Statistické testování

Uvedeme obecný postup statistického testování:

A. Postup testování statistické hypotézy

1. Formulace nulové H_0 a alternativní hypotézy H_A.
2. Volba hladiny významnosti α.
3. Volba testační statistiky, např. t.
4. Určení kritického oboru testové charakteristiky.
5. Vyčíslení testační statistiky a jejích kvantilů.
6. Rozhodnutí, zda
   a) Zamítnout hypotézu H_0 a přijmout H_A, jestliže testační statistika padne do kritického oboru,
   b) Nezamítnout hypotézu H_0, jestliže testační statistika nepadne do kritického oboru.

Výsledek testování:

a) Zamítnutí hypotézy H_0 neznamená, že testovaná nulová hypotéza *neplatí*, ale znamená, že její platnosti nevěříme, protože výsledek testu poskytl objektivní důvod. V dalším pak budeme uvažovat, že H_0 neplatí a H_A platí.

b) Nezamítneme-li hypotézu H_0, neznamená to její přijetí. Výsledek testu neukázal tak velkou neshodu mezi zjištěnou skutečností a testovanou hypotézou, která by dala dostatečný důvod k zamítnutí hypotézy.

Dva případy chybného rozhodnutí při testování:

a) Testační statistika padne mimo obor přijetí nulové H_0 hypotézy O_p, tj. mimo interval

$$u_{\alpha/2} \le u_s \le u_{1-\alpha/2}$$

a hypotéza H_0 přitom platí. Platí-li H_0, je pravděpodobnost padnutí u_s mimo obor O_p rovna právě hladině významnosti α. Velikost α určuje velikost *chyby I. druhu*, tj. nesprávného zamítnutí správné nulové hypotézy H_0.

b) Testační statistika padne do oboru, O_P, tj. mimo interval

$$u_s < u_{\alpha/2} \quad \text{resp.} \quad u_s > u_{1-\alpha/2}.$$

a přitom platí alternativní hypotéza H_A. Pravděpodobnost, že u_s padne do oboru přijetí O_p, i když H_0 neplatí, představuje velikost *chyby II. druhu*, ß.

B. Testy střední hodnoty ("testy správnosti")

a) **100(1 - α)%ní interval spolehlivosti:** vypočteme intervalový odhad parametru μ (tj. polohy či rozptýlení). Padne-li zadaná hodnota μ_0 parametru μ do tohoto intervalu, nezamítá se hypotéza $H_0: \mu = \mu_0$. Padne-li μ_0 mimo tento interval, zamítá se H_0.

b) **Studentův *t*-test:** ze základního souboru s rozdělením $N(\mu, \sigma^2)$ provedeme náhodný výběr rozsahu n a vypočteme výběrový průměr $\bar{x}$ a směrodatnou odchylku s. Jako testovou statistiku zvolíme náhodnou veličinu

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{n}$$

Kritické obory testů polohy hypotézy H_0: $\mu = \mu_0$ proti různým alternativám H_A pro hladinu významnosti α jsou uvedeny v tabulce. Hraniční body kritického oboru představují 100α%ní kvantily známých rozdělení. Místo formálního testování, zda jsou tyto kvantily větší než testové statistiky, je možné přímo vyčíslit velikost pravděpodobnosti (1 - α) (u oboustranného testu (1 - α/2)).

Nulová hypotéza H_0	**Alternativní hypotéza H_A**	**Testační charakteristika**	**Kritický obor**
	$\mu > \mu_0$		$t \geq t_{(1-\alpha)}(n-1)$
$\mu = \mu_0$	$\mu < \mu_0$	$t = (x - \mu_0)\sqrt{n}/s$	$t < t_{\alpha}(n-1)$
	$\mu \neq \mu_0$		$\|t\| \geq t_{(1-\alpha/2)}(n-1)$

C. Testy shody středních hodnot ("testy shodnosti")

Porovnání dvou výběrů *{x_i}*, $i = 1, ..., n_1$, a *{y_j}*, $j = 1, ..., n_2$, patří k častým úlohám v přírodních i technických vědách, a to při

(a) porovnání výsledků z různých instrumentálních metod nebo laboratoře,

(b) ověřování nutnosti dělení heterogenních výběrů do homogenních podskupin,

(c) hodnocení rozdílu mezi rozličnými materiály a přístroji.

