POROVNANIE DVOCH TESTOV KOVARIANČNEJ ŠTRUKTÚRY V MULTIVARIÁTNOM MODELI

Viktor Witkovský a Marta Bognárová
Ústav Merania SAV, Bratislava

Abstrakt: V príspevku je navrhnutý nový prístup k testovaniu lineárnej hypotézy, $R\vartheta = c$, v modeli s varianč-nými a kovariančnými komponentami. Navrhnutá testovacia štatistika je založená na známych odhadoch variančných komponentov typu MINQE. Významnú úlohu má predovšetkým nový, tzv. lineárny prístup k odhadovaniu variančných komponentov.

Nový test má lokálny charakter, teda závisí od apriórnej voľby neznámeho vektorového parametra variančných a kovariančných komponentov ϑ_0. Štatistické vlastnosti odhadu sa skúmajú porovnaním so známym exaktným testom pomerom vierohodnosti v multivariátnom modeli.

Kľúčové slová: Test lineárnej hypotézy o variančných komponentoch; Model s variančnými a kovariančnými komponentami; MINQE; MINQE s podmienkami; Lineárny prístup k odhadovaniu variančných komponentov.

1. Úvod

Lineárny model s variančnými a kovariančnými komponentami je obyčajne daný nasledovne:

$$y = X\beta + \varepsilon, \tag{1}$$

kde y je n-rozmerný náhodný vektor pozorovaní, X je $(n\times k)$-rozmerná známa matica plánu, β je k-rozmerný vektor nenáhodných neznámych parametrov (v literatúre o analýze rozptylu sú známe ako pevné efekty) a náhodný vektor ε ukrýva v sebe štruktúru, vytvorenú napr. náhodnými efektmi. Teda stredná hodnota $E(\varepsilon) = 0$ a kovariančná matica $V(\vartheta) = E(\varepsilon\varepsilon') = \sum_{i=1}^{p} \vartheta_i V_i$. Vektor parametrov $\vartheta = (\vartheta_1, \ldots, \vartheta_p)' \in \Theta$. Kvôli jednoduchosti budeme tu o parametrickom priestore Θ predpokladať, že $\Theta = \{\vartheta : V(\vartheta) = \sum_{i=1}^{p} \vartheta_i V_i$ je $p.d.\}$. V_i sú známe $(n \times n)$-rozmerné symetrické matice.

Tento model pôvodne vznikol ako zovšeobecnenie modelov analýzy rozptylu (ANOVA modelov). Avšak vo svojej všeobecnosti značne presahuje oblasť analýzy rozptylu. Totiž ako špeciálny prípad tohto modelu možno získať:

1. ANOVA modely: Všeobecný model pevnými aj náhodnými efektmi (tzv. zmiešaný lineárny model). Z hľadiska štruktúry modely s vyváženými (balanced) aj nevyváženými dátami (unbalanced).
2. Multivariátny model.
3. Model rastových kriviek (Growth Curve Model).
4. Špeciálne štruktúry modelov: Epochové, etapové modely, atď.
5. Niektoré typy modelov časových radov.
6. Mnohé ďalšie, ktoré majú štrukturovanú kovariančnú maticu.

V týchto partikulárnych modeloch je zaujímavé testovať hypotézy o variančných komponentoch. Špeciálne v ANOVA modeli je to hypotéza, či i-ta komponenta vektora $\vartheta = (\vartheta_1, \ldots, \vartheta_p)'$ má významný vplyv alebo nie (t.j. $\vartheta_i = 0$), poprípade, či platí rovnosť $\vartheta_i = \vartheta_j$ pre $i \neq j$. V ostatných modeloch je zaujímavé testovať hypotézu, či možno predpokladať jednoduchšiu štruktúru kovariančnej matice (napr. identickú maticu — t.j. hypotézu o nezávislosti, popr. blokovú, trojdiagonálnu alebo tzv. Raovu jednoduchú štruktúru kovariančnej matice, atď.).

Tieto problémy sú široko študované v literatúre a v mnohých špeciálnych typoch modelov boli nájdené testy, ktoré sú v určitom zmysle optimálne. V ANOVA modeli s náhodnými efektami a s vyváženými dátami sú optimálne testy známe len v špeciálnych modeloch (v modeli jednoduchého triedenia a v modeli dvojného triedenia s interakciami a bez interakcií). V ANOVA modeloch s vyváženými dátami sú testy niektorých špeciálnych hypotéz (nelineárne funkcie variančných komponentov) o náhodných efektoch optimálne (UMPUI - rovnomerne najsilnejšie nevychýlené a invariantné) a zhodujú so známym F testom za predpokladu normality rozdelenia vektora y (pozri [10], [2]). Situácia je podstatne horšia v prípade, že ANOVA model má dáta nevyvážené, a je predmetom ďalšieho výskumu (pozri [10], [8] a [3]).

