Katedra aplikovanej matematiky, Strojnícka fakulta, UNIZA

Testovanie štatistických hypotéz a korelačná analýza

Pavol ORŠANSKÝ

18. októbra 2023

Testovanie štatistických hypotéz

Pod pojmom **štatistická hypotéza** rozumieme **určité tvrdenie o rozdelení základného štatistického súboru, resp. o jeho parametroch** (v prípade parametrických testov).

Testovanie štatistických hypotéz

Pod pojmom **štatistická hypotéza** rozumieme **určité tvrdenie o rozdelení základného štatistického súboru, resp. o jeho parametroch** (v prípade parametrických testov).

Overovanie pravdivosti takýchto tvrdení na základe vlastností náhodného výberu označujeme **testovanie hypotéz**.

Testovanie štatistických hypotéz

Pod pojmom **štatistická hypotéza** rozumieme **určité tvrdenie o rozdelení základného štatistického súboru, resp. o jeho parametroch** (v prípade parametrických testov).

Overovanie pravdivosti takýchto tvrdení na základe vlastností náhodného výberu označujeme **testovanie hypotéz**.

V ďalšom sa obmedzíme výlučne na parametrické testovanie.

Testovanie štatistických hypotéz

Pod pojmom **štatistická hypotéza** rozumieme **určité tvrdenie o rozdelení základného štatistického súboru, resp. o jeho parametroch** (v prípade parametrických testov).

Overovanie pravdivosti takýchto tvrdení na základe vlastností náhodného výberu označujeme **testovanie hypotéz**.

V ďalšom sa obmedzíme výlučne na parametrické testovanie.

Testovať budeme parametre strednej hodnoty $E(\xi) = \mu$ a rozptylu $D(\xi) = \sigma^2$ (resp. smerodajnej odchýlky σ) náhodnej premennej ξ s normálnym rozdelením $N\left(\mu, \sigma^2\right)$.

Vo všeobecnosti parameter označíme Θ.

Testovanie štatistických hypotéz

Proti sebe postavíme dve disjunktné hypotézy:
nulovú hypotézu

$$H_0 : \Theta = \Theta_0,$$

alternatívnu hypotézu (opačné tvrdenie k nulovej hypotéze)

$$H_1 : \Theta \neq \Theta_0 \qquad \text{(obojstranný test)}.$$

[1] Správnejšia by bola formulácia nulovej hypotézy v tvare $H_0 : \Theta \geq \Theta_0$ pre pravostranný test, resp. $H_0 : \Theta \leq \Theta_0$ pre ľavostranný test, avšak pre jednoduchosť unifikácie ponecháme v pôvodnom tvare.

Testovanie štatistických hypotéz

Proti sebe postavíme dve disjunktné hypotézy:
nulovú hypotézu

$$H_0 : \Theta = \Theta_0,$$

alternatívnu hypotézu (opačné tvrdenie k nulovej hypotéze)

$$H_1 : \Theta \neq \Theta_0 \qquad \text{(obojstranný test)}.$$

v prípade jednostranného testu staviame proti nulovej hypotéze $H_0 : \Theta = \Theta_0$[1] alternatívnu hypotézu tvaru:

$$H_1 : \Theta < \Theta_0 \qquad \text{(pre pravostranný test)},$$
$$H_1 : \Theta > \Theta_0 \qquad \text{(pre ľavostranný test)}.$$

[1] Správnejšia by bola formulácia nulovej hypotézy v tvare $H_0 : \Theta \geq \Theta_0$ pre pravostranný test, resp. $H_0 : \Theta \leq \Theta_0$ pre ľavostranný test, avšak pre jednoduchosť unifikácie ponecháme v pôvodnom tvare.

Parametrické testovanie jedného súboru

K rozhodovaniu použijeme vhodnú funkciu náhodnej veličiny, ktorú nazveme **testovacia štatistika** (kritérium).

Obor hodnôt testovacej štatistiky rozdelíme na dve disjunktné (jedna odporuje druhej) časti:

- **obor prijatia** hypotézy H_0,
- **obor zamietnutia (neprijatia)** hypotézy H_0.

Deliacimi bodmi oboru prijatia a oboru zamietnutia sú tzv. **kritické hodnoty**, ktoré budú zhodné s hodnotami príslušného rozdelenia na danej hladine významnosti.

Parametrické testovanie jedného súboru

Pri testovaní štatistických hypotéz budeme postupovať nasledujúcim spôsobom:

1. určíme nulovú hypotézu H_0 a alternatívnu hypotézu H_1,

Parametrické testovanie jedného súboru

Pri testovaní štatistických hypotéz budeme postupovať nasledujúcim spôsobom:

1. určíme nulovú hypotézu H_0 a alternatívnu hypotézu H_1,
2. vyberieme testovaciu štatistiku,

Parametrické testovanie jedného súboru

Pri testovaní štatistických hypotéz budeme postupovať nasledujúcim spôsobom:

1. určíme nulovú hypotézu H_0 a alternatívnu hypotézu H_1,
2. vyberieme testovaciu štatistiku,
3. určíme hladinu významnosti α a k nej príslušnú oblasť zamietnutia hypotézy H_0,

Parametrické testovanie jedného súboru

Pri testovaní štatistických hypotéz budeme postupovať nasledujúcim spôsobom:

1. určíme nulovú hypotézu H_0 a alternatívnu hypotézu H_1,
2. vyberieme testovaciu štatistiku,
3. určíme hladinu významnosti α a k nej príslušnú oblasť zamietnutia hypotézy H_0,
4. vypočítame hodnotu testovacej štatistiky,

Parametrické testovanie jedného súboru

Pri testovaní štatistických hypotéz budeme postupovať nasledujúcim spôsobom:

1. určíme nulovú hypotézu H_0 a alternatívnu hypotézu H_1,
2. vyberieme testovaciu štatistiku,
3. určíme hladinu významnosti α a k nej príslušnú oblasť zamietnutia hypotézy H_0,
4. vypočítame hodnotu testovacej štatistiky,
5. rozhodnutie o prijatí, resp. zamietnutí hypotézy H_0.

