

Сводная таблица по математической статистике

Python для всех пунктов:

1. Двусторонний тест: **p-value** = $2 \cdot \min \left\{ \text{r.cdf}(v_p), 1 - \text{r.cdf}(v_p) \right\}$, где r – распределение статистики, v_p – значение расчётной статистики.
2. $\overline{X} = \text{np.mean}(x)$
3. $S_0 = \text{np.std}(x, ddof = 1)$, $S_0^2 = \text{np.var}(x, ddof = 1)$

1. Одно распределение

Название	Предпосылки	H_0	H_1	Статистика	Выводы	Python (numpy, scipy.stats)
Гипотеза о матожидании	<div>1. $X \sim N(\mu, \sigma^2)$</div> <div>2. σ^2 - известно</div>	$\mu = \mu_0$	$\mu \neq \mu_0$	$z_p = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$	Не отвергаем на уровне значимости α , если <div>1. $z_p \in \left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$,</div> <div>2. $\mu_0 \in \left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$</div> <div>3. p-value $> \alpha$</div>	<div>1. $z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(q = 1 - \alpha/2)$,</div> <div>2. p-value = $2 \cdot \left(1 - \text{norm.cdf}(\text{abs}(z_p))\right)$</div>
Гипотеза о матожидании	<div>1. $X \sim N(\mu, \sigma^2)$</div> <div>2. σ^2 - неизвестно</div>	$\mu = \mu_0$	$\mu \neq \mu_0$	$t_p = t^{(n-1)} = \frac{\overline{X} - \mu_0}{S_0 / \sqrt{n}} \sim T_{n-1}$	Не отвергаем на уровне значимости α , если <div>1. $t_p \in \left(-t_{1-\frac{\alpha}{2}}^{(n-1)}, t_{1-\frac{\alpha}{2}}^{(n-1)}\right)$,</div> <div>2. $\mu_0 \in \left(\overline{X} - t_{1-\frac{\alpha}{2}}^{(n-1)} \frac{S_0}{\sqrt{n}}, \overline{X} + t_{1-\frac{\alpha}{2}}^{(n-1)} \frac{S_0}{\sqrt{n}}\right)$</div> <div>3. p-value $> \alpha$</div>	<div>1. $t_{1-\frac{\alpha}{2}}^{(n-1)} = \text{t.ppf}(df = n - 1, q = 1 - \alpha/2)$,</div> <div>2. p-value = $2 \cdot \left(1 - \text{t.cdf}(\text{abs}(t_p), df = n - 1)\right)$</div>
Гипотеза о дисперсии	<div>1. $X \sim N(\mu, \sigma^2)$</div> <div>2. μ - неизвестно</div>	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$C_p = C^{(n-1)} = \frac{S_0^2(n-1)}{\sigma_0^2} \sim \chi_{n-1}^2$	Не отвергаем на уровне значимости α , если <div>1. $C_p \in \left(C_{\frac{\alpha}{2}}^{(n-1)}, C_{1-\frac{\alpha}{2}}^{(n-1)}\right)$,</div> <div>2. $\sigma_0^2 \in \left(\frac{(n-1)S_0^2}{C_{1-\frac{\alpha}{2}}^{(n-1)}}, \frac{(n-1)S_0^2}{C_{\frac{\alpha}{2}}^{(n-1)}}\right)$</div> <div>3. p-value $> \alpha$</div>	<div>1. $C_{\frac{\alpha}{2}}^{(n-1)} = \text{chi2.ppf}(df = n - 1, q = \alpha/2)$,</div> <div>2. $C_{1-\frac{\alpha}{2}}^{(n-1)} = \text{chi2.ppf}(df = n - 1, q = 1 - \alpha/2)$,</div> <div>3. p-value</div>
Гипотеза о дисперсии	<div>1. $X \sim N(\mu, \sigma^2)$</div> <div>2. μ - известно</div>	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$C_p = C^{(n)} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} \sim \chi_n^2$	Не отвергаем на уровне значимости α , если <div>1. $C_p \in \left(C_{\frac{\alpha}{2}}^{(n)}, C_{1-\frac{\alpha}{2}}^{(n)}\right)$,</div> <div>2. $\sigma_0^2 \in \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{C_{1-\frac{\alpha}{2}}^{(n-1)}}, \frac{\sum_{i=1}^n (x_i - \mu)^2}{C_{\frac{\alpha}{2}}^{(n-1)}}\right)$</div> <div>3. p-value $> \alpha$</div>	<div>1. $C_{\frac{\alpha}{2}}^{(n-1)} = \text{chi2.ppf}(df = n - 1, q = \alpha/2)$,</div> <div>2. $C_{1-\frac{\alpha}{2}}^{(n-1)} = \text{chi2.ppf}(df = n - 1, q = 1 - \alpha/2)$,</div> <div>3. p-value</div>
Асимптотическая гипотеза о матожидании	<div>1. $X \sim \mathcal{F}$</div> <div>2. $D(x) = \sigma^2$ - известно</div> <div>3. $n \rightarrow \infty$ ($n \gg 0$)</div>	$\mu = \mu_0$	$\mu \neq \mu_0$	$z_p = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$	Не отвергаем на уровне значимости α , если <div>1. $z_p \in \left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$,</div> <div>2. $\mu_0 \in \left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$</div> <div>3. p-value $> \alpha$</div>	<div>1. $z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(q = 1 - \alpha/2)$,</div> <div>2. p-value = $2 \cdot \left(1 - \text{norm.cdf}(\text{abs}(z_p))\right)$</div>
Асимптотическая гипотеза о матожидании	<div>1. $X \sim \mathcal{F}$</div> <div>2. $D(x) = \sigma^2$ - неизвестно</div> <div>3. $n \rightarrow \infty$ ($n \gg 0$)</div>	$\mu = \mu_0$	$\mu \neq \mu_0$	$z_p = \frac{\overline{X} - \mu_0}{S_0 / \sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$	Не отвергаем на уровне значимости α , если <div>1. $z_p \in \left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$,</div> <div>2. $\mu_0 \in \left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{S_0}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{S_0}{\sqrt{n}}\right)$</div> <div>3. p-value $> \alpha$</div>	<div>1. $z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(q = 1 - \alpha/2)$,</div> <div>2. p-value = $2 \cdot \left(1 - \text{norm.cdf}(\text{abs}(z_p))\right)$</div>
Bootstrap	<div>1. $X \sim \mathcal{F}$</div> <div>2. n - небольшое</div>	$\mu = \mu_0$ или $\sigma^2 = \sigma_0^2$	$\mu \neq \mu_0$ или $\sigma^2 \neq \sigma_0^2$	Генерируем много выборок из данной одинаковой длины. Считаем для каждой них нужную статистику $\left(\overline{X}_i \text{ или } \hat{\sigma}_i^2\right)$. Считаем квантили $q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}$ для выборки этих статистик.	Не отвергаем на уровне значимости α , если <div>1. $\mu_0 \left(\sigma_0^2\right) \in \left(q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}\right)$</div>	<div>1. scipy.stats.bootstrap</div> <div>2. numpy.random.choice</div>