Úplne nový prístup k testovaniu hypotézy o variančných komponentoch v ANOVA modeli s náhodnými efektami navrhol B. Seifert v práci [9]. Zaviedol nové *lokálne* kritérium optimality, a odvodil nový test - LBLUIT (Locally Best Locally Unbiased Invariant Test), ktorý spĺňa nové kritérium optimality. Navyše dokázal, že tento test je len funkciou odhadovacej štatistiky, ktorá vystupuje v odhade MINQE(U,I). Explicitný tvar tohoto odhadu však doteraz nebol známy, pretože určenie koeficientov, ktoré vystupujú v testovacej štatistike vyžaduje riešenie pomerne zložitého systému integrálnych rovníc.

Iný prístup sa používa v multivariátnom modeli a v modeli rastových kriviek, kde sa na testovanie hypotéz o štruktúre kovariančnej matice používajú testy pomerom vierohodnosti, poprípade testy založené na princípe invariantnosti. Tieto testy sú odvodené len pre niektoré špeciálne typy hypotéz a vlastnosti týchto testov sú známe obyčajne iba asymptoticky. V mnohých prípadoch odvodenie (asymptotického) rozdelenia je komplikované, takže je doposiaľ neznáme (pozri [11]).

2. Test lineárnej hypotézy o variančných komponentoch vo všeobecnom modeli

V tomto príspevku sa pokúsime ukázať alternatívny prístup k riešeniu problému testovania lineárnej hypotézy o variančných a kovariančných komponentoch vo všeobecnom lineárnom modeli s variančnými komponentami, ktorý bude viesť k explicitnému tvaru testovacej štatistiky. V modeli (1) budeme testovať hypotézu:

$$\begin{aligned} H_0 &: \vartheta \in \Theta_0 \quad \text{oproti alternatíve} \quad H_1 : \vartheta \in \Theta \setminus \Theta_0 \\ \text{resp. } H_0 &: R\vartheta = c \quad \text{oproti alternatíve} \quad H_1 : R\vartheta \neq c, \end{aligned} \tag{2}$$

kde R je daná $r \times p$-rozmerná matica a c je daný vektor koeficientov. Navyše budeme pre testovanie predpokladať normalitu rozdelenia vektora y.

V práci [15] boli na základe Seifertovej myšlienky zavedené nové lokálne kritéria optimality pre test takejto všeobecnej lineárnej hypotézy v lineárnom modeli s variančnými a kovariančnými komponentami. Taktiež bol odvodený test, ktorý spĺňa tieto kritéria, tzv. d-LBLUIT. Avšak podobne ako v prípade Seifertovho testu LBLUIT v ANOVA modeli, aj tu sú koeficienty testu určené len implicitne ako riešenie systému integrálnych rovníc.

Nová navrhovaná testovacia štatistika je založená na nevychýlených invariantných odhadoch variančných komponentov — tzv. MINQE(U,I) odhadoch (Minimum Norm Quadratic Unbiased and Invariant Estimator), ktoré zaviedol C. R. Rao vo svojich prácach [5] a [6]. Podrobný popis problematiky odhadovania variančných komponentov možno nájsť v monografii C. R. Rao a J. Kleffe [7].

Avšak až nový prístup k odhadovaniu variančných komponentov, tzv. *lineárny prístup*, umožnil nájsť MINQE(U,I) odhad, ktorý spĺňa systém lineárnych podmienok typu $R\vartheta = c$. Tento odhad publikovali

J. Volaufová a V. Witkovský v práci [14]. Ďalšie podrobnosti o lineárnom prístupe k odhadovaniu varianč-ných komponentov možno nájsť v [13], [12] a [16].

Odhady MINQE(U,I) aj MINQE(U,I) s podmienkami závisia od voľby apriórnej voľby parametra ϑ_0 — sú to odhady lokálne. Je ďalej známe, že za predpokladu normality rozdelenia sú to najlepšie odhady (v zmysle minimalizácie variancie).

2.1. Lineárny prístup k odhadovaniu variančných komponentov

Hlavnou myšlienkou nového prístupu je vhodne transformovať pôvodný model (1) na nový model, v ktorom už vektor variančných a kovariančných komponentov ϑ, je parametrom strednej hodnoty nového transformovaného náhodného vektora.

Označme ako ϑ_0 pevne zvolený vektor parametrov, a nech $V_0 = V(\vartheta_0) = \sum_{i=1}^{p} \vartheta_{0i} V_i$. Nech ďalej $V_0^{-1/2}$ označuje takú maticu, že $V_0^{-1} = V_0^{-1/2} V_0^{-1/2'}$. Keďže sa budeme zaujímať len o invariantné odhady, budeme pracovať len s funkciami maximálneho invariantu $z = M_0 V_0^{-1/2} y$, kde

$$M = I - V_0^{-1/2} X (X' V_0^{-1} X)^{-} X' V_0^{-1/2'}$$

označuje maticu projektora na ortogonálny komplement priestoru generovaného stĺpcami matice $V_0^{-1/2} X$.

