

Odhad regresných parametrov lineárneho zmiešaného modelu pre longitudinálne dáta

Seminář PAST
Matematicko-fyzikální fakulta
Univerzita Karlova v Praze

Gejza Wimmer jr.

Praha
18. apríl 2012

Úvod

- longitudinálne experimenty zohrávajú dôležitú úlohu v rôznych bio-medicínskych, sociologických, psychologických, či behaviorálnych výskumoch
 - opakované merania určitej vlastnosti na viacerých subjektoch v rôznych časových okamihoch
 - na základe získaných pozorovaní odhaliť vplyv času na meranú veličinu
 - popísať zmenu meranej veličiny počas sledovaného časového obdobia
- pri analýze sa v posledných rokoch vo veľkej miere využíva lineárny zmiešaný model
 - k odhadovaniu neznámych parametrov modelu je vhodné použiť REML funkciu vierohodnosti
 - presné rozdelenie takto získaného odhadu pevných efektov (regresných parametrov) nie je vo všeobecnosti známe
 - štatistické inferencie ohľadne nich sú prevažne založené na asymptotických vlastnostiach odhadu regresných parametrov v prípade známych kovariančných parametrov modelu

Glykemický test

- údaje predstavené v článku od Zerbe, G., O. (1979)

Zerbe, G., O. (1979), *Randomization Analysis of the Completely Randomized Design Extended to Growth and Response Curves*, Journal of the American Statistical Association, Vol. 74, No. 365, 215 – 221

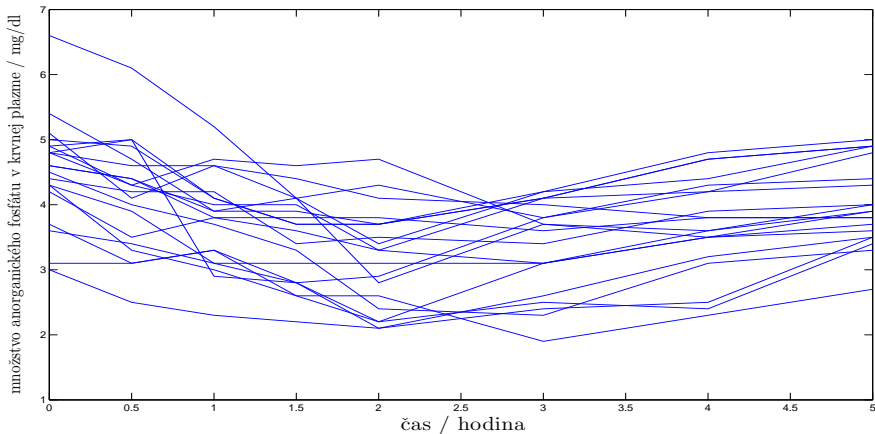
- štandardným glykemickým testom na 20 obéznych pacientoch
 - pre každý sledovaný subjekt bolo merané množstvo anorganického fosfátu v krvnej plazme (mg/dl) v časoch 0, 0.5, 1, 1.5, 2, 3, 4, 5 hodiny po orálnom požití roztoku glukózy
- snaha popísať časový priebeh množstva anorganického fosfátu v krvnej plazme obéznych pacientov počas sledovaného časového úseku

Glykemický test

<i>hodina</i>	<i>0</i>	<i>0.5</i>	<i>1</i>	<i>1.5</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
<i>subjekt</i>								
1	4.3	3.3	3.0	2.6	2.2	2.5	2.4	3.4
2	5.0	4.9	4.1	3.7	3.7	4.1	4.7	4.9
3	4.6	4.4	3.9	3.9	3.7	4.2	4.8	5.0
4	4.3	3.9	3.1	3.1	3.1	3.1	3.6	4.0
5	3.1	3.1	3.3	2.6	2.6	1.9	2.3	2.7
6	4.8	5.0	2.9	2.8	2.2	3.1	3.5	3.6
7	3.7	3.1	3.3	2.8	2.9	3.8	4.3	4.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tabuľka: Množstvo anorganického fosfátu v krvnej plazme (mg/dl) pre 7 vybraných sledovaných subjektov počas obdobia 0, 0.5, 1, 1.5, 2, 3, 4, 5 hodiny po orálnom požití roztoku glukózy.

Glykemický test



Množstvo anorganického fosfátu v krvnej plazme (mg/dl) pre 20 sledovaných subjektov počas obdobia 0, 0.5, 1, 1.5, 2, 3, 4, 5 hodiny po orálnom požití roztoku glukózy.

