Содержание

1 Параметрические критерии

1.1 Одновыборочные критерии Стьюдента

1.1.1 z-критерий

$$\begin{aligned}
\text{выборка:}\quad & X^n = (X_1, \dots, X_n) \\
& X \sim N(\mu, \sigma^2), \sigma \text{ известна} \\
\text{нулевая гипотеза:}\quad & H_0:\ \mu = \mu_0 \\
\text{альтернатива:}\quad & H_1:\ \mu <\neq> \mu_0 \\
\text{статистика:}\quad & Z(X^n) = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \\
\text{нулевое распределение:}\quad & Z(X^n) \sim N(0,1)
\end{aligned}$$

$$p = \begin{cases} F_{N(0,1)}(z) & H_1\colon \mu < \mu_0 \\ 1 - F_{N(0,1)}(z) & H_1\colon \mu > \mu_0 \\ 2(1 - F_{N(0,1)}(|z|)) & H_1\colon \mu \neq \mu_0 \end{cases}$$

1.1.2 t-критерий

$$\begin{aligned}
\text{выборка:}\quad & X^n = (X_1, \dots, X_n) \\
& X \sim N(\mu, \sigma^2), \sigma \text{ неизвестна} \\
\text{нулевая гипотеза:}\quad & H_0:\ \mu = \mu_0 \\
\text{альтернатива:}\quad & H_1:\ \mu <\neq> \mu_0 \\
\text{статистика:}\quad & T(X^n) = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \\
\text{нулевое распределение:}\quad & T(X^n) \sim St(n-1)
\end{aligned}$$

$$p = \begin{cases} F_{St(n-1)}(t) & H_1\colon \mu < \mu_0 \\ 1 - F_{St(n-1)}(t) & H_1\colon \mu > \mu_0 \\ 2(1 - F_{St(n-1)}(|t|)) & H_1\colon \mu \neq \mu_0 \end{cases}$$

1.2 Двухвыборочные критерии Стьюдента

1.2.1 z-критерий

$$
\begin{aligned}
\text{выборки:}\quad & X_1^{n_1} = (X_{11}, \ldots, X_{1n_1}) \\
& X_2^{n_2} = (X_{21}, \ldots, X_{2n_2}) \\
& X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), \\
& \sigma_1, \sigma_2 \text{ известны} \\
\text{нулевая гипотеза:}\quad & H_0: \ \mu_1 = \mu_2 \\
\text{альтернатива:}\quad & H_1: \ \mu_1 <\neq> \mu_2 \\
\text{статистика:}\quad & Z\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\
\text{нулевое распределение:}\quad & Z\left(X_1^{n_1}, X_2^{n_2}\right) \sim N(0,1)
\end{aligned}
$$

1.2.2 t-критерий

$$
\begin{aligned}
\text{выборки:}\quad & X_1^{n_1} = (X_{11}, \ldots, X_{1n_1}) \\
& X_2^{n_2} = (X_{21}, \ldots, X_{2n_2}) \\
& X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), \\
& \sigma_1, \sigma_2 \text{ неизвестны} \\
\text{нулевая гипотеза:}\quad & H_0: \ \mu_1 = \mu_2 \\
\text{альтернатива:}\quad & H_1: \ \mu_1 <\neq> \mu_2 \\
\text{статистика:}\quad & T\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \\
\text{нулевое распределение:}\quad & T\left(X_1^{n_1}, X_2^{n_2}\right) \approx\sim St(\nu)
\end{aligned}
$$

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

Нулевое распределение приближённое, а не точное. Точного решения не существует! (Проблема Беренца-Фишера). Приближение достаточно при $n_1 = n_2$ или при $[n_1 > n_2] = [\sigma_1 > \sigma_2]$

1.3 Критерий Стьюдента для связанных выборок

1.3.1 t-критерий для связанных выборок

$$
\begin{aligned}
\text{выборки:}\quad & X_1^n = (X_{11}, \ldots, X_{1n}), X_1 \sim N(\mu_1, \sigma_1^2) \\
& X_2^n = (X_{21}, \ldots, X_{2n}), X_2 \sim N(\mu_2, \sigma_2^2) \\
\text{нулевая гипотеза:}\quad & H_0: \ \mu_1 = \mu_2 \\
\text{альтернатива:}\quad & H_1: \ \mu_1 <\neq> \mu_2 \\
\text{статистика:}\quad & T(X_1^n, X_2^n) = \frac{\overline{X_1} - \overline{X_2}}{S/\sqrt{n}} \\
& S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(D_i - \overline{D}\right)^2, D_i = X_{1i} - X_{2i} \\
\text{нулевое распределение:}\quad & T(X_1^n, X_2^n) \sim St(n-1)
\end{aligned}
$$

