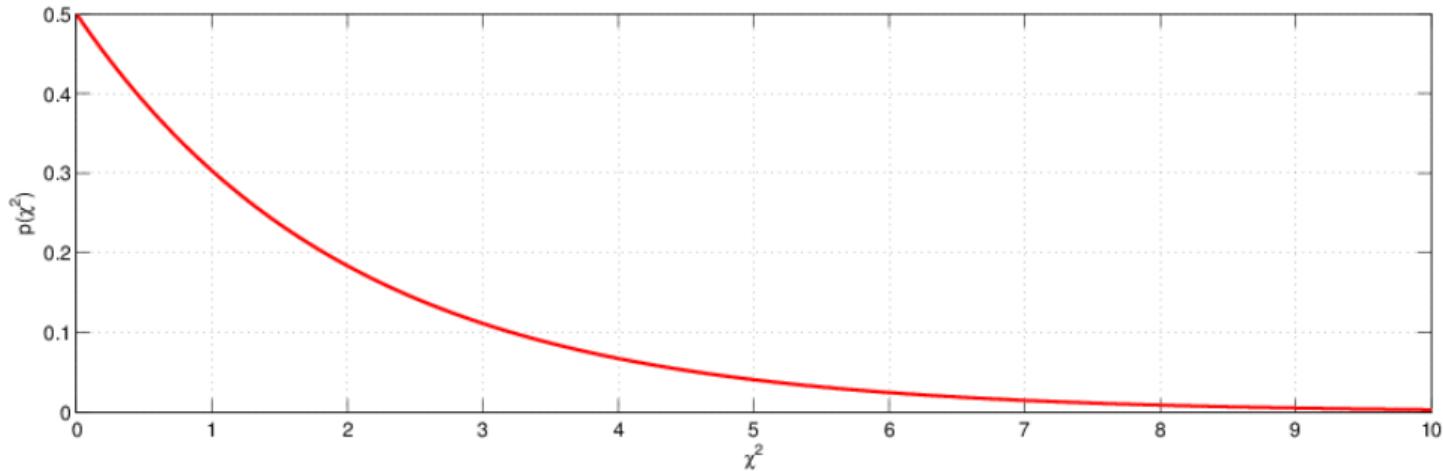
- коэффициент эксцесса (excess, без вычитания тройки — kurtosis):

$$\gamma_2 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{(\mathbb{D}X)^2} - 3$$



# Критерий Харке-Бера

- выборка:  $X^n = (X_1, \dots, X_n)$   
нулевая гипотеза:  $H_0: X \sim N(\mu, \sigma^2)$   
альтернатива:  $H_1: H_0$  неверна  
статистика:  $\chi^2(X^n) = \frac{n}{6} (\gamma_1^2 + \frac{1}{4}\gamma_2^2)$   
нулевое распределение:  $\chi_2^2$



достигаемый уровень значимости:

$$p(\chi^2) = 1 - F_{\chi_2^2}(\chi^2)$$

# Критерий согласия Пирсона (хи-квадрат)

выборка:  $X^n = (X_1, \dots, X_n)$

нулевая гипотеза:  $H_0: X \sim N(\mu, \sigma^2)$

альтернатива:  $H_1: H_0$  неверна

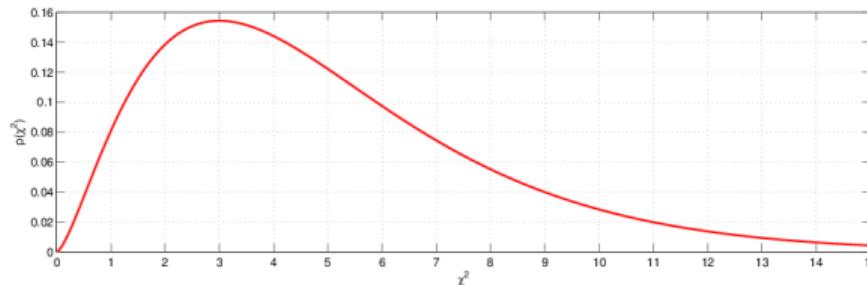
статистика:  $\chi^2(X^n) = \sum_{i=1}^K \frac{(n_i - np_i)^2}{np_i}$

$[a_i, a_{i+1}], i = 1, \dots, K$  — интервалы гистограммы

$n_i$  — число элементов выборки в  $[a_i, a_{i+1}]$

$p_i = F(a_{i+1}) - F(a_i)$  — вероятность попадания  
в  $i$ -й интервал при  $H_0$

нулевое распределение:

$$\begin{cases} \chi^2_{K-1}, & \mu, \sigma \text{ заданы,} \\ \chi^2_{K-3}, & \mu, \sigma \text{ оцениваются,} \end{cases}$$


## Недостатки:

- разбиение на интервалы неоднозначно
- требует больших выборок ( $np_i > 5$  в 80% ячеек)

# Критерии, основанные на эмпирической функции распределения

Ряд критериев согласия основаны на различиях между  $F(x)$  и  $F_n(x)$ :

- Джини:

$$\int |F_n(x) - F(x)| dx$$

- Крамера-фон Мизеса:

$$\int (F_n(x) - F(x))^2 dx$$

- Колмогорова (одновыборочный Колмогорова-Смирнова):

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)|$$

- Смирнова-Крамера-фон Мизеса:

$$\int (F_n(x) - F(x))^2 dF(x)$$

## Критерии, основанные на эмпирической функции распределения

- Андерсона-Дарлинга:

$$\int \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x)$$

- Купера:

$$\sup_{-\infty < x < \infty} (F_n(x) - F(x)) + \sup_{-\infty < x < \infty} (F(x) - F_n(x))$$

- Ватсона:

$$\int \left( F_n(x) - F(x) - \int (F_n(x) - F(x)) dF(x) \right) dF(x)$$

- Фроцини:

$$\int |F_n(x) - F(x)| dF(x)$$

Предполагается, что  $F(x)$  известна с точностью до параметров (если они оцениваются по выборке, нулевое распределение корректируется).