Hmotnosť prasíat

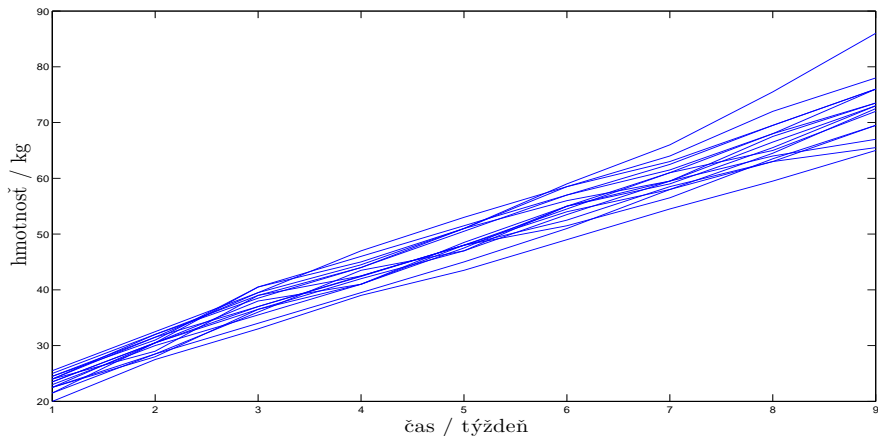
- údaje predstavené v knihe od Diggle, P., J., Heagerty, P., Liang, K., Y., Zeger, S., L. (2002)
Diggle, P., J., Heagerty, P., Liang, K., Y., Zeger, S., L. (2002), *Analysis of Longitudinal Data*, Oxford University Press
- týždenné meranie hmotnosti 16 prasíat počas 9 týždňov
- z nameraných hodnôt popísať tento spoločný trend rastúcej hmotnosti zvierat v danom časovom období

Hmotnosť prasíat

týždeň	1	2	3	4	5	6	7	8	9
<i>prasa</i>									
1	24.0	32.0	39.0	42.5	48.0	54.5	61.0	65.0	72.0
2	22.5	30.5	40.5	45.0	51.0	58.5	64.0	72.0	78.0
3	22.5	28.0	36.5	41.0	47.5	55.0	61.0	68.0	76.0
4	24.0	31.5	39.5	44.5	51.0	56.0	59.5	64.0	67.0
5	24.5	31.5	37.0	42.5	48.0	54.0	58.0	63.0	65.5
6	23.0	30.0	35.5	41.0	48.0	51.5	56.5	63.5	69.5
7	22.5	28.5	36.0	43.5	47.0	53.5	59.5	67.5	73.5
8	23.5	30.5	38.0	41.0	48.5	55.0	59.5	66.5	73.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Tabuľka: Hmotnosti 8 vybraných prasíat merané počas 9 týždňov v kilogramoch.

Hmotnosť prasíat



Hmotnosti 16 prasíat merané počas 9 týždňov v kilogramoch.

Možné "komplikácie" pri analýze

- vzniká množstvo možných komplikácií pri ich ďalšom spracovávaní
 - pozorujeme rôzne subjekty v nejakých časových intervaloch
 - zrejme je nevhodné uvažovať, že pozorovania získané na danom subjekte budú navzájom nezávislé
 - naopak, jednotlivé vektory pozorovaní na rôznych subjektoch budú zrejme navzájom nezávislé
 - vzájomná korelácia jednotlivých opakovaných meraní na každom subjekte
 - praxi takmer nikdy nepoznáme
 - je potrebné ju odhadovať
- problém nevyváženého dizajnu
- využívajú zložitejšie modely
 - špeciálne lineárne modely
 - povoľujú rôzne tvary
 - stredné hodnoty pozorovaných vektorov meraní
 - kovariančné štruktúry vektorov meraní

Longitudinálne dáta a ich analýza

- hlavná črta longitudinálnych experimentov je, že jednotlivé subjekty sú merané opakovane v čase
 - pre i -ty ($i = 1, \dots, I$) subjekt dostaneme n_i -rozmerný vektor (\mathbf{y}_i) opakovaných meraní získaných v rôznych časových okamihoch
- primárny cieľ analýzy longitudinálnych dát je popísať spoločnú črtu pozorovaných subjektov
- navyše je možné zachytiť aj individuálne efekty jednotlivých subjektov
- predpoklady
 - jednotlivé vektory pozorovaní na rôznych subjektoch sú navzájom nezávislé
 - pozorovania získané na danom subjekte vykazujú (zväčša kladnú) koreláciu
 - korelácia medzi opakovanými meraniami musí byť zahrnutá v analýze
 - spôsobuje (značné) komplikácie pri ďalšom spracovaní