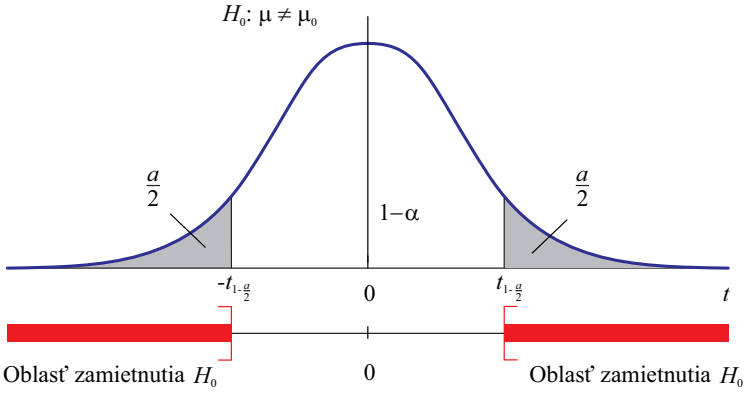
Testovanie parametra strednej hodnoty μ, ak je súbor malý ($n < 30$)

Nulová hypotéza: $H_0 : \mu = \mu_0$.
Testovacia štatistika:

$$T = \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n}, \qquad \text{resp.} \qquad T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n}.$$

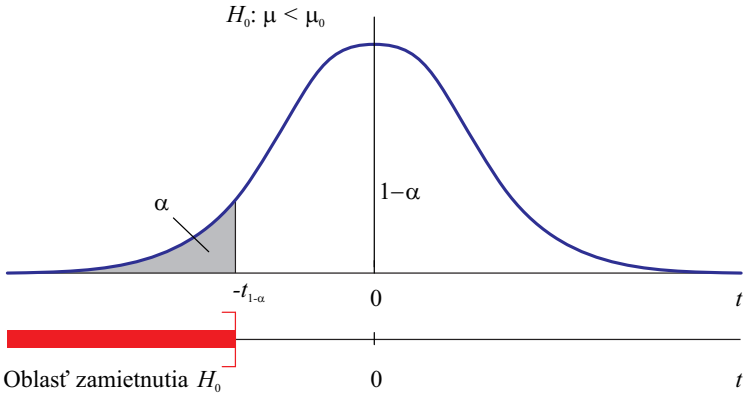
1. alternatívna hypotéza: $H_1 : \mu \neq \mu_0$,
   oblasť zamietnutia H_0: $|T| \geq t_{1-\frac{\alpha}{2}}^{(n-1)}$,
2. alternatívna hypotéza: $H_1 : \mu < \mu_0$,
   oblasť zamietnutia H_0: $T \leq -t_{1-\alpha}^{(n-1)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu > \mu_0$,
   oblasť zamietnutia H_0: $T \geq t_{1-\alpha}^{(n-1)}$ (pravostranný test).

Testovanie parametra strednej hodnoty (obojstranný test)



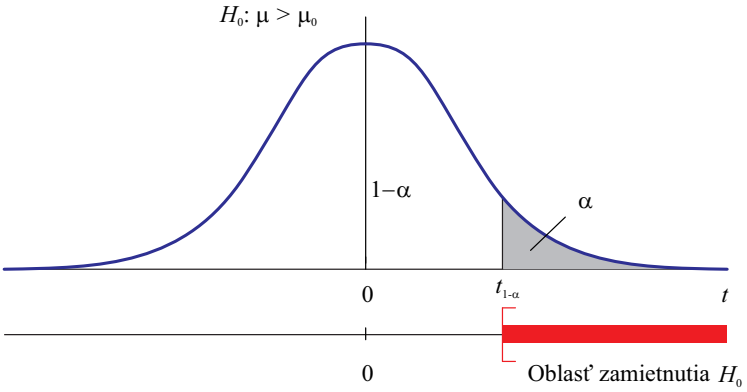
Obr.: Obojstranný test parametra strednej hodnoty

Testovanie parametra strednej hodnoty (ľavostranný test)



Obr.: Ľavostranný test parametra strednej hodnoty

Testovanie parametra strednej hodnoty (pravostranný test)



Obr.: Pravostranný test parametra strednej hodnoty

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Príklad (1.)

Podľa Japonskej národnej agentúry vzrástla priemerná cena pozemkov centrálnej časti Tokia za prvých šesť mesiacov roku 1986 o 49% *. Predpokladajme, že medzinárodná realitná spoločnosť chce overiť, či tvrdenie agentúry je pravdivé alebo nie. Spoločnosť náhodne vybrala* 18 *vlastníkov v centre Tokia, u ktorých boli známe ceny pozemkov na začiatku a v polovici roku 1986. Na základe údajov, ktoré mala k dispozícii zistila, že priemerný vzrast ceny pozemkov u vybraných* 18 *vlastníkov predstavoval za prvý polrok sledovaného roku* 38% *so štandardnou výberovou odchýlkou* 14*.*
Na hladine významnosti $\alpha = 0.01$ *overme pravdivosť tvrdenia agentúry.*

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Riešenie: Nulovú hypotézu a alternatívnu hypotézu môžeme zapísať:

$$H_0 \quad : \quad \mu = 49,$$
$$H_1 \quad : \quad \mu \neq 49.$$

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Riešenie: Nulovú hypotézu a alternatívnu hypotézu môžeme zapísať:

$$\begin{aligned} H_0 &: \mu = 49, \\ H_1 &: \mu \neq 49. \end{aligned}$$

Testovacia štatistika ($n = 18 < 30, \sigma = ? \Rightarrow \sigma \approx S_x = 14$) je rovná

$$T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n} = \frac{38 - 49}{14} \cdot \sqrt{18} = -3.33.$$

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Riešenie: Nulovú hypotézu a alternatívnu hypotézu môžeme zapísať:

$$\begin{aligned} H_0 &: \mu = 49, \\ H_1 &: \mu \neq 49. \end{aligned}$$

Testovacia štatistika ($n = 18 < 30, \sigma = ? \Rightarrow \sigma \approx S_x = 14$) je rovná

$$T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n} = \frac{38 - 49}{14} \cdot \sqrt{18} = -3.33.$$