## Критерий Колмогорова (Лиллиефорса)

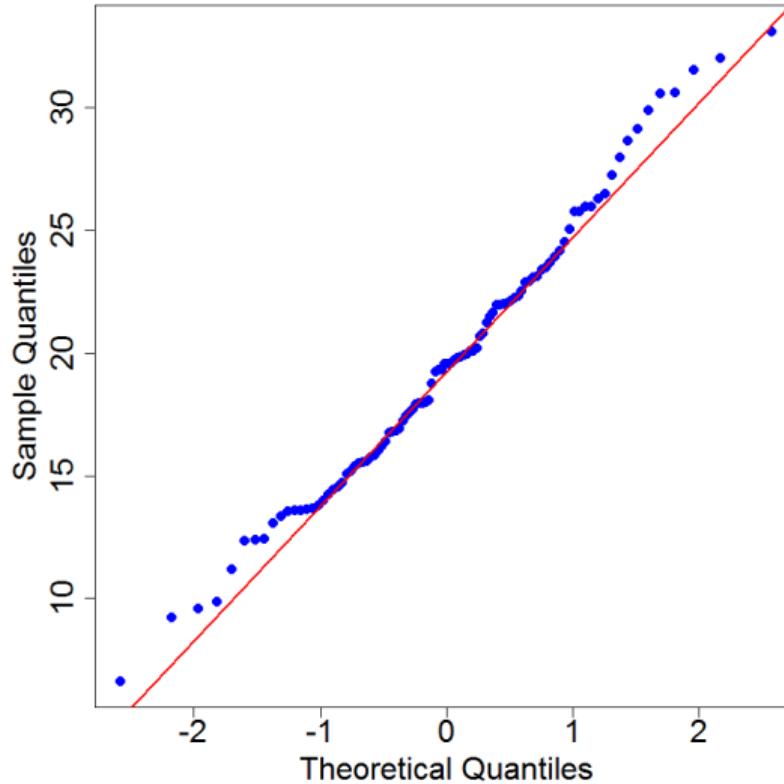
- выборка:  $X^n = (X_1, \dots, X_n)$   
нулевая гипотеза:  $H_0: X \sim N(\mu, \sigma^2)$   
альтернатива:  $H_1: H_0$  неверна  
статистика:  $D(X^n) = \sup_{-\infty < x < \infty} |F_n(x) - \Phi(x)|$   
нулевое распределение: табличное

Недостатки:

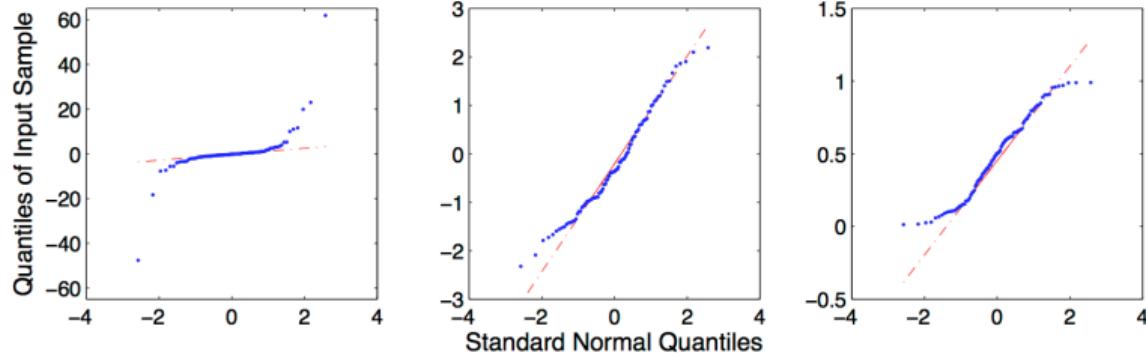
- имеет низкую мощность
- не чувствителен к различиям на хвостах распределений

## Q-Q plot

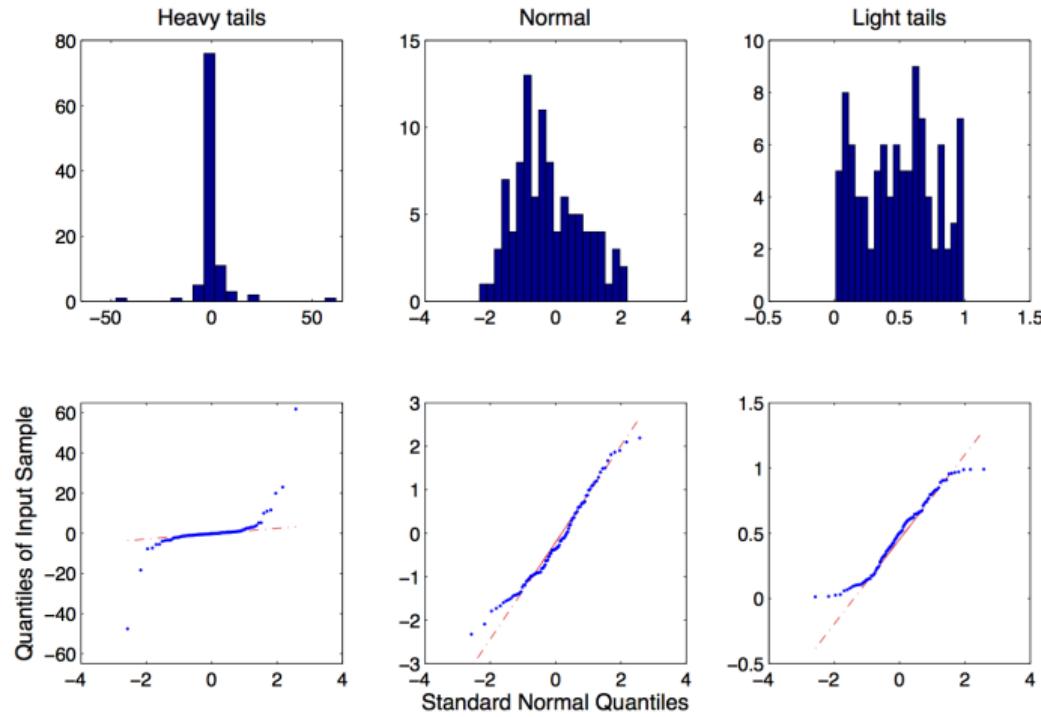
Визуальный метод проверки согласия выборки и распределения — q-q plot (для нормального распределения называется также normal probability plot)



