Popisná štatistika

Modus

$$\widehat{x} = x_i + \frac{h}{2}\,\frac{n_{i+1} - n_{i-1}}{2n_i - n_{i+1} - n_{i-1}}$$

Medián

$$\widetilde{x} = a + h\,\frac{\frac{n}{2} - (n_1 + n_2 + ... + n_{i-1})}{n_i}$$

Momenty

$$m_3 = m_3^* - 3m_2^* m_1^* + 2{m_1^*}^3$$

$$m_4 = m_4^* - 4m_3^* m_1^* + 6m_2^* {m_1^*}^2 - 3{m_1^*}^4$$

Náhodný výber z normálneho rozdelenia

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\sim\; N\left(\mu,\, \frac{\sigma^2}{n}\right)$$

$$Y = \frac{n-1}{\sigma^2} S_X^2 \sim \chi^2(n-1)$$

$$U = \frac{\bar{X} - \mu}{\sigma}\sqrt{n} \;\sim\; N(0,\, 1)$$

$$T = \frac{\bar{X} - \mu}{S_X}\sqrt{n} \;\sim\; t(n-1)$$

$$Z = \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2,\, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

$$T = \frac{\bar{X} - \bar{Y}(\mu_1 - \mu_2)}{S^*} \;\sim\; t(n+m-2),\; S^* = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}\sqrt{\frac{n+m}{nm}}$$

$$F = \frac{S_X^2}{\sigma_1^2} : \frac{S_Y^2}{\sigma_2^2} \;\sim\; F(n-1,\, m-1)$$

Intervaly spoľahlivosti pre parametre normálneho rozdelenia

1.

a) Interval spoľahlivosti pre strednú hodnotu $\boldsymbol{\mu}$, rozptyl $\boldsymbol{\sigma}^2$ známy

$$\left\langle \bar{X} - k_\alpha \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + k_\alpha \frac{\sigma}{\sqrt{n}} \right\rangle$$

b) Interval spoľahlivosti pre strednú hodnotu $\boldsymbol{\mu}$, rozptyl $\boldsymbol{\sigma}^2$ neznámy

$$\left\langle \bar{X} - t_\alpha(n-1)\frac{S_X}{\sqrt{n}}, \quad \bar{X} + t_\alpha(n-1)\frac{S_X}{\sqrt{n}} \right\rangle$$

$$\left\langle \bar{X} - t_{2\alpha}(n-1)\frac{S_X}{\sqrt{n}}, \quad \infty \right), \quad \left(-\infty, \quad \bar{X} + t_{2\alpha}(n-1)\frac{S_X}{\sqrt{n}} \right\rangle$$

c) Interval spoľahlivosti pre rozptyl $\boldsymbol{\sigma}^2$, stredná hodnota $\boldsymbol{\mu}$ známa

$$\left\langle \frac{n\,S_0^2}{\chi^2_{\frac{\alpha}{2}}(n)}, \quad \frac{n\,S_0^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \right\rangle$$

$$\left\langle \frac{n\,S_0^2}{\chi^2_{\alpha}(n)}, \quad \infty \right), \quad \left(-\infty, \quad \frac{n\,S_0^2}{\chi^2_{1-\alpha}(n)} \right\rangle$$

d) Interval spoľahlivosti pre rozptyl σ^2, stredná hodnota μ neznáma

$$\left\langle \frac{(n-1)\,S_X^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \quad \frac{(n-1)\,S_X^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \right\rangle$$

$$\left\langle \frac{(n-1)\,S_X^2}{\chi^2_{\alpha}(n-1)}, \quad \infty \right), \quad \left(-\infty, \quad \frac{(n-1)\,S_X^2}{\chi^2_{1-\alpha}(n-1)} \right\rangle$$

2.

a) Interval spoľahlivosti pre rozdiel stredných hodnôt $(\mu_1 - \mu_2)$, rozptyly σ_1^2 a σ_2^2 známe

$$\left\langle \bar{X} - \bar{Y} - k_\alpha \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \quad \bar{X} - \bar{Y} + k_\alpha \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right\rangle$$

b) Interval spoľahlivosti pre rozdiel stredných hodnôt $(\mu_1 - \mu_2)$, rozptyly $\sigma_1^2 \neq \sigma_2^2$ neznáme

$$\left\langle \bar{X} - \bar{Y} - t_\alpha(n+m-2) \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}, \quad \bar{X} - \bar{Y} + t_\alpha(n+m-2) \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}} \right\rangle$$