Lineárny Zmiešaný Model

- pre longitudinálne dáta začali používať Laird a Ware (1982)
Laird, N., M., Ware, J., H. (1982), *Random-Effects Models for Longitudinal Data*, Biometrics, Vol. 38, No. 4, 963–974
- výhodný, pretože už v samotnom zápise zohľadňuje vplyv jednotlivých subjektov na svoje opakované merania
 - tieto individuálne efekty sa môžu pre jednotlivé subjekty navzájom odlišovať
 - odhadovať celkovú zmenu spoločnú pre všetky subjekty
 - navyše pre každý subjekt predikovať jeho odchýlku od spoločného priemeru
- pre vektor pozorovaní \mathbf{y}_i na i -tom subjekte platí

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

\mathbf{X}_i sú známe ($n_i \times p$)-rozmerné matice plánu

$\boldsymbol{\beta}$ je neznámy p -rozmerný vektor parametrov

- sú pre všetky subjekty rovnaké

- pevné efekty

\mathbf{Z}_i sú známe ($n_i \times r$)-rozmerné matice

$\boldsymbol{\eta}_i$ sú r -rozmerné vektory neznámych parametrov

- pre každý subjekt sa navzájom odlišujú

- považujeme ich za náhodné, navzájom nezávislé s rozdelením $N(\mathbf{0}, \mathbf{D})$

- individuálne náhodné efekty jednotlivých subjektov na svoje opakované merania

$\boldsymbol{\varepsilon}_i$ je n_i -rozmerný vektor náhodných chýb i -teho subjektu

- predpokladáme jeho nezávislosť od $\boldsymbol{\eta}_i$

- rozdelenie $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_i \\ \vdots \\ \mathbf{Y}_I \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_i \\ \vdots \\ \mathbf{X}_I \end{bmatrix} \beta + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_i & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{Z}_I \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_i \\ \vdots \\ \eta_I \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_I \end{bmatrix}$$

$$\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2, \dots, \mathbf{Y}'_I)' \text{ a } \mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_I)'$$

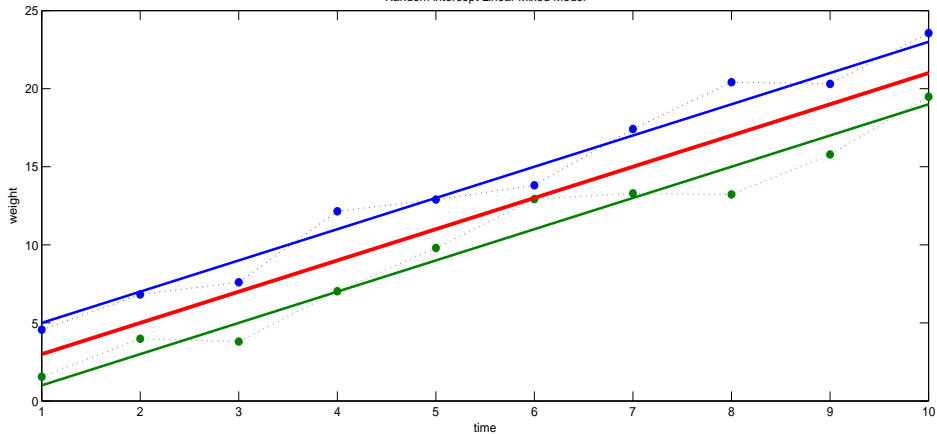
kovariančná matica náhodného vektora pozorovaní \mathbf{Y}

$$\text{Var}(\mathbf{Y}) = \Sigma = \begin{bmatrix} \Sigma_1 & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Sigma_2 & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Sigma_i & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \Sigma_I \end{bmatrix},$$

kde jednotlivé Σ_i predstavujú kovariančnú maticu náhodného vektora pozorovaní i -teho subjektu \mathbf{Y}_i