Někdy lze tuto úlohu převést na testování jednoho výběru. To je totiž případ, kdy mezi prvky obou výběrů existuje jistá logická vazba. Představují-li prvky x_i vlastnosti před úpravou materiálu a prvky y_i stejné vlastnosti po úpravě materiálu ***těchže*** vzorků ($n_1 = n_2$), lze utvořit jednorozměrný výběr, $D_i = x_i - y_i$, pro který lze užít klasickou statistickou analýzu. Pokud se střední hodnota μ_D významně neliší od nuly, znamená to, že $\mu_x = \mu_y$ a efekt zpracování materiálu není pro sledovanou vlastnost statisticky významný (t.zv. *párový test)*. V obecnějším případě dvou výběrů lze zjistit, zda pocházejí ze stejného rozdělení pravděpodobnosti a zda se neliší v parametrech polohy a rozptýlení.

Postup testu shodnosti středních hodnot dvou souborů:

1. Ověření normálního rozdělení obou souborů: testy a statistické diagnostiky k ověření předpokladů o výběru.

2. Test shody rozptylů:

2.1 Klasický Fisher-Snedecorovým *F*-test,

2.2 Modifikovaný Fisher-Snedecorův *F*-test,

2.3 Robustní Jackknife test F_J.

3. Test shody středních hodnot dvou souborů:

3.1 Klasický Studentův *t*-test T_1 pro homoskedasticitu,

3.2 Klasický Studentův *t*-test T_2 pro heteroskedasticitu,

3.3 Modifikovaný Studentův *t*-test T_3 pro výběry, odchýlené od normálního rozdělení.

3.4 Robustní Jackknife test polohy T_4 pro homoskedasticitu,

3.5 Robustní Jackknife test polohy T_5 pro heteroskedasticitu.

1. krok: **Ověření normálního rozdělení obou výběrů**

Klasické testy vycházejí z předpokladů:

a) výběry *{x_i}*, $i = 1, ..., n_1$, a *{y_j}*, $j = 1, ..., n_2$ jsou vzájemně nezávislé;

b) rozdělení obou výběrů je normální, $x_i \sim N(\mu_x, \sigma_x^2)$ a $y_j \sim N(\mu_y, \sigma_y^2)$.

Existuje řada různých metod, které jsou použitelné i v případech, kdy jsou tyto dva předpoklady narušeny. Před vlastní statistickou analýzou je výhodné vyšetřit nejprve metodami průzkumové analýzy chování obou výběrů.

2. krok: **Testy shody rozptylů**

(a) Klasický *F*-test: umožňuje ověření nulové hypotézy H_0: $\sigma_x^2 = \sigma_y^2$ proti alternativní H_A: $\sigma_x^2 \neq \sigma_y^2$. Vychází se z předpokladu, že oba výběry jsou nezávislé a pocházejí z normálního rozdělení. Testovací kritérium má tvar

$$F = \max \left\| \frac{s_x^2}{s_y^2}, \frac{s_y^2}{s_x^2} \right\| .$$

Platí-li hypotéza H_0 a $s_x^2 > s_y^2$, má *F* kritérium *F*-rozdělení s $\nu_1 = n_1 - 1$ a $\nu_2 = n_2 - 1$ stupni volnosti. V opačném případě se pořadí stupňů volnosti zamění. Je-li $F > F_{1-\alpha}(\nu_1, \nu_2)$, je nulová hypotéza H_0 o shodnosti rozptylů zamítnuta.

(b) Modifikovaný *F*-test: předchozí klasický F-test je značně citlivý na předpoklad normality. Mají-li obě výběrová rozdělení jinou špičatost než odpovídá normálnímu, je třeba užít kvantil $F_{1-\alpha}(\nu_1, \nu_2)$ se stupni volnosti ν_1 a ν_2, vyčíslenými podle vztahů

$$\nu_1 = \frac{n_1 - 1}{1 + \frac{\hat{g}_{2c}}{2}}, \quad \nu_2 = \frac{n_2 - 1}{1 + \frac{\hat{g}_{2c}}{2}}$$

$$\text{kde} \quad \hat{g}_{2c} = \frac{2(n_1 + n_2)\left[\sum_{i=1}^{n_1}(x_i - \bar{x})^4 + \sum_{i=1}^{n_2}(y_i - \bar{y})^4\right]}{\left[\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{i=1}^{n_2}(y_i - \bar{y})^2\right]^2} - 3$$

