Vedecká rada Fakulty matematiky, fyziky a informatiky
Univerzity Komenského v Bratislave

Mgr. Gejza Wimmer

Autoreferát dizertačnej práce

Modely, metódy a algoritmy pre analýzu longitudinálnych dát

pre získanie akademickej hodnosti *Philosophiae doctor*

v odbore doktorandského štúdia
9-1-9 aplikovaná matematika

Bratislava, 2010

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Matematickom ústave Slovenskej akadémie vied v Bratislave.

Predkladateľ: Mgr. Gejza Wimmer
Matematický ústav SAV
Štefánikova 49
814 73 Bratislava

Školiteľ: doc. RNDr. Viktor Witkovský, CSc.
Ústav merania SAV, Bratislava

Oponenti: prof. RNDr. Ing. Lubomír Kubáček, DrSc., Dr. h. c.
Katedra matematické analýzy a aplikací matematiky
Přírodovědecká fakulta Univerzity Palackého v Olomouci

doc. PaedDr. RNDr. Stanislav Katina, PhD.
Katedra aplikovanej matematiky a štatistiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského

RNDr. Martina Hančová, PhD.
Ústav matematických vied
Univerzita Pavla Jozefa Šafárika v Košiciach

Autoreferát bol rozoslaný dňa 15. júla 2010.

Obhajoba dizertačnej práce sa koná dňa 27. augusta 2010 o 13.00 hod. pred komisiou pre obhajobu dizertačnej práce v odbore doktorandského štúdia 9-1-9 aplikovaná matematika, vymenovanou predsedom odborovej komisie dňa 9.júla 2010.

Predseda odborovej komisie:

prof. RNDr. Marek Fila, DrSc.
predseda odborovej komisie pre obhajoby dizertačných prác vo vednom odbore 9-1-9 aplikovaná matematika

1 Úvod

Longitudinálne experimenty, ktoré sa vyznačujú opakovaným meraním určitej vlastnosti na viacerých subjektoch v rôznych časových okamihoch, zohrávajú dôležitú úlohu v rôznych bio-medicínskych, sociologických, psychologických, či behaviorálnych výskumoch. Na analýzu takto získaných údajov sa v posledných rokoch vo veľkej miere využíva lineárny zmiešaný model [12], pričom k odhadovaniu neznámych parametrov modelu je vhodné použiť REML funkciu vierohodnosti [6]. Bohužiaľ, presné rozdelenie takto získaného odhadu pevných efektov (regresných parametrov) nie je vo všeobecnosti známe a štatistické inferencie ohľadne nich sú prevažne založené na asymptotických vlastnostiach odhadu regresných parametrov v prípade známych kovariančných parametrov modelu [5]. Avšak, je všeobecne známou skutočnosťou, že v prípade malého počtu pozorovaní je takýto prístup nevhodný. Mnoho autorov sa venovalo tejto problematike, pričom v záujme odstránenia nedostatkov tohto obvyklého prístupu ponúkajú jeho modifikácie ([4], [10], [11]). V dizertačnej práci sa podrobnejšie venujeme konštrukcii približných konfidenčných oblastí známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta. Na základe práce [11] je navrhnutá nová modifikácia odhadu kovariančnej matice vektora pevných efektov lineárneho zmiešaného modelu v prípade neznámych kovariančných parametrov modelu, ak tieto sú odhadované pomocou REML funkcie vierohodnosti. Na základe [1] a [10] je odvodený postup pre konštrukciu približných konfidenčných oblastí známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta v prípade použitia navrhnutého odhadu kovariančnej matice odhadu vektora pevných efektov. V simulačnej štúdii sme porovnali vlastnosti navrhnutej približnej konfidenčnej oblasti známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta s vlastnosťami niektorých používaných približných konfidenčných oblastí známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta ([4], [8], [10]), ako aj s vlastnosťami konfidenčnej oblasti založenej na asymptotických výsledkoch.

2 Lineárny zmiešaný model pre longitudinálne dáta

Lineárny zmiešaný model je pre analýzu longitudinálnych dát výhodný, pretože už v samotnom zápise zohľadňuje vplyv jednotlivých subjektov na svoje opakované merania, pričom tieto individuálne efekty sa môžu pre jednotlivé subjekty navzájom odlišovať. Vsunutím týchto efektov do modelu je potom možné odhadovať nielen celkovú zmenu spoločnú pre všetky subjekty, ale aj jednotlivé zmeny pre každý subjekt, ktoré vyjadrujú ich odchýlku od všeobecného priemeru. Pre túto vlastnosť lineárnych zmiešaných modelov sa tento prístup často využíva v mnohých praktických aplikáciách.

Pre vektor pozorovaní $\mathbf{Y}_i$ na i-tom subjekte, $i = 1, 2, \ldots, I$, v tomto prípade platí vzťah

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

kde $\mathbf{X}_i$ sú známe $(n_i \times p)$-rozmerné matice plánu pre neznámy p-rozmerný vektor parametrov $\boldsymbol{\beta}$. Tieto nazývame pevné efekty a sú pre všetky subjekty rovnaké. Naopak, q-rozmerné vektory neznámych parametrov $\boldsymbol{\eta}_i$ sa pre každý subjekt navzájom odlišujú, pričom ich považujeme za náhodné, navzájom nezávislé pochádzajúce z $N(\mathbf{0}, \underline{\mathbf{D}})$ rozdelenia. Sú to teda individuálne náhodne efekty jednotlivých subjektov na svoje opakované merania. K nim prislúchajúce $(n_i \times q)$-rozmerné matice $\mathbf{Z}_i$ sú, podobne ako $\mathbf{X}_i$, známe. Pre n_i-rozmerný vektor náhodných chýb i-teho subjektu $\boldsymbol{\varepsilon}_i$ predpokladáme jeho nezávislosť od $\boldsymbol{\eta}_i$, nezávislosť od $\boldsymbol{\varepsilon}_j$ pre $i \neq j$ a rozdelenie $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$.