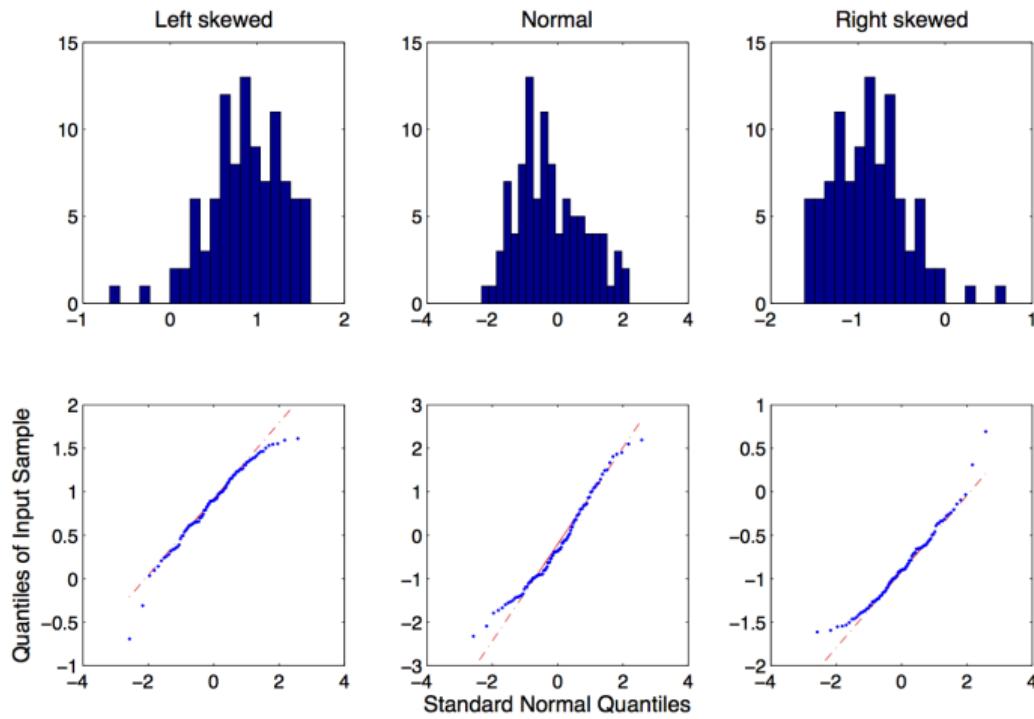
# Q-Q plot



# Q-Q plot



# Q-Q plot



# Критерий Шапиро-Уилка

выборка:  $X^n = (X_1, \dots, X_n)$

нулевая гипотеза:  $H_0: X \sim N(\mu, \sigma^2)$

альтернатива:  $H_1: H_0$  неверна

статистика:

$$W(X^n) = \frac{\left( \sum_{i=1}^n a_i X_{(i)} \right)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

$m = (m_1, \dots, m_n)^T$  — матожидания порядковых  
статистик  $N(0, 1)$ ,  $V$  — их ковариационная матрица  
табличное

нулевое распределение:

Значения  $a_i$  также табулированы.

## Итого о проверке нормальности

- **выбросы:** сильно влияют на выборочные коэффициенты асимметрии и эксцесса
- **критерий Лиллиефорса:** представляет только исторический интерес
- **критерий хи-квадрат:** слишком общий, не самый мощный, потеря информации из-за разбиения на интервалы

# Итого о проверке нормальности

Сравнение критериев проверки  
нормальности распределения случайных величин

Наименование критерия (раздел)	Характер альтернативного распределения					Ранг
	асимметричное		симметричное		$\approx$ нормальное	
	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 \approx 3$	
Критерий Шапиро–Уилка (3.2.2.1)	1	1	3	2	2	1
Критерий $K^2$ (3.2.2.16)	7	8	10	6	4	2
Критерий Дарбина (3.1.2.7)	11	7	7	15	1	3
Критерий Д'Агостино (3.2.2.14)	12	9	4	5	12	4
Критерий $\alpha_4$ (3.2.2.16)	14	5	2	4	18	5
Критерий Васичека (3.2.2.2)	2	14	8	10	10	6
Критерий Дэвида–Хартли–Пирсона (3.2.2.10)	21	2	1	9	1	7
Критерий $\chi^2$ (3.1.1.1)	9	20	9	8	3	8
Критерий Андерсона–Дарлинга (3.1.2.4)	18	3	5	18	7	9
Критерий Филлибена (3.2.2.5)	3	12	18	1	9	10
Критерий Колмогорова–Смирнова (3.1.2.1)	16	10	6	16	5	11
Критерий Мартинеса–Иглевича (3.2.2.14)	10	16	13	3	15	12
Критерий Лина–Мудхолкара (3.2.2.13)	4	15	12	12	16	13
Критерий $\alpha_3$ (3.2.2.16)	8	6	21	7	19	14
Критерий Шпигельхальтера (3.2.2.11)	19	13	11	11	8	15
Критерий Саркади (3.2.2.12)	5	18	15	14	13	16
Критерий Смирнова–Крамера–фон Мизеса (3.1.2.2)	17	11	20	17	6	17
Критерий Локка–Спурье (3.2.2.7)	13	4	19	21	17	18
Критерий Оя (3.2.2.8)	20	17	14	13	14	19
Критерий Хегази–Грина (3.2.2.3)	6	19	16	19	21	20
Критерий Муроты–Такеучи (3.2.2.17)	15	21	17	20	20	21

Кобзарь, 3.2.2.19, табл. 80.

## Итого о проверке нормальности

- **очень маленькие выборки:** любой критерий может пропустить отклонения от нормальности, графические методы бесполезны
- **очень большие выборки:** любой критерий может выявлять небольшие статистически, но не практически значимые отклонения от нормальности; значительная часть методов, предполагающих нормальность, демонстрируют устойчивость к отклонениям от неё

## Итого о проверке нормальности

- если *данные явно ненормальны* (например, бинарны или дискретны), нужно выбрать метод, специфичный для такого распределения
- если *на ку-ку графике не видно существенных отклонений от нормальности*, можно сразу использовать методы, устойчивые к небольшим отклонениям (например, критерии Стьюдента)
- если метод *чувствителен к отклонениям от нормальности* (например, критерий Фишера), проверять её рекомендуется критерием Шапиро-Уилка
- если *нормальность отвергается*, чувствительные методы, предполагающие нормальность, использовать нельзя!

