

Прикладной статистический анализ данных.
2. Параметрическая проверка гипотез.

Рябенко Евгений
riabenko.e@gmail.com

12 сентября 2014 г.

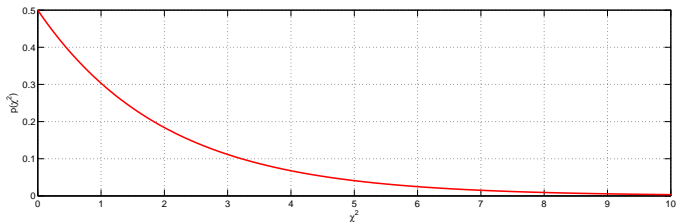
О нормальном распределении

Благодаря центральной предельной теореме и удобству вывода критериев для нормально распределённых выборок методы, основанные на предположении о нормальности данных, наиболее широко распространены.

- Если принять предположение о нормальности, то можно применять более мощные критерии. Зачастую они также чувствительны к небольшим отклонениям от нормальности.
- Перед использованием методов, предполагающих нормальность, стоит проверить нормальность.
- Если гипотеза нормальности отвергается, следует использовать непараметрические методы.

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: $\chi^2(X^n) = \frac{n}{6}(g_1^2 + \frac{1}{4}g_2^2)$;
 $\chi^2(X^n) \sim \chi_2^2$ при H_0 ;



достигаемый уровень значимости:

$$p(\chi^2) = 1 - \text{chi2cdf}(\chi^2, 2).$$

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

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

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

$W(X^n)$ при H_0 имеет табличное распределение.

Значения a_i также табулированы.

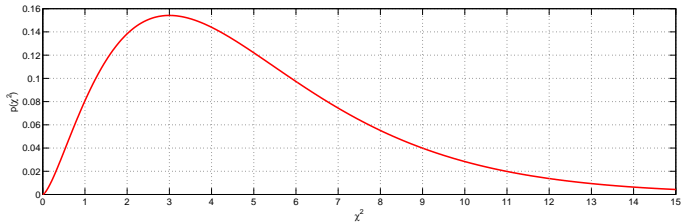
Критерий согласия Пирсона (хи-квадрат)

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

$$\chi^2(X^n) \sim \begin{cases} \chi_{K-1}^2, & \mu, \sigma \text{ заданы,} \\ \chi_{K-3}^2, & \mu, \sigma \text{ оцениваются} \end{cases} \text{ при } H_0;$$

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



Недостатки:

- требует больших выборок;
- разбиение на интервалы неоднозначно.

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).

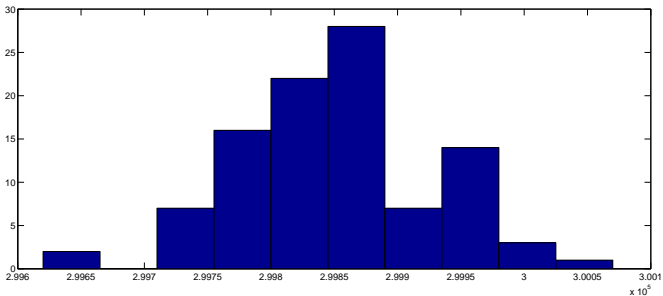
Критерий Колмогорова (Лиллиефорса)

- 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: $D(X^n) = \sup_{-\infty < x < \infty} |F_n(x) - \Phi(x)|$;
- $D(X^n)$ при H_0 имеет табличное распределение.

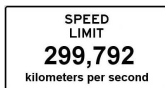
Имеет низкую мощность, не чувствителен к различиям на хвостах распределений.

Измерения скорости света

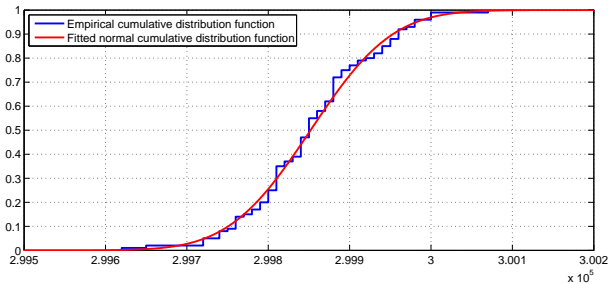
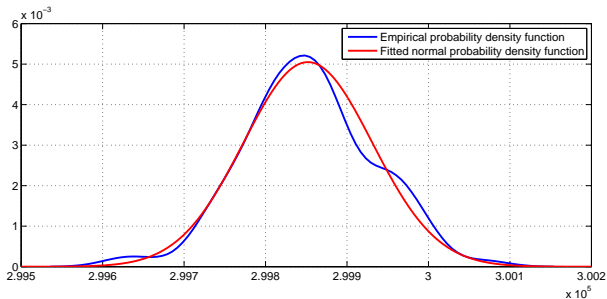
Пример (Michelson, 1880): данные классического эксперимента Майкельсона по измерению скорости света с помощью вращающегося зеркала, 100 наблюдений.