Potom pre strednú hodnotu transformovaného vektora vec zz', (kde „vec" označuje operáciu, ktorá vytvorí z ľubovoľnej matice vektor, a to tak, že usporiada stĺpce matice pod seba) platí:

$$\begin{aligned} E_\vartheta(\text{vec } zz') &= \left(\text{vec } M_0' V_0^{-1/2} V_1 V_0^{-1/2'} M_0, \ldots, \text{vec } M_0' V_0^{-1/2} V_p V_0^{-1/2'} M_0\right) \vartheta \\ &= Q\vartheta, \end{aligned} \tag{3}$$

kde Q označuje $n^2 \times p$-rozmernú maticu, ktorej i-ty stĺpec je vec $M_0' V_0^{-1/2} V_i V_0^{-1/2'} M_0$, $i = 1, \ldots, p$.

Označme ako Σ kovariančnú maticu transformovaného vektora vec zz'. Táto kovariančná matica vo všeobecnosti závisí od parametra strednej hodnoty ϑ. Nech Σ_0 zodpovedá matici, ktorá je funkciou pevne zvoleného ϑ_0. Model

$$(\text{vec } zz', Q\vartheta, \Sigma_0) \tag{4}$$

budeme považovať za *linearizovaný model* lokálne v okolí ϑ_0.

Potom odhad $\widehat{Q\vartheta}$ metódou najmenších štvorcov vektora stredných hodnôt $Q\vartheta$ je daný vzťahom:

$$\widehat{Q\vartheta} = Q(Q'Q)Q' \text{ vec } zz',$$

a odhad $\widehat{\widehat{Q\vartheta}}$ metódou najmenších štvorcov, ktorý spĺňa podmienku $R\vartheta = c$ je daný vzťahom:

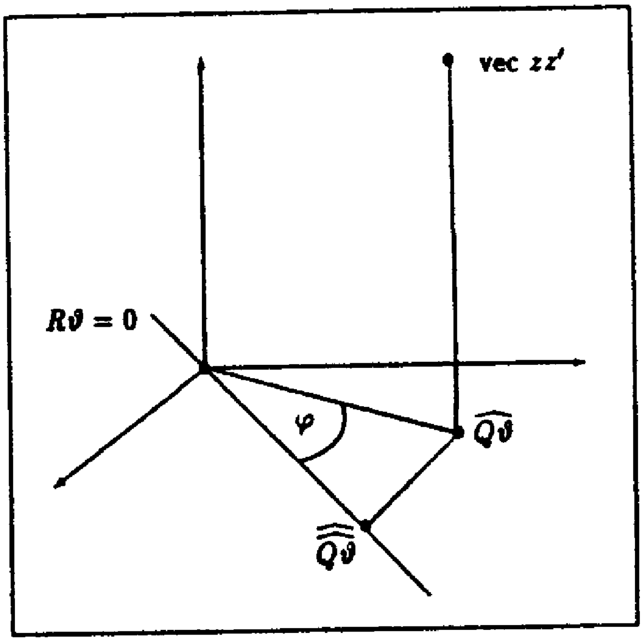
$$\widehat{\widehat{Q\vartheta}} = \widehat{Q\vartheta} + Q(Q'Q)^{-} R' \left[R(Q'Q)^{-} R'\right]^{-} \left(c - R(Q'Q)^{-} Q' \text{ vec } zz'\right).$$

Nech $K = Q'Q$ označuje $p \times p$ rozmernú „kriteriálnu" maticu a $q = Q'$ vec zz' označuje p-rozmerný vektor. S využitím platnosti všeobecného vzťahu $(\text{vec } A)'(\text{vec } B) = \text{tr } AB$, kde „tr" označuje operáciu stopy, ľahko možno ukázať, že pre i, j-ty prvok kriteriálnej matice K a i-ty prvok vektora q platia vzťahy:

$$\begin{aligned} \{K\}_{i,j} &= \text{tr}(MV_0M)^{+} V_i (MV_0M)^{+} V_j, \\ \{q\}_i &= y'(MV_0M)^{+} V_i (MV_0M)^{+} y, \end{aligned} \tag{5}$$

pričom $M = I - XX^{+}$ a vo všeobecnosti A^{+} označuje Moorovu-Penroseovu g-inverziu matice A.

Ako poznámku na záver tejto časti možno uviesť, že v literatúre sa práve vektor q označuje za MINQE(U,I) svojej strednej hodnoty. Navyše odhady $\widehat{Q\vartheta}$ a $\widehat{\widehat{Q\vartheta}}$ sú ϑ_0-lokálne najlepšie kvadratické invariantné a nevychýlené odhady za predpokladu normality rozdelenia vektora y. (Pozri napr. [14] a [12]).