## Критерии на основе правдоподобия

$$X^n = (X_1, \dots, X_n), \quad X \sim f(x, \theta).$$

ОМП для  $\theta$ :

$$\log L(X^n, \theta) = \sum_{i=1}^n \log f(X_i, \theta),$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \log L(X^n, \theta),$$

$$I(\theta) = -\frac{\partial^2}{\partial \theta^2} \log L(\theta),$$

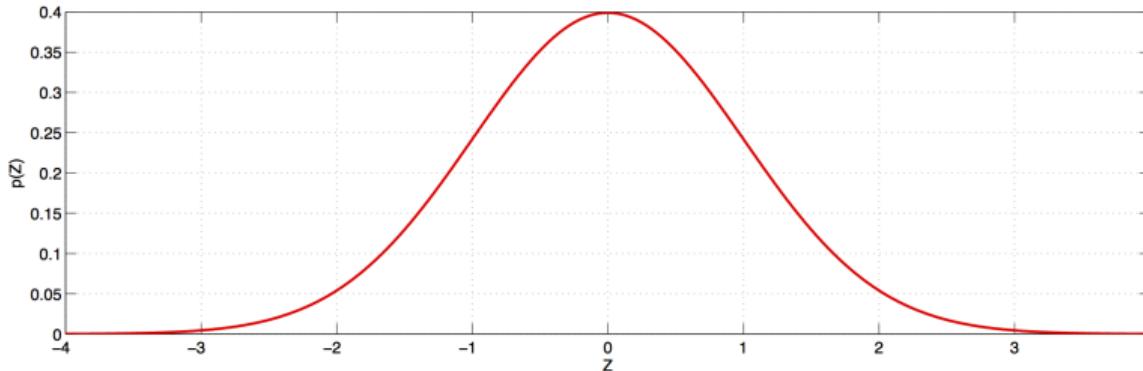
$$\mathbb{D}\hat{\theta}_{MLE} = I^{-1}(\hat{\theta}_{MLE}),$$

$$S(\theta) = \frac{\partial}{\partial \theta} \log L(\theta).$$

$\hat{\theta}_{MLE}$  и  $S(\hat{\theta}_{MLE})$  асимптотически нормально распределены.

## Критерий Вальда

- выборка:  $X^n = (X_1, \dots, X_n), X \sim F(x, \theta)$
- нулевая гипотеза:  $H_0: \theta = \theta_0$
- альтернатива:  $H_1: \theta < \neq > \theta_0$
- статистика:  $Z_W(X^n) = \frac{\hat{\theta}_{MLE} - \theta_0}{\sqrt{\text{Д}\hat{\theta}_{MLE}}}$
- нулевое распределение:  $Z_W(X^n) \sim N(0, 1)$  при  $H_0$



достигаемый уровень значимости:

$$p(Z_W) = \begin{cases} 1 - F_{N(0,1)}(Z_W), & H_1: \theta > \theta_0, \\ F_{N(0,1)}(Z_W), & H_1: \theta < \theta_0, \\ 2(1 - F_{N(0,1)}(|Z_W|)), & H_1: \theta \neq \theta_0. \end{cases}$$

## Критерий отношения правдоподобия

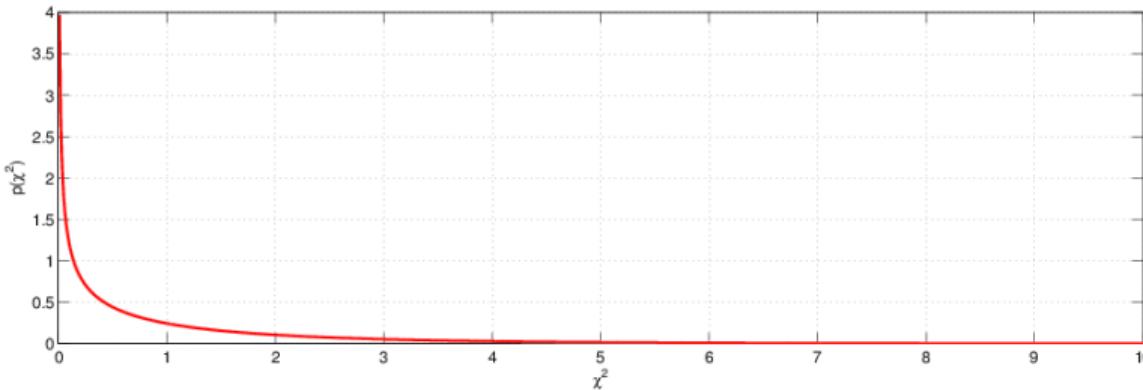
выборка:  $X^n = (X_1, \dots, X_n), X \sim F(x, \theta)$

нулевая гипотеза:  $H_0: \theta = \theta_0$

альтернатива:  $H_1: \theta \neq \theta_0$

статистика:  $LR(X^n) = -2 \log \frac{L(X^n, \theta_0)}{L(X^n, \hat{\theta}_{MLE})}$

нулевое распределение:  $LR(X^n) \sim \chi^2_1$  при  $H_0$



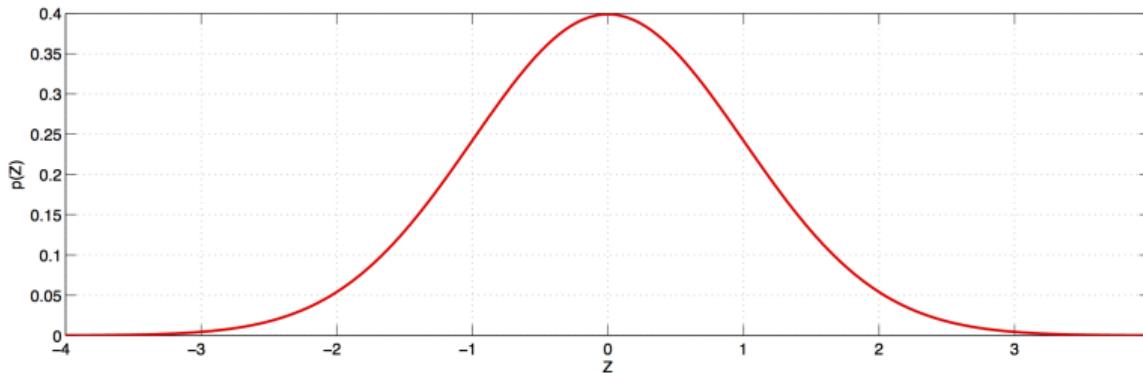
достигаемый уровень значимости:

$$p(LR) = 1 - F_{\chi^2_1}(LR).$$

Если  $\theta$  — вектор размерности  $k$ , то нулевое распределение критерия —  $\chi_k^2$ .