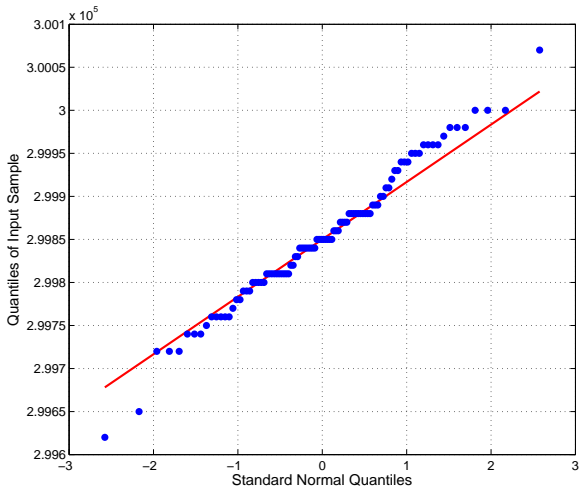
Можно ли считать измерения нормально распределёнными?



Измерения скорости света



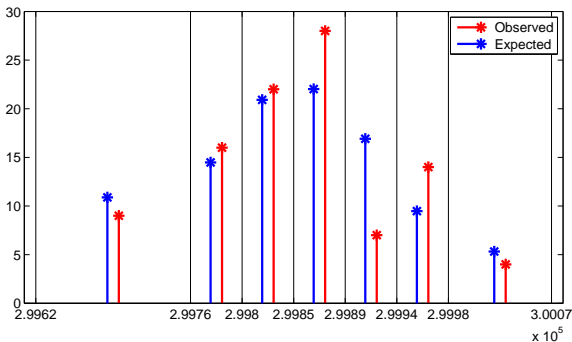
Измерения скорости света



Q-Q plot (для нормального распределения называется также normal probability plot)

Измерения скорости света

Критерий хи-квадрат: $p = 0.0333$.



Чему равно число степеней свободы?

Критерий Колмогорова (Лиллиефорса): $p = 0.0860$.

Критерий Харке-Бера: $p = 0.5$ / $p = 0.8533$.

Критерий Крамера-фон Мизеса: $p = 0.2227$.

Критерий Шапиро-Уилка: $p = 0.5138$.

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;
- **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.

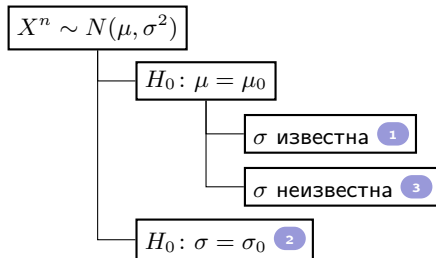
Итого о проверке нормальности

Сравнение критериев проверки нормальности распределения случайных величин

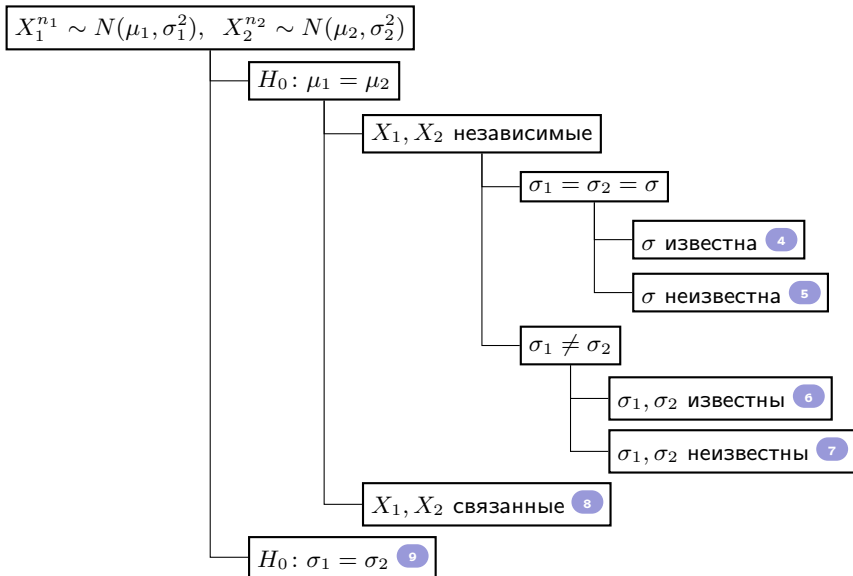
Наименование критерия (раздел)	Характер альтернативного распределения					Ранг
	асимметричное		симметричное		≈ нормальное	
	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 \approx 3$	
Критерий Шапиро-Уилка (3.2.2.1)	1	1	3	2	2	1
Критерий K^2 (3.2.2.16)	7	8	10	6	4	2
Критерий Дарбина (3.1.2.7)	11	7	7	15	1	3
Критерий Д'Агостино (3.2.2.14)	12	9	4	5	12	4
Критерий α_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 χ^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 α_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.