$$\left\langle \bar{X} - \bar{Y} - t_{2\alpha}(n+m-2) \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}, \quad \infty \right)$$

$$\left(-\infty, \quad \bar{X} - \bar{Y} + t_{2\alpha}(n+m-2) \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}} \right\rangle$$

c) Interval spoľahlivosti pre rozdiel stredných hodnôt $(\mu_1 - \mu_2)$, rozptyly $\sigma_1^2 = \sigma_2^2$ neznáme

$$\left\langle \bar{X} - \bar{Y} - t_\alpha(n+m-2) S^*, \quad \bar{X} - \bar{Y} + t_\alpha(n+m-2) S^* \right\rangle$$

$$\left(-\infty, \quad \bar{X} - \bar{Y} + t_{2\alpha}(n+m-2) S^* \right\rangle, \quad \left\langle \bar{X} - \bar{Y} - t_{2\alpha}(n+m-2) S^*, \quad \infty \right)$$

d) Interval spoľahlivosti pre podiel rozptylov $\dfrac{\sigma_1^2}{\sigma_2^2}$

$$\left\langle \frac{S_X^2}{S_Y^2\, F_{\frac{\alpha}{2}}(n-1,\, m-1)}, \quad \frac{S_X^2}{S_Y^2\, F_{1-\frac{\alpha}{2}}(n-1,\, m-1)} \right\rangle$$

$$\left\langle \frac{S_X^2}{S_Y^2\, F_{\alpha}(n-1,\, m-1)}, \quad \infty \right), \quad \left(-\infty, \quad \frac{S_X^2}{S_Y^2\, F_{1-\alpha}(n-1,\, m-1)} \right\rangle$$

Testovanie štatistických hypotéz

Testy hypotéz o parametroch normálneho rozdelenia

1. Testy hypotéz o parametroch μ a σ^2 normálneho rozdelenia

a) Test hypotézy o μ, rozptyl σ^2 známy

$$\boldsymbol{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

Testovacia štatistika

$$\boldsymbol{U} = \frac{\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0}{\boldsymbol{\sigma}}\sqrt{\boldsymbol{n}}$$

H_0 zamietame ak $|\boldsymbol{U}| \geq \boldsymbol{k}_\alpha$.

$$\boldsymbol{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu} > \boldsymbol{\mu}_0 \quad \text{resp.} \quad \boldsymbol{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu} < \boldsymbol{\mu}_0$$

H_0 zamietame ak $\boldsymbol{U} > \boldsymbol{k}_{2\alpha}$ resp. $\boldsymbol{U} < -\boldsymbol{k}_{2\alpha}$

b) Test hypotézy o μ, rozptyl σ^2 neznámy.

$$\boldsymbol{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

Testovacia štatistika

$$\boldsymbol{T} = \frac{\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0}{\boldsymbol{S}_X}\sqrt{\boldsymbol{n}}$$

H_0 zamietame ak $|\boldsymbol{T}| \geq \boldsymbol{t}_\alpha(\boldsymbol{n}-1)$.

$$\boldsymbol{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu} > \boldsymbol{\mu}_0 \quad \text{resp.} \quad \boldsymbol{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu} < \boldsymbol{\mu}_0$$

H_0 zamietame ak $\boldsymbol{T} > \boldsymbol{t}_{2\alpha}(\boldsymbol{n}-1)$ resp. $\boldsymbol{T} < -\boldsymbol{t}_{2\alpha}(\boldsymbol{n}-1)$.

c) Test hypotézy o σ^2, stredná hodnota μ neznáma

$$\boldsymbol{H}_0 : \boldsymbol{\sigma}^2 = \boldsymbol{\sigma}_0^2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\sigma}^2 \neq \boldsymbol{\sigma}_0^2$$

Testovacia štatistika

$$\boldsymbol{\chi}^2 = \frac{(\boldsymbol{n}-1)\boldsymbol{S}_X^2}{\boldsymbol{\sigma}_0^2}$$

H_0 zamietame ak $\boldsymbol{\chi}^2 \leq \boldsymbol{\chi}^2_{1-\frac{\alpha}{2}}(\boldsymbol{n}-1)$ alebo $\boldsymbol{\chi}^2 \geq \boldsymbol{\chi}^2_{\frac{\alpha}{2}}(\boldsymbol{n}-1)$.