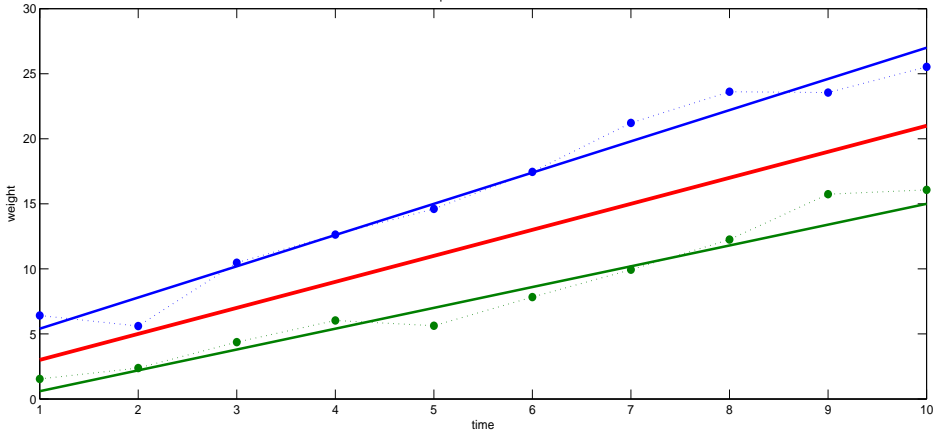
$$\text{Var}(\mathbf{Y}_i) = \Sigma_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i + \mathbf{R}_i$$

Random intercept Linear Mixed Model



$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \eta_{0i} + \varepsilon_{ij}$$

Random intercept and trend Linear Mixed Model



$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \eta_{0i} + \eta_{1i} t_{ij} + \varepsilon_{ij}$$

Odhady neznámých regresných parametrov

predpokladajme, že kovariančná matica náhodného vektora pozorovaní \mathbf{Y} je známou funkciou r -rozmerného vektora parametrov θ , a teda $\text{Var}(\mathbf{Y}) = \mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$

- θ je známy vektor parametrov, a teda aj $\mathbf{\Sigma}$ je známa matica a za predpokladu existencie matice k nej inverznej, je potom odhad pevných efektov

$$\tilde{\beta} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{Y}$$

- predpoklad existencie matice $(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} = \Phi$
- takto získaný odhad zodpovedá riešeniu problému odhadovania neznámých parametrov váženou metódou najmenších štvorcov (VMNS) a za daných predpokladov platí

$$(\tilde{\beta} - \beta) \sim N\left(\mathbf{0}, (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\right)$$

- najlepší lineárny nevychýlený odhad (*BLUE* - best linear unbiased estimator) vektora pevných efektov β

Odhad neznámych kovariančných parametrov

- z hľadiska analýzy longitudinálnych dát nás primárne zaujímajú odhady neznámych regresných parametrov β
- na ich odhadovanie je často potrebné odhadnúť aj neznáme kovariančné parametre modelu θ - nežiaduce
 - vo veľkej miere využíva REML funkcia vierohodnosti, pričom jej logaritmus je v tvare

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

$$\text{kde } n = \sum_{i=1}^I n_i$$

- pre odhady kovariančných parametrov $\hat{\theta}$
 - asymptotické rozdelenie (pre $n \rightarrow \infty$); (napr. Searle, S., R., Casella, G., McCulloch, Ch., E. (1992), *Variance Components*, JOHN WILEY AND SONS)

$$\left(\hat{\theta} - \theta_0 \right) \sim N(\mathbf{0}, \mathbf{W}),$$

kde \mathbf{W} je inverzia Fisherovej informačnej matice

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

a θ_0 je skutočný vektor kovariančných parametrov modelu

- pomocou funkcie vierohodnosti - ML odhad
- maximálne invariantný odhad - REML odhad
 - REML funkcia vierohodnosti je hustota pre $\mathbf{A}'\mathbf{y}$, kde $\mathbf{A}'\mathbf{X} = \mathbf{0}$
 - REML odhad kovariačných parametrov θ , ozn. $\hat{\theta}$, maximalizuje REML funkciu vierohodnosti, resp. jej logaritmus
- prečo REML a nie ML
 - ML odhad nie je nevychýlený
 - REML “redukuje (zjemňuje)” odchýlku
 - ML odhaduje regresné aj kovariančné parametre modelu súčasne
 - REML odhaduje len kovariančné parametre modelu

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

Štruktúry pre kovariančnú maticu v lineárnom zmiešanom modeli pre longitudinálne dáta

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

- $\text{Var}(\mathbf{Y}) = \mathbf{ZD}^*\mathbf{Z}' + \mathbf{R}^* = \boldsymbol{\Sigma} = \text{diag}[\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_i, \dots, \boldsymbol{\Sigma}_I]$
- $\text{Var}(\mathbf{Y}_i) = \boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R}_i$
- $\mathbf{D}^* = \text{diag}[\mathbf{D}, \mathbf{D}, \dots, \mathbf{D}]$
- $\mathbf{R}^* = \text{diag}[\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_i, \dots, \mathbf{R}_I]$
- $\boldsymbol{\Sigma}$ je zväčša neznáma a treba odhadovať i jej hodnotu
- tieto odhady ovplyvňujú aj efektívnosť odhadov neznámych regresných parametrov
- výhodné uvažovať rôzne štruktúry pre kovariančnú maticu vektora opakovaných pozorovaní