Obrázok 1: Geometria vo výberovom priestore

2.2. Návrh testovacej štatistiky

Návrh novej testovacej štatistiky vychádza z jednoduchej geometrickej interpretácie v n^2-rozmernom výberovom priestore (obrázok 1). Ako testovaciu štatistiku, ktorá by bola citlivá vzhľadom na porušenie nulovej hypotézy $H_0 : R\vartheta = c$ navrhujeme štatistiku, ktorá priamo úmerne závisí od veľkosti uhla φ (napr. tangens uhla φ) podľa obrázku 1.

Test založený na tejto štatistike by mal zamietať nulovú hypotézu pre „veľké" hodnoty uhla φ. Teda

$$\operatorname{tg}(\varphi) = \frac{\|\widehat{Q\vartheta} - \widehat{\widehat{Q\vartheta}}\|}{\|\widehat{\widehat{Q\vartheta}}\|} = \frac{\|QH_2q\|}{\|QH_1q\|}, \tag{6}$$

kde $H_1 = (QQ')^-$ a $H_2 = (Q'Q)^- R'[R(Q'Q)^- R']^- (QQ')^-$.

V ďalšom sa budeme sústreďovať len na obmedzenú nulovú hypotézu typu $H_0 : P\vartheta = 0$, pričom $(r \times p)$-rozmerná matica P má špeciálnu blokovú štruktúru $P = (I \vdots 0)$, kde I je identická matica typu $r \times r$. Dá sa ukázať, pozri [16], že takéto obmedzenie nie je na újmu všeobecnosti, pretože každý linearizovaný model (4) možno regulárne pretransformovať tak, že podmienka $R\vartheta = c$ nadobudne tvar $P\vartheta_* = 0$.

Keďže $H_1 = (QQ')^-$ a $H_2 = (Q'Q)^- R'[R(Q'Q)^- R']^- (QQ')^-$, potom ľahko vidieť, že platia vzťahy:

$$QH_1(Q'Q)H_1Q' = QH_1Q' \quad \text{a} \quad QH_2(Q'Q)H_2Q' = QH_2Q'.$$

Potom pre štatistiku (6) platí:

$$\operatorname{tg}^2(\varphi) = \frac{q'H_2q}{q'H_1q}.$$

Táto testovacia štatistika závisí podobne ako MINQE(U,I) od apriórne zvolenej hodnoty ϑ_0, a teda má lokálny charakter. Avšak rozdelenie testovacej štatistiky $\operatorname{tg}(\varphi)$ nie je doposiaľ známe ani za platnosti nulovej hypotézy a za predpokladu normality rozdelenia vektora y. (Vzhľadom na (5) vidieť, že testovacia štatistika je funkciou „štvrtých mocnín" vektora y). Je preto vhodné hľadať modifikáciu testovacej štatistiky so známou distribúciou za platnosti H_0.

2.3. Modifikácia testovacej štatistiky

Nech d a l označujú pevne zvolené nenáhodné n^2-rozmerné vektory, a také, že $\|d\| = \|l\| = 1$. Za predpokladu normality rozdelenia vektora y podľa [14] pre varianciu náhodných vektorov QH_2q a QH_1q lokálne v bode ϑ_0 platí:

$$\begin{aligned} \mathrm{var}_{\vartheta_0}(QH_2q) &= QH_2Q'\Sigma_0QH_2Q' = 2QH_2(Q'Q)H_2Q = 2QH_2Q \\ \mathrm{var}_{\vartheta_0}(QH_1q) &= QH_1Q'\Sigma_0QH_1Q' = 2QH_1(Q'Q)H_1Q = 2QH_1Q. \end{aligned}$$

Zvoľme za d vlastný vektor matice QH_2Q', ktorý zodpovedá jej maximálnemu vlastnému číslu, a za l zvoľme vlastný vektor matice QH_1Q', ktorý zodpovedá maximálnemu vlastnému číslu matice QH_1Q'. Potom testovaciu štatistiku $\mathrm{tg}(\varphi)$ môžeme aproximovať takto:

$$\mathrm{tg}^2(\varphi) \approx \frac{|d'QH_2q|}{|l'QH_1q|} = |Fidl|. \tag{7}$$

Ihneď vidieť, že modifikovaná testovacia štatistika $Fidl$ je podielom dvoch kvadratických foriem, a teda je typu $\frac{y'A_1y}{y'A_2y}$, kde matice A_1 a A_2 sú jednoznačne určené zo vzťahu (7).

Za predpokladu normality rozdelenia vektora y, je známe rozdelenie štatistiky $Fidl$ lokálne v bode ϑ_0, (pozri [1]). V prípade, že $\vartheta_0 \in \Theta_0$, teda ϑ_0 je z nulovej hypotézy, potom poznáme rozdelenie testovacej štatistiky $Fidl$ za nulovej hypotézy lokálne v bode ϑ_0.