Kritická hodnota Studentovho rozdelenia je

$$t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{1-\frac{0.01}{2}}^{(17)} = t_{0.995}^{(17)} = 2.898.$$

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Riešenie: Nulovú hypotézu a alternatívnu hypotézu môžeme zapísať:

$$\begin{aligned} H_0 &: \mu = 49, \\ H_1 &: \mu \neq 49. \end{aligned}$$

Testovacia štatistika ($n = 18 < 30, \sigma = ? \Rightarrow \sigma \approx S_x = 14$) je rovná

$$T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n} = \frac{38 - 49}{14} \cdot \sqrt{18} = -3.33.$$

Kritická hodnota Studentovho rozdelenia je

$$t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(17)}_{1-\frac{0.01}{2}} = t^{(17)}_{0.995} = 2.898.$$

Oblasť zamietnutia H_0 je $|T| \geq t^{(n-1)}_{1-\frac{\alpha}{2}}$,

$$|-3.33| \geq 2.898.$$

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Riešenie: Nulovú hypotézu a alternatívnu hypotézu môžeme zapísať:

$$\begin{aligned} H_0 &: \mu = 49, \\ H_1 &: \mu \neq 49. \end{aligned}$$

Testovacia štatistika ($n = 18 < 30, \sigma = ? \Rightarrow \sigma \approx S_x = 14$) je rovná

$$T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n} = \frac{38 - 49}{14} \cdot \sqrt{18} = -3.33.$$

Kritická hodnota Studentovho rozdelenia je

$$t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{1-\frac{0.01}{2}}^{(17)} = t_{0.995}^{(17)} = 2.898.$$

Oblasť zamietnutia H_0 je $|T| \geq t_{1-\frac{\alpha}{2}}^{(n-1)}$,

$$|-3.33| \geq 2.898.$$

H_0 zamietame, lebo -3.33 je v oblasti zamietnutia hypotézy H_0, a teda platí H_1 že cena pozemkov nezvrástla o 49%.♠

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Príklad (2.)

Výrobca uvádza, že priemerná životnosť ním vyrábaných reflektorov je 70 *hodín. Konkurenčná firma sa domnieva, že je v skutočnosti nižšia, preto sa rozhodla dokázať, že výrobcovo tvrdenie nie je správne. Náhodne vybrala* 20 *reflektorov a zistila, že ich priemerná životnosť bola* 67 *hodín a štandardná odchýlka bola* 5 *hodín. Na hladine významnosti* $\alpha = 0.05$ *overme, či výrobcovo tvrdenie je skutočne nesprávne.*

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Príklad (2.)

Výrobca uvádza, že priemerná životnosť ním vyrábaných reflektorov je 70 hodín. Konkurenčná firma sa domnieva, že je v skutočnosti nižšia, preto sa rozhodla dokázať, že výrobcovo tvrdenie nie je správne. Náhodne vybrala 20 reflektorov a zistila, že ich priemerná životnosť bola 67 hodín a štandardná odchýlka bola 5 hodín. Na hladine významnosti $\alpha = 0.05$ overme, či výrobcovo tvrdenie je skutočne nesprávne.

Riešenie: $H_0 : \mu = 70$ proti $H_1 : \mu < 70$.

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Príklad (2.)

Výrobca uvádza, že priemerná životnosť ním vyrábaných reflektorov je 70 hodín. Konkurenčná firma sa domnieva, že je v skutočnosti nižšia, preto sa rozhodla dokázať, že výrobcovo tvrdenie nie je správne. Náhodne vybrala 20 reflektorov a zistila, že ich priemerná životnosť bola 67 hodín a štandardná odchýlka bola 5 hodín. Na hladine významnosti $\alpha = 0.05$ overme, či výrobcovo tvrdenie je skutočne nesprávne.

Riešenie: $H_0 : \mu = 70$ proti $H_1 : \mu < 70$.

$$T = \frac{\overline{x} - \mu_0}{S_x}\sqrt{n} = \frac{67 - 70}{5} \cdot \sqrt{20} = -2.683 < -1.73 = -t_{0.95}^{(19)} = -t_{1-\alpha}^{(n-1)}.$$

Testovanie parametra str. hodnoty μ, ak je súbor malý ($n < 30$) - príklad

Príklad (2.)

Výrobca uvádza, že priemerná životnosť ním vyrábaných reflektorov je 70 hodín. Konkurenčná firma sa domnieva, že je v skutočnosti nižšia, preto sa rozhodla dokázať, že výrobcovo tvrdenie nie je správne. Náhodne vybrala 20 reflektorov a zistila, že ich priemerná životnosť bola 67 hodín a štandardná odchýlka bola 5 hodín. Na hladine významnosti $\alpha = 0.05$ overme, či výrobcovo tvrdenie je skutočne nesprávne.

Riešenie: $H_0 : \mu = 70$ proti $H_1 : \mu < 70$.

$$T = \frac{\overline{x} - \mu_0}{S_x}\sqrt{n} = \frac{67 - 70}{5} \cdot \sqrt{20} = -2.683 < -1.73 = -t_{0.95}^{(19)} = -t_{1-\alpha}^{(n-1)}.$$

na základe toho môžeme hypotézu H_0 zamietnuť, nakoľko $T < -t_{0.95}^{(19)}$, a teda prijímame hypotézu H_1. Čiže domnienka konkurenčnej firmy sa potvrdila a životnosť je nižšia.♠

Testovanie parametra strednej hodnoty μ, ak je súbor veľký ($n > 30$)

Nulová hypotéza: $H_0 : \mu = \mu_0$.
Testovacia štatistika:

$$T = \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n}, \qquad \text{resp.} \qquad T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n}.$$

1. alternatívna hypotéza: $H_1 : \mu \neq \mu_0$,
   oblasť zamietnutia H_0: $|T| \geq u_{1-\frac{\alpha}{2}}$,
2. alternatívna hypotéza: $H_1 : \mu < \mu_0$,
   oblasť zamietnutia H_0: $T \leq -u_{1-\alpha}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu > \mu_0$,
   oblasť zamietnutia H_0: $T \geq u_{1-\alpha}$ (pravostranný test).