$$\boldsymbol{H}_0 : \boldsymbol{\sigma}^2 = \boldsymbol{\sigma}_0^2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\sigma}^2 > \boldsymbol{\sigma}_0^2 \quad \text{resp.} \quad \boldsymbol{H}_0 : \boldsymbol{\sigma}^2 = \boldsymbol{\sigma}_0^2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\sigma}^2 < \boldsymbol{\sigma}_0^2$$

H_0 zamietame ak $\boldsymbol{\chi}^2 \geq \boldsymbol{\chi}^2_\alpha(\boldsymbol{n}-1)$ resp. $\boldsymbol{\chi}^2 \leq \boldsymbol{\chi}^2_{1-\alpha}(\boldsymbol{n}-1)$.

2. Párový test

$$\boldsymbol{H}_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \Delta \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \Delta$$

Testovacia štatistika

$$\boldsymbol{T} = \frac{\bar{\boldsymbol{X}} - \Delta}{\boldsymbol{S}_X}\sqrt{\boldsymbol{n}}.$$

H_0 zamietame ak $|\boldsymbol{T}| \geq \boldsymbol{t}_\alpha(\boldsymbol{n}-1)$.

3. Test rovnosti parametrov dvoch normálnych rozdelení

a) Test rovnosti rozptylov, stredné hodnoty neznáme. F-test

$$\boldsymbol{H}_0 : \boldsymbol{\sigma}_1^2 = \boldsymbol{\sigma}_2^2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\sigma}_1^2 \neq \boldsymbol{\sigma}_2^2$$

Testovacia štatistika

$$\boldsymbol{F} = \frac{\boldsymbol{S}_X^2}{\boldsymbol{S}_Y^2}$$

H_0 zamietame ak $\boldsymbol{F} \leq \boldsymbol{F}_{1-\frac{\alpha}{2}}(\boldsymbol{n}-1, \boldsymbol{m}-1)$ alebo $\boldsymbol{F} \geq \boldsymbol{F}_{\frac{\alpha}{2}}(\boldsymbol{n}-1, \boldsymbol{m}-1)$.

$$\boldsymbol{H}_0 : \boldsymbol{\sigma}_1^2 = \boldsymbol{\sigma}_2^2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\sigma}_1^2 > \boldsymbol{\sigma}_2^2 \quad \text{resp.} \quad \boldsymbol{H}_0 : \boldsymbol{\sigma}_1^2 = \boldsymbol{\sigma}_2^2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\sigma}_1^2 < \boldsymbol{\sigma}_2^2$$

H_0 zamietame ak $\boldsymbol{F} \geq \boldsymbol{F}_\alpha(\boldsymbol{n}-1, \boldsymbol{m}-1)$ resp. $\boldsymbol{F} \leq \boldsymbol{F}_{1-\alpha}(\boldsymbol{n}-1, \boldsymbol{m}-1))$.

b) Test rovnosti stredných hodnôt, rozptyly rovnaké. Dvojvýberový Studentov test

$$\boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Testovacia štatistika

$$\boldsymbol{T} = \frac{\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\sqrt{\dfrac{(\boldsymbol{n}-1)\boldsymbol{S}_X^2 + (\boldsymbol{m}-1)\boldsymbol{S}_Y^2}{\boldsymbol{n}+\boldsymbol{m}-2}}} \sqrt{\frac{\boldsymbol{nm}}{\boldsymbol{n}+\boldsymbol{m}}}$$

H_0 zamietame ak $|\boldsymbol{T}| \geq \boldsymbol{t}_\alpha(\boldsymbol{n}+\boldsymbol{m}-2)$.

$$\boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_1 > \boldsymbol{\mu}_2 \quad \text{resp.} \quad \boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_1 < \boldsymbol{\mu}_2$$

H_0 zamietame ak $\boldsymbol{T} > \boldsymbol{t}_{2\alpha}(\boldsymbol{n}+\boldsymbol{m}-2)$ resp. $\boldsymbol{T} < -\boldsymbol{t}_{2\alpha}(\boldsymbol{n}+\boldsymbol{m}-2)$.

c) Test rovnosti stredných hodnôt, rozptyly rôzne. Cochranov- Coxov t-test

$$\boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Testovacia štatistika je

$$\boldsymbol{T} = \frac{\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\sqrt{\dfrac{\boldsymbol{S}_X^2}{\boldsymbol{n}} + \dfrac{\boldsymbol{S}_Y^2}{\boldsymbol{m}}}}.$$