2.4. Rozdelenie kvadratických foriem

V tejto časti sú veľmi stručne zhrnuté niektoré výsledky z práce J. P. Imhofa [1], ktoré dávajú explicitný a numericky vhodný tvar pre výpočet pravdepodobnosti $\Pr(y'Ay > x)$, pre ľubovoľnú kvadratickú formu $y'Ay$ za predpokladu o normalite rozdelenia vektora y.

Nech $y \sim \mathcal{N}(X\beta, V_0)$. Označme ako λ m-rozmerný vektor nenulových vlastných čísel matice AV_0, nech ν označuje vektor zodpovedajúci násobnosti jednotlivých vlastných čísel vo vektore λ. Nech ďalej δ označuje vektor s parametrami necentrality. Potom pre $x \geq 0$ platí:

$$\Pr(y'Ay > x) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin\theta(u)}{u\varrho(u)}\,du, \tag{8}$$

kde

$$\begin{aligned} \theta(u) &= \frac{1}{2}\sum_{i=1}^{m}\left(\nu_i \,\mathrm{arctg}(\lambda_i u) + \delta_i^2\lambda_i u(1+\lambda_i^2u^2)^{-1}\right) - \frac{1}{2}xu, \\ \varrho(u) &= \prod_{i=1}^{m}(1+\lambda_i^2u^2)^{\frac{1}{4}\nu_i}\exp\left(\frac{1}{2}\sum_{i=1}^{m}\frac{(\delta_i\lambda_i u)^2}{1+\lambda_i^2u^2}\right). \end{aligned}$$

Jednoducho možno situáciu zovšeobecniť na počítanie pravdepodobnosti podielu dvoch kvadratických foriem, platí totiž:

$$\Pr\left(\frac{y'A_1y}{y'A_2y} > x\right) = \Pr(y'A_1y - x(y'A_2y) > 0) = \Pr(y'(A_1 - xA_2)y > 0).$$

Algoritmus pre numerický výpočet tejto pravdepodobnosti a jeho analýzu z numerického hľadiska možno nájsť v práci [4].

3. Simulačná štúdia

Obraz o kvalite navrhovaného testu $Fidl$, o jeho štatistických vlastnostiach a o citlivosti na voľbu polohy parametra ϑ_0, získame porovnaním so známym exaktným testom v špeciálnom modeli. Test $Fidl$ budeme porovnávať simulačne s testom pomerom vierohodnosti v multivariátnom modeli.

3.1. Test hypotézy o nezávislosti v multivariátnom modeli

Uvažujme tento multivariátny model:

$$Y = 1' \otimes \mu + \epsilon, \tag{9}$$

kde Y je $(m \times N)$-rozmerná náhodná matica, 1 označuje N-rozmerný stĺpec jednotiek, μ je m-rozmerný neznámy vektorový parameter strednej hodnoty. O matici náhodných chýb ϵ budeme predpokladať:

$$E(\text{vec } \epsilon) = 0, \qquad E\left((\text{vec } \epsilon)(\text{vec } \epsilon)'\right) = I \otimes \Sigma,$$

kde I je $N \times N$ identická matica a Σ je typu $m \times m$. Navyše predpokladajme normalitu rozdelenia. Teda pre každý stĺpec matice Y platí:

$$Y_i \sim \mathcal{N}(\mu, \Sigma), \qquad i = 1, \ldots, N.$$

Rozdeľme maticu Σ na bloky, teda

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix},$$

pričom Σ_{11} je $m_1 \times m_1$ a Σ_{22} je $m_2 \times m_2$ matica, kde $m_1 + m_2 = m$. Navyše $\Sigma_{12} = \Sigma_{21}'$. Potom hypotézu o nezávislosti možno zapísať nasledovne:

$$H_0 : \Sigma_{12} = 0 \quad \text{oproti alternatíve} \quad H_1 : \Sigma_{12} \neq 0.$$

M. S. Srivastava a C. G. Khatri v práci [11] odvodili test pomerom vierohodnosti (LR) pre test hypotézy o nezávislosti. Platí:

$$LR = (\lambda^*)^{\frac{1}{2}N}, \quad \text{kde} \quad \lambda^* = \frac{1}{|V_{11}||V^{11}|},$$

pričom V_{11} označuje prvý blok výberovej matice V a V^{11} označuje prvý blok inverznej matice V^{-1}. Výberová matica V je daná vzťahom:

$$V = \sum_{i=1}^{N}(Y_i - \bar{Y})(Y_i - \bar{Y})', \qquad \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

V špeciálnom prípade, keď $m_1 = 1$, t.j. keď

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12}' \\ \sigma_{12} & \Sigma_{22} \end{pmatrix},$$

je za platnosti nulovej hypotézy známe exaktné rozdelenie štatistiky λ^*, ktorá má v tom prípade beta rozdelenie s parametrami:

$$\lambda^* \sim \text{Beta}\left(\frac{1}{2}(N - m_2 - 1), \frac{1}{2}m_2\right).$$

Nulovú hypotézu test zamieta pre „malé" hodnoty štatistiky λ^*.