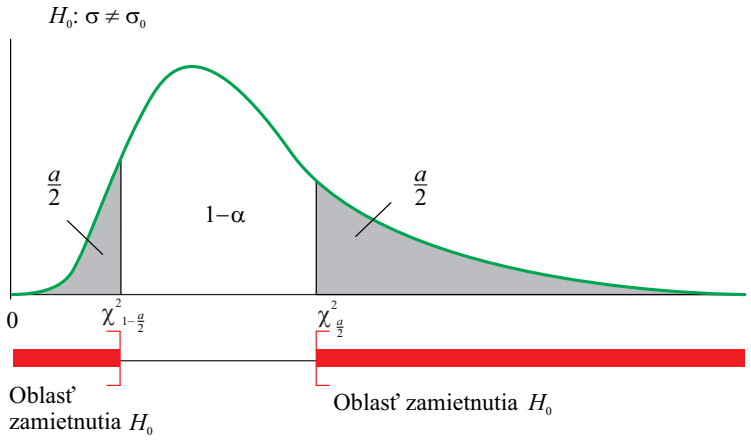
Testovanie parametra rozptylu σ^2

Nulová hypotéza: $H_0 : \sigma = \sigma_0$.
Testovacia štatistika:

$$\boxed{\chi^2 = \frac{(n-1)\, S_x^2}{\sigma_0^2}.}$$

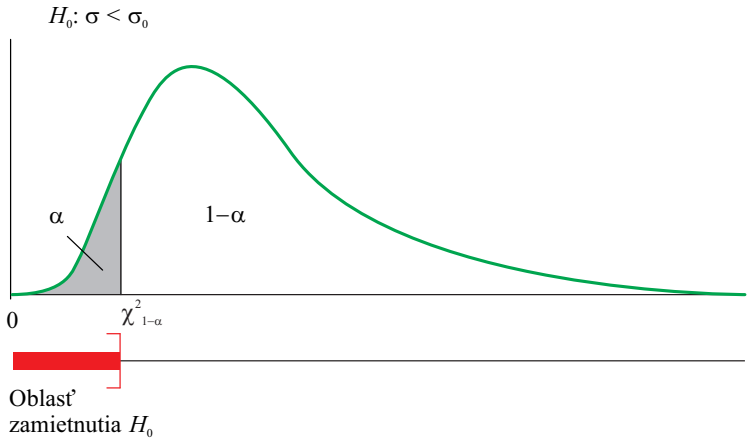
1. alternatívna hypotéza: $H_1 : \sigma \neq \sigma_0$,
   oblasť zamietnutia H_0: $\chi^2 \leq \chi^2_{1-\frac{\alpha}{2};(n-1)}$ alebo $\chi^2 \geq \chi^2_{\frac{\alpha}{2};(n-1)}$,
2. alternatívna hypotéza: $H_1 : \sigma < \sigma_0$,
   oblasť zamietnutia H_0: $\chi^2 \leq \chi^2_{1-\alpha;(n-1)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \sigma > \sigma_0$,
   oblasť zamietnutia H_0: $\chi^2 \geq \chi^2_{\alpha;(n-1)}$ (pravostranný test).

Testovanie parametra rozptylu σ^2 (obojstranný test)



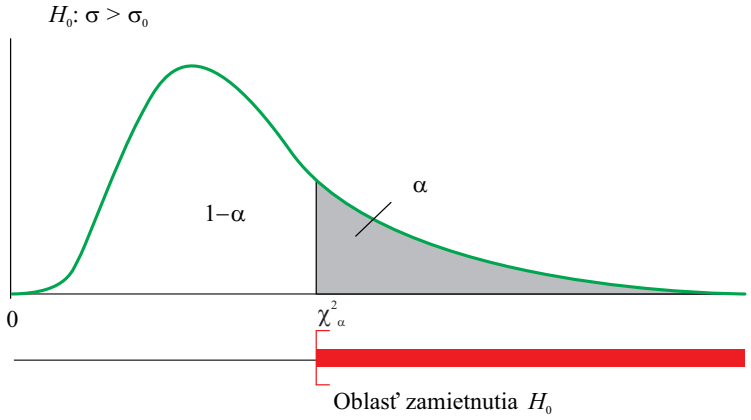
Obr.: Obojstranný test parametra rozptylu σ^2

Testovanie parametra rozptylu σ^2 (ľavostranný test)



Obr.: Ľavostranný test parametra rozptylu σ^2

Testovanie parametra rozptylu σ^2 (pravostranný test)



Obr.: Pravostranný test parametra rozptylu σ^2

Testovanie parametra rozptylu σ^2 - príklad

Príklad (3.)

Štandardná odchýlka obsahu určitej látka v tabletách, ktoré vyrába farmaceutický podnik, nesmie prekročiť hodnotu 0.45 *miligramu. Ak prekročí túto hodnotu, musí sa urobiť korekcia v nastavení výrobnej linky.*
Kontrolór náhodne vybral 25 *tabliet a zistil, že rozptyl obsahu sledovanej látky je* 0.3383. *Aký má urobiť záver, ak pripúšťa pravdepodobnostnú chybu I. druhu* 5% *(hladina významnosti* $\alpha = 0.05$*)?*

Testovanie parametra rozptylu σ^2 - príklad

Riešenie: $n = 25, \quad \alpha = 0.05, \quad \sigma_0 = 0.45, \quad S_x^2 = 0.3383,$
$H_0 : \sigma = 0.45, \quad$ proti $H_1 : \sigma > 0.45.$

Testovanie parametra rozptylu σ^2 - príklad

Riešenie: $n = 25, \qquad \alpha = 0.05, \qquad \sigma_0 = 0.45, \qquad S_x^2 = 0.3383,$
$H_0 : \sigma = 0.45, \qquad$ proti $H_1 : \sigma > 0.45.$
Určíme testovaciu štatistiku

$$\chi^2 = \frac{(n-1)\, S_x^2}{\sigma_0^2} = \frac{(25-1) \cdot 0.3383}{(0.45)^2} = 40.095.$$