Kovariančná matica náhodného vektora pozorovaní i-teho subjektu $\mathbf{Y}_i$, $i = 1, 2, \ldots, I$, je potom

$$Var(\mathbf{Y}_i) = \boldsymbol{\Sigma}_i = \mathbf{Z}_i\underline{\mathbf{D}}\mathbf{Z}_i^{'} + \mathbf{R}_i. \tag{2}$$

Pre prehľadnejší zápis budeme model (1) uvádzať vo všeobecnejšom tvare lineárneho zmiešaného modelu nasledovne

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_i \\ \vdots \\ \mathbf{Y}_I \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_i \\ \vdots \\ \mathbf{X}_I \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & \mathbf{Z}_2 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{Z}_i & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \ldots & \mathbf{Z}_I \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_i \\ \vdots \\ \boldsymbol{\eta}_I \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_i \\ \vdots \\ \boldsymbol{\varepsilon}_I \end{bmatrix}, \tag{3}$$

pričom $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_I')^{'}$ a $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_I')^{'}$. V tomto zápise je potom kovariančná matica náhodného vektora pozorovaní $\mathbf{Y}$ v tvare

$$Var(\mathbf{Y}) = \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \ldots & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \ldots & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Sigma}_i & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \ldots & \boldsymbol{\Sigma}_I \end{bmatrix},$$

kde jednotlivé $\mathbf{\Sigma}_i$ sú dané vzťahom (2). Navyše predpokladajme, že kovariančná matica náhodného vektora pozorovaní $\mathbf{Y}$ je známou funkciou r-rozmerného vektora parametrov $\boldsymbol{\theta}$, a teda $Var(\mathbf{Y}) = \mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$.

Za predpokladu, že $\boldsymbol{\theta}$ je známy vektor parametrov, a teda aj $\mathbf{\Sigma}$ je známa matica a existencie matice k nej inverznej, je potom možné odhad pevných efektov zapísať v tvare

$$\widetilde{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{Y}, \tag{4}$$

kde predpokladáme existenciu matice $(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$. Takto získaný odhad zodpovedá riešeniu problému odhadovania neznámych parametrov váženou metódou najmenších štvorcov (VMNŠ) a za daných predpokladov platí

$$\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \sim N\left(\mathbf{0}, \left(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\right)^{-1}\right). \tag{5}$$

Z Gauss-Markovovej vety je zrejmé, že $\widetilde{\boldsymbol{\beta}}$ je najlepší lineárny nevychýlený odhad (*BLUE* - best linear unbiased estimator) vektora pevných efektov $\boldsymbol{\beta}$.

3 Odhad neznámych kovariančných parametrov

Napriek tomu, že v analýze longitudinálnych dát nás primárne zaujímajú odhady neznámych pevných efektov (regresných parametrov), respektíve predikcie lineárnej kombinácie neznámych pevných a náhodných efektov, k ich získaniu je v praktických úlohách častokrát potrebné odhadovať aj kovariančné parametre modelu. Keďže z hľadiska analýzy longitudinálnych dát nás tieto primárne nezaujímajú, označujeme ich aj ako nežiaduce parametre. Je však potrebné podotknúť, že ich odhady ovplyvňujú efektívnosť odhadov parametrov, ktoré sú v prvoradom záujme našej analýzy. Na získanie týchto odhadov sa vo veľkej miere využíva REML funkcia vierohodnosti (restricted maximum likelihood), pričom jej logaritmus je v tvare

$$\begin{aligned} l_{REML}(\boldsymbol{\theta}; \mathbf{Y}) &= -\frac{1}{2}(n-r)\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}\ln\left|\mathbf{X}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right| \\ &-\frac{1}{2}\mathbf{Y}'\left\{\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}) - \mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\left[\mathbf{X}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\right\}\mathbf{Y}, \end{aligned} \tag{6}$$

kde $n = \sum_{i=1}^{I} n_i$. Maximalizáciou uvedeného logaritmu REML funkcie vierohodnosti (6) vzhľadom na neznáme parametre $\boldsymbol{\theta}$ získame REML odhad $\widehat{\boldsymbol{\theta}}$ neznámych kovariančných parametrov modelu $\boldsymbol{\theta}$.

Pre odhady kovariančných parametrov $\widehat{\boldsymbol{\theta}}$ získaných pomocou uvedenej REML funkcie vierohodnosti (6) potom platí (pozri napr. [16]), že ich asymptotické rozdelenie (pre $n \to \infty$) je

$$\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \sim N\left(\mathbf{0}, \mathbf{W}\right), \tag{7}$$

kde $\mathbf{W}$ je inverzia Fisherovej informačnej matice

$$\mathbf{W} = \left\{ E\left[\frac{\partial l_{REML}}{\partial \boldsymbol{\theta}} \frac{\partial l_{REML}}{\partial \boldsymbol{\theta}'} \right] \right\}^{-1} = \left\{ -E\left[\frac{\partial^2 l_{REML}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right\}^{-1}, \tag{8}$$

kde $\boldsymbol{\theta}_0$ je skutočný vektor kovariančných parametrov modelu.

4 Vlastnosti odhadu pevných efektov lineárneho zmiešaného modelu

Predpokladajme, že vektor pozorovaní $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_I')'$ spĺňa lineárny zmiešaný model pre longitudinálne dáta (3), resp. (1), v ktorom sú okrem vektora pevných efektov neznáme navyše aj kovariančné parametre modelu $\boldsymbol{\theta}$. Je zrejmé, že v tomto prípade je kovariančná matica náhodného vektora pozorovaní $\boldsymbol{\Sigma}$ neznáma. Máme však k dispozícii jej REML odhad, označme $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$. Prirodzeným postupom k získaniu odhadu neznámeho vektora pevných efektov $\boldsymbol{\beta}$ je v tomto prípade nahradenie známej matice $\boldsymbol{\Sigma}$ vo vzťahu (4) jej REML odhadom $\widehat{\boldsymbol{\Sigma}}$. Takto získaný odhad vektora pevných efektov $\boldsymbol{\beta}$ nazveme empirický najlepší lineárny nevychýlený odhad (*EBLUE* - empirical best linear unbiased estimator)

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{Y}. \tag{9}$$

Tento je podľa [2] nevychýleným odhadom vektora pevných efektov modelu $\boldsymbol{\beta}$, a teda

$$E\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}. \tag{10}$$

Kovariančná matica $\widehat{\boldsymbol{\beta}}$ žiaľ vo všeobecnosti nie je známa a v prípade neznámych kovariančných parametrov modelu je zväčša odhadovaná ako

$$\widehat{\boldsymbol{\Phi}} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{X}\right)^{-1}. \tag{11}$$

Autori [10] uvádzajú, že pre $n \to \infty$ je asymptotická kovariančná matica $\widehat{\boldsymbol{\beta}}$ rovná práve $\boldsymbol{\Phi}$. Bohužiaľ, pre „malé“ rozsahy výberov je uvedená aproximácia nevhodná, keďže v sebe nezahŕňa neistotu spojenú s odhadovaním neznámych kovariančných parametrov modelu $\boldsymbol{\theta}$ a navyše $\widehat{\boldsymbol{\Phi}}$ nie je nevychýleným odhadom $\boldsymbol{\Phi}$.