## Критерий меток

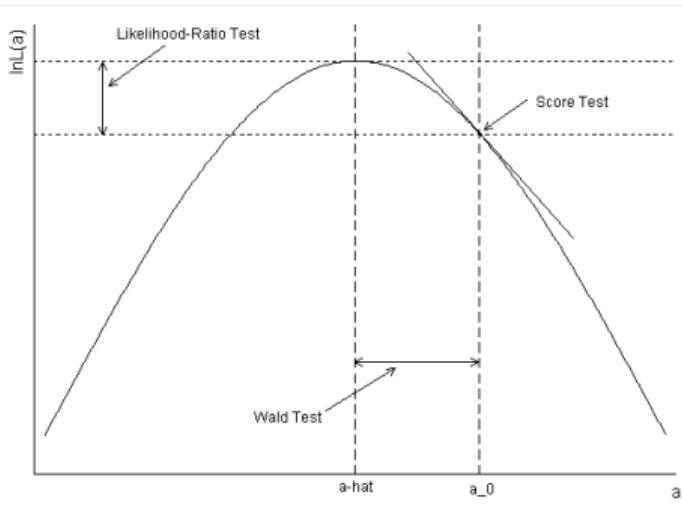
- выборка:  $X^n = (X_1, \dots, X_n), X \sim F(x, \theta)$   
нулевая гипотеза:  $H_0: \theta = \theta_0$   
альтернатива:  $H_1: \theta < \neq > \theta_0$   
статистика:  $Z_S(X^n) = \frac{S(\theta_0)}{\sqrt{I(\theta_0)}}$   
нулевое распределение:  $N(0, 1)$



достигаемый уровень значимости:

$$p(Z_S) = \begin{cases} 1 - F_{N(0,1)}(Z_S), & H_1: \theta > \theta_0, \\ F_{N(0,1)}(Z_S), & H_1: \theta < \theta_0, \\ 2(1 - F_{N(0,1)}(|Z_S|)), & H_1: \theta \neq \theta_0. \end{cases}$$

# Три критерия



- критерий Вальда использует информацию о правдоподобии только в  $\hat{\theta}_{MLE}$ , критерий меток — только в  $\theta_0$ , критерий отношения правдоподобия — в обеих точках
- все три критерия асимптотические; на конечных выборках хуже всего критерий Вальда

## Распределение Бернуlli

$$X^n = (X_1, \dots, X_n), \quad X \sim Ber(p), \quad T = \sum_{i=1}^n X_i.$$

ОМП для  $p$ :

$$L(p) = p^T (1-p)^{(n-T)},$$

$$\log L(p) = T \ln p + (n - T) \ln (1 - p),$$

$$\hat{p} = \frac{T}{n},$$

$$I(p) = -\frac{\partial^2 \log L(p)}{\partial p^2} = \frac{n}{p(1-p)},$$

$$\mathbb{D}\hat{p} = \frac{\hat{p}(1-\hat{p})}{n},$$

$$S(p) = \frac{T}{p} - \frac{n-T}{1-p}$$

## Распределение Бернулли

$$LR = -2 \log \frac{L(p_0)}{L(\hat{p})} \sim \chi^2_1$$

$$Z_W = \frac{\hat{p} - p_0}{\sqrt{1/I(\hat{p})}} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0, 1)$$

$$Z_S = \frac{S(p_0)}{\sqrt{I(p_0)}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$$

## Z-критерий меток для доли

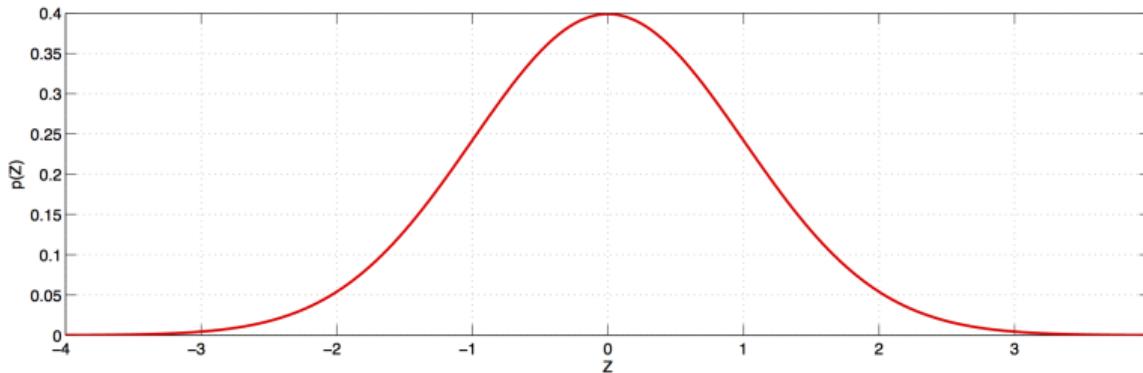
выборка:  $X^n = (X_1, \dots, X_n), X \sim Ber(p)$

нулевая гипотеза:  $H_0: p = p_0$

альтернатива:  $H_1: p < \neq > p_0$

статистика:  $Z_S(X^n) = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

нулевое распределение:  $N(0, 1)$



## Биномиальный критерий

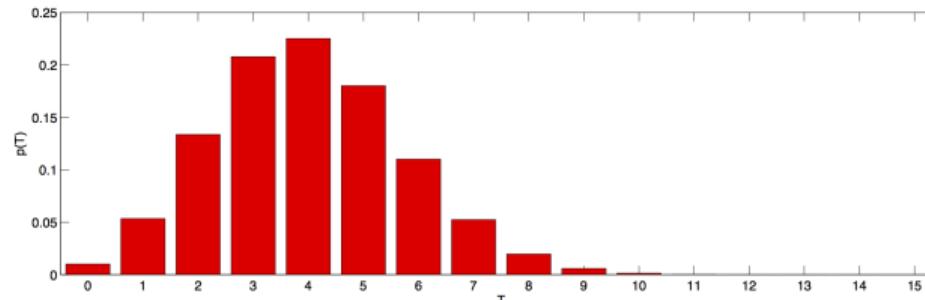
выборка:  $X^n = (X_1, \dots, X_n), X \sim Ber(p)$