Testovanie parametra rozptylu σ^2 - príklad

Riešenie: $n = 25, \quad \alpha = 0.05, \quad \sigma_0 = 0.45, \quad S_x^2 = 0.3383,$
$H_0 : \sigma = 0.45, \quad$ proti $H_1 : \sigma > 0.45.$
Určíme testovaciu štatistiku

$$\chi^2 = \frac{(n-1)\, S_x^2}{\sigma_0^2} = \frac{(25-1) \cdot 0.3383}{(0.45)^2} = 40.095.$$

Kritická hodnota je $\chi^2_{\alpha;(n-1)} = \chi^2_{0.05;(24)} = 36.415.$

Testovanie parametra rozptylu σ^2 - príklad

Riešenie: $n = 25, \quad \alpha = 0.05, \quad \sigma_0 = 0.45, \quad S_x^2 = 0.3383,$
$H_0 : \sigma = 0.45, \quad$ proti $H_1 : \sigma > 0.45.$
Určíme testovaciu štatistiku

$$\chi^2 = \frac{(n-1)\, S_x^2}{\sigma_0^2} = \frac{(25-1) \cdot 0.3383}{(0.45)^2} = 40.095.$$

Kritická hodnota je $\chi^2_{\alpha;(n-1)} = \chi^2_{0.05;(24)} = 36.415.$

$$40.095 > 36.415.$$

Testovanie parametra rozptylu σ^2 - príklad

Riešenie: $n = 25, \quad \alpha = 0.05, \quad \sigma_0 = 0.45, \quad S_x^2 = 0.3383,$
$H_0 : \sigma = 0.45, \quad$ proti $H_1 : \sigma > 0.45.$
Určíme testovaciu štatistiku

$$\chi^2 = \frac{(n-1)\, S_x^2}{\sigma_0^2} = \frac{(25-1)\cdot 0.3383}{(0.45)^2} = 40.095.$$

Kritická hodnota je $\chi^2_{\alpha;(n-1)} = \chi^2_{0.05;(24)} = 36.415.$

$$40.095 > 36.415.$$

Hodnota testovacej štatistiky sa nachádza v oblasti zamietnutia hypotézy H_0, na hladine významnosti 0.05 môžeme tvrdiť, že variabilita obsahu danej látky v tabletkách je vyššia ako prípustná norma hodnoty, a preto je nutné vykonať korekciu nastavenia výrobnej linky.♠

Porovnávanie dvoch súborov

Často sa v praxi stretávame so situáciou, keď chceme porovnať dva súbory. Tým máme na mysli **porovnať parameter strednej hodnoty** μ týchto dvoch súborov, t. j. **či je jeden väčší alebo menší ako druhý** resp. **či sa rovnajú**. Zisťujeme, či sú tieto súbory reprezentatami toho istého normálneho rozdelenia pravdepodobnosti $N(\mu, \sigma)$, alebo nie.

Pri porovnávaní priemerov dvoch súborov môžeme testovať niekoľko hypotéz. Zaujímať nás môže, či sa priemery rovnajú alebo nie (obojstranný test), alebo či jeden menší resp. väčší ako druhý (jednostranný test).

Pre uvedené situácie môžeme sformulovať nasledovnú nulovú hypotézu:

$$H_0 : \mu_1 = \mu_2$$

proti alternatívnej hypotéze

$$H_1 : \mu_1 \neq \mu_2, \qquad (\textbf{obojstranný test}).$$

resp.

$$H_1 : \mu_1 < \mu_2, \text{ alebo, } H_1 : \mu_1 > \mu_2 \qquad (\textbf{jednostranné testy}).$$

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory veľké, t. j. $n = n_1 + n_2 > 30$, a rozptyly poznáme

Nulová hypotéza: $H_0 : \mu_1 = \mu_2$.

Testovacia štatisika:

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2$,
   oblasť zamietnutia H_0: $|U| \geq u_{1-\frac{\alpha}{2}}$,
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2$,
   oblasť zamietnutia H_0: $U \leq -u_{1-\alpha}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2$,
   oblasť zamietnutia H_0: $U \geq u_{1-\alpha}$ (pravostranný test).

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly poznáme

Nulová hypotéza: $H_0 : \mu_1 = \mu_2$,

Testovacia štatistika:

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2$,
   oblasť zamietnutia H_0: $|U| \geq t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$,
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2$,
   oblasť zamietnutia H_0: $U \leq -t_{1-\alpha}^{(n_1+n_2-2)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2$,
   oblasť zamietnutia H_0: $U \geq t_{1-\alpha}^{(n_1+n_2-2)}$ (pravostranný test).

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory veľké, t. j. $n = n_1 + n_2 > 30$, a rozptyly nepoznáme

Nulová hypotéza: $H_0 : \mu_1 = \mu_2,$

Testovacia štatistika:

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S_{x_1}^2}{n_1} + \frac{S_{x_2}^2}{n_2}}}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2,$
   oblasť zamietnutia H_0: $|U| \geq u_{1-\frac{\alpha}{2}},$
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2,$
   oblasť zamietnutia H_0: $U \leq -u_{1-\alpha}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2,$
   oblasť zamietnutia H_0: $U \geq u_{1-\alpha}$ (pravostranný test).

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory veľké, t. j. $n = n_1 + n_2 > 30$, a rozptyly nepoznáme - príklad

Príklad (4.)