Виды задач: одновыборочные



Виды задач: двухвыборочные



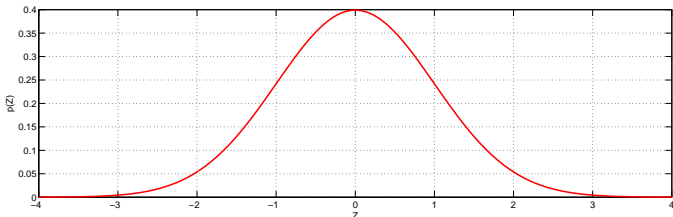
1 Z-критерий

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

нулевая гипотеза: $H_0: \mu = \mu_0$;

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

статистика: $Z(X^n) = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$;
 $Z(X^n) \sim N(0, 1)$ при H_0 ;



достигаемый уровень значимости:

$$p(z) = \begin{cases} 1 - \text{ncdf}(z, 0, 1), & H_1: \mu > \mu_0, \\ \text{ncdf}(z, 0, 1), & H_1: \mu < \mu_0, \\ 2(1 - \text{ncdf}(|z|, 0, 1)), & H_1: \mu \neq \mu_0. \end{cases}$$

1 Z-критерий

Пример (Капji, 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 $\Rightarrow p = 0.0359$, односторонний нижний 95% доверительный предел для среднего веса — 4.05 г.

2 Критерий хи-квадрат

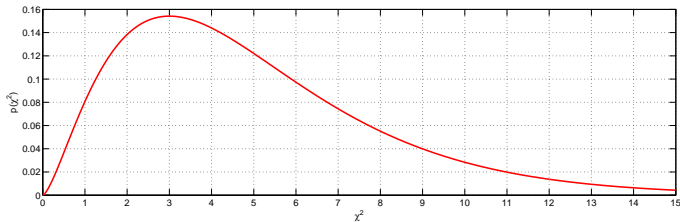
выборка: $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$;

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

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

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

$\chi^2(X^n) \sim \chi_{n-1}^2$ при H_0 ;



достигаемый уровень значимости:

$$p(\chi^2) = \begin{cases} 1 - \text{chi2cdf}(\chi^2, n - 1), & H_1: \sigma > \sigma_0, \\ \text{chi2cdf}(\chi^2, n - 1), & H_1: \sigma < \sigma_0, \\ 2 \min(1 - \text{chi2cdf}(\chi^2, n - 1), \text{chi2cdf}(\chi^2, n - 1)), & H_1: \sigma \neq \sigma_0. \end{cases}$$

2 Критерий хи-квадрат

Пример (Капji, 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 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

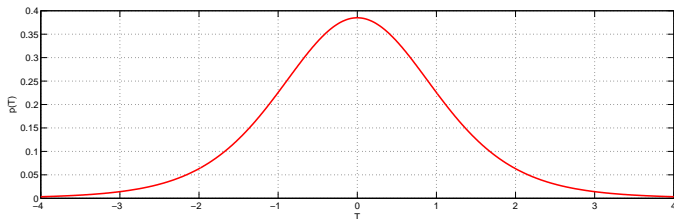
выборка: $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$,
 σ неизвестна;

нулевая гипотеза: $H_0: \mu = \mu_0$;

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

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

$T(X^n) \sim St(n-1)$ при H_0 .



3 t-критерий Стьюдента

Пример: в 1975 году с помощью лазерного интерферометра была получена оценка скорости света 299792458 м/с.

Насколько этому значению соответствуют данные эксперимента Майкельсона?

H_0 : оценки Майкельсона являются несмещёнными.

H_1 : оценки Майкельсона смещены $\Rightarrow p = 1.8 \times 10^{-11}$, 95%

доверительный интервал для смещения — [44.2, 75.6] км/с.