V prípade konečného počtu pozorovaní je možné na základe práce [9] ukázať, že približná kovariančná matica *EBLUE* pre pevné efekty modelu $\boldsymbol{\beta}$ je

$$Var\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\Phi} + \boldsymbol{\Lambda}_1, \tag{12}$$

kde

$$\boldsymbol{\Lambda}_1 \approx \boldsymbol{\Phi}\left\{\sum_{k=1}^{r}\sum_{l=1}^{r}\{\mathbf{W}\}_{kl}\left(\mathbf{Q}_{kl} - \mathbf{P}_k\boldsymbol{\Phi}\mathbf{P}_l\right)\right\}\boldsymbol{\Phi}, \tag{13}$$

$$\mathbf{P}_k = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\mathbf{X}, \tag{14}$$

$$\mathbf{Q}_{kl} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}\boldsymbol{\Sigma}^{-1}\mathbf{X}. \tag{15}$$

Kovariančná matica $\widehat{\boldsymbol{\beta}}$ daná vzťahom (12) je zrejme funkciou kovariančných parametrov modelu, ktoré však nie sú známe. Preto je potrebné ju pre ďalšiu analýzu odhadnúť. Z predchádzajúceho možno usúdiť, že „rozumným“ odhadom $Var(\widehat{\boldsymbol{\beta}})$ by mohlo byť

$$\widehat{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\boldsymbol{\Phi}} + \widehat{\boldsymbol{\Lambda}}_1. \tag{16}$$

Bohužiaľ, ako už bolo spomenuté, $\widehat{\boldsymbol{\Phi}}$ nie je nevychýleným odhadom $\boldsymbol{\Phi}$. Preto autori [10] navrhujú, na základe prác [7] a [14] aproximáciu tejto matice pomocou Taylorovho rozvoja druhého rádu $\widehat{\boldsymbol{\Phi}}$ okolo skutočného vektora kovariančných parametrov $\boldsymbol{\theta}_0$, čím dostávajú odhad kovariančnej matice empirického najlepšieho lineárneho nevychýleného odhadu pevných efektov modelu (3), resp. modelu (1),

$$\widehat{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\boldsymbol{\Phi}} + 2\widehat{\boldsymbol{\Phi}}\left\{\sum_{k=1}^{r}\sum_{l=1}^{r}\left\{\widehat{\mathbf{W}}\right\}_{kl}\left(\widehat{\mathbf{Q}}_{kl} - \widehat{\mathbf{P}}_k\widehat{\boldsymbol{\Phi}}\widehat{\mathbf{P}}_l - \frac{1}{4}\widehat{\mathbf{R}}_{kl}\right)\right\}\widehat{\boldsymbol{\Phi}} \equiv \widehat{\boldsymbol{\Phi}}_{mod}, \tag{17}$$

kde

$$\mathbf{R}_{kl} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\right]\boldsymbol{\Sigma}^{-1}\mathbf{X} \tag{18}$$

a $\widehat{\mathbf{W}}$, $\widehat{\mathbf{P}}_k$, $\widehat{\mathbf{Q}}_{kl}$ a $\widehat{\mathbf{R}}_{kl}$ sú postupne (8), (14), (15) a (18) s dosadenými REML odhadmi kovariančných parametrov modelu $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$.

Uvedený postup k získaniu odhadu kovariančnej matice $\widehat{\boldsymbol{\beta}}$ predpokladá nevychýlenosť odhadu kovariančných parametrov modelu $\boldsymbol{\theta}$. REML odhady kovariančných parametrov však vo všeobecnosti nie sú nevychýlené, pričom táto odchýlka sa vo väčšej miere prejavuje najmä v prípadoch, ktoré predpokladajú nelineárne štruktúry kovariančnej matice náhodného vektora pozorovaní $\mathbf{Y}$. Preto autori v [11] odvodzujú odchýlku REML odhadu $\widehat{\boldsymbol{\theta}}$ kovariančných parametrov modelu, ktorá je

$$Bias(\widehat{\theta}_k) = E\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)_k\right] = -\frac{1}{4}\sum_{l=1}^{r}\sum_{i=1}^{r}\sum_{j=1}^{r}\{\mathbf{W}\}_{ij}\{\mathbf{W}\}_{kl}\, trace\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\mathbf{C}_k\right], \quad (19)$$

kde

$$\mathbf{C}_k = \left(\left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\left(\mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\right)\right)\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\left(\mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\right)\right)'\right)$$

a $trace\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\mathbf{C}_k\right]$ je stopa matice $\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\mathbf{C}_k\right]$. Tento vzťah následne zahrnuli do odhadu kovariančnej matice odhadu vektora pevných efektov, čím dosiahli modifikovaný odhad kovariančnej matice odhadu vektora pevných efektov

$$\widehat{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\boldsymbol{\Phi}}_{mod} + \widehat{\mathbf{B}} \equiv \widehat{\boldsymbol{\Phi}}^*_{mod1}, \quad (20)$$

kde

$$\mathbf{B} = \frac{1}{4}\sum_{k=1}^{r}\sum_{l=1}^{r}\sum_{i=1}^{r}\sum_{j=1}^{r}\{\mathbf{W}\}_{ij}\{\mathbf{W}\}_{kl}\, trace\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\mathbf{C}_k\right]\left[\frac{\partial\boldsymbol{\Phi}}{\partial\theta_k}\right] = - \sum_{k=1}^{r} Bias(\widehat{\theta}_k)\left[\frac{\partial\boldsymbol{\Phi}}{\partial\theta_k}\right] \quad (21)$$

a $\widehat{\mathbf{B}}$ je odhad $\mathbf{B}$, kde vo vzťahu (21) je namiesto skutočného vektora kovariančných parametrov použitý jeho REML odhad $\widehat{\boldsymbol{\theta}}$.