Všeobecne u ľudí s vyššími príjmami a výdavkami prevažovala tendencia vlastniť kartu American Express, kým ľudia s nižšími príjmami a výdavkami využívali viac VISA karty. Z tohto dôvodu VISA zintenzívnila svoje úsilie viac preniknúť i do skupiny obyvateľstva s vyššími príjmami a pomocou reklám v časopisoch a televízií sa snažila vytvoriť u ľudí väčší dojem. Po určitom čase spoločnosť urobila prieskum, v ktorom náhodne vybrala 1200 *držiteľov kariet Preffered VISA a zistila, že ich priemerné mesačné platby boli* 452*$ s výberovou štandardnou odchýlkou* 212*$. Nezávisle od tohto výberu náhodne vybrala* 800 *vlastníkov kariet Gold Card, ktorých priemerné mesačné platby predstavovali* 523*$ s výberovou štandardnou odchýlkou* 185*$. Držitelia oboch kariet boli z prieskumu vylúčení. Potvrdzujú výsledky výberového zisťovania rozdiel výšok platieb realizovaných kartami Gold Card a VISA Prefferd, overme tento predpoklad na hladine významnosti* 0.01*.*

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory veľké, t. j. $n = n_1 + n_2 > 30$, a rozptyly nepoznáme - príklad

Riešenie: Sformulujme hypotézy

$$\begin{aligned} H_0 &: \mu_1 = \mu_2, \\ H_1 &: \mu_1 \neq \mu_2. \end{aligned}$$

Testovacia štatistika je

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S_{x_1}^2}{n_1} + \frac{S_{x_2}^2}{n_2}}} = \frac{452 - 523}{\sqrt{\frac{212^2}{1200} + \frac{185^2}{800}}} = -7.927.$$

Pre $\alpha = 0.01$ je $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$.

Oblasť zamietnutia H_0 je $|U| \geq u_{1-\frac{\alpha}{2}}$, t. j. $7.927 \geq 2.58$. Vidíme, že testovacia štatistika do nej patrí. Na základe čoho môžeme tvrdiť, že medzi priemernými platbami realizovanými prostredníctvom uvedených dvoch kreditných kariet sú štatisticky významné rozdiely.♠

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly nepoznáme

Ak štandardnú odchýlku nepoznáme, porovnávanie μ_1, μ_2 pomocou nezávislých náhodných výberov malého rozsahu vyžaduje okrem nezávislosti výberov a normality rozdelenia základných súborov aj ďalšiu podmienku a to, že rozptyly oboch základných súborov sú si rovné ($\sigma_1^2 = \sigma_2^2$). Označme tento spoločný rozptyl oboch základných súborov σ^2. Jeho hodnotu samozrejme taktiež nepoznáme, ale vieme ho odhadnúť pomocou **spoločného výberové rozptylu** S_p^2 výberových rozptylov. Vzťah pre výpočet spoločného výberové rozptylu má tvar

$$S_p^2 = \frac{(n_1 - 1) \cdot S_{x_1}^2 + (n_2 - 1) \cdot S_{x_2}^2}{n_1 + n_2 - 2}.$$

Odhad štandardnej chyby rozdielu priemerov $(\overline{x}_1 - \overline{x}_2)$ je daný vzťahom

$$S_{(\overline{x}_1 - \overline{x}_2)} = \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}.$$

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly nepoznáme

Testovacia štatistika pre test zhody priemerov dvoch základných súborov, za predpokladu rovnosti ich rozptylov, pri malých výberoch súborov, pre nulovú hypotézu $H_0 : \mu_1 = \mu_2$, potom je

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2$,
   oblasť zamietnutia H_0: $|U| \geq t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$,
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2$,
   oblasť zamietnutia H_0: $U \leq -t_{1-\alpha}^{(n_1+n_2-2)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2$,
   oblasť zamietnutia H_0: $U \geq t_{1-\alpha}^{(n_1+n_2-2)}$ (pravostranný test).

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov, ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly nepoznáme - príklad

Príklad (5.)

Výrobca prehrávačov kompaktných diskov chce zistiť, či ním navrhované malé zníženie cien jeho výrobkov je dostatočné pre zvýšenie objemu ich predaja. Náhodným výberom údajov o 14 týždenných tržbách v jednom obchode pred znížením cien zistil, že priemerné týždenná tržba bola 39 600 *korún so štandardnou odchýlkou* 5 060 *korún. Náhodným výberom 11 týždenných tržieb za jeho výrobky po znížení ich cien zistil, že priemerná týždenná tŕžba bola* 41 200 *korún so štandardnou odchýlkou* 4 010 *korún. Dokazujú uvedené údaje, že malé zníženie cien je dostatočné pre zvýšenie predaja CD prehrávačov, ak použijeme hladinu významnosti* 0.05*?*

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly nepoznáme - príklad

Riešenie: Ide o ľavostranný test, kde nulová a alternatívna hypotéza sú formulované

$$\begin{aligned} H_0 &: \mu_1 = \mu_2, \\ H_1 &: \mu_1 < \mu_2. \end{aligned}$$

Testovacia štatistika je

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{39600 - 41200}{\sqrt{S_p^2 \left(\frac{1}{14} + \frac{1}{11} \right)}} = \dots$$

kde spoločný výberový rozptyl je rovný

$$S_p^2 = \frac{13 \cdot 5060^2 + 10 \cdot 4010^2}{23} = 2.146 \times 10^7.$$

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly nepoznáme - príklad

Po dosadení dostávame

$$U = \ldots = \frac{39600 - 41200}{\sqrt{2.146 \times 10^7 \cdot \left(\frac{1}{14} + \frac{1}{11}\right)}} = -0.857.$$

Oblasť zamietnutia nulovej hypotézy H_0 na hladine významnosti $\alpha = 0.05$ je

$$U \leq -t_{1-\alpha}^{(n_1+n_2-2)} = -t_{0.95}^{(14+11-2)} = -t_{0.95}^{(23)} = -1.714.$$

Hodnota testovacej štatistiky sa nachádza v oblasti prijatia hypotézy H_0, preto môžeme na hladine významnosti $\alpha = 0.05$ tvrdiť, že výrobcom navrhnuté zníženie cien CD prehrávačov neprinieslo vzrast objemu predaja.♠

Korelačná analýza - Lineárny regresný model

Koreláciou rozumieme vzájomný lineárny vzťah (závislosť) dvoch náhodných premenných X a Y[2]. Tento vzťah môže byť priamy, t. j. s rastúcimi hodnotami jednej premennej rastú hodnoty druhej premennej, resp. nepriamy, t. j. s rastúcimi hodnotami jednej premennej klesajú hodnoty druhej.

Zadefinujeme si:

- **koeficient kovariancie** - určujúci mieru priamej resp. nepriamej lineárnej závislosti
- **koeficient korelácie** - určujúci mieru vzájomnej lineárnej závislosti
- **koeficient determinácie** - určujúci percentil vzájomnej lineárnej závislosti

[2]Nateraz upúšťame od označenia náhodných premenných gréckymi písmenami ξ a namiesto označenia $\xi_1, \xi_2, \ldots$ budeme v ďalšom používať označenie $X, Y, \ldots$.