- nezávislé náhodné chyby $\mathbf{R}_i = \sigma^2\mathbf{I}_{n_i}$
- MA(p) model

- MA(1) model - pre náhodné chyby platí, že $\varepsilon_{i,j} = \tau_{i,j} + \theta\tau_{i,j-1}$, kde $\tau_{i,j} \sim N(0, \sigma^2)$ a θ je koeficient

$$\text{MA(1) modelu } \mathbf{R} = \sigma^2 \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 + \theta^2 & -\theta & 0 & \dots & 0 \\ -\theta & 1 + \theta^2 & -\theta & \dots & 0 \\ 0 & -\theta & 1 + \theta^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 + \theta^2 \end{bmatrix}$$

- AR(p) model
- ARMA(p,q) model

- ARMA(1, 1) - chyby spĺňajú autoregresný-moving average model prvého rádu

$$\varepsilon_{i,j} = \rho\varepsilon_{i,j-1} + \tau_{i,j} + \theta\tau_{i,j-1}, \text{ kde } \tau_{i,j} \sim N(0, \sigma^2), \rho \text{ je AR koeficient a } \theta \text{ je MA koeficient}$$

Lineárny zmiešaný model pre Longitudinálne Dáta s AR(1) chybami

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

- nech pre i -ty subjekt platí
 - $\boldsymbol{\eta}_i \sim N_r(\mathbf{0}, \mathbf{D})$
 - $\varepsilon_{i,j} = \rho\varepsilon_{i,j-1} + \tau_{i,j}$
 - $\tau_{i,j} \sim N(0, \sigma^2)$ a ρ je autoregresný koeficient
 - stacionárny prípad, teda $|\rho| < 1$
 - kovariančná matica vektora chýb má tvar

$$\sigma^2\mathbf{R}_i = \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n_i-1} \\ \rho & 1 & \rho & \dots & \rho^{n_i-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n_i-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n_i-1} & \rho^{n_i-2} & \rho^{n_i-3} & \dots & 1 \end{bmatrix}$$

- parametrizácia matice \mathbf{D}

- $\mathbf{D} = \sigma_b^2\mathbf{I}_r$, $\mathbf{D} = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \sigma_2^2 & \dots & \sigma_2^2 \\ \sigma_2^2 & \sigma_1^2 & \sigma_2^2 & \dots & \sigma_2^2 \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_2^2 & \sigma_2^2 & \dots & \sigma_2^2 & \sigma_1^2 \end{bmatrix}$, všeobecný tvar

- potom $\boldsymbol{\theta} = (d_{11}, d_{12}, \dots, d_{22}, \dots, d_{rr}, \sigma^2, \rho)'$ je vektor všetkých kovariančných komponentov

Vlastnosti odhadu pevných efektov

- okrem vektora pevných efektov sú neznáme navyše aj kovariančné parametre modelu θ
 - Σ neznáma
- máme k dispozícii jej REML odhad, označme $\hat{\Sigma} = \Sigma(\hat{\theta})$
- prirodzeným postupom k získaniu odhadu neznámeho vektora pevných efektov β

$$\hat{\beta} = \left(\mathbf{X}' \Sigma^{-1}(\hat{\theta}) \mathbf{X} \right)^{-1} \mathbf{X}' \Sigma^{-1}(\hat{\theta}) \mathbf{Y}$$

- empirický najlepší lineárny nevychýlený odhad (*EBLUE* - empirical best linear unbiased estimator)
- podľa Demidenko (2004) nevychýlený odhad vektora pevných efektov modelu β

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

$$E(\hat{\beta}) = \beta$$

- kovariančná matica $\hat{\beta}$ žiaľ vo všeobecnosti nie je známa
- zväčša je odhadovaná ako

$$\hat{\Phi} = \left(\mathbf{X}' \Sigma^{-1}(\hat{\theta}) \mathbf{X} \right)^{-1}$$

- pre „malé“ rozsahy výberov je uvedená aproximácia $\text{Var}(\hat{\beta})$ pomocou $\hat{\Phi}$ nevhodná
 - nezahŕňa v sebe neistotu spojenú s odhadovaním neznámych kovariančných parametrov modelu θ
 - $\hat{\Phi}$ nie je nevychýleným odhadom Φ
- Kackar, Harville (1984) ukázali, že približná kovariančná matica *EBLUE* pre pevné efekty modelu β je