3.2. Multivariátny model ako model s variančnými a kovariančnými komponentami

Mulrivariátny model (9) možno prepísať do tvaru všeobecného lineárneho modelu s variančnými a kovariančnými komponentami, t.j.

$$y = X\beta + \epsilon, \qquad E(\epsilon\epsilon') = \sum_{i=1}^{p} \vartheta_i V_i.$$

K tomu najskôr preznačme prvky matice Σ:

$$\Sigma = \begin{pmatrix} \vartheta_{11} & & & \\ \vartheta_{12} & \vartheta_{22} & & \\ \vdots & & \ddots & \\ \vartheta_{1m} & & & \vartheta_{mm} \end{pmatrix}.$$

Potom

$$E(\mathrm{vec}\, Y) = (1 \otimes I)\mu \quad E\left((\mathrm{vec}\,\epsilon)(\mathrm{vec}\,\epsilon)'\right) = I \otimes \Sigma = \sum_{i=j} \vartheta_{ij} V_{ij} + \sum_{i<j} \vartheta_{ij} V_{ij},$$

kde $V_{ii} = (I \otimes e_i e_i')$ a pre $i \neq j$ je $V_{ij} = (I \otimes (e_i e_j' + e_j e_i'))$. Pričom e_i označuje m-rozmerný jednotkový vektor, v ktorom jednotka je na i-tom mieste, t.j. $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$.

Nulovú hypotézu o nezávislosti potom možno zapísať v tvare:

$$H_0 : \vartheta_{12} = \vartheta_{13} = \cdots = \vartheta_{1m} = 0,$$

alebo v maticovom zápise

$$H_0 : \begin{pmatrix} 0 & 1 & 0 & & 0 & 0 & \ldots & 0 \\ \vdots & & 1 & & & & & \vdots \\ \vdots & & & \ddots & & & & \vdots \\ 0 & & & & 1 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \vartheta_{11} \\ \vdots \\ \vdots \\ \vartheta_{mm} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix},$$

a stručnejšie $R\vartheta = 0$. Vektor ϑ je pritom $p = \frac{m(m+1)}{2}$ rozmerný, a obsahuje všetky prvky matice Σ, ktoré sú na alebo pod hlavnou diagonálou.

3.3. Parametre pre simulácie

Bolo nagenerovaných 20 000 realizácií náhodnej $(p \times N)$-rozmernej matice Y, (pričom $p = 3$ a $N = 10$), podľa modelu (9), t.j. $Y_i \sim \mathcal{N}(\mu, \Sigma)$. Vektorový parameter strednej hodnoty μ mal pre generovanie pevne zvolenú hodnotu $(12, 5.634, 18.56)'$.

Apriórna hodnota parametra ϑ, pomocou ktorej sme počítali hodnoty testovacích štatistík a ich kritické hodnoty je $\vartheta_0 = (21, 0, 0, 16, 4, 24)'$, t.j. $R\vartheta_0 = 0$, teda ϑ_0 patrí do nulovej hypotézy H_0. V maticovom zápise teda máme

$$\Sigma_0 = \begin{pmatrix} 21 & 0 & 0 \\ 0 & 16 & 4 \\ 0 & 4 & 24 \end{pmatrix}.$$

Realizácie matice Y sme generovali z rozdelení, ktoré spĺňajú nulovú hypotézu $R\vartheta = 0$, ako aj z rozdelení, kde je nulová hypotéza porušená. Aby sme zistili citlivosť na porušenie predpokladu o polohe, (lokálny charakter testov), generovali sme niekoľko rôznych rozdelení v okolí apriórne zvoleného bodu ϑ_0. Presnejšie, boli generované realizácie z rozdelení s parametrami kovariančnej matice, ktoré sú uvedené v nasledujúcej tabuľke.