Keďže matica $\widehat{\boldsymbol{\Phi}}^*_{mod1}$ vo svojom zápise zahŕňa odchýlku odhadu kovariančných parametrov modelu $\boldsymbol{\theta}$ len pomocou členu $\widehat{\mathbf{B}}$ vystupujúceho vo vzťahu (20), navrhujeme túto navyše uvažovať aj pre postupy na odvodenie kovariančnej matice $\widehat{\boldsymbol{\beta}}$, konkrétne pri výpočte korekčného člena $\boldsymbol{\Lambda}_1$ daného vzťahom (13), ako aj jeho zahrnutím do odhadu kovariančnej matice $\widehat{\boldsymbol{\beta}}$ nielen pomocou korekčného člena $\widehat{\mathbf{B}}$ matice $\widehat{\boldsymbol{\Phi}}^*_{mod1}$ vzhľadom na maticu $\widehat{\boldsymbol{\Phi}}_{mod}$, ale aj do člena

$$2\widehat{\boldsymbol{\Phi}}\left\{\sum_{k=1}^{r}\sum_{l=1}^{r}\left\{\widehat{\mathbf{W}}\right\}_{kl}\left(\widehat{\mathbf{Q}}_{kl} - \widehat{\mathbf{P}}_k\widehat{\boldsymbol{\Phi}}\widehat{\mathbf{P}}_l - \frac{1}{4}\widehat{\mathbf{R}}_{kl}\right)\right\}\widehat{\boldsymbol{\Phi}}$$

vystupujúceho vo vzťahu (17).

Upustením od predpokladu nevychýlenosti REML odhadu kovariančných parametrov modelu $\boldsymbol{\theta}$ dostávame korekčný člen kovariančnej matice $\widehat{\boldsymbol{\beta}}$ vzhľadnom na kovariančnú maticu $\widetilde{\boldsymbol{\beta}}$

$$\boldsymbol{\Lambda} \approx \boldsymbol{\Phi} \left\{ \sum_{k=1}^{r} \sum_{l=1}^{r} w_{kl} \left(\mathbf{Q}_{kl} - \mathbf{P}_k \boldsymbol{\Phi} \mathbf{P}_l \right) \right\} \boldsymbol{\Phi}, \tag{22}$$

kde

$$w_{kl} = E\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right)_k \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right)_l \right] = \{\mathbf{W}\}_{kl} + Bias(\widehat{\theta}_k) Bias(\widehat{\theta}_l). \tag{23}$$

Je zrejmé, že tento sa od pôvodného korekčného člena $\boldsymbol{\Lambda}_1$ kovariančnej matice $\widehat{\boldsymbol{\beta}}$ vzhľadom na kovariančnú maticu $\widetilde{\boldsymbol{\beta}}$ daného vzťahom (13) odlišuje použitím w_{kl} namiesto $\{\mathbf{W}\}_{kl}$.

Podobne, upustením od predpokladu nevychýlenosti REML odhadu kovariančných parametrov modelu $\boldsymbol{\theta}$, dostávame odhad kovariančnej matice $\widehat{\boldsymbol{\beta}}$

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{\Phi}} + \widehat{\mathbf{B}} + 2\widehat{\boldsymbol{\Phi}} \left\{ \sum_{k=1}^{r} \sum_{l=1}^{r} \widehat{w}_{kl} \left(\widehat{\mathbf{Q}}_{kl} - \widehat{\mathbf{P}}_k \widehat{\boldsymbol{\Phi}} \widehat{\mathbf{P}}_l - \frac{1}{4} \widehat{\mathbf{R}}_{kl} \right) \right\} \widehat{\boldsymbol{\Phi}} \equiv \widehat{\boldsymbol{\Phi}}_{mod1}, \tag{24}$$

kde $\widehat{w}_{kl}$ je (23) s dosadeným REML odhadom vektora kovariančných parametrov modelu $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$.

5 Konfidenčné oblasti pevných efektov

Uvažujme teraz o štatistických inferenciách ohľadne l lineárnych kombinácií prvkov $\boldsymbol{\beta}$, konkrétne zostrojením konfidenčnej oblasti pre $\mathbf{L}'\boldsymbol{\beta}$, kde $\mathbf{L}$ je známa $(p \times l)$-rozmerná matica plnej hodnosti $l \leq p$. V prípade, že kovariančné parametre modelu $\boldsymbol{\theta}$ sú známe, je podľa (5)

$$(\mathbf{L}'\widetilde{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}) \sim N(\mathbf{0}, \mathbf{L}'\boldsymbol{\Phi}\mathbf{L}) \ ,$$

a teda

$$\chi^2 = (\mathbf{L}'\widetilde{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})'(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\widetilde{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}) \tag{25}$$

má chi-kvadrát rozdelenie s l stupňami voľnosti, χ^2_l. $(1-\alpha)\cdot 100\%$-nú konfidenčnú oblasť ($\alpha \in (0,1)$) tvorí množina tých vektorov pevných efektov (regresných parametrov) $\boldsymbol{\beta}^+$, pre ktoré je

$$(\mathbf{L}'\widetilde{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+)'(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\widetilde{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+) < \chi^2_l(1-\alpha), \tag{26}$$

kde $\chi^2_l(1-\alpha)$ je $(1-\alpha)$-kvantil chi-kvadrát rozdelenia s l stupňami voľnosti.

Takáto situácia je však v praxi veľmi zriedkavá, keďže kovariančné parametre modelu sú taktiež prevažne neznáme. „Naivným“ prístupom by sme v tomto prípade mohli postupovať nasledovne. Nahradiť vo vzťahu (25) neznámu hodnotu vektora kovariančných parametrov modelu $\boldsymbol{\theta}$ ich REML odhadmi $\widehat{\boldsymbol{\theta}}$ a predpokladať, že

$$\chi^2_* = (\mathbf{L}^{'}\widehat{\boldsymbol{\beta}} - \mathbf{L}^{'}\boldsymbol{\beta})^{'}(\mathbf{L}^{'}\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}^{'}\widehat{\boldsymbol{\beta}} - \mathbf{L}^{'}\boldsymbol{\beta}) \tag{27}$$

má taktiež χ^2 rozdelenie s l stupňami voľnosti. Tento prístup využíva väčšina základnej literatúry o longitudinálnych dátach (pozri napr. [3], [5]). $(1-\alpha)\cdot$ 100%-nú konfidenčnú oblasť ($\alpha \in (0,1)$) tvorí množina tých vektorov pevných efektov $\boldsymbol{\beta}^+$, pre ktoré je

$$(\mathbf{L}^{'}\widehat{\boldsymbol{\beta}} - \mathbf{L}^{'}\boldsymbol{\beta}^+)^{'}(\mathbf{L}^{'}\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}^{'}\widehat{\boldsymbol{\beta}} - \mathbf{L}^{'}\boldsymbol{\beta}^+) < \chi^2_l(1-\alpha). \tag{28}$$

V skutočnosti však presné rozdelenie $(\mathbf{L}^{'}\widetilde{\boldsymbol{\beta}} - \mathbf{L}^{'}\boldsymbol{\beta}$ nepoznáme a navyše, ako bolo uvedené v predchádzajúcej časti, taktiež použitie $\widehat{\boldsymbol{\Phi}}$ ako odhadu kovariančnej matice $\widehat{\boldsymbol{\beta}}$ nie je správne. Oba tieto nedostatky sa prejavujú najmä v prípade „malého“ počtu pozorovaní.