H_0 zamietame ak $|\boldsymbol{T}| \geq \boldsymbol{t}_\alpha^*$, kde $t_\alpha^* = \dfrac{\dfrac{S_X^2}{n} t_\alpha(n-1) + \dfrac{S_Y^2}{m} t_\alpha(m-1)}{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}$.

$$\boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_1 > \boldsymbol{\mu}_2 \quad \text{resp.} \quad \boldsymbol{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \textbf{proti} \quad \boldsymbol{H}_1 : \boldsymbol{\mu}_1 < \boldsymbol{\mu}_2$$

H_0 zamietame ak $\boldsymbol{T} > \boldsymbol{t}_{2\alpha}^*$ resp. $\boldsymbol{T} < -\boldsymbol{t}_{2\alpha}^*$.

Regresná a korelačná analýza

Regresná priamka

Systém normálnych rovníc pre priamku

$$\begin{aligned} \beta_0 n &+ \beta_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n Y_i \\ \beta_0 \sum_{i=1}^n x_i &+ \beta_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i Y_i \end{aligned}$$

$$b_0 = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n Y_i - \sum_{i=1}^n x_i \sum_{i=1}^n Y_i x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}, \qquad b_1 = \frac{n \sum_{i=1}^n Y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$b_0 = \bar{Y} - b_1 \bar{x}.$$

$$s_e^2 = \frac{\sum_{i=1}^n Y_i^2 - b_0 \sum_{i=1}^n Y_i - b_1 \sum_{i=1}^n x_i Y_i}{n-2}.$$

Rozptyly odhadov

$$s^2(b_0) = \frac{s_e^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}, \qquad s^2(b_1) = \frac{s_e^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Testy hypotéz o parametroch

$$H_0 : \quad \beta_1 = 0 \quad \text{proti} \quad H_1 : \quad \beta_1 \neq 0$$

Testovacia štatistika

$$T_1 = \frac{(b_1 - 0)}{s(b_1)}$$

H_0 zamietame na hladine významnosti α, ak $|T_1| \geq t_\alpha(n-2)$.

$$H_0 : \quad \beta_0 = 0 \quad \text{proti} \quad H_1 : \quad \beta_0 \neq 0$$

Testovacia štatistika

$$T_0 = \frac{(b_0 - 0)}{s(b_0)}$$

H_0 zamietame na hladine významnosti α, ak $|T_0| \geq t_\alpha(n-2)$.

Intervaly spoľahlivosti

$100.(1-\alpha)\%$ interval spoľahlivosti pre parameter β_1

$$\langle b_1 - t_\alpha(n-2)\, s(b_1); \quad b_1 + t_\alpha(n-2)\, s(b_1)\rangle$$

$100.(1-\alpha)\%$ interval spoľahlivosti pre parameter β_0

$$\langle b_0 - t_\alpha(n-2)\, s(b_0); \quad b_0 + t_\alpha(n-2)\, s(b_0)\rangle$$

Test adekvátnosti regresného modelu, analýza rozptylu pre regresnú priamku

$$S_C = \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$$
$$S_M = b_0 \sum_{i=1}^{n} Y_i + b_1 \sum_{i=1}^{n} x_i\, Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$$
$$S_R = \sum_{i=1}^{n} Y_i^2 - b_0 \sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} x_i\, Y_i$$

Test adekvátnosti regresného modelu

$$H_0: \quad \text{model nie je adekvátny} \quad \text{proti} \quad H_1: \quad \text{model je adekvátny}$$

Testovacie kritérium

$$F = \frac{S_M(n-2)}{S_R}.$$

H_0 zamietame na hladine významnosti α, ak $F \geq F_\alpha(1, n-2)$.

Združené regresné priamky

$$\widehat{Y} - \bar{Y} = r_{XY}\frac{s_y}{s_x}(x - \bar{x})$$
$$\widehat{X} - \bar{x} = r_{XY}\frac{s_x}{s_y}(y - \bar{y})$$

Koeficient korelácie

$$r_{XY} = \frac{\sum\limits_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}}{\sqrt{\left(\sum\limits_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum\limits_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right)}}$$

Index korelácie

$$I_{YX} = \sqrt{\frac{\sum\limits_{i=1}^{n} \widehat{Y}_i^2 - n\bar{Y}^2}{\sum\limits_{i=1}^{n} Y_i^2 - n\bar{Y}^2}}$$