нулевая гипотеза:  $H_0: p = p_0$

альтернатива:  $H_1: p < \neq > p_0$

статистика:  $T(X^n) = \sum_{i=1}^n X_i$

нулевое распределение:  $Bin(n, p_0)$



достигаемый уровень значимости:

$$p(T) = \begin{cases} 1 - F_{Bin(n, p_0)}(T - 1), & H_1: p > p_0, \\ F_{Bin(n, p_0)}(T), & H_1: p < p_0, \\ \text{через бета-распределение,} & H_1: p \neq p_0. \end{cases}$$

Поскольку нулевое распределение дискретно, нельзя добиться, чтобы вероятность ошибки первого рода была равна в точности  $\alpha$ .

## Примеры

### Королёв, задача 7.2.2

Бенджамин Спок, знаменитый педиатр и автор большого количества книг по воспитанию детей, был арестован за участие в антивоенной демонстрации в Бостоне. Его дело должен был рассматривать суд присяжных. Присяжные назначаются с помощью многоступенчатой процедуры, на очередном этапе которой было отобрано 300 человек. Однако среди них оказалось только 90 женщин. Адвокаты доктора Спока подали протест на предвзятость отбора.

$H_0$ : процедура отбора была беспристрастной, женщины попадали в выборку с вероятностью  $1/2$ .

$H_1$ : предпочтение отдавалось кандидатам-мужчинам.

Критерий	$p$
Z-критерий меток	$2.3 \times 10^{-12}$
Z-критерий Вальда	$2.1 \times 10^{-12}$
Биномиальный	$1.6 \times 10^{-12}$

## Примеры

### Кобзарь, задача 227

Нормируемый уровень дефектных изделий в партии  $p_0 = 0.05$ . Среди 20 изделий партии проверка обнаружила 2 дефектных.

$H_0$ : доля дефектных изделий в партии не выше нормы.

$H_1$ : доля дефектных изделий в партии выше нормы.

**Обратите внимание:** если  $H_0: p = p_0$  проверяется против  $H_1: p > p_0$ , ничего не изменится от замены нулевой гипотезы на  $H_0: p \leq p_0$ .

Критерий	$p$
Z-критерий меток	0.1524
Z-критерий Вальда	0.2280
Биномиальный	0.2642

## Примеры

Нормируемый уровень дефектных изделий в партии  $p_0 = 0.05$ . Среди 24 изделий партии проверка обнаружила 4 дефектных.

$H_0$ : доля дефектных изделий в партии не выше нормы.

$H_1$ : доля дефектных изделий в партии выше нормы.

**Обратите внимание:** если  $H_0: p = p_0$  проверяется против  $H_1: p > p_0$ , ничего не изменится от замены нулевой гипотезы на  $H_0: p \leq p_0$ .

Критерий	$p$
Z-критерий меток	0.0067
Z-критерий Вальда	0.0625
Биномиальный	0.0297

## Доверительные интервалы для доли

Доверительный интервал Вальда основывается на аппроксимации нормальным распределением:

$$F_X(x) \approx \Phi\left(\frac{x - Np}{\sqrt{Np(1-p)}}\right).$$

100  $(1 - \alpha)\%$  доверительный интервал Вальда:

Метод	Пример 1	Пример 2
Вальда	[0.2481, 0.3519]	[-0.0315, 0.2315]

**Недостатки:**

- доверительные пределы могут выходить за границы  $[0, 1]$  (вообще, при  $\hat{p} \in (0, 1)$  нежелательно даже  $C_L = 0$  и  $C_U = 1$ )
- при  $\hat{p} = 0$  и  $\hat{p} = 1$  вырождается в точку
- антиконсервативен — накрывает  $p$  реже, чем в  $100(1 - \alpha)\%$  случаев

## Доверительные интервалы для доли

Более точный доверительный интервал Уилсона (основан на критерии меток):

$$\frac{\hat{p} + z_{1-\frac{\alpha}{2}}/2}{n + z_{1-\frac{\alpha}{2}}^2} \pm \frac{z_{1-\frac{\alpha}{2}}\sqrt{n}}{n + z_{1-\frac{\alpha}{2}}^2} \sqrt{\hat{p}(1-\hat{p}) + \frac{z_{1-\frac{\alpha}{2}}^2}{4n}}$$

Метод	Пример 1	Пример 2
Вальда	[0.2481, 0.3519]	[-0.0315, 0.2315]
Уилсона	[0.2509, 0.3541]	[0.0279, 0.3010]

## Доверительные интервалы для доли

Доверительный интервал Клоппера-Пирсона (основан на биномиальном критерии) определяется квантилями бета-распределения.

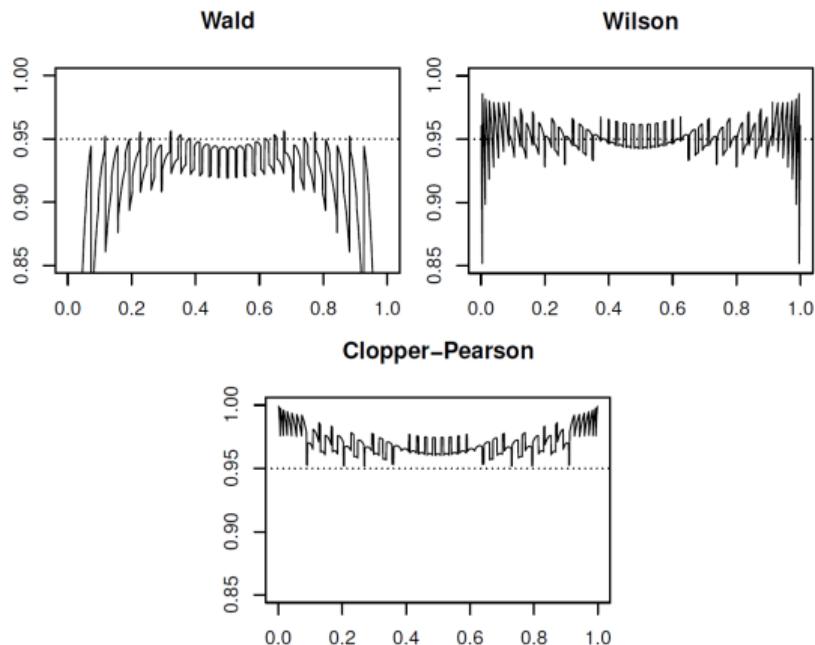
Метод	Пример 1	Пример 2
Вальда	[0.2481, 0.3519]	[-0.0315, 0.2315]
Уилсона	[0.2509, 0.3541]	[0.0279, 0.3010]
Клоппера-Пирсона	[0.2486, 0.3553]	[0.0123, 0.3170]