2. Два распределения

Название	Предпосылки	H_0	H_1	Статистика	Выводы	Python (numpy, scipy.stats)
Гипотеза о разности матожиданий связанных пар	<div>1. $X \sim N(\mu_x, \sigma_x^2)$</div> <div>2. $Y \sim N(\mu_y, \sigma_y^2)$</div> <div>3. $n = n_x = n_y$</div> <div>4. σ_x^2, σ_y^2 – известно</div>	$\mu_x - \mu_y = \mu_0$	$\mu_x - \mu_y \neq \mu_0$	$\Delta = X - Y,$ $z_p = \frac{\overline{\Delta} - \mu_0}{D(\overline{\Delta})} \sim N(0, 1),$ $z_p = \frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2 + \sigma_y^2}{n}}} \sim N(0, 1)$	Не отвергаем на уровне значимости α , если <div>1. $z_p \in \left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$,</div> <div>2. $\mu_0 \in \left(\overline{\Delta} - z_{1-\frac{\alpha}{2}} D(\overline{\Delta}), \overline{\Delta} + z_{1-\frac{\alpha}{2}} D(\overline{\Delta})\right)$</div> <div>3. p-value $> \alpha$</div>	См. аналогичное выше, p-value
Гипотеза о разности матожиданий связанных пар	<div>1. $X \sim N(\mu_x, \sigma_x^2)$</div> <div>2. $Y \sim N(\mu_y, \sigma_y^2)$</div> <div>3. $n = n_x = n_y$</div> <div>4. σ_x^2, σ_y^2 – неизвестно</div>	$\mu_x - \mu_y = \mu_0$	$\mu_x - \mu_y \neq \mu_0$	$\Delta = X - Y,$ $t_p = \frac{\overline{\Delta} - \mu_0}{S_0(\Delta) / \sqrt{n}} \sim T_{n-1}$	Не отвергаем на уровне значимости α , если <div>1. $t_p \in \left(-t_{1-\frac{\alpha}{2}}^{(n-1)}, t_{1-\frac{\alpha}{2}}^{(n-1)}\right)$,</div> <div>2. $\mu_0 \in \left(\overline{\Delta} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \frac{S_0(\Delta)}{\sqrt{n}}\right)$</div> <div>3. p-value $> \alpha$</div>	См. аналогичное выше, p-value
Гипотеза о разности матожиданий	<div>1. $X \sim N(\mu_x, \sigma_x^2)$</div> <div>2. $Y \sim N(\mu_y, \sigma_y^2)$</div> <div>3. σ_x^2, σ_y^2 – известно</div>	$\mu_x - \mu_y = \mu_0$	$\mu_x - \mu_y \neq \mu_0$	$z_p = \frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim N(0, 1)$	Не отвергаем на уровне значимости α , если <div>1. $z_p \in \left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$,</div> <div>2. $(\mu_x - \mu_y) \in \left(\overline{X} - \overline{Y} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)$</div> <div>3. p-value $> \alpha$</div>	См. аналогичное выше, p-value