Preto je vhodnejšie pri konštrukcii konfidenčných oblastí pre známu lineárnu kombináciu pevných efektov modelu $\mathbf{L}^{'}\boldsymbol{\beta}$ využiť štatistiku v tvare

$$F = \frac{1}{l}(\mathbf{L}^{'}\widehat{\boldsymbol{\beta}} - \mathbf{L}^{'}\boldsymbol{\beta})^{'}(\mathbf{L}^{'}\widehat{\boldsymbol{\Phi}}^{*}\mathbf{L})^{-1}(\mathbf{L}^{'}\widehat{\boldsymbol{\beta}} - \mathbf{L}^{'}\boldsymbol{\beta}), \tag{29}$$

kde $\widehat{\boldsymbol{\Phi}}^{*}$ je buď „naivný“ odhad kovariančnej matice vektora pevných efektov modelu $\widehat{\boldsymbol{\Phi}}$, alebo jej modifikovaný odhad $\widehat{\boldsymbol{\Phi}}_{mod}$. Pre túto potom predpokladáme, že má Fisherovo-Snedecorovo rozdelenie s l a m stupňami voľnosti, $F_{l,m}$. Hodnota stupňov voľnosti m však vo všeobecnosti nie je známa a preto sa k jej získaniu využívajú rôzne aproximácie.

5.1 Faiova-Corneliusova metóda

Jednou z možností odhadnutia počtu stupňov voľnosti m je tzv. Faiova-Corneliusova metóda uvedená v [4]. Tento postup predpokladá, že

$$F_{FC} = \frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})^{'}(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}) \tag{30}$$

má približne $F_{l,m}$ rozdelenie. Poznamenajme, že uvedená metóda používa ako odhad kovariančnej matice *EBLUE*-u pevných efektov maticu $\widehat{\boldsymbol{\Phi}}$. Využitím spektrálnej dekompozície matice $(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}$ je potom náhodnú premennú $M = l \cdot F_{FC}$ možné zapísať ako súčet kvadrátov l nezávislých približne

Studentovo t-rozdelených náhodných premenných $t^2_{\nu_s}$ s ν_s, $s = 1, 2, \ldots, l$, stupňami voľnosti

$$M = \sum_{s=1}^{l} \frac{\left(\mathbf{v}_s^{'}\mathbf{L}'\widehat{\boldsymbol{\beta}}\right)^2}{v_s} = \sum_{s=1}^{l} t^2_{\nu_s}, \tag{31}$$

kde $\mathbf{v}_s$ je s-tý vlastný vektor matice $(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}$ a v_s je príslušná vlastná hodnota. Porovnaním strednej hodnoty náhodnej premennej M/l so strednou hodnotou $F_{l,m}$ rozdelenia dostávame približný počet stupňov voľnosti m pomocou Faiovej-Corneliusovej metódy

$$m = \frac{2E\left[M\right]}{E\left[M\right] - l}. \tag{32}$$

Odhad počtu stupňov voľnosti m pomocou Faiovej-Corneliusovej metódy je potom $\widehat{m}$ definované vzťahom (32), kde jednotlivé ν_s, $s = 1, 2, \ldots, l$ sú odhadované pomocou Satterthwaiteovej aproximácie (pozri [15]).

Faiovu-Corneliusovu aproximatívnu $(1-\alpha)\cdot 100\%$-nú konfidenčnú oblasť ($\alpha \in (0,1)$) tvorí množina tých vektorov pevných efektov $\boldsymbol{\beta}^+$, pre ktoré je

$$\frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+)^{'}(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+) < F_{l,\widehat{m}}(1-\alpha), \tag{33}$$

kde $F_{l,\widehat{m}}(1-\alpha)$ je $(1-\alpha)$-kvantil Fisherovho-Snedecorovho rozdelenia s l a $\widehat{m}$ stupňami voľnosti.

5.2 Kenwardova-Rogerova metóda

Ďalšia metóda na odhad počtu stupňov voňosti m, prezentovaná v [10], je založená na predpoklade, že štatistika

$$F_{KR} = \lambda\frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})^{'}(\mathbf{L}'\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}) \tag{34}$$

má približne $F_{l,m}$ rozdelenie. Tento vzťah využíva ako odhad kovariančnej matice $EBLUE$-u pevných efektov maticu $\widehat{\boldsymbol{\Phi}}_{mod}$ definovanú vzťahom (17). Na rozdiel od predchádzajúcej Faiovej-Corneliusovej metódy tu navyše vystupuje aj „škálovacia“ konštanta λ, ktorej hodnota je podobne ako počet stupňov voľnosti m neznáma a je potrebné ju taktiež odhadovať. Pomocou Taylorovho rozvoja druhého rádu matice $(\mathbf{L}'\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L})^{-1}$ okolo skutočného vektora kovariančných parametrov $\boldsymbol{\theta}_0$ a jeho následným využitím pri výpočte prvých dvoch momentov náhodnej premennej $F = \frac{1}{\lambda}F_{KR}$, za predpokladu, že $\widehat{\boldsymbol{\theta}}$ je nevychýleným odhadom vektora kovariančných parametrov modelu $\boldsymbol{\theta}$, dostávame

$$E\left[F\right] = \frac{l + A_2^*}{l}, \qquad Var\left[F\right] = \frac{2 + 2B^*}{l}, \tag{35}$$

kde

$$A_1^* = \sum_{k=1}^{r}\sum_{s=1}^{r} \{\mathbf{W}\}_{ks}\, trace\left((\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_k\mathbf{\Phi}\mathbf{L}\right) trace\left((\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_s\mathbf{\Phi}\mathbf{L}\right), \quad (36)$$

$$A_2^* = \sum_{k=1}^{r}\sum_{s=1}^{r} \{\mathbf{W}\}_{ks}\, trace\left((\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_k\mathbf{\Phi}\mathbf{L}(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_s\mathbf{\Phi}\mathbf{L}\right), \quad (37)$$

$$B^* = \frac{A_1^* + 6A_2^*}{2l}. \quad (38)$$

Pomocou uvedených vzťahov a porovnaním $E\,[F_{KR}] = \lambda\, E\,[F]$ a $Var\,[F_{KR}] = \lambda^2\; Var\,[F]$ s prvými dvoma momentami $F_{l,m}$ rozdelenia dostávame

$$\lambda = \frac{m}{E[F](m-2)}, \qquad m = 4 + \frac{l+2}{l\rho - 1}, \quad (39)$$

kde

$$\rho = \frac{Var[F]}{2\,\{E[F]\}^2}.$$

Odhady „škálovacej“ konštanty λ a počtu stupňov voľnosti m sú potom také $\widehat{\lambda}$ a $\widehat{m}$, kde v (39) (respektíve vo vzťahoch (36), (37) a (38)) je skutočný vektor kovariančných parametrov nahradený jeho REML odhadom $\widehat{\boldsymbol{\theta}}$.