H_1 : оценки Майкельсона завышены $\Rightarrow p = 9.1 \times 10^{-12}$, односторонний нижний 95% доверительный предел для смещения — 46.8 км/с.

4 Z-критерий

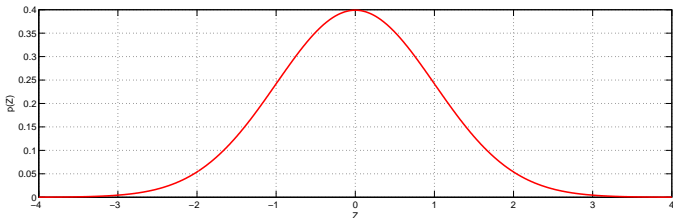
выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma^2),$
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma^2),$
 σ известна;

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

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

статистика: $Z(X_1^{n_1}, X_2^{n_2}) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}};$

$Z(X_1^{n_1}, X_2^{n_2}) \sim N(0, 1)$ при $H_0.$



4 Z-критерий

Пример (Канжі, критерий 2): два отдела сбыта сравниваются по коэффициенту результативности при выполнении схожих операций. В первом отделе на 9 операциях среднее значение коэффициента результативности составило 1.2, во втором на 16 операциях — 1.7. Дисперсии коэффициента результативности в обоих отделах равны 2.075.

H_0 : средняя результативность в обоих отделах одинакова.

H_1 : средняя результативность в двух отделах различается $\Rightarrow p = 0.405$,
95% доверительный интервал для разности — $[-1.7, 0.7]$.

5 t-критерий Стьюдента

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

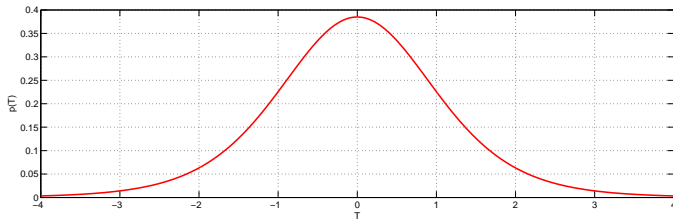
нулевая гипотеза: $H_0: \mu_1 = \mu_2;$

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

статистика: $T(X_1^{n_1}, X_2^{n_2}) = \frac{\bar{X}_1 - \bar{X}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}};$$

$T(X_1^{n_1}, X_2^{n_2}) \sim St(n_1 + n_2 - 2)$ при $H_0.$



5 t-критерий Стьюдента

Пример (Капji, критерий 8): чипсы продаются в тридцатиграммовых пакетах двух разновидностей. В выборке из 12 пачек каждого вида средние веса равны 31.75 г и 28.67 г, выборочные дисперсии — 112.25 г² и 66.64 г².

H_0 : количество чипсов в упаковках двух разновидностей совпадает.

H_1 : количество чипсов в упаковках двух разновидностей различается

$\Rightarrow p = 0.433$, 95% доверительный интервал для разности — $[-4.9, 11.1]$.

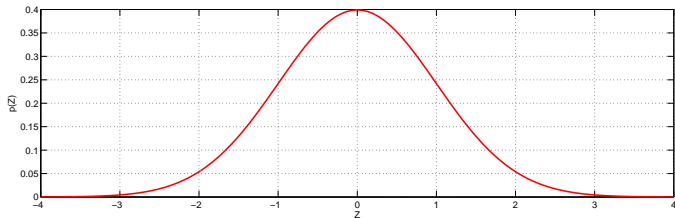
6 Z-критерий

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2),$
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2),$
 σ_1, σ_2 известны;

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

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

статистика: $Z(X_1^{n_1}, X_2^{n_2}) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}};$
 $Z(X_1^{n_1}, X_2^{n_2}) \sim N(0, 1)$ при $H_0.$



6 Z-критерий

Пример (Капји, критерий 3): известно, что одна из линий по расфасовке чипсов даёт упаковки с более вариабельным весом продукта, чем вторая. Дисперсии равны 0.000576 г^2 и 0.001089 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 — $[0.039, 0.041]$.