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

$$\text{Var}(\hat{\beta}) = \Phi + \Lambda_1,$$

kde

$$\Lambda_1 \approx \Phi \left\{ \sum_{k=1}^r \sum_{l=1}^r \{W\}_{kl} (\mathbf{Q}_{kl} - \mathbf{P}_k \Phi \mathbf{P}_l) \right\} \Phi,$$

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

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

- je funkciou kovariančných parametrov modelu, ktoré však nie sú známe
 - je potrebné ju pre ďalšiu analýzu odhadnúť

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\Phi} + \hat{\Lambda}_1$$

- Kenward, Roger (1997) odvodili, na základe prác Harville, Jeske (1992) a Prasad, Rao (1990), aproximáciu $\widehat{\text{Var}}(\widehat{\beta})$ tak, aby v sebe zahrнула aj odchýlku $\widehat{\Phi}$

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

kde

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

- uvedený postup predpokladá nevychýlenosť odhadu kovariančných parametrov modelu θ
- REML odhady kovariančných parametrov však vo všeobecnosti nie sú nevychýlené
 - odchýlka sa vo väčšej miere prejavuje najmä v prípadoch, ktoré predpokladajú nelineárne štruktúry kovariančnej matice náhodného vektora pozorovaní \mathbf{Y}

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

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

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

- Kenward, Roger (2009) odvodzujú odchýlku REML odhadu $\hat{\theta}$ kovariančných parametrov modelu

$$\text{Bias}(\hat{\theta}_k) = E \left[\left(\hat{\theta} - \theta_0 \right)_k \right] = -\frac{1}{4} \sum_{l=1}^r \sum_{i=1}^r \sum_{j=1}^r \{ \mathbf{W} \}_{ij} \{ \mathbf{W} \}_{kl} \text{trace} \left[\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_k \partial \theta_l} \mathbf{C}_k \right],$$

kde

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

- zahrnuli do postupu k získaniu odhadu kovariančnej matice odhadu vektora pevných efektov z ich predchádzajúceho článku, čím dosiahli modifikovaný odhad kovariančnej matice odhadu vektora pevných efektov

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

kde

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

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

Ešte maličké vylepšenie

- zahrnúť znalosť o vzťahu pre výpočet odchýlky REML odhadu $\hat{\theta}$ už do postupu uvedeného v Kackar, Harville (1984)
 - pri výpočte korekčného člena $\mathbf{\Lambda}_1$
 - do odhadu kovariančnej matice $\hat{\beta}$ nielen pomocou korekčného člena $\hat{\mathbf{B}}$ matice $\hat{\Phi}_{mod1}^*$ vzhľadom na maticu $\hat{\Phi}_{mod}$, ale aj do člena

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

- upustením od predpokladu nevychýlenosti REML odhadu kovariančných parametrov modelu θ

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

kde

$$w_{kl} = E \left[\left(\hat{\theta} - \theta_0 \right)_k \left(\hat{\theta} - \theta_0 \right)_l \right] = \left\{ \mathbf{W} \right\}_{kl} + Bias(\hat{\theta}_k) Bias(\hat{\theta}_l)$$

Konfidenčné oblasti pevných efektov

- inferencie ohľadne / lineárnych kombinácií prvkov β
 - zostrojenie konfidenčnej oblasti pre $\mathbf{L}'\beta$, kde \mathbf{L} je známa ($p \times l$) rozmerná matica

Kovariančné parametre modelu θ sú známe

- poznáme rozdelenie $\tilde{\beta}$ a teda

$$\chi^2 = (\mathbf{L}'\tilde{\beta} - \mathbf{L}'\beta)'(\mathbf{L}'\Phi\mathbf{L})^{-1}(\mathbf{L}'\tilde{\beta} - \mathbf{L}'\beta)$$

- asymptoticky presné konfidenčné oblasti pre $\mathbf{L}\beta$

Kovariančné parametre modelu θ sú neznáme

- nahradiť θ ich REML odhadmi $\hat{\theta}$
 - $\Phi = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$ sa nahradí $\hat{\Phi} = (\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}$
- predpokladať, že $\chi_*^2 = (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)'(\mathbf{L}'\hat{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)$ má taktiež χ^2 rozdelenie s l stupňami voľnosti
 - prevažná časť základnej literatúry o longitudinálnych dátach
 - použiteľné pre „veľké“ rozsahy výberov

- v prípade dostatočného počtu pozorovaní je predchádzajúci prístup prijateľný, pretože odhady kovariančných parametrov modelu θ (resp. Σ_i) vnášajú do inferencií ohľadne prvkov β navyše len zanedbateľnú neistotu

