Testy hypotéz na základě více než 2 výběrů [1]

[1]Tyto materiály byly vytvořeny za pomoci grantu FRVŠ číslo 1145/2004.

Testy hypotéz na základě více než 2 výběrů

Na analýzu rozptylu lze pohlížet v podstatě jako na zobecnění dvouvýběrového t testu pro k souborů. Klasický t-test v takovém případě nelze použít, neboť bychom se mohli dopustit s mnohem větší pravděpodobností chyby I. druhu a to tak, jak bychom zvětšovali počet porovnávaných skupin. Situaci lze zachytit v následující tabulce:

Tabulka 1: Schéma výchozí situace

číslo výběru	počet prvků	zjištěné hodnoty sledovaného znaku	průměr	rozptyl
1	n_1	$y_{11}, y_{12}, \cdots, y_{1j}, \cdots, y_{1n_1}$	$\bar{y}_{1.}$	s_1^2
2	n_2	$y_{21}, y_{22}, \cdots, y_{2j}, \cdots, y_{2n_2}$	$\bar{y}_{2.}$	s_2^2
$\vdots$	$\vdots$		$\vdots$	$\vdots$
i	n_i	$y_{i1}, y_{i2}, \cdots, y_{ij}, \cdots, y_{in_i}$	$\bar{y}_{i.}$	s_i^2
$\vdots$	$\vdots$		$\vdots$	$\vdots$
k	n_k	$y_{k1}, y_{k2}, \cdots, y_{kj}, \cdots, y_{kn_k}$	$\bar{y}_{k.}$	s_k^2

V průběhu následujících výpočtů využijeme některé vzorce, definujme je tedy:

Průměrná úroveň i-tého výběru:

$$\bar{y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \ . \tag{1}$$

Celkový počet pozorování:

$$n = \sum_{i=1}^{k} n_i \ . \tag{2}$$

Celkový průměr:

$$\bar{y}_{..} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} \ . \tag{3}$$

Předpoklady použití analýzy rozptylu

Při aplikaci analýzy rozptylu je nutno zhodnotit, zda jsou splněny předpoklady pro její použití.

Je nutné zajistit, např. vhodným uspořádáním pokusu, aby byly jednotlivé výběry navzájem nezávislé a pocházely z populací s normálním rozdělením.

Dalším důležitým předpokladem pro využití analýzy rozptylu je shodnost rozptylů u jednotlivých výběrů. Předpoklad homoskedasticity se dá pochopitelně testovat. K tomuto účelu se využíval např. Bartlettův test.

Bartlettův test

Bartlettův test je univerzálním testem v tom smyslu, že jej lze využít k hodnocení homoskedasticity u vyvážených i nevyvážených souborů. Bartlettův test využíváme tedy k testování hypotéz:

$$\begin{aligned} &\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 \ , \\ &\mathrm{H}_A : \text{non } \mathrm{H}_0 \ . \end{aligned}$$

Testovým kritériem Bartlettova testu je veličina B, která je definována jako

$$B = [(n-k)\ln s^2 - \sum_{i=1}^{k}(n_i - 1)\ln s_i^2]/C \quad , \tag{4}$$

mající za předpokladu platnosti H_0 a je-li $n_i \geq 6$, přibližně $\chi^2(k-1)$. Testovanou hypotézu zamítáme pokud platí

$$B \geq \chi^2_{1-\alpha}(k-1) \ . \tag{5}$$

Jednotlivé symboly využité při výpočtu testové statistiky lze definovat takto:

$$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2 \qquad i = 1, \cdots, k \quad , \tag{6}$$

celkový rozptyl jako

$$s = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2 \quad , \tag{7}$$

a konstantu C

$$C = 1 + \left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{n-k}\right)/3(k-1) \ . \tag{8}$$

Bartlettův test je však poměrně slabý a dosti citlivý na porušení normality souborů. To může být velký problém především u souborů s malým počtem pozorování. Z tohoto důvodu se dnes spíše používá tzv. Levenův test.

Hartleyův test pro testování homogenity k rozptylů

Jak plyne z názvu testujeme hypotézu

$$\begin{aligned} &\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 \\ &\mathrm{H}_A : \text{non } \mathrm{H}_0 \ . \end{aligned}$$

Pokud bychom testovali všechny dvojice rozptylů, bylo by jich $k(k-1)/2$. To si lze ověřit, neboť se lze logicky domnívat, že pokud zjistíme maximální a minimální hodnotu rozptylů (tedy identifikujeme max s_i^2 a min s_i^2) pak nebude-li se tato dvojice statisticky významně lišit, nebudou se lišit ani ostatní kombinace dvojic. Testovací statistika má v případě Hartleyova testu tvar

$$F_{\max} = \frac{\max\ s_i^2}{\min\ s_i^2}\ . \tag{9}$$

Ke stanovení kritického oboru je nutno využít speciálně sestrojených tabulek, neboť testovaná dvojice rozptylů není náhodně zvolena. Nulovou hypotézu o shodě rozptylů zamítáme na hladině významnosti α, pokud testovací statistika $F_{\max}$ překročí jistou kritickou hodnotu.