Por.	H_0	H_1	Por.	H_0	H_1
0	(21,0,0,16,4,24)'	(21,6,-7.3,16,4,24)'	5	(25,0,0,16,4,24)'	(25,6,-7.3,16,4,24)'
1	(17,0,0,16,4,24)'	(17,6,-7.3,16,4,24)'	6	(21,0,0,20,4,24)'	(21,6,-7.3,20,4,24)'
2	(21,0,0,12,4,24)'	(21,6,-7.3,12,4,24)'	7	(21,0,0,16,8,24)'	(21,6,-7.3,16,8,24)'
3	(21,0,0,16,0,24)'	(21,6,-7.3,16,0,24)'	8	(21,0,0,16,4,28)'	(21,6,-7.3,16,4,28)'
4	(21,0,0,16,4,20)'	(21,6,-7.3,16,4,20)'			

3.4. Výsledky simulačnej štúdie

Simuláciami sme porovnávali tieto testy resp. testovacie štatistiky:

1. LR: Test pomerom vierohodnosti s rozdelením beta za platnosti nulovej hypotézy H_0.
2. tg^2: Test motivovaný jednoduchou geometrickou interpretáciou.

3. *Fid*1: Modifikácia testu tg^2, ($tg^2 \approx |Fid1|$). Smer $d1$ zodpovedá vlastnému vektoru prislúchajúcemu maximálnemu vlastnému číslu, l je pevne zvolený vektor v smere $Q\vartheta_0$.
4. *Fid*2: Modifikácia testu tg^2, smer $d2$ zodpovedá vlastnému vektoru prislúchajúcemu druhému najväčšiemu vlastnému číslu, l je pevne zvolený vektor v smere $Q\vartheta_0$.
5. *Fi*: Kombinácia testov *Fid*1 a *Fid*2. Presnejšie $Fi = \max(|Fid1|, |Fid2|)$.

Typ rozd.	α	H_0 *LR*	tg^2	*Fi*	*Fid*1	*Fid*2	H_1 *LR*	tg^2	*Fi*	*Fid*1	*Fid*2
	0.1	.1015	.1143	.0992	.1033	.1028	.2538	.2792	.2636	.2941	.1644
0	*0.05*	.0526	.0589	.0524	.0505	.0522	.1499	.1734	.1643	.1861	.0957
	0.01	.0102	.0100	.0117	.0106	.0125	.0406	.0430	.0488	.0568	.0277
	0.1	.1015	.0988	.0859	.0922	.0905	.3298	.3235	.3065	.3512	.1722
1	*0.05*	.0526	.0490	.0436	.0440	.0446	.2069	.2104	.1961	.2287	.1020
	0.01	.0102	.0068	.0084	.0083	.0100	.0620	.0578	.0610	.0736	.0312
	0.1	.1015	.1222	.1014	.1143	.0922	.2538	.2609	.2480	.2817	.1438
2	*0.05*	.0526	.0636	.0539	.0583	.0455	.1499	.1563	.1487	.1745	.0842
	0.01	.0102	.0117	.0114	.0131	.0098	.0406	.0368	.0488	.0499	.0225
	0.1	.1015	.1078	.1045	.1296	.083	.2538	.3275	.3228	.3706	.1663
3	*0.05*	.0526	.0557	.0550	.0688	.0411	.1499	.2154	.2170	.2609	.0950
	0.01	.0102	.0095	.0121	.0162	.0093	.0406	.0644	.0755	.0992	.0251
	0.1	.1015	.1192	.1094	.0958	.1239	.2538	.2596	.2570	.2691	.1921
4	*0.05*	.0526	.0625	.0582	.0468	.0665	.1499	.1576	.1590	.1635	.1156
	0.01	.0102	.0114	.0136	.0096	.0168	.0406	.0379	.0466	.0447	.0371
	0.1	.1015	.1247	.1097	.1097	.1114	.2077	.2447	.2321	.2543	.1592
5	*0.05*	.0526	.0654	.0583	.0555	.0580	.1175	.1443	.1421	.1553	.0917
	0.01	.0102	.0112	.0133	.0123	.0139	.0312	.0332	.0400	.0457	.0256
	0.1	.1015	.1045	.0976	.0918	.1101	.2538	.2741	.2644	.2889	.1817
6	*0.05*	.0526	.0535	.0502	.0447	.0568	.1499	.1730	.1654	.1819	.1085
	0.01	.0102	.0085	.0109	.0084	.0141	.0406	.0453	.0504	.0547	.0330
	0.1	.1015	.1131	.1049	.0640	.1419	.2538	.1939	.1953	.1613	.1987
7	*0.05*	.0526	.0591	.0566	.0288	.0784	.1499	.1091	.1150	.0795	.1256
	0.01	.0102	.0104	.0137	.0043	.0208	.0406	.0240	.0327	.0149	.0414
	0.1	.1015	.1084	.0912	.1069	.0854	.2538	.2862	.2670	.3103	.1404
8	*0.05*	.0526	.0548	.0478	.0528	.0412	.1499	.1799	.1651	.2023	.0801
	0.01	.0102	.0095	.0100	.0111	.0084	.0406	.0488	.0505	.0657	.0214

Výsledky sú zhrnuté v tabuľke. Pre tri rôzne hodnoty hladiny významnosti α, (0.1, 0.05 a 0.01), boli počítané relatívne početnosti zamietnutia nulovej hypotézy. Poradové číslo v prvom stĺpci hovorí o parametroch podľa ktorých boli generované realizácie náhodnej matice Y, (pozri predchádzajúcu tabuľku). Keďže rozdelenie a teda ani kritické hodnoty testovacej štatistiky tg^2 nie sú známe, realizácie tejto štatistiky sme porovnávali vzhľadom na teoretické kritické hodnoty testu *Fi*, (predpokladali sme že majú približne rovnaké rozdelenie). Vo všetkých prípadoch bola voľba apriórneho bodu $\vartheta_0 = (21, 0, 0, 16, 4, 24)'$.