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

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2),$
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2),$
 σ_1, σ_2 неизвестны;

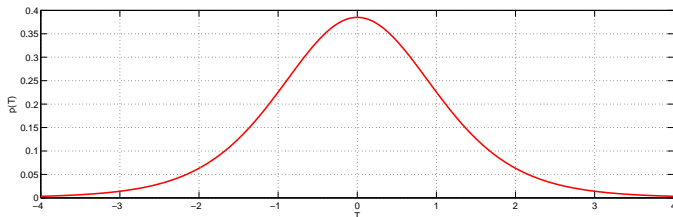
нулевая гипотеза: $H_0: \mu_1 = \mu_2;$

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

статистика: $T(X_1^{n_1}, X_2^{n_2}) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}};$$

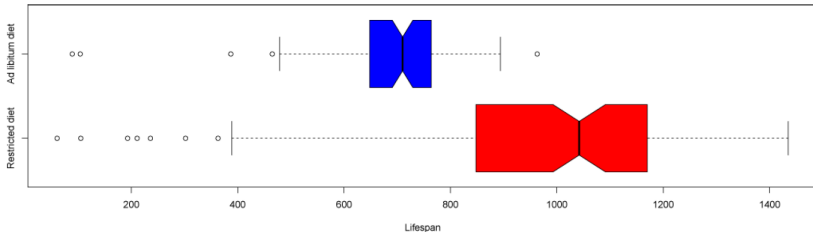
$T(X_1^{n_1}, X_2^{n_2}) \approx \sim St(\nu)$ при $H_0.$



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

Пример (Walford, Weindruch, 1988): в исследовании принимало участие 195 крыс. 106 из них держали на строгой диете, оставшиеся 89 — на диете ad libitum.

Имеющиеся данные: продолжительность жизни крыс в каждой из групп. Влияет ли диета на продолжительность жизни?



Ящик с усами (boxplot), от центра к краям:

- медиана;
- 95% доверительный интервал для медианы;
- квартили;
- точки данных, ближайшие (изнутри) к концу отрезка длиной $1.5 \times IQR$;
- точки, не попадающие в этот интервал.

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

H_0 : продолжительность жизни крыс не меняется при ограничении диеты.

H_1 : крысы на строгой диете живут дольше $\Rightarrow p = 2 \times 10^{-15}$;
односторонний нижний 95% доверительный предел для увеличения продолжительности жизни — 227.

H_1 : средняя продолжительность жизни крыс меняется при ограничении диеты $\Rightarrow p = 4 \times 10^{-15}$; 95% доверительный интервал для изменения продолжительности жизни — [217, 344].

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

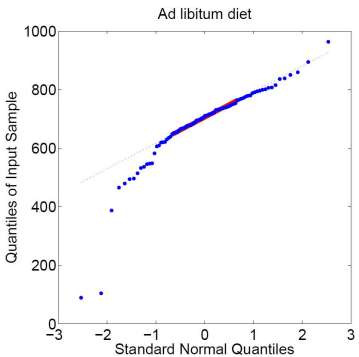
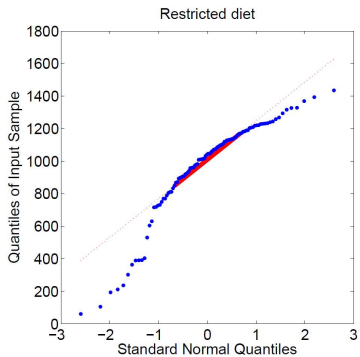
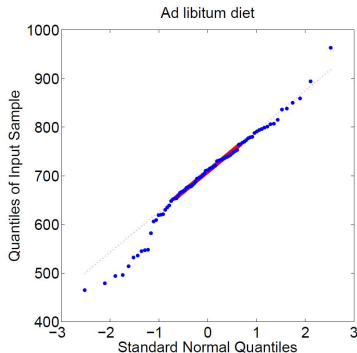
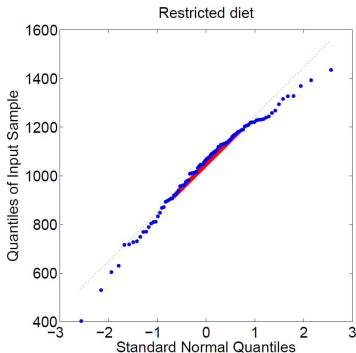


График Q-Q (квантиль-квантиль)

Критерий Шапиро-Уилка отклоняет гипотезы нормальности:
 $p_1 = 1.7 \times 10^{-6}$, $p_2 = 1.5 \times 10^{-7}$.