θ sú neznáme; "malý" počet pozorovaní

- neistota spojená s odhadovaním kovariančných parametrov modelu θ musí byť zahrnutá do inferencií ohľadne regresných parametrov β
 - použitím vhodnejšieho odhadu kovariančnej matice pre $\hat{\beta}$
 - použitím F rozdelenia namiesto klasického χ^2 rozdelenia
- rozdelenie pre

$$F = \frac{1}{l}(\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)'(\mathbf{L}'\hat{\Phi}^*\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)$$

- $\hat{\Phi}^*$ je buď „naivný“ odhad kovariančnej matice vektora pevných efektov modelu $\hat{\Phi}$, alebo jej modifikovaný odhad $\hat{\Phi}_{mod}$
- pre $l = 1$ - Satterthwaite (1941)
- pre $l > 1$ - zložitejšie
 - zohľadniť náhodnú štruktúru $\mathbf{L}\hat{\Phi}_{mod}\mathbf{L}'$
 - s výberovou kovariančnou maticou - Hotellingovo T^2 rozdelenie
 - vzťah medzi rozdeleniami T^2 a F
$$F_{l,m-l+1} = ((m-l+1)/(ml)) \cdot T^2(l, m)$$
- jednoduché simulačné štúdie F rozdelenia pri pevnom l

Faiova-Corneliusova metóda

- odhadnúť počet stupňov voľnosti m
- predpokladá sa, že

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

má približne $F_{l,m}$ rozdelenie

- uvedená metóda používa ako odhad kovariančnej matice *EBLUE*-u pevných efektov maticu $\hat{\boldsymbol{\Phi}}$
- využitím spektrálnej dekompozície matice $(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}$
- pomocou Satterthwaiteovej aproximácie

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

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

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

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

Kenwardova-Rogerova metóda

- odhadnúť počet stupňov voľnosti m
- založená na predpoklade, že štatistika

$$F_{KR} = \lambda \frac{1}{T} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)' (\mathbf{L}'\hat{\Phi}_{mod}\mathbf{L})^{-1} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)$$

má približne $F_{l,m}$ rozdelenie

- využíva ako odhad kovariančnej matice *EBLUE*-u pevných efektov maticu $\hat{\Phi}_{mod}$ z článku Kenward, Roger (1997)
- navyše tu vystupuje aj „škálovacia“ konštanta λ - je potrebné ju taktiež odhadovať
- pomocou Taylorovho rozvoja druhého rádu matice $(\mathbf{L}'\hat{\Phi}_{mod}\mathbf{L})^{-1}$ okolo skutočného vektora kovariančných parametrov θ_0
- jeho využitím pri výpočte prvých dvoch momentov náhodnej premennej $F = \frac{1}{\lambda} F_{KR}$

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

$$\hat{\lambda} \frac{1}{T} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta^+)' (\mathbf{L}'\hat{\Phi}_{mod}\mathbf{L})^{-1} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta^+) < F_{l,\hat{m}}(1 - \alpha),$$

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

$$E[F] = \frac{l+A_2^*}{l}, \quad \text{Var}[F] = \frac{2+2B^*}{l},$$

kde

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

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

$$B^* = \frac{A_1^* + 6A_2^*}{2l}$$

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

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

kde

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

Modifikovaná Kenwardova-Rogerova metóda

- predpokladajme, že

$$F^* = \lambda \frac{1}{7} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)' (\mathbf{L}'\hat{\Phi}_{mod1}\mathbf{L})^{-1} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta)$$

má približne Fisherovo-Snedecorovo rozdelenie s l a m stupňami voľnosti

- odhad kovariančnej matice odhadu pevných efektov maticu $\hat{\Phi}_{mod1}$
- na základe postupov uvedených v Alnosaier (2007) a Kenward, Roger (1997) sú odvodené vzťahy na približný výpočet „škálovacej“ konštanty λ a počtu stupňov voľnosti m

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

$$\hat{\lambda} \frac{1}{7} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta^+)' (\mathbf{L}'\hat{\Phi}_{mod1}\mathbf{L})^{-1} (\mathbf{L}'\hat{\beta} - \mathbf{L}'\beta^+) < F_{l, \hat{m}}(1 - \alpha),$$

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

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

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

kde

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

v ktorých

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

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

a

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

$$w_{kl} = E \left[\left(\hat{\theta} - \theta_0 \right)_k \left(\hat{\theta} - \theta_0 \right)_l \right] = \{\mathbf{W}\}_{kl} + \text{Bias}(\hat{\theta}_k) \text{Bias}(\hat{\theta}_l)$$