Ako vidieť podľa tabuľky, test *LR*, pomerom vierohodnosti, sa za platnosti nulovej hypotézy chová rovnako, bez ohľadu na typ rozdelenia (ktorý spĺňa nulovú hypotézu). Ostatné testy sa chovajú tiež pomerne veľmi dobre (nevychýlenosť), ale cítiť závislosť od polohy parametra ϑ. Najväčšia odchýlka (testy *Fid*1 a*Fid*2) je v prípade rozdelenia č. 7, t.j. $\vartheta = (21, 0, 0, 16, 8, 24)'$, keď sa silne porušila (oproti ϑ_0) korelačná väzba medzi druhým a tretím riadkom matice Y. Kombinovaný test *Fi* však vykazuje pomerne veľkú robustnosť vzhľadom na porušenie predpokladu o polohe parametra ϑ.

Ako vidieť ďalej, sila všetkých testov je pomerne malá (možno to odôvodniť malým rozsahom výberu, t.j. N je príliš malé na to aby aby sa dalo odhaliť porušenie nulovej hypotézy o štruktúre kovariančnej matice). Test LR je už tiež citlivý na polohu parametra ϑ (pozri riadky č. 1 a č. 5). Ostatné testy (okrem testu $Fid2$, ktorý má silu systematicky menšiu) majú približne rovnakú silu ako test LR. Najväčšia odchýlka v sile opäť nastala v prípade č. 7.

Použitá literatúra

[1] J.P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3 and 4):419–426, 1961.

[2] T. Mathew and B. K. Sinha. Optimum test for fixed effects and variance components in balanced models. *Journal of the American Statistical Association*, 83(401):133–135, March 1988.

[3] T. Mathew and B. K. Sinha. Optimum tests in unbalanced two-way models without interaction. *The Annals of Statistics*, 16(4):1727–1740, 1988.

[4] A. Michalski. *Intro- and Intra-class Estimation of Variance Components in Mixed Linear Models with Block Design*. PhD dissertation, Akademia Rolnicza, Wroclaw, Dept. of Mathematics, November 1989. (In Polish).

[5] C.R. Rao. Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association Theory and Methods Section*, 329(63):161–172, 1970.

[6] C.R. Rao. Estimation of variance and covariance components — MINQUE theory. *Journal of Multivariate Analysis*, 1:257–275, 1971.

[7] C.R. Rao and J. Kleffe. *Estimation of Variance Components and Applications*, volume 3 of *Statistics and probability*. North-Holland, Amsterdam, New York, Oxford, Tokyo, first edition, 1988.

[8] J. F. Seely and Y. El-Bassiouni. Applying Wald's variance components test. *The Annals of Statistics*, 11(1):197–201, 1983.

[9] B. Seifert. Optimal tests in unbalanced mixed analysis of variance. Preprint P — Math — 13/90, AdW der DDR, Karl - Weierstraß - Institut für Mathematik, Berlin, February 1990.

[10] E. Spjøtvoll. Optimum invariant tests in unbalanced variance components models. *Annals of Mathematical Statistics*, 38:422–428, 1967.

[11] M.S. Srivastava and C.G. Khatri. *An Introduction to Multivariate Statistics*. Elsevier North Holland, New York, Oxford, first edition, 1979.

[12] J. Volaufová. A brief survey on the linear methods in variance-covariance components model. In Werner Müller, editor, *Proceedings MODA-3*, St. Petersburg, Russia, May 1992. Physica Verlag Vienna. Submitted for publication.

[13] J. Volaufová. MINQUE of variance components in replicated and multivariate linear model with linear restrictions. *QÜESTIIÓ*, Submitted for publication.

[14] J. Volaufová and V. Witkovský. Estimation of variance components in mixed linear model. *Applications of Mathematics*, 37(2):139–148, 1992.

[15] V. Witkovský. On the test of linear hypothesis in mixed linear model. In A. Pázman and J. Volaufová, editors, *PROBASTAT'91, International Conference on Probability and Mathematical Statistics. Proceedings*, Bratislava, August 1991.

[16] V. Witkovský. Testovanie lineárnej hypotézy v zmiešanom lineárnom modeli. Minimová práca, Ústav merania SAV, Bratislava, November 1991.