Особенности:

- всегда точен — уровень доверия никогда не ниже номинального
- почти всегда консервативен — уровень доверия часто выше номинального

# Доверительные интервалы для доли

Эксперименты при  $n = 40$ :



## Z-критерий для разности двух долей, независимые выборки

выборки:  $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim Ber(p_1)$   
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim Ber(p_2)$   
выборки независимы

Исход	Выборка	
	$X_1^{n_1}$	$X_2^{n_2}$
1	$a$	$b$
0	$c$	$d$
$\sum$	$n_1$	$n_2$

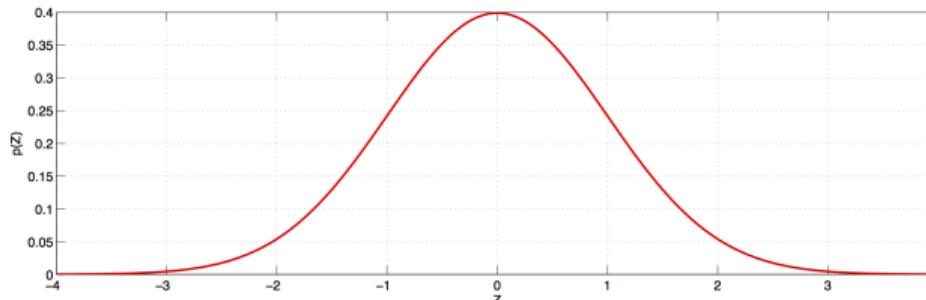
нулевая гипотеза:  $H_0: p_1 = p_2$

альтернатива:  $H_1: p_1 \neq p_2$

статистика:  $Z(X_1^{n_1}, X_2^{n_2}) = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$P = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}, \quad \hat{p}_1 = \frac{a}{n_1}, \quad \hat{p}_2 = \frac{b}{n_2}$$

нулевое распределение:  $N(0, 1)$



## Z-критерий для разности двух долей, независимые выборки

### Кобзарь, задача 226

В двух партиях объёмами  $n_1 = 100$  шт. и  $n_2 = 200$  шт. обнаружено соответственно  $t_1 = 3$  и  $t_2 = 5$  дефектных приборов. Необходимо проверить гипотезу о равенстве долей дефектных приборов в партиях.

Наличие дефекта	Номер партии	
	1	2
Есть	$a = 3$	$b = 5$
Нет	$c = 97$	$d = 195$
Всего	$n_1 = 100$	$n_2 = 200$

$H_0$ : доли дефектных изделий в партиях равны.

$H_1$ : доли дефектных изделий в партиях различаются  $\Rightarrow p = 0.8$ .

$H_1$ : доля дефектных изделий в первой партии выше  $\Rightarrow p = 0.4$ .

$H_1$ : доля дефектных изделий в первой партии ниже  $\Rightarrow p = 0.6$ .

## Доверительный интервал для разности двух долей

Доверительный интервал Уилсона:

$$[C_L, C_U] = [\hat{p}_1 - \hat{p}_2 - \delta, \hat{p}_1 - \hat{p}_2 + \varepsilon],$$

$$\delta = z_{1-\frac{\alpha}{2}} \sqrt{\frac{l_1(1-l_1)}{n_1} + \frac{u_2(1-u_2)}{n_2}},$$

$$\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{u_1(1-u_1)}{n_1} + \frac{l_2(1-l_2)}{n_2}},$$

$l_1, u_1$  — корни уравнения  $|x - \hat{p}_1| = z_{1-\frac{\alpha}{2}} \sqrt{\frac{x(1-x)}{n_1}}$ ,

$l_2, u_2$  — корни уравнения  $|x - \hat{p}_2| = z_{1-\frac{\alpha}{2}} \sqrt{\frac{x(1-x)}{n_2}}$ .

В примере 95% доверительный интервал —  $[-0.0331, 0.0616]$ , минимальное значение  $\alpha$ , при котором интервал не содержит нуля — 0.8003.

## Z-критерий для разности двух долей, связанные выборки

выборки:  $X_1^n = (X_{11}, \dots, X_{1n}), X_1 \sim Ber(p_1)$   
 $X_2^n = (X_{21}, \dots, X_{2n}), X_2 \sim Ber(p_2)$   
выборки связанные

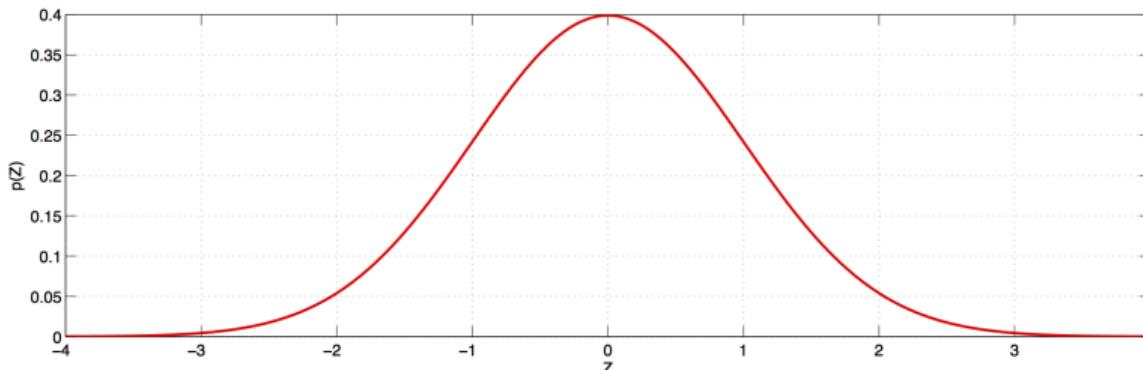
$X_1^n$	$X_2^n$	1	0
1		$e$	$f$
0		$g$	$h$