1.4 Нормальность выборок

1.4.1 Критерей χ^2

$$
\begin{aligned}
\text{выборка:}\quad & X^n = (X_1, \ldots, X_n) \\
\text{нулевая гипотеза:}\quad & H_0: \ X \sim N(\mu, \sigma^2) \\
\text{альтернатива:}\quad & H_1: \ H_0 \text{ неверна} \\
\text{статистика:}\quad & \chi^2(X^n) = \sum_{i=1}^{K} \frac{(n_i - np_i)^2}{np_i} \\
\text{нулевое распределение:}\quad & \chi^2(X^n) = \begin{cases} \chi^2_{K-1}, & \mu,\sigma \text{ заданы} \\ \chi^2_{K-3}, & \mu,\sigma \text{ оцениваются} \end{cases} \\
& n_i \text{ — число элементов выборки в } [a_i, a_{i+1}] \\
& p_i = F_{N(\mu,\sigma^2)}(a_{i+1}) - F_{N(\mu,\sigma^2)}(a_i)
\end{aligned}
$$

1.4.2 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:	$W(X^n) = \dfrac{\left(\sum\limits_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum\limits_{i=1} n(X_i - \overline{X})^2}$
нулевое распределение:	табличное

a_i основаны на матожиданиях порядковых статистик нормального распределения и также табулированы.

Если нормальность отвергается, чувствительные методы, предполагающие нормальность, использовать **нельзя**!

1.5 Гипотезы о долях

1.5.1 z-критерий для доли

$$
\begin{aligned}
\text{выборка:}\quad & X^n = (X_1, \dots, X_n) \\
& X \sim Ber(p) \\
\text{нулевая гипотеза:}\quad & H_0:\ p = p_0 \\
\text{альтернатива:}\quad & H_1:\ p <\neq> p_0 \\
\text{статистика:}\quad & Z\left(X^n\right) = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \hat{p} = \overline{X_n} \\
\text{нулевое распределение:}\quad & Z\left(X^n\right) \sim N(0,1)
\end{aligned}
$$

1.5.2 z-критерий для двух долей, не связанные выборки

$$
\begin{aligned}
\text{выборка:}\quad & X_1^{n_1} = \left(X_{11}, \dots, X_{1n_1}\right), X_1 \sim Ber(p_1) \\
& X_2^{n_2} = \left(X_{21}, \dots, X_{2n_2}\right), X_2 \sim Ber(p_2) \\
& \text{выборки независимы} \\
\text{нулевая гипотеза:}\quad & H_0:\ p_1 = p_2 \\
\text{альтернатива:}\quad & H_1:\ p_1 <\neq> p_2 \\
\text{статистика:}\quad & Z\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\
& P = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2} \\
\text{нулевое распределение:}\quad & Z\left(X_1^{n_1}, X_2^{n_2}\right) \sim N(0,1)
\end{aligned}
$$

При независимых выборках Z-критерий использует только первую строку таблицы:

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

	X_1	X_2
1	a	b
0	c	d
$\sum$	n_1	n_2

1.5.3 z-критерий для двух долей, связанные выборки

выборка:	$X_1^n = (X_{11}, \dots, X_{1n})\,, X_1 \sim Ber(p_1)$
	$X_2^n = (X_{21}, \dots, X_{2n})\,, X_2 \sim Ber(p_2)$
	выборки связанные
нулевая гипотеза:	$H_0:\ p_1 = p_2$
альтернатива:	$H_1:\ p_1 <\neq> p_2$
статистика:	$Z\left(X_1^n, X_2^n\right) = \dfrac{f-g}{\sqrt{f+g-\dfrac{(f-g)^2}{n}}}$
нулевое распределение:	$Z\left(X_1^n, X_2^n\right) \sim N(0,1),$ при $H_0 = 0$

X_1^n / X_2^n	1	0	$\sum$
1	e	f	$e+f$
0	g	h	$g+h$
$\sum$	$e+g$	$f+h$	n

При связанных выборках Z-критерий использует только внедиагональные элементы таблицы.

2 Непараметрические критерии

2.1 Критерий знаков

2.1.1 Одновыборочный критерий знаков

$$\begin{aligned}
\text{выборка:}\quad & X^n = (X_1, \ldots, X_n), X_i \neq m_0 \\
\text{нулевая гипотеза:}\quad & H_0:\ med\ X = m_0 \\
\text{альтернатива:}\quad & H_1:\ med\ X <\neq> m_0 \\
\text{статистика:}\quad & T(X^n) = \sum_{i=1}^{n} [X_i > m_0] \\
\text{нулевое распределение:}\quad & T(X^n) \sim Bin\left(n, \tfrac{1}{2}\right)
\end{aligned}$$

2.1.2 Двухвыборочный критерий знаков

$$\begin{aligned}
\text{выборка:}\quad & X_1^n = (X_{11}, \ldots, X_{1n}) \\
& X_2^n = (X_{21}, \ldots, X_{2n}) \\
& X_{1i} \neq X_{2i}, \text{ выборки связанные} \\
\text{нулевая гипотеза:}\quad & H_0:\ P(X_1 > X_2) = \frac{1}{2} \\
\text{альтернатива:}\quad & H_1:\ P(X_1 > X_2) <\neq> \frac{1}{2} \\
\text{статистика:}\quad & T(X_1^n, X_2^n) = \sum_{i=1}^{n} [X_{1i} > X_{2i}] \\
\text{нулевое распределение:}\quad & T(X_1^n, X_2^n) \sim Bin\left(n, \tfrac{1}{2}\right)
\end{aligned}$$