Название	Предпосылки	H_0	H_1	Статистика	Выводы	Python (numpy, scipy.stats)
Гипотеза о разности матожиданий	1. $X \sim N(\mu_x, \sigma^2)$ 2. $Y \sim N(\mu_y, \sigma^2)$ 3. $\sigma^2 = \sigma_x^2 = \sigma_y^2$ – неизвестно	$\mu_x - \mu_y = \mu_0$	$\mu_x - \mu_y \neq \mu_0$	$\hat{\sigma}^2 = \frac{S_0^2(X)(n_x - 1) + S_0^2(Y)(n_y - 1)}{n_x + n_y - 2}$ $t_p = \frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\hat{\sigma} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim T_{n_x + n_y - 2}$	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> $t_p \in \left(-t_{1-\frac{\alpha}{2}}^{(n_x+n_y-2)}, t_{1-\frac{\alpha}{2}}^{(n_x+n_y-2)}\right)$, $(\mu_x - \mu_y) \in \left(\overline{X} - \overline{Y} \pm t_{1-\frac{\alpha}{2}}^{(n_x+n_y-2)} \hat{\sigma}\right)$ p-value $> \alpha$ 	<ol style="list-style-type: none"> $t_{1-\frac{\alpha}{2}}^{(n_x+n_y-2)} = \text{t.ppf}(q = 1 - \alpha/2, df = n_x + n_y - 2)$, p-value $= 2 \cdot (1 - \text{t.cdf}(\text{abs}(t_p), df = n_x + n_y - 2))$, scipy.stats.ttest_ind
Гипотеза о равенстве матожиданий. Тест Уэлча	1. $X \sim N(\mu_x, \sigma_x^2)$ 2. $Y \sim N(\mu_y, \sigma_y^2)$ 3. σ_x^2, σ_y^2 – неизвестно	$\mu_x - \mu_y = 0$	$\mu_x - \mu_y \neq 0$	$t_p = -\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}}} \sim T_{\hat{d}}$ $\hat{d} = \frac{\left(\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}\right)^2}{\frac{\hat{\sigma}_x^4}{n_x^2(n_x - 1)} + \frac{\hat{\sigma}_y^4}{n_y^2(n_y - 1)}}$	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> $t_p \in \left(-t_{1-\frac{\alpha}{2}}^{(\hat{d})}, t_{1-\frac{\alpha}{2}}^{(\hat{d})}\right)$, $0 \in \left(\overline{X} - \overline{Y} \pm t_{1-\frac{\alpha}{2}}^{(\hat{d})} \sqrt{\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}}\right)$ p-value $> \alpha$ 	<ol style="list-style-type: none"> $t_{1-\frac{\alpha}{2}} = \text{t.ppf}(q = 1 - \alpha/2, df = d)$, p-value $= 2 \cdot (1 - \text{t.cdf}(\text{abs}(t_p), df = d))$ scipy.stats.ttest_ind(equal_var=False)
Гипотеза об отношении дисперсий	1. $X \sim N(\mu_x, \sigma_x^2)$ 2. $Y \sim N(\mu_y, \sigma_y^2)$ 3. σ_x^2, σ_y^2 – неизвестно	$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} \neq 1$	$f_p = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} \sim F_{n_x-1, n_y-1}$	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> $f_p \in \left(f_{\frac{\alpha}{2}}^{(n_x-1, n_y-1)}, f_{1-\frac{\alpha}{2}}^{(n_x-1, n_y-1)}\right)$, p-value $> \alpha$ 	<ol style="list-style-type: none"> $f_{\frac{\alpha}{2}}^{(n_x-1, n_y-1)} = \text{f.ppf}(df\,n = n_x - 1, df\,d = n_y - 1, q = \alpha/2)$, $f_{1-\frac{\alpha}{2}}^{(n_x-1, n_y-1)} = \text{f.ppf}(df\,n = n_x - 1, df\,d = n_y - 1, q = 1 - \alpha/2)$, p-value