Kenwardovu-Rogerovu aproximatívnu $(1-\alpha)\cdot 100\%$-nú konfidenčnú oblasť ($\alpha \in (0,1)$) tvorí množina tých vektorov pevných efektov $\boldsymbol{\beta}^+$, pre ktoré je

$$\widehat{\lambda}\frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+)^{'}(\mathbf{L}'\widehat{\mathbf{\Phi}}_{mod}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+) < F_{l,\widehat{m}}(1-\alpha), \quad (40)$$

kde $F_{l,\widehat{m}}(1-\alpha)$ je $(1-\alpha)$-kvantil Fisherovho-Snedecorovho rozdelenia s l a $\widehat{m}$ stupňami voľnosti.

5.3 Modifikovaná Kenwardova-Rogerova metóda

Predpokladajme teraz, že

$$F^* = \lambda\frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})^{'}(\mathbf{L}'\widehat{\mathbf{\Phi}}_{mod1}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}) \quad (41)$$

má približne Fisherovo-Snedecorovo rozdelenie s l a m stupňami voľnosti. Táto testovacia štatistika uvažuje, na rozdiel od vzťahu (34), ako odhad kovariančnej matice odhadu pevných efektov maticu $\widehat{\mathbf{\Phi}}_{mod1}$ danú vzťahom (24). V dizertačnej práci sú na základe postupov uvedených v [1] a [10] odvodené vzťahy na približný výpočet „škálovacej“ konštanty λ a počtu stupňov voľnosti m pre uvažovanú situáciu, pričom tieto sú

$$\lambda = \frac{m}{E[F](m-2)}, \qquad m = 4 + \frac{l+2}{l\rho - 1}, \quad (42)$$

kde

$$\rho = \frac{Var[F]}{2\left\{E[F]\right\}^2}, \quad E\left[F\right] = 1 + \frac{A_2}{l}, \quad Var\left[F\right] = \frac{2}{l}\left(1 + B\right),$$

v ktorých

$$A_1 = \sum_{k=1}^{r}\sum_{s=1}^{r} w_{ks}\, trace\left((\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_k\mathbf{\Phi}\mathbf{L}\right) trace\left((\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_s\mathbf{\Phi}\mathbf{L}\right), \tag{43}$$

$$A_2 = \sum_{k=1}^{r}\sum_{s=1}^{r} w_{ks}\, trace\left((\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_k\mathbf{\Phi}\mathbf{L}(\mathbf{L}'\mathbf{\Phi}\mathbf{L})^{-1}\mathbf{L}^{'}\mathbf{\Phi}\boldsymbol{P}_s\mathbf{\Phi}\mathbf{L}\right) \tag{44}$$

a

$$B = \frac{A_1 + 6A_2}{2l}. \tag{45}$$

Odhady „škálovacej“ konštanty λ a počtu stupňov voľnosti m sú také $\widehat{\lambda}$ a $\widehat{m}$, kde v (42) (resp. v (43), (44) a (45)) je nahradený skutočný vektor kovariančných parametrov jeho REML odhadom $\widehat{\boldsymbol{\theta}}$.

Modifikovanú Kenwardovu-Rogerovu aproximatívnu $(1-\alpha)\cdot 100\%$-nú konfidenčnú oblasť ($\alpha \in (0,1)$) potom tvorí množina tých vektorov pevných efektov $\boldsymbol{\beta}^+$, pre ktoré

$$\widehat{\lambda}\frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+)^{'}(\mathbf{L}'\widehat{\mathbf{\Phi}}_{mod1}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+) < F_{l,\widehat{m}}(1-\alpha), \tag{46}$$

kde $F_{l,\widehat{m}}(1-\alpha)$ je $(1-\alpha)$-kvantil Fisherovho-Snedecorovho rozdelenia s l a $\widehat{m}$ stupňami voľnosti.

Poznámka 1: Všimnime si, že uvedené výsledky sú takmer totožné s výsledkami ponúknutými v článku [10] prezentovanými v časti 5.2, až na použitie w_{kl} vo vzťahoch pre výpočet A_1 a A_2 namiesto použitia (k,l)-tého prvku inverzie Fisherovej informačnej matice $\{\mathbf{W}\}_{kl}$ vo vzťahoch pre výpočet A_1^* a A_2^* daných v (36) a (37).

Poznámka 2: V prípade použitia matice $\widehat{\mathbf{\Phi}}_{mod1}^*$ v uvedenej kvadratickej forme (41), tzn. v prípade, že pre štatistiku danú vzťahom

$$F_1^* = \lambda\frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})^{'}(\mathbf{L}'\widehat{\mathbf{\Phi}}_{mod1}^*\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})$$

predpokladáme, že má $F_{l,m}$ rozdelenie dostaneme za predpokladu nevychýlenosti REML odhadu $\widehat{\boldsymbol{\theta}}$ vzťahy pre približný výpočet „škálovacej“ konštanty λ a počtu stupňov voľnosti m zhodné s tými, ktoré boli uvedené v časti 5.2.

6 Simulačné štúdie

V simulačnej štúdii sme sa zamerali na adekvátnosť použitia metód pre konštrukciu konfidenčných oblastí uvedených v predchádzajúcich častiach pre lineárny zmiešaný model (ale aj pre lineárny model) v prípade, že chyby nie sú korelované alebo spĺňajú autoregresný proces prvého rádu AR(1). Keďže tieto metódy vo všeobecnosti (najmä Kenwardova-Rogerova metóda popísaná v časti 5.2) neboli dostatočne preverené v prípade „malého“ počtu meraní pre longitudinálne dáta (pozri napr. [5], [13]), považovali sme za vhodné overiť ich vlastnosti na základe simulácii a porovnať s vlastnosťami modifikovanej Kenwardovej-Rogerovej metódy odvodenej v dizertačnej práci uvedenej v časti 5.3. Keďže longitudinálne dáta sú charakteristické opakovanými meraniami určitej vlastnosti počas sledovaného časového obdobia na viacerých subjektoch, bolo v simulačnej štúdii potrebné uvažovať „malý“ počet subjektov, „malý“ rozsah vektorov pozorovaní na jednotlivých subjektoch, ale i „malý“ rozsah vektorov pozorovaní na „malom“ počte subjektov.