2.2 Ранговые критерии

2.2.1 Критерий ранговых знаков

$$\begin{aligned}
\text{выборка:}\quad & X^n = (X_1, \ldots, X_n), X_i \neq m_0 \\
& F_X \text{ симметрично относительно медианы} \\
\text{нулевая гипотеза:}\quad & H_0:\ med\ X = m_0 \\
\text{альтернатива:}\quad & H_1:\ med\ X <\neq> m_0 \\
\text{статистика:}\quad & W(X^n) = \sum_{i=1}^{n} rank(|X_i - m_0|) \cdot sign(X_i - m_0) \\
\text{нулевое распределение:}\quad & \text{табличное}
\end{aligned}$$

Апроксимация для $n > 20$:

$$W \approx\sim N\left(0, \frac{n(n+1)(2n+1)}{6}\right)$$

2.2.2 Критерий ранговых знаков, связанные выборки

выборка:	$X_1^n = (X_{11}, \ldots, X_{1n})$
	$X_2^n = (X_{21}, \ldots, X_{2n})$
	$X_{1i} \neq X_{2i}$, выборки связанные
нулевая гипотеза:	$H_0: \ med(X_1 - X_2) = 0$
альтернатива:	$H_1: \ med(X_1 - X_2) <\neq> 0$
статистика:	$W\left(X_1^n, X_2^n\right) = \sum\limits_{i=1}^{n} rank(\lvert X_{1i} - X_{2i}\rvert) \cdot sign(X_{1i} - X_{2i})$
нулевое распределение:	табличное

2.3 Критерий Манна-Уитни

выборка:	$X_1^{n_1} = (X_{11}, \ldots, X_{1n_1})$
	$X_2^{n_2} = (X_{21}, \ldots, X_{2n_2})$
нулевая гипотеза:	$H_0: \ F_{X_1}(x) = F_{X_2}(x)$
альтернатива:	$H_1: \ F_{X_1}(x) = F_{X_2}(x + \Delta), \Delta <\neq> 0$
статистика:	$X_{(1)} \leqslant \ldots \leqslant X_{(n_1+n_2)}$ - вариационный ряд
	объеденённой выборки
	$X = X_1^{n_1} \cup X_2^{n_2}$
	$R\left(X_1^{n_1}, X_2^{n_2}\right) = \sum\limits_{i=1}^{n_1} rank(X_{1i})$
нулевое распределение:	табличное

Апроксимация для $n_1, n_2 > 10$:

$$R_1 \sim N\left(\frac{n_1(n_1+n_2+1)}{2}, \frac{n_1 n_2(n_1+n_2+1)}{12}\right)$$

2.4 Перестановочные критерии

2.4.1 Одновыборочный критерий

выборка: $X^n = (X_1, \ldots, X_n)$
F_X симметрично относительно матожидания
нулевая гипотеза: $H_0: \ \mathsf{E}X = m_0$
альтернатива: $H_1: \ \mathsf{E}X <\neq> m_0$
статистика: $T\left(X^n\right) = \sum\limits_{i=1}^{n}(X_i - m_0)$
нулевое распределение: порождается перебором 2^n знаков
перед слагаемыми $X_i - m_0$

2.4.2 Для связанных выборок

выборка: $X_1^n = (X_{11}, \ldots, X_{1n})$
$X_2^n = (X_{21}, \ldots, X_{2n})$
выботки связанные
нулевая гипотеза: $H_0: \ \mathsf{E}(X_1 - X_2) = m_0$
альтернатива: $H_1: \ \mathsf{E}(X_1 - X_2) <\neq> m_0$
статистика: $D^n = (X_{1i} - X_{2i})$
$T\left(X_1^n, X_2^n\right) = T(D^n) = \sum\limits_{i=1}^{n} D_i$
нулевое распределение: порождается перебором 2^n знаков
перед слагаемыми D_i

2.4.3 Для независимых выборок

выборка: $X_1^{n_1} = (X_{11}, \ldots, X_{1n_1})$
$X_2^{n_2} = (X_{21}, \ldots, X_{2n_2})$
нулевая гипотеза: $H_0: \ F_{X_1}(x) = F_{X_2}(x)$
альтернатива: $H_1: \ F_{X_1}(x) = F_{X_2}(x + \Delta), \Delta <\neq> 0$
статистика: $T\left(X_1^{n_1}, X_2^{n_2}\right) = \dfrac{1}{n_1}\sum\limits_{i=1}^{n_1} X_{1i} - \dfrac{1}{n_2}\sum\limits_{i=1}^{n_2} X_{2i}$
нулевое распределение: порождается перебором $C_{n_1+n_2}^{n_1}$
размещений объеденённой выборки