(c) Robustní Jackknife test: jsou-li v datech navíc odlehlé hodnoty, jeví se užitečný robustní Jackknife test. Testovací kritérium má tvar

$$F_J = \frac{n_1(\bar{z}_1 - \bar{z})^2 + n_2(\bar{z}_2 - \bar{z})^2}{\dfrac{\sum_{i=1}^{n_1}(z_{1i} - \bar{z}_1)^2 + \sum_{i=1}^{n_2}(z_{2i} - \bar{z}_2)^2}{n_1 + n_2 - 2}}$$

kde $$\bar{z} = \frac{n_1\bar{z}_1 + n_2\bar{z}_2}{n_1 + n_2}, \quad \bar{z}_j = \frac{\sum_{i=1}^{n_j} z_{ji}}{n_j}, \qquad j = 1, 2$$

Veličiny z_{1i} se počítají podle vztahu $z_{1i} = n_1 \ln s_x^2 - (n_1 - 1)\ln s_{1(i)}^2$,

kde $$s_{1(i)}^2 = \frac{1}{n_1 - 2}\sum_{j \neq i}^{n_1}(x_j - \bar{x}_{(i)})^2 .$$

Ve vztahu se vyskytuje průměr s vynechanou i-tou hodnotou, pro který platí

$$\bar{x}_{(i)} = \frac{1}{n_1 - 1}\sum_{j \neq i}^{n_1} x_j .$$

Při výpočtu z_{2i} se ve výše uvedených vztazích dosazují hodnoty $\{y_j\}$, $j = 1, ..., n_2$, rozptyl s_y^2 a rozsah výběru n_2.

Platí-li nulová hypotéza H_0, má testovací kritérium F_J přibližně F rozdělení s $\nu_1 = 2$, $\nu_2 = n_1 + n_2 - 2$ stupni volnosti. Vyjde-li, že $F_J > F_{1-\alpha}(\nu_1, \nu_2)$, je nutné zamítnout hypotézu H_0 o shodnosti obou výběrových rozptylů na hladině významnosti.

3. krok: **Testy shody středních hodnot dvou souborů**

Studentův t-test umožňuje testování hypotézy H_0: $\mu_x = \mu_y$ proti alternativní H_A : $\mu_x \neq \mu_y$ i při splnění obou uvedených předpokladů o výběrech:

(a) Klasický Studentův *t*-test T_1 pro shodné rozptyly: pro $\sigma_x^2 = \sigma_y^2$ a když obě rozdělení vykazují Gaussovo rozdělení, má testovací kritérium tvar

$$T_1 = \frac{|\bar{x} - \bar{y}|}{\sqrt{(n_1 - 1)\, s_x^2 + (n_2 - 1)\, s_y^2}} \sqrt{\frac{n_1\, n_2\, (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Platí-li, že $T_1 > t_{1-\alpha/2}(n_1 + n_2 - 2)$, je hypotéza H_0 o shodě středních hodnot na hladině významnosti α zamítnuta.

(b) Klasický Studentův *t*-test T_2 pro různé rozptyly: pro $\sigma_x^2 \neq \sigma_y^2$ a když obě rozdělení vykazují Gaussovo rozdělení, má testovací kritérium tvar

$$T_2 = \frac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}}$$

Platí-li hypotéza H_0, má tato testová statistika Studentovo rozdělení s ***"ekvivalentními"*** stupni volnosti ν

$$\nu = \frac{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}{\dfrac{s_x^4}{n_1^2\,(n_1 - 1)} + \dfrac{s_y^4}{n_2^2\,(n_2 - 1)}}$$

Platí-li, že $T_2 > t_{1-\alpha/2}(\nu)$, je hypotéza H_0 o shodě středních hodnot na hladině významnosti α zamítnuta.

Testovací kritérium T_1 není robustní vůči heteroskedasticitě, tj. případu, kdy data jsou ve výběrech měřena s různou přesností. V této situaci je správnější užít testovacího kritéria T_2, které je vůči heteroskedasticitě robustnější. Na druhé straně však ekvivalentní stupně volnosti ν vycházejí menší než $n_1 + n_2 - 2$, takže síla testu T_2 je nižší než síla T_1.

(c) Modifikovaný Studentův *t*-test T_3 pro výběry, odchýlené od normálního rozdělení: jestliže jedno z rozdělení se odchyluje od normality nebo se významně liší v šikmosti od druhého, je vhodné použít modifikované testovací kritérium T_3

$$T_3 = \frac{|\bar{x} - \bar{y}| + C + D\,(\bar{x} - \bar{y})^2}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}},$$

kde

$$C = \frac{1}{6}\, \frac{\dfrac{\hat{g}_{1x}}{n_1^2}\dfrac{s_x^3}{\sqrt{n_1}} - \dfrac{\hat{g}_{1y}}{n_2^2}\dfrac{s_y^3}{\sqrt{n_2}}}{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}} \quad \text{a} \quad D = \frac{1}{3}\, \frac{\dfrac{\hat{g}_{1x}}{n_1^2}\dfrac{s_x^3}{\sqrt{n_1}} - \dfrac{\hat{g}_{1y}}{n_2^2}\dfrac{s_y^3}{\sqrt{n_2}}}{\left[\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}\right]^2}$$