Na overenie oprávnenosti použitia uvažovaných metód v prípade „malého“ počtu pozorovaní bol v programovacom jazyku *MATLAB* vytvorený program, pomocou ktorého sme na základe simulácií spočítali empirické pravdepodobnosti pokrytia skutočnej hodnoty vektora pevných efektov pre uvažované metódy.

Na základe výsledkov simulačnej štúdie môžeme usudzovať, že empirické konfidenčné oblasti známej lineárnej kombinácie vektora pevných efektov zostrojené pomocou Kenwardovej-Rogerovej metódy s použitím modifikovaného odhadu kovariančnej matice odhadu vektora pevných efektov $\widehat{\Phi}_{mod}$ z článku [10] v prípade menšieho počtu subjektov so zvyšujúcim sa počtom pozorovaní na jednotlivých subjektoch ponúkajú v prípade lineárneho zmiešaného modelu (ale aj lineárneho modelu) s AR(1) chybami takmer porovnateľné výsledky ako empirické konfidenčné oblasti zostrojené pomocou „naivnej“ metódy (pomocou vzťahu (28)), avšak v prípade lineárneho zmiešaného modelu s nekorelovanými chybami ponúka táto metóda o niečo prijateľnejšie výsledky ako „naivná“ metóda, teda empirická hodnota konfidencie je bližšie k jej nominálnej hodnote. Napriek tomu túto vo všetkých uvažovaných prípadoch podhodnocuje. V prípade nižšieho počtu opakovaných meraní na jednotlivých subjektoch sa pre lineárny zmiešaný model s nekorelovanými chybami so zvyšujúcim počtom subjektov empirická hodnota konfidencie približuje požadovanej nominálnej hodnote rýchlejšie ako „naivná“ metóda, a teda požadovanú nominálnu hodnotu konfidencie dosahuje už pre menší počet subjektov. Pre lineárny zmiešaný model (ale aj lineárny model) s AR(1) chybami vykazuje uvedená metóda v prípade nižšieho počtu opakovaných meraní na zvyšujúcom sa počte subjektov takmer porovnateľné výsledky ako „naivná“ metóda.

Empirické konfidenčné oblasti známej lineárnej kombinácie vektora pevných efektov zostrojené pomocou modifikovanej Kenwardovej-Rogerovej metódy s použitím navrhnutého modifikovaného odhadu kovariančnej matice odhadu vektora pevných efektov $\widehat{\Phi}_{mod1}$ dosahujú nominálnu hladinu konfidencie už pre menší počet subjektov, ako aj pre menší rozsah opakovaných meraní na jednotlivých subjektoch ako empirické konfidenčné oblasti zostrojené pomocou Kenwardovej-Rogerovej metódy s použitím modifikovaného odhadu kovariančnej matice odhadu vektora pevných efektov $\widehat{\Phi}_{mod}$ z článku [10] len v prípade nelineárnej štruktúry kovariančných matíc náhodných vektorov $\mathbf{Y}_i$ (v dizertačnej práci sme uvažovali, že chyby spĺňajú AR(1) proces). V prípade lineárnych štruktúr kovariančných matíc náhodných vektorov $\mathbf{Y}_i$ sa odhad kovariančnej matice $\widehat{\boldsymbol{\beta}}$ navrhnutý v dizertačnej práci ($\widehat{\mathbf{\Phi}}_{mod1}$ daný vzťahom (24)) redukuje na odhad kovariančnej matice $\widehat{\boldsymbol{\beta}}$ navrhnutý v článku [10] ($\widehat{\mathbf{\Phi}}_{mod}$ daný vzťahom (17)), a taktiež výraz w_{kl} daný vzťahom (23) sa rovná (k, l)-tému prvku inverzie Fisherovej informačnej matice $\{\mathbf{W}\}_{kl}$. Modifikovaná Kenwardova-Rogerova metóda odvodená v dizertačnej práci ponúka v týchto prípadoch výsledky totožné s Kenwardovou-Rogerovou metódou z [10].

Empirické konfidenčné oblasti známej lineárnej kombinácie vektora pevných efektov zostrojené pomocou Faiovej-Corneliusovej metódy ponúkajú takmer porovnateľné výsledky ako navrhnutá modifikovaná Kenwardova-Rogerova metóda, pričom v niektorých prípadoch dosahujú požadovanú nominálnu hladinu konfidencie už pre nižší počet subjektov. V prípade zvyšujúceho sa počtu pozorovaní na malom počte subjektov, ktoré má na hodnotu empirických pravdepodobností pokrytia konštruovaných pomocou modifikovanej Kenwardovej-Rogerovej metódy len nepatrný vplyv sú hodnoty empirických pravdepodobností pokrytia zostrojené pomocou Faiovej-Corneliusovej metódy s ňou porovnateľné, resp. sú bližšie požadovanej nominálnej konfidenčnej hodnote.

Taktiež je možné konštatovať, že pre lineárny zmiešaný model pre longitudinálne dáta s AR(1) chybami nemá skutočná hodnota autokorelačného koeficientu na vlastnosti konfidenčných oblastí skonštruovanými pomocou „naivnej“, Kenward-Roger, modifikovanej Kenwardovej-Rogerovej a Faiovej-Corneliusovej metódy takmer žiadny vplyv.

7 Záver

V dizertačnej práci bol v kapitole 3 uvedený stručný prehľad dvoch lineárnych modelov pre analýzu longitudinálnych dát spolu s praktickou ukážkou štatisticky správnej voľby modelu pre namerané údaje. Kapitola 4 ponúkla prehľad súčasných metód na odhadovanie kovariančnej matice odhadu vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta spolu s najpoužívanejšími postupmi pre zostrojenie približnej konfidenčnej oblasti známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta v prípade „malého“ počtu pozorovaní. Na základe výsledkov z [11] bol v kapitole 5 navrhnutý nový odhad kovariančnej matice odhadu vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta v prípade neznámych kovariančných parametrov modelu, ak tieto sú odhadované pomocou REML funkcie vierohodnosti, pričom je vynechaná podmienka nevychýlenosti týchto odhadov. Následne bol odvodený postup pre konštrukciu približných konfidenčných oblastí známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta v prípade použitia navrhnutého odhadu kovariančnej matice odhadu vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta za vynechania podmienky nevychýlenosti REML odhadu neznámych kovariančných parametrov modelu. V simulačnej štúdii boli porovnané vlastnosti niektorých približných konfidenčných oblastí v prípade „malého“ počtu pozorovaní, pričom tieto sú porovnané aj s vlastnosťami konfidenčnej oblasti založenej na asymptotických výsledkoch.