Koeficient kovariancie

$$\boxed{\mathrm{cov}xy = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y}) = \overline{x \cdot y} - \overline{x} \cdot \overline{y}.}$$

Kladná (resp. záporná) hodnota kovariancie, indikuje priamy (resp. nepriamy) lineárny vzťah medzi premennými X a Y.

Poznámka

Platí, že

$$\mathrm{cov}xy = E(XY) - E(X) \cdot E(Y).$$

Poznámka

Okrem toho tiež platí, že

$$\mathrm{cov}xx = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}) \cdot (x_i - \overline{x}) = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2 = D(x) = \sigma_x^2.$$

Koeficient korelácie

Sila vzájomnej lineárnej závislosti premenných X a Y meraná **koeficientom korelácie** súboru r_{XY} je definovaná

$$r_{XY} = \frac{\mathrm{cov}xy}{\sigma_x \cdot \sigma_y}.$$

Použitím odhadov dostávame

$$r_{XY} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2} \cdot \sqrt{\overline{y^2} - \overline{y}^2}},$$

a po úplnom vyjadrení bude mať vzťah, tzv. **Pearsonov koeficient korelácie**

$$r_{XY} = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}.$$

Koeficient determinácie

Stupeň príčinnej závislosti premennej Y od premennej X vyjadruje **koeficient determinácie**, definovaný ako druhá mocnina koeficientu korelácie r. Vo výberovom súbore ho označujeme r^2.

Interpretácia koeficienta determinácie vychádza z analýzy variability (rozptylu) závisle premennej Y, ktorú by mala do značnej miery vysvetliť variabilita nezávisle premennej X za predpokladu, že od nej lineárne závisí veľkosť hodnôt Y.

Ak napr. $r = 0.7$, potom $r^2 = 0.49$, čo znamená, že iba 49% variability premennej Y sa dá vysvetliť lineárnym vzťahom s premennou X (regresnou priamkou). Pretože 51% variability premennej Y zostalo nevysvetlenej lineárnym vzťahom s premennou X je zrejmé, že model bol zvolený nevhodne (namiesto lineárnej závislosti sa mala uvažovať nelineárna závislosť).

Korelačná analýza - príklad

Príklad (6.)

Pracovník personálneho oddelenia určitého podniku má domnienku, že existuje vzťah medzi počtom dní absencie v práci a vekom pracovníka. Náhodne vyberie pracovné záznamy 10 pracovníkov a získa údaje o ich veku v rokoch (náhodná premenná X v rokoch) a počte dní, v ktorých nenastúpili do práce počas kalendárneho roka (náhodná premenná Y).
Údaje sú uvedené v tabuľke

x_i :	27	61	37	23	46	58	29	36	64	40
y_i :	15	6	10	18	9	7	14	11	5	8

.

Za predpokladu, že medzi počtom dní absencie a vekom pracovníka je lineárna závislosť, posúďte, či je priama alebo nepriama.
Vypočítajte koeficient korelácie a koeficient determinácie.

Korelačná analýza - príklad

Medzivýsledky získame z tabuľky, ktorú mierne modifikujeme

n	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	27	15	729	225	405
2	61	6	3721	36	366
3	37	10	1690	100	370
4	23	18	529	324	414
5	46	9	2116	81	414
6	58	7	3364	49	406
7	29	14	841	196	406
8	36	11	1296	121	396
9	64	5	4096	25	320
10	40	8	1600	64	320
$\sum$	421	103	19661	1221	3817

Korelačná analýza - príklad

Pre výpočet kovariancie je najvhodnejšie použiť vzťah

$$\begin{aligned} \mathrm{cov}xy &= \overline{x \cdot y} - \overline{x} \cdot \overline{y} = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{n} - \frac{\sum_{i=1}^{n} x_i}{n} \cdot \frac{\sum_{i=1}^{n} y_i}{n} \\ &= \frac{3817}{10} - \frac{421}{10} \cdot \frac{103}{10} = -\frac{5193}{100} = -51.93. \end{aligned}$$

Medzi počtom dní absencie v roku a vekom pracovníka je nepriama lineárna závislosť (s rastúcim vekom počet dní v roku, v ktorých pracovník nenastúpi do práce bez udania dôvodu, klesá).

Korelačná analýza - príklad

Dosadením do vzťahu

$$r_{XY} = \frac{\mathrm{cov}xy}{\sigma_x \cdot \sigma_y} \approx \frac{-51.93}{13.917 \cdot 4.001} = -0.932\,62.$$

kde σ_x a σ_x sme odhadli ako

$$\sigma_x^2 = \overline{x \cdot x} - \overline{x} \cdot \overline{x} = \overline{x^2} - \overline{x}^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2 = 193.69$$

a teda $\sigma_x = \sqrt{193.69} = 13.917$,

$$\sigma_y^2 = \overline{y^2} - \overline{y}^2 = \frac{\sum_{i=1}^n y_i^2}{n} - \left(\frac{\sum_{i=1}^n y_i}{n}\right)^2 = \frac{1221}{10} - \left(\frac{103}{10}\right)^2 = 16.01,$$

a teda $\sigma_y^2 = \sqrt{16.01} = 4.001$.

Korelačná analýza - príklad

Ďalšia možnosť je dosadiť priamo do vzťahu Pearsonovho koeficientu korelácie

$$
\begin{aligned}
r_{XY} &= \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left(n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right) \cdot \left(n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}} \\
&= \frac{10 \cdot 3817 - 421 \cdot 103}{\sqrt{(10 \cdot 19661 - 421^2) \cdot (10 \cdot 1221 - 103^2)}} = -0.932\,54.
\end{aligned}
$$

Koeficient korelácie $r = -0.93$ interpretujeme ako vysokú nepriamu lineárnu závislosť medzi počtom dní absencie v roku a vekom pracovníka. Koeficient determinácie $r^2 = (-0.93)^2 = 0.8649$ znamená, že 87% variability počtu dní absencie v roku je vysvetlená vplyvom veku pracovníka a 13% variability počtu dní absencie v roku možno vysvetliť inými príčinami ako je lineárnosť medzi premennými X, Y.