V těchto vztazích jsou $\hat{g}_{1x}$ a $\hat{g}_{1y}$ výběrové šikmosti. Aby bylo možné užít kvantilů Studentova rozdělení pro předepsanou hladinu významnosti α, je třeba přeformulovat testovací kritérium T_3 do tvaru

$$T_3 = T_2 + B_x - B_y,$$

kde

$$B_x = \frac{\dfrac{\hat{g}_{1x}\, s_x^3}{6\, n_1^2 \sqrt{n_1}\left[\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}\right]} + \dfrac{\hat{g}_{1x}\, s_x^2\, (\bar{x} - \bar{y})^2}{3\, n_2^2 \sqrt{n_2}\left[\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}\right]}}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}}$$

a B_y se vyčíslí analogicky, pouze šikmost $\hat{g}_{1x}$ se nahradí hodnotou $\hat{g}_y$, rozptyl σ_x^2 hodnotou σ_y^2 a rozsah n_1 hodnotou n_2. Za předpokladu platnosti hypotézy H_0 má testovací kritérium T_3 Studentovo rozdělení s počtem stupňů volnosti ν. Test založený na kritériu T_3 je robustní vůči sešikmení výběrových rozdělení i vůči heteroskedasticitě a není u něho požadována ani shoda rozptylů, $\sigma_x^2 \neq \sigma_y^2$. Vůči odchylkám rozdělení od normality ve špičatosti jsou uvedené *t*-testy T_1, T_2 a T_3 dostatečně robustní. Je možné použít i korekcí na špičatost, což však nepřináší výrazné zlepšení.

(d) Robustní Jackknife test polohy T_4 pro homoskedasticitu: jsou-li ve výběrech přítomna vybočující měření, lze pro test hypotézy H_0: $\mu_1 = \mu_2$, a $\sigma_1^2 = \sigma_2^2$ upravit testovací kritérium založené na uřezaném průměru na tvar

$$T_4 = \frac{(\bar{x}(\vartheta) - \bar{y}(\vartheta))}{\sqrt{S_{w,x}(\vartheta) + S_{w,y}(\vartheta)}}$$

kde $S_{w,x}(\vartheta)$ a $S_{w,y}(\vartheta)$ se vyčíslí pro výběry $\{x_i\}$, $i = 1, ..., n_1$, a $\{y_j\}$, $j = 1, ..., n_2$. Je-li $n_1 = n_2$, má náhodná veličina T_4 přibližně Studentovo rozdělení s 2 (k - 1) stupni volnosti. Test T_4 lze použít jen pro rozsah $n \geq 7$.

(e) Robustní Jackknife test polohy T_5 pro heteroskedasticitu: pro případ nestejných rozptylů $\sigma_1^2 \neq \sigma_2^2$ a nestejných rozsahů $n_1 \neq n_2$ a s využitím kritéria T_2 lze formulovat robustní kritérium T_5 pro test hypotézy H_0: $\mu_x = \mu_y$

$$T_5 = \frac{\bar{x}(\vartheta) - \bar{y}(\vartheta)}{\sqrt{\dfrac{s_{w,x}^2}{h_1} + \dfrac{s_{w,y}^2}{h_2}}}, \text{ kde } s_{w,x}^2 = \frac{S_{w,x}(\vartheta)}{h_1 - 1}, \quad s_{w,y}^2 = \frac{S_{w,y}(\vartheta)}{h_2 - 1}$$

$$h_i = n_i - 2\, int\left(\frac{\vartheta\, n_i}{100}\right) \quad \text{pro} \quad i = 1, 2.$$

Testovací kritérium T_5 má přibližně Studentovo rozdělení s ν stupni volnosti, pro které platí

$$\frac{1}{\nu} = \frac{z^2}{h_1 - 1} + \frac{(1 - z)^2}{h_2 - 1}, \text{ kde } z = \frac{\dfrac{s_{w,x}^2}{h_1}}{\dfrac{s_{w,x}^2}{h_1} + \dfrac{s_{w,y}^2}{h_2}}.$$

Robustní testy T_4 a T_5 jsou výhodné také pro rozdělení s dlouhými konci, když je špičatost větší než 3. V případě normálního rozdělení však mají menší sílu než testy T_1 a T_2.