Hlavné prínosy dizertačnej práce

- Návrh nového odhadu kovariančnej matice odhadu vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta, vzťah (24).

- Odvodenie postupu pre zostrojenie približnej konfidenčnej oblasti známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta v prípade „malého“ počtu pozorovaní, vzťahy (42),(43), (44), (45) a (46).

- Porovnanie vlastností známych približných konfidenčných oblastí známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta s vlastnosťami navrhnutej približnej konfidenčnej oblasti známej lineárnej kombinácie vektora pevných efektov lineárneho zmiešaného modelu pre longitudinálne dáta v prípade „malého“ počtu pozorovaní, ako aj ich porovnanie s vlastnosťami konfidenčnej oblasti založnej na základe asymptotických výsledkoch.

8 Summary

The thesis mainly deals with the construction of approximate confidence regions for known linear combination of the vector of fixed effects in the linear mixed model for longitudinal data. Following [11] a new modification of the estimator of covariance matrix for the estimator of fixed effects in the linear mixed model is proposed for the case of unknown covariance parameters of the model, where these are estimated using the REML likelihood function. Using this modified estimator of covariance matrix, based on [1] and [10], a method to construct an approximate confidence region for known linear combination of the vector of fixed effects in the linear mixed model for longitudinal data is derived. A simulation study compares the quality of the proposed confidence region with qualities of the commonly used approximate confidence regions presented in [4], [8] and [10] in the case of "small" sample sizes. Properties of all these confidence regions are compared also with properties of confidence region based on asymptotic results.

9 Zoznam publikácií autora dizertačnej práce

Články uverejnené v časopisoch

- Wimmer, G. (2009), *Niektoré vlastnosti odhadu regresných parametrov v lineárnom zmiešanom modeli*, Forum Statisticum Slovacum, No. 7, 199 – 202
- Wawruch, M., Dukat, A., Murin, J., Wsolova, L., Kuzelova, M., Macugova, A., **Wimmer, G.**, Shah, R. (2009), *The effect of selected patient's characteristics on the choice of antihypertensive medication in the elderly in Slovakia*, Pharmacoepidemiology and Drug Safety, Vol. 18, Issue 12, 1199 – 1205
- Wimmer, G. (2008), *Statistical method based on confidence and prediction regions for analysis of volatile organic compounds in human breath gas*, Measurement Science Review 2008, Vol. 8, 111 – 113

Príspevky uverejnené v zborníkoch konferencií

- Wimmer, G. (2009): *Confidence Region in Linear Mixed Model for Longitudinal Data*, MEASUREMENT 2009: 7th International Conference on Measurement, 37 – 40
- Wimmer, G. (2009): *Algoritmus výpočtu približných konfidenčných intervalov parametra polohy z digitalizovaných meraní*, ROBUST'2008, Sborník prací 15. zimní školy JČMF, J. Antoch & G. Dohnal (eds.), Praha, JČMF, 505 -– 512
- Wimmer, G. (2009): *Vyhodnocovanie mnohorozmerných údajov pomocou lineárneho zmiešaného modelu*, Firma a konkurenční prostředí

2009: Sborník z mezinárodní vědecké konference, 1. Část, P. Žufan (eds.), MSD, spol. s.r.o., 355 -– 357

- Wimmer, G. (2007): *Confidence and prediction regions for statistical analysis of volatile organic compounds in human breath gas*, MEASUREMENT 2007: 6th International Conference on Measurement, 102 – 105

10 Literatúra

[1] Alnosaier, W., S. (2007), *Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models*, A DISSERTATION, Oregon State University

[2] Demidenko, E. (2004), *Mixed Models Theory and Applications*, Wiley Series in Probability and Statistics, New Jersey

[3] Diggle, P., J., Heagerty, P., Liang, K., Y., Zeger, S., L. (2002), *Analysis of Longitudinal Data*, Oxford University Press

[4] Fai, A., H., T., Cornelius, P., L. (1996), *Approximate F-tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-Plot Experiments*, Journal of Statistical Computation and Simulation, Vol. 54, Issue 4, 363,-378

[5] Fitzmaurice, G.M., Laird, N.M., Ware, J.H. (2004), *Applied Longitudinal Analysis*, John Wiley & Sons, New Jersey

[6] Harville, D., A. (1977), *Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems*, Journal of the American Statistical Association, Vol. 72, No. 358, 320 – 338

[7] Harville, D., A., Jeske, D., R. (1992), *Mean Squared Error of Estimation or Prediction Under a General Linear Model*, Journal of the American Statistical Association, Vol. 87, No. 419, 724 – 731

[8] Chatterjee, S., Lahiri, P., Li, H. (2008), *Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Linear Mixed Models*, The Annals of Statistics, Vol. 36, No. 3., 1221 – 1245

[9] Kackar, R.N., Harville, D.A. (1984), *Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models*, Journal of the American Statistical Association, Vol. 79, No. 388, 853 – 862

[10] Kenward, M.G., Roger, J.H. (1997), *Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood*, Biometrics, Vol.53, No. 3, 983 – 997

[11] Kenward, M.G., Roger, J.H. (2009), *An Improved Approximation to the Precision of Fixed Effects from Restricted Maximum Likelihood*, Computational Statistics and Data Analysis, doi:10.1016/j.csda.2008.12.013

[12] Laird, N., M., Ware, J., H. (1982), *Random-Effects Models for Longitudinal Data*, Biometrics, Vol. 38, No. 4, 963 – 974

[13] Manor, O., Zucker, D., M. (2004), *Small Sample Inference for the Fixed Effects in the Mixed Linear Model*, Computational Statistics & Data Analysis, No. 46., 801 – 817

[14] Prasad, N., G., N., Rao, J., N., K. (1990) *The Estimation of the Mean Squared Error of Small-Area Estimators*, Journal of the American Statistical Association, Vol. 85, No. 409., 163 – 171

[15] Satterthwaite, F., E. (1941), *Synthesis of Variance*, Psychometrika, Vol. 6. No. 5., 309 – 316

[16] Searle, S.R., Casella, G., McCulloch, Ch., E. (1992), *Variance Components*, JOHN WILEY AND SONS