Cochranův test pro testování homogenity k rozptylů

Dalším testem pro ověření homoskedasticity je tzv. Cochranův test. V případě jeho použití zamítáme H_0, hypotézu pokud hodnota testového kritéria

$$C = \frac{s_{max}^2}{s_1^2 + s_2^2 + \ldots + s_k^2} \tag{10}$$

překročí kritickou hodnotu Cochranovy statistiky. Jinými slovy, pokud hodnota C bude náležet do kritického oboru, který je definován jako

$$K = \{C \geq C_{1-\alpha}(k, n-1)\}\ .$$

zamítáme hypotézu o shodě rozptylů.

Levenův test homogenity rozptylů

Levenův test v podstatě provádí analýzu rozptylu na reziduích. Využívá přitom proměnnou $z_{ij} = |y_{ij} - \bar{y}_{i.}|$ pro $i = 1, 2, \cdots, k$ a $j = 1, 2, \cdots, n_i$. F-statistika je následně porovnávána s kritickou hodnotou F-rozdělení s $(k-1)$ a $(n-k)$ stupni volnosti. Pro jisté případy jsou navrženy i modifikace Levenova testu. V případě šikmosti souboru lze využít místo $\bar{y}_{i.}$ mediánu. V případě výrazné špičatosti souboru je pak místo $\bar{y}_{i.}$ doporučován 10 % ořezaný průměr.

ANOVA

Jednofaktorová analýza variance s pevnými efekty

Pro další postup předpokládejme, že se jednotlivé výběry pocházejí z normálního rozdělení, jsou nezávislé a mají shodné rozptyly. Jinými slovy, nepodařilo se nám prokázat platnost alternativní hypotézy, tj. heteroskedasticity, některým z výše uvedených testů.

Nulovou hypotézu lze zapsat, v případě jednofaktorové analýzy rozptylu s pevnými efekty, následovně:

$$\begin{array}{l} \mathrm{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_k\ , \\ \mathrm{H}_A : \text{non } \mathrm{H}_0. \end{array} \tag{11}$$

Nebo ekvivalentně pomocí rovnice

$$y_i = \mu + \alpha_i + \epsilon_i\ , \tag{12}$$

testovanou hypotézou je pak shodnost efektů α_i pro všech k úrovní. Pokud bychom chtěli interpretovat symboly v uvedené rovnici, pak symbol μ představuje průměrnou úroveň všech faktorů, symbol α_i efekt i-tého faktoru.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k \tag{13}$$

Je nutné si uvědomit, že úrovně faktoru jsou nenáhodné, neboť jsou dány experimentátory. Příkladem může být ošetření pozemku určitou dávkou hnojiva, či různá reklamní kampaň.

Rozklad celkové variability

Celkovou variabilitu SS_T lze rozdělit na dva sčítance, ty představují variabilitu meziskupinovou SS_A a variabilitu vnitroskupinovou SS_r [1].

$$SS_T = SS_A + SS_r$$

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2}_{SS_T} = \underbrace{\sum_{i=1}^{k} n_i(\bar{y}_{i.} - \bar{y}_{..})^2}_{SS_A} + \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}_{SS_r} \quad . \tag{14}$$

Proti testované hypotéze svědčí případy, ve kterých se statistiky výrazně liší od $\bar{y}_{..}$. Proto se při posuzování platnosti nulové hypotézy zajímáme o variabilitu mezi výběry. Zatímco variabilita uvnitř výběrů, tedy reziduální součet čtverců, nám umožňuje odhadnout rozptyl σ^2 a zároveň slouží jako míra velikosti rozdílu variability mezi výběry.

Nulovou hypotézu pak zamítáme na zvolené hladině významnosti α, pokud testovací statistika F

$$F = \frac{SS_A/(k-1)}{SS_r/(n-k)} \tag{15}$$

překročí příslušný kvantil Fischerova - Snedecorova rozdělení. Formálně zapsáno:

$$F \geq F_{1-\alpha}(k-1, n-k) \ . \tag{16}$$

V podstatě je tato testovací statistika založena na poměru průměrných meziskupinových a vnitroskupinových součtů čtverců.

Výsledky analýzy rozptylu se zapisují do tzv. tabulky analýzy rozptylu. Ta měla v minulosti svůj význam z hlediska výpočtů. V nejjednodušším případě má následující podobu, viz tabulka 2.