3. Критерии сравнения

Название	Предпосылки	H_0	H_1	Статистика	Выводы	Python (numpy, scipy.stats)
Критерий Пирсона (χ^2) о согласии	1. $X \sim \mathcal{F}_x$ 2. \mathcal{F}_0 – дискретное.	$\mathcal{F}_x = \mathcal{F}_0$	$\mathcal{F}_x \neq \mathcal{F}_0$	Для каждого значения a_i имеем частоту/количество (ν_i) в данной выборке и теоретическую вероятность p_i . $\rho = \sum_{i=1}^k \frac{(\nu_i - np_i)^2}{np_i} \xrightarrow[n \rightarrow \infty]{H_0} \chi_{k-1}^2$	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> p-value $> \alpha$ 	<ol style="list-style-type: none"> p-value $= 2 \cdot \text{chi2.cdf}(\rho, df = n - 1)$
Критерий Колмогорова о согласии	1. $X \sim \mathcal{F}_x$ 2. \mathcal{F}_0 – непрерывное.	$\mathcal{F}_x = \mathcal{F}_0$	$\mathcal{F}_x \neq \mathcal{F}_0$	$\widehat{F}_n(x)$ – эмпирическая функция распределения, $F_0(x)$ – функция распределения \mathcal{F}_0 . $D_n = \sup_x \left \widehat{F}_n(x) - F_0(x) \right ,$ $k_p = \sqrt{n} D_n \xrightarrow[n \rightarrow \infty]{d} \eta \sim \mathcal{K}(y)$ – функция распределения Колмогорова.	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> $k_p \leq K_{1-\alpha}$, p-value $> \alpha$ 	<ol style="list-style-type: none"> scipy.stats.ksone scipy.stats.ks_1samp scipy.stats.kstest
Критерий Колмогорова-Смирнова об однородности	1. $X \sim \mathcal{F}_x$ 2. $Y \sim \mathcal{F}_y$	$\mathcal{F}_x = \mathcal{F}_y$	$\mathcal{F}_x \neq \mathcal{F}_y$	$\widehat{F}_{n_x}(x), \widehat{F}_{n_y}(x)$ – эмпирические функции распределения. $ks_p = \sqrt{\frac{n_x n_y}{n_x + n_y}} \sup_x \left \widehat{F}_{n_x}(x) - \widehat{F}_{n_y}(x) \right $ $ks_p \xrightarrow[n_x, n_y \rightarrow \infty]{d} \eta \sim \mathcal{K}(y)$ – функция распределения Колмогорова.	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> $ks_p \leq K_{1-\alpha}$, p-value $> \alpha$ 	<ol style="list-style-type: none"> scipy.stats.ksone scipy.stats.ks_2samp
Критерий Пирсона (χ^2) о независимости	Объекты имеют пары из категорий (x_i, y_i) . Всего X имеет s категорий, Y имеет k категорий.	X, Y - независимые	X, Y - зависимые	ν_{ij} - частоты пары категорий $(a_i, b_j) \sim (X, Y)$. $n_{i\bullet} = \sum_{j=1}^k \nu_{ij}, \quad n_{\bullet j} = \sum_{i=1}^s \nu_{ij}$ $\gamma = \sum_{i=1}^s \sum_{j=1}^k \frac{\left(\nu_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}} \sim \chi_{(s-1)(k-1)}^2$	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> $\gamma \in (0, C_{1-\alpha}^{(s-1)(k-1)})$, p-value $> \alpha$ 	<ol style="list-style-type: none"> scipy.stats.contingency.crosstab pandas.crosstab scipy.stats.chi2_contingency (correction = False)
Коэффициент корреляции Спирмена	Объекты имеют пары из порядковых (ранговых) переменных (r_i, k_i) .	X, Y - независимые	X, Y - зависимые	$S = \sum_{i=1}^n (r_i - k_i)^2 \in \left[0, \frac{n^3 - n}{3}\right]$ $\rho = 1 - \frac{6S}{n^3 - n} \in [-1, 1]$ $\rho_p = \sqrt{n - 1} \rho \xrightarrow[n \rightarrow \infty]{H_0} N(0, 1)$	Не отвергаем на уровне значимости α , если <ol style="list-style-type: none"> $\rho_p \in \left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$, p-value $> \alpha$ 	<ol style="list-style-type: none"> scipy.stats.spearmanr p-value

Красный текст – ссылка в этом документе. Синий текст – ссылка на страницу в интернете.