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

Возьмём усечённую выборку:



$n_1 = 96, n_2 = 86.$

Критерий Шапиро-Уилка: $p_1 = 0.0443, p_2 = 0.0960.$

Критерий Стьюдента:

- для односторонней альтернативы $p = 4 \times 10^{-32}, C_L = 298;$
- для двусторонней альтернативы $p = 9 \times 10^{-32}, [C_L, C_U] = [290, 382].$

8 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: $H_0: \mu_1 = \mu_2;$

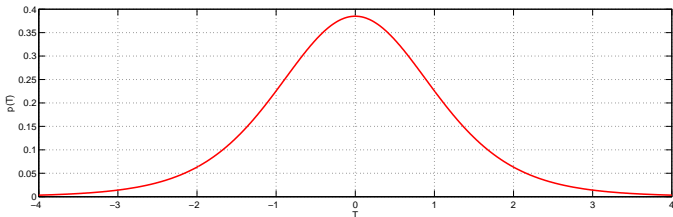
альтернатива: $H_1: \mu_1 < \neq > \mu_2;$

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

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

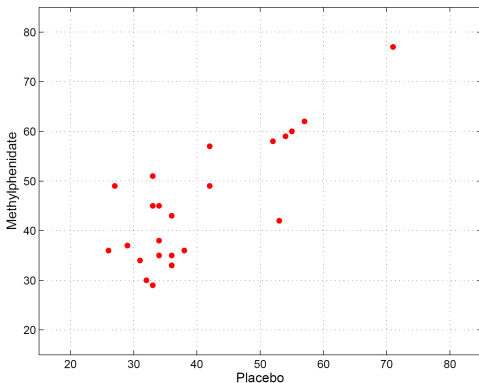
$$D_i = X_{1i} - X_{2i};$$

$$T(X_1^n, X_2^n) \sim St(n-1) \text{ при } H_0.$$



8 t-критерий Стьюдента для связанных выборок

Пример (Pearson, 2003): 24 умственно отсталых ребёнка с СДВГ в течение недели принимали либо метилфенидат, либо плацебо, а в конце недели проходили тест на способность к подавлению импульсивных поведенческих реакций. На втором этапе плацебо и препарат менялись, после недельного курса каждый испытуемые проходили второй тест. Каков эффект препарата?



8 t-критерий Стьюдента для связанных выборок

H_0 : терапия неэффективна, способность к подавлению импульсивных поведенческих реакций не меняется.

H_1 : в результате терапии способность к подавлению импульсивных поведенческих реакций увеличивается $\Rightarrow p = 0.0019$, односторонний верхний 95% доверительный предел для увеличения — -2.32 ;

H_1 : в результате терапии способность к подавлению импульсивных поведенческих реакций меняется $\Rightarrow p = 0.0038$, 95% доверительный интервал для изменения — $[-8.14, -1.78]$.

Игнорируем связь между выборками и применим обычный двухвыборочный критерий Стьюдента:

- для односторонней альтернативы $p = 0.0766$, $C_U = 0.77$;
- для двусторонней альтернативы $p = 0.1532$, $[C_L, C_U] = [-11.83, 1.91]$.

• F-критерий Фишера

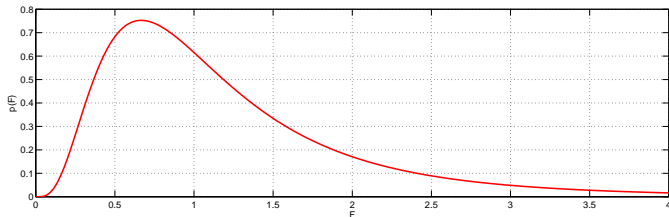
выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2),$
 $X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2);$

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

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

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

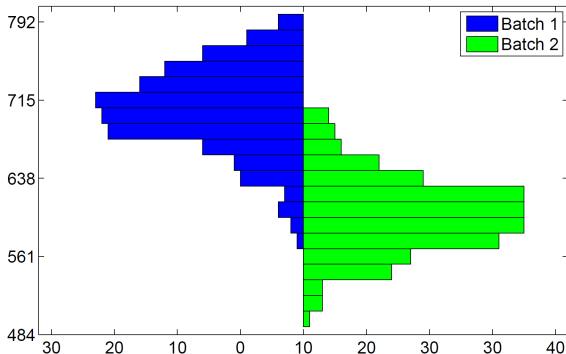
$F(X_1^{n_1}, X_2^{n_2}) \sim F(n_1 - 1, n_2 - 1)$ при $H_0.$



Ф-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.

Одинакова ли дисперсия прочности в разных партиях?

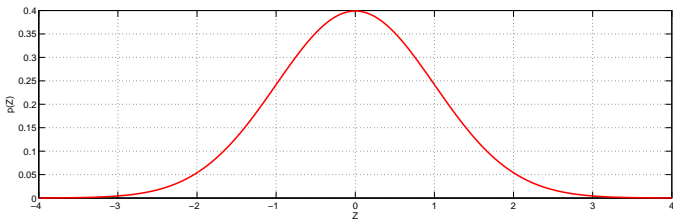


Гипотезы нормальности не отклоняются критерием Шапиро-Уилка ($p_1 = 0.2062, p_2 = 0.7028$).