Zhrnutie I.

Testovanie parametra strednej hodnoty ak je súbor malý ($n < 30$)

Nulová hypotéza: $H_0 : \mu = \mu_0$.

Testovacia štatistika:

$$T = \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n}, \qquad \text{resp.} \qquad T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n}.$$

1. alternatívna hypotéza: $H_1 : \mu \neq \mu_0$,
   oblasť zamietnutia H_0: $|T| \geq t_{1-\frac{\alpha}{2}}^{(n-1)}$,
2. alternatívna hypotéza: $H_1 : \mu < \mu_0$,
   oblasť zamietnutia H_0: $T \leq -t_{1-\alpha}^{(n-1)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu > \mu_0$,
   oblasť zamietnutia H_0: $T \geq t_{1-\alpha}^{(n-1)}$ (pravostranný test).

Zhrnutie II.

Testovanie parametra strednej hodnoty ak je súbor veľký ($n > 30$)

Nulová hypotéza: $H_0 : \mu = \mu_0$.

Testovacia štatistika:

$$T = \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n}, \qquad \text{resp.} \qquad T = \frac{\overline{x} - \mu_0}{S_x} \cdot \sqrt{n}.$$

1. alternatívna hypotéza: $H_1 : \mu \neq \mu_0$,
   oblasť zamietnutia H_0: $|T| \geq u_{1-\frac{\alpha}{2}}$,
2. alternatívna hypotéza: $H_1 : \mu < \mu_0$,
   oblasť zamietnutia H_0: $T \leq -u_{1-\alpha}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu > \mu_0$,
   oblasť zamietnutia H_0: $T \geq u_{1-\alpha}$ (pravostranný test).

Zhrnutie III.

Testovanie parametra rozptylu

Nulová hypotéza: $H_0 : \sigma = \sigma_0$.

Testovacia štatistika:

$$\chi^2 = \frac{(n-1)\, S_x^2}{\sigma_0^2}.$$

1. alternatívna hypotéza: $H_1 : \sigma \neq \sigma_0$,
   oblasť zamietnutia H_0: $\chi^2 \leq \chi^2_{1-\frac{\alpha}{2};(n-1)}$ alebo $\chi^2 \geq \chi^2_{\frac{\alpha}{2};(n-1)}$,
2. alternatívna hypotéza: $H_1 : \sigma < \sigma_0$,
   oblasť zamietnutia H_0: $\chi^2 \leq \chi^2_{1-\alpha;(n-1)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \sigma > \sigma_0$,
   oblasť zamietnutia H_0: $\chi^2 \geq \chi^2_{\alpha;(n-1)}$ (pravostranný test).

Zhrnutie IV.

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov ak sú súbory veľké, t. j. $n = n_1 + n_2 > 30$, a rozptyly poznáme

Nulová hypotéza: $H_0 : \mu_1 = \mu_2$.

Testovacia štatisika:

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2$,
   oblasť zamietnutia H_0: $|U| \geq u_{1-\frac{\alpha}{2}}$,
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2$,
   oblasť zamietnutia H_0: $U \leq -u_{1-\alpha}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2$,
   oblasť zamietnutia H_0: $U \geq u_{1-\alpha}$ (pravostranný test).

Zhrnutie V.

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly poznáme

Nulová hypotéza: $H_0 : \mu_1 = \mu_2$,

Testovacia štatistika:

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2$,
   oblasť zamietnutia H_0: $|U| \geq t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$,
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2$,
   oblasť zamietnutia H_0: $U \leq -t_{1-\alpha}^{(n_1+n_2-2)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2$,
   oblasť zamietnutia H_0: $U \geq t_{1-\alpha}^{(n_1+n_2-2)}$ (pravostranný test).

Zhrnutie VI.

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov ak sú súbory veľké, t. j. $n = n_1 + n_2 > 30$, a rozptyly nepoznáme

Nulová hypotéza: $H_0 : \mu_1 = \mu_2,$

Testovacia štatistika:

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S_{x_1}^2}{n_1} + \frac{S_{x_2}^2}{n_2}}}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2,$
   oblasť zamietnutia H_0: $|U| \geq u_{1-\frac{\alpha}{2}},$
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2,$
   oblasť zamietnutia H_0: $U \leq -u_{1-\alpha}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2,$
   oblasť zamietnutia H_0: $U \geq u_{1-\alpha}$ (pravostranný test).

Zhrnutie VII.

Porovnávanie dvoch súborov - test o zhode priemerov dvoch súborov ak sú súbory malé, t. j. $n = n_1 + n_2 < 30$, a rozptyly nepoznáme

Nulová hypotéza $H_0 : \mu_1 = \mu_2$,

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \qquad \text{kde} \qquad S_p^2 = \frac{(n_1 - 1) \cdot S_{x_1}^2 + (n_2 - 1) \cdot S_{x_2}^2}{n_1 + n_2 - 2}.$$

1. alternatívna hypotéza: $H_1 : \mu_1 \neq \mu_2$,
   oblasť zamietnutia H_0: $|U| \geq t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$,
2. alternatívna hypotéza: $H_1 : \mu_1 < \mu_2$,
   oblasť zamietnutia H_0: $U \leq -t_{1-\alpha}^{(n_1+n_2-2)}$ (ľavostranný test),
3. alternatívna hypotéza: $H_1 : \mu_1 > \mu_2$,
   oblasť zamietnutia H_0: $U \geq t_{1-\alpha}^{(n_1+n_2-2)}$ (pravostranný test).

Zhrnutie VIII.

Koeficient kovariancie

$$\mathrm{cov}xy = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y}) = \overline{x \cdot y} - \overline{x} \cdot \overline{y}.$$

Koeficient korelácie

$$r_{XY} = \frac{\mathrm{cov}xy}{\sigma_x \cdot \sigma_y}.$$

Pearsonov koeficient korelácie

$$r_{XY} = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}.$$