Pro úplnost lze dodat, že pokud bychom měli pouze dva výběry, tj. $k = 2$ a

[1] Složka SS_T obsahuje při n pozorování n sčítanců. Ty nejsou zcela libovolné, neboť výrazy uvnitř závorek dají v součtu nulu. Z tohoto důvodu má součet celkové variability SS_T právě $n-1$ stupňů volnosti. Těchto $n-1$ stupňů volnosti lze rozložit na dvě složky. První složkou jsou stupně volnosti f_A příslušející meziskupinové variabilitě. Těch je $k-1$, neboť se zde sice vyskytuje k sčítanců, ale musí být dodržena podmínka $\sum_i^k = n_i(\bar{y}_{i.} - \bar{y}_{..}) = 0$. Druhou složkou jsou tzv. reziduální stupně volnosti. Na ty logicky zůstává $(n-1)-(k-1) = n-k$ stupňů volnosti.

Tabulka 2: Tabulka analýzy rozptylu

Zdroj	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F	Dosažená hladina p
Výběr	SS_A	$k-1$	$MSS_A = \frac{SS_A}{k-1}$	$F = \frac{MSS_A}{MSS_r}$	p
Reziduální	SS_r	$n-k$	$MSS_r = \frac{SS_r}{n-k}$		
Celkový	SS_T	$n-1$			

uplatnili bychom analýzu rozptylu, získali bychom identický výsledek, jako v případě použití klasického dvoustranného t-testu. Tímto postupem však nelze testovat jednostranné hypotézy, které t-testem studovat lze.

Testy mnohonásobného srovnávání

V případě nezamítnutí nulové hypotézy testování končí. Pokud však zamítáme H_0 ve prospěch H_A, obvykle si klademe další otázky. Asi nás bude zajímat mezi kterými dvěma soubory existují statisticky významné rozdíly a jaká je tedy struktura nehomogenity středních hodnot. K těmto účelům slouží testy mnohonásobného srovnávání.

Bonferroniho metoda mnohonásobného porovnání

Bonferroniho metoda nám odpovídá na otázku proč byla zamítnuta nulová hypotéza při analýze rozptylu. Tato metoda považuje za různé ty skupiny, u nichž populační průměry, např. μ_i a μ_h, splňují následující nerovnici

$$|\bar{y}_{i.} - \bar{y}_{h.}| \geq t_{\frac{\alpha}{m}}(n-k)\sqrt{\frac{SS_r}{n-k}\left(\frac{1}{n_i} + \frac{1}{n_h}\right)}\ . \tag{17}$$

Symbol m zde představuje počet všech možných porovnávaných dvojic, tedy $m = k(k-1)/2$.

Schefféého metoda mnohonásobného srovnávání

Předpokladem pro využití Schéffeho metody je normalita všech k souborů. Testujeme hypotézu

$$\mathrm{H}_0: \quad \mu_I = \mu_J\ ,$$
$$\mathrm{H}_A: \quad \mu_I \neq \mu_J\ .$$

Symboly μ_I, μ_J zde mohou představovat střední hodnoty skupin I, J, přičemž jedna skupina může obsahovat pouze jeden výběr. Naopak druhá z nich může obsahovat maximálně $k-1$ výběrů. Tato metoda považuje za různé ty skupiny, u nichž populační průměry, např. μ_i a μ_h, splňují následující nerovnici

$$|\bar{y}_{i.} - \bar{y}_{h.}| \geq \sqrt{(k-1) \cdot SS_r \cdot F_{1-\alpha}(k-1, n-k) \cdot \left(\frac{1}{n_i} + \frac{1}{n_h}\right)}\,. \tag{18}$$

Tukeyova metoda

Ta označí na hladině významnosti α za rozdílné takové populační průměry μ_i a μ_h, které splňují nerovnici

$$|\bar{y}_{i.} - \bar{y}_{h.}| \geq q_\alpha(k, n-k)\sqrt{\frac{SS_r}{2(n-k)}\left(\frac{n_i + n_h}{n_i n_h}\right)}\,, \tag{19}$$

kde $q_\alpha(k, n-k)$ je kritická hodnota studentizovaného rozpětí. Tento test lze použít pouze v případě vyvážených souborů.

Modifikovaná LSD metoda

Posup při této metodě je jednoduchý. Jednotlivé statistiky $\bar{y}_{1.}, \bar{y}_{2.}, \ldots, \bar{y}_{k.}$ seřadíme sestupně dle velikosti. Vypočteme rozdíl mezi dvěma sousedními statistikami $\bar{y}_{i.}$ a $\bar{y}_{h.}$ a ten porovnáváme s tzv. nejmenší signifikantní diferencí. Tu zjistíme ze vztahu

$$LSD_{ih} = t_{1-\frac{\alpha}{2}(n-k)}\sqrt{SS_r\frac{n_i + n_h}{n_i n_h}} \tag{20}$$

Je-li sledovaná diference $\Delta_{ih} > LSD_{ih}$, zamítáme hypotézu o shodě středních hodnot μ_i a μ_h.