нулевая гипотеза:  $H_0: p_1 = p_2$

альтернатива:  $H_1: p_1 \neq p_2$

статистика:  $Z(X_1^n, X_2^n) = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{f+g}{n^2} - \frac{(f-g)^2}{n^3}}} = \frac{f-g}{\sqrt{f+g - \frac{(f-g)^2}{n}}}$

нулевое распределение:  $N(0, 1)$



## Z-критерий для разности двух долей, связанные выборки

### Пример, Agresti, табл. 10.1

Из опрошенных 1600 граждан Великобритании, имеющих право голоса, 944 высказали одобрение деятельности премьер-министра. Через 6 месяцев эти же люди были опрошены снова, на этот раз одобрение высказали только 880 опрошенных.

		II	+	-	$\Sigma$
		I			
I	+	$e = 794$	$f = 150$	944	
	-	$g = 86$	$h = 570$	656	
$\Sigma$		880	720	1600	

$H_0$ : рейтинг премьер-министра не изменился.

$H_1$ : рейтинг премьер-министра изменился  $\Rightarrow p = 2.8 \times 10^{-5}$ .

$H_1$ : рейтинг премьер-министра снизился  $\Rightarrow p = 1.4 \times 10^{-5}$ .

$H_1$ : рейтинг премьер-министра повысился  $\Rightarrow p = 0.99999$ .

## Z-критерий для разности двух долей, связанные выборки

Без учёта информации о связи между выборками:

Результат	Опрос		
		I	II
+		$a = 944$	$b = 880$
-		$c = 656$	$d = 720$
$\sum$		$n_1 = 1600$	$n_2 = 1600$

$H_0$ : рейтинг премьер-министра не изменился.

$H_1$ : рейтинг премьер-министра изменился  $\Rightarrow p = 0.0222$ .

$H_1$ : рейтинг премьер-министра снизился  $\Rightarrow p = 0.0112$ .

$H_1$ : рейтинг премьер-министра повысился  $\Rightarrow p = 0.9889$ .

## Доверительный интервал для разности двух долей

Доверительный интервал Уилсона:

$$[C_L, C_U] = [\hat{p}_1 - \hat{p}_2 - \delta, \hat{p}_1 - \hat{p}_2 + \varepsilon],$$

$$\delta = \sqrt{dl_1^2 - 2\hat{\phi}dl_1du_2 + du_2^2},$$

$$\varepsilon = \sqrt{du_1^2 - 2\hat{\phi}du_1dl_2 + dl_2^2},$$

$$\hat{\phi} = \begin{cases} \frac{eh-fg}{(e+f)(g+h)(e+h)(f+h)}, & \text{если знаменатель не равен нулю,} \\ 0, & \text{иначе;} \end{cases}$$

$$dl_1 = \hat{p}_1 - l_1,$$

$$du_1 = u_1 - \hat{p}_1,$$

$$dl_2 = \hat{p}_2 - l_2,$$

$$du_2 = u_2 - \hat{p}_2,$$

$l_1, u_1$  — корни уравнения  $|x - \hat{p}_1| = z_{1-\frac{\alpha}{2}} \sqrt{\frac{x(1-x)}{n}}$ ,

$l_2, u_2$  — корни уравнения  $|x - \hat{p}_2| = z_{1-\frac{\alpha}{2}} \sqrt{\frac{x(1-x)}{n}}$ .

В примере 95% доверительный интервал —  $[2.14, 5.90]\%$ , минимальное значение  $\alpha$ , при котором интервал не содержит нуля —  $3.1 \times 10^{-5}$ .

# Литература

Критерии нормальности:

- Харке-Бера (Jarque–Bera) — Кобзарь, 3.2.2.16
- Шапиро-Уилка (Shapiro-Wilk) — Кобзарь, 3.2.2.1
- хи-квадрат (chi-square) — Кобзарь, 3.1.1.1, 3.2.1.1
- согласия (goodness-of-fit), основанные на эмпирической функции распределения — Кобзарь, 3.1.2, 3.2.1.2

Для нормальных распределений:

- Z-критерии (Z-tests) — Kanji, №№ 1, 2, 3
- t-критерии Стьюдента (t-tests) — Kanji, №№ 7, 8, 9
- критерий хи-квадрат (chi-square test) — Kanji, №15
- критерий Фишера (F-test) — Kanji, №16

Критерии, основанные на правдоподобии: Bilder, раздел В.5

# Литература

Для распределения Бернулли:

- всё про одновыборочную задачу — Agresti, 1.3, 1.4
- Z-критерии (Z-tests) — Kanji, №№ 4, 5
- точный критерий (exact binomial test) — McDonald,  
<http://www.biostathandbook.com/exactgof.html>
- доверительные интервалы Уилсона (score confidence intervals) — Newcombe, 1998a, 1998b, 1998c

Agresti A. *Categorical Data Analysis*, 2013.

Bilder C.R., Loughin T.M. *Analysis of Categorical Data with R*, 2013.

Kanji G.K. *100 statistical tests*, 2006.

McDonald J.H. *Handbook of Biological Statistics*, 2008.

Newcombe R.G. (1998). *Two-sided confidence intervals for the single proportion: comparison of seven methods*. Statistics in Medicine, 17, 857–872.

Newcombe R.G. (1998). *Interval estimation for the difference between independent proportions: comparison of eleven methods*. Statistics in Medicine, 17, 873–890.

Newcombe R.G. (1998). *Improved confidence intervals for the difference between binomial proportions based on paired data*. Statistics in Medicine, 17, 2635–2650.

О сравнении асимптотических критериев, ссылка

# Литература

Кобзарь А.И. *Прикладная математическая статистика*, 2006.

Королёв В.Ю. *Теория вероятностей и математическая статистика*, 2008.

NIST/SEMATECH. *e-Handbook of Statistical Methods*. <http://www.itl.nist.gov/div898/handbook/>