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

Z-критерий для доли

- выборка: $X^n = (X_1, \dots, X_n), X \sim \text{Ber}(p)$;
 нулевая гипотеза: $H_0: p = p_0$;
 альтернатива: $H_1: p < \neq > p_0$;
 статистика: $Z(X^n) = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$,
 $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$;
 $Z(X^n) \sim N(0, 1)$ при H_0 .



Z-критерий для доли

Пример 1 (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 $\Rightarrow p = 2.3 \times 10^{-12}$.

Z-критерий для доли

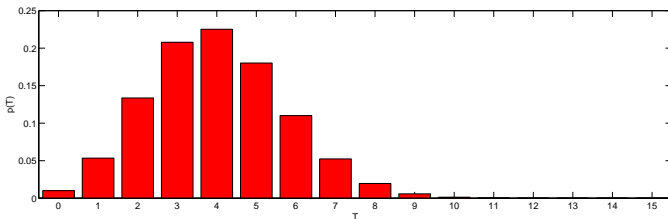
Пример 2 (Кобзарь, задача 227): нормируемый уровень дефектных изделий в партии $p_0 = 0.05$. Из партии извлечена выборка $n = 20$ изделий, в которой обнаружены при проверке $t = 2$ дефектных.

H_0 : доля дефектных изделий в партии не превосходит нормируемого значения.

H_1 : доля дефектных изделий в партии превышает нормируемое значение
 $\Rightarrow p = 0.15$.

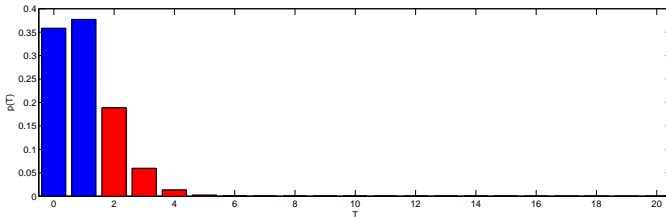
Точный критерий для доли

- 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: $T(X^n) = \sum_{i=1}^n X_i$;
 $T(X^n) \sim Bin(n, p_0)$ при H_0 .



Точный критерий для доли

Пример 2:



H_0 : доля дефектных изделий в партии не превосходит нормируемого значения.

H_1 : доля дефектных изделий в партии превышает нормируемое значение
 $\Rightarrow p = 0.26$.

Доверительные интервалы для доли

$$\frac{t - np}{\sqrt{np(1-p)}} \sim N(0, 1).$$

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{t - np}{\sqrt{np(1-p)}} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha,$$

$$P\left(\frac{t}{n} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{t}{n} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right) \approx 1 - \alpha;$$

подставим вместо p оценку $\hat{p} = \frac{t}{n}$, получим

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

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

В примере 1 95% доверительный интервал Вальда — $[0.2481, 0.3519]$.

В примере 2 — $[-0.0315, 0.2315]$.

Доверительные интервалы для доли

Недостатки доверительного интервала Вальда:

- 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 — покрывает p реже, чем в $100(1 - \alpha)\%$ случаев.

Доверительные интервалы для доли

Более точный доверительный интервал Уилсона (score confidence interval):

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

В примере 1 95% доверительный интервал Уилсона — [0.2509, 0.3541].
 В примере 2 — [0.0279, 0.3010].

```

В R:
> library("PropCIs")
> scoreci(2,20,0.95)
95 percent confidence interval:
0.0279 0.3010
    
```

Минимальный уровень доверия, при котором интервал не содержит 0.05 — 0.305.

Z-критерий для разности двух долей, независимые выборки

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

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

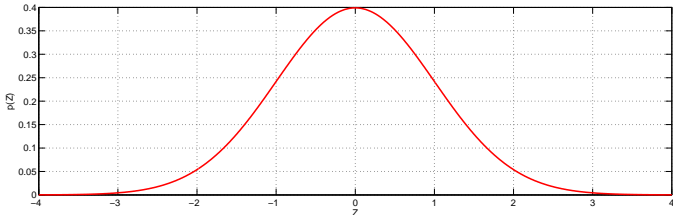
$$p_1 = \frac{\mathbb{E}A}{n_1}, \hat{p}_1 = \frac{a}{n_1}, \quad p_2 = \frac{\mathbb{E}B}{n_2}, \hat{p}_2 = \frac{b}{n_2};$$

нулевая гипотеза: $H_0: p_1 = p_2;$

альтернатива: $H_1: p_1 < \neq > p_2;$

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

$$Z(X_1^{n_1}, X_2^{n_2}) \sim N(0, 1) \text{ при } H_0.$$



Z-критерий для разности двух долей, независимые выборки

Пример (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.

Номер партии	1	2
Наличие дефекта		
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 R:

```
> library("PropCIs")
> diffscoreci(3, 100, 5, 200, 0.95)
95 percent confidence interval:
-0.03313566 0.06158107
```

Минимальный уровень доверия, при котором интервал не содержит нуля — 0.8003.

Z-критерий для разности двух долей, связанные выборки

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

$X_1^n \backslash X_2^n$	1	0	Σ
1	e	g	$e + g$
0	f	h	$f + h$
Σ	$e + f$	$g + h$	n

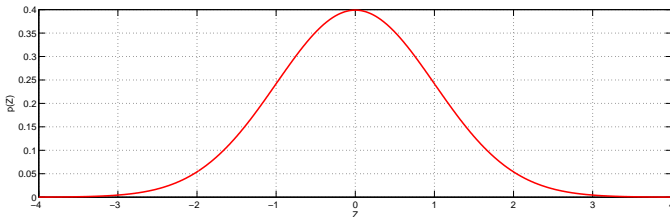
$$p_1 = \frac{\mathbb{E}(E+F)}{n}, \quad \hat{p}_1 = \frac{e+f}{n}, \quad p_2 = \frac{\mathbb{E}(E+G)}{n}, \quad \hat{p}_2 = \frac{e+g}{n};$$

нулевая гипотеза: $H_0: p_1 = p_2;$

альтернатива: $H_1: p_1 < \neq > p_2;$

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

$$Z(X_1^n, X_2^n) \sim N(0, 1) \text{ при } H_0.$$



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.

I \ II	+	-	Σ
+	$e = 794$	$g = 150$	944
-	$f = 86$	$h = 570$	656
Σ	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 $\Rightarrow p = 0.999986$.

Z-критерий для разности двух долей, связанные выборки

Без учёта информации о связи между выборками:

Результат \ Опрос	I	II
+	$a = 994$	$b = 880$
-	$c = 606$	$d = 720$
Σ	$n_1 = 1600$	$n_2 = 1600$

- H_0 : рейтинг премьер-министра не изменился.
- H_1 : рейтинг премьер-министра изменился $\Rightarrow p = 4.3 \times 10^{-5}$.
- H_1 : рейтинг премьер-министра снизился $\Rightarrow p = 2.1 \times 10^{-5}$.
- H_1 : рейтинг премьер-министра повысился $\Rightarrow p = 0.999979$.

Доверительный интервал для разности двух долей

Доверительный интервал Уилсона:

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

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

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

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

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

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

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

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

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

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

Доверительный интервал для разности двух долей

В R:

```
> library("PropCIs")  
> scoreci.mp(86, 150, 1600, 0.95)  
95 percent confidence interval:  
0.02136157 0.05899805
```

Минимальный уровень доверия, при котором интервал не содержит нуля — 3.1×10^{-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) — Канжи, №№ 1, 2, 3;
- t-критерии Стьюдента (t-tests) — Канжи, №№ 7, 8, 9;
- критерий хи-квадрат (chi-square test) — Канжи, №15;
- критерий Фишера (F-test) — Канжи, №16.

Литература

Для распределения Бернулли:

- 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.

Кобзарь А.И. *Прикладная математическая статистика*. — М.: Физматлит, 2006.

Kanji G.K. *100 statistical tests*. — London: SAGE Publications, 2006.

McDonald J.H. *Handbook of Biological Statistics*. — Baltimore: Sparky House Publishing, 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). *Improved confidence intervals for the difference between binomial proportions based on paired data*. *Statistics in Medicine*, 17, 2635–2650.

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

Литература

- Королёв В.Ю. *Теория вероятностей и математическая статистика*. — М.: Проспект, 2008.
- Agresti A. *Categorical Data Analysis*. — Hoboken: John Wiley & Sons, 2002.
- Michelson A.A. *Experimental determination of the velocity of light made at the U.S. Naval Academy, Annapolis*. *Astronomical Papers*, 1, 109-145, 1880.
- NIST/SEMATECH. *e-Handbook of Statistical Methods*.
<http://www.itl.nist.gov/div898/handbook/>
- Pearson D.A, Santos C.W., Casat C.D., et al. (2004). *Treatment effects of methylphenidate on cognitive functioning in children with mental retardation and ADHD*. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43(6), 677–685.
- Weindruch R., Walford R.L. *The Retardation of Aging and Disease by Dietary Restriction*. — Springfield: Charles C. Thomas, 1988.