- uvedené výsledky sú takmer totožné s výsledkami ponúknutými v článku Kenward, Roger (1997), 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^*

Simulačné štúdie

- adekvátnosť použitia uvedených metód pre konštrukciu konfidenčných oblastí pre lineárny zmiešaný model v prípade, že chyby nie sú korelované alebo splňajú autoregresný proces prvého rádu AR(1)
 - tieto metódy vo všeobecnosti (najmä Kenwardova-Rogerova metóda) neboli dostatočne preverené v prípade „malého“ počtu meraní pre longitudinálne dáta
 - overiť ich vlastnosti na základe simulácii a porovnať s vlastnosťami modifikovanej Kenwardovej-Rogerovej metódy
 - v simulačnej štúdii bolo potrebné uvažovať
 - „malý“ počet subjektov
 - „malý“ rozsah vektorov pozorovaní na jednotlivých subjektoch
 - „malý“ rozsah vektorov pozorovaní na „malom“ počte subjektov
- na základe simulácií sme spočítali empirické pravdepodobnosti pokrytia skutočnej hodnoty vektora pevných efektov pre rôzne uvažované metódy

Výsledky simulačnej štúdie

- Kenwardova-Rogerova metóda
 - v prípade lineárneho zmiešaného modelu s AR(1) chybami takmer porovnateľné výsledky ako empirické konfidenčné oblasti zostrojené pomocou „naivnej“ metódy
 - v prípade lineárneho zmiešaného modelu s nekorelovanými chybami ponúka táto metóda o niečo prijateľnejšie výsledky ako „naivná“ metóda
- modifikovaná Kenwardova-Rogerova metóda
 - v prípade nelineárnej štruktúry kovariančných matíc náhodných vektorov \mathbf{Y}_i 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
- Faiova-Corneliusova metóda
 - ponúka takmer porovnateľné výsledky ako modifikovaná Kenwardova-Rogerova metóda, pričom v niektorých prípadoch dosahuje požadovanú nominálnu hladinu konfidencie už pre nižší počet subjektov, ako aj pre menší rozsah opakovaných meraní na jednotlivých subjektoch

Dodatok

Aproximačné Konfidenčné Oblasti pre Regresné Parametre ak θ sú Neznáme - Parametrický Bootstrap metóda

- predpokladáme, že $\mathbf{y} \sim N(\mathbf{X}\beta, \Sigma^*(\theta))$
- odhadnúť neznáme parametre modelu β a θ
- generovať $\bar{\mathbf{y}} \sim N(\mathbf{X}\hat{\beta}, \Sigma^*(\hat{\theta}))$
- odhadnúť $\hat{\beta}^*$ a $\hat{\theta}^*$ z $\bar{\mathbf{y}}$
- na základe získaných odhadov $\hat{\beta}^*$ a $\hat{\theta}^*$ vyčísliť príslušný hodnotu

$$\Gamma_b = (\mathbf{L}\hat{\beta}^* - \mathbf{L}\hat{\beta})' (\mathbf{L}\hat{\Phi}^* \mathbf{L}')^{-1} (\mathbf{L}\hat{\beta}^* - \mathbf{L}\hat{\beta})$$

- posledné tri kroky opakovať B -krát
- spočítať príslušný empirický kvantil $Kv_{0.95}$ z $\Gamma = (\Gamma_1, \dots, \Gamma_B)$
- zostrojiť príslušnú konfidenčnú oblasť pomocou

$$(\mathbf{L}\hat{\beta} - \mathbf{L}\beta)' (\mathbf{L}\hat{\Phi} \mathbf{L}')^{-1} (\mathbf{L}\hat{\beta} - \mathbf{L}\beta) < Kv_{0.95}$$

- založené na predpoklade, že ak máme rozdelenie s hustotou $F_\zeta(\cdot)$ závislou od nejakého parametra ζ a máme jeho konzistentný odhad $\hat{\zeta}$, tak $\left\| F_{\hat{\zeta}}(\cdot) - F_\zeta(\cdot) \right\|_\infty \xrightarrow{pp} 0$

Simulačná štúdia A - Spoločné črty

- Príklad 1 - Glykemický test
- množstvo anorganického fosfátu v krvnej plazme (mg/dl) v časoch 0, 0.5, 1, 1.5, 2, 3, 4, 5 hodiny po orálnom požití roztoku glukózy
- lineárny zmiešaný model

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\eta}_i + \varepsilon_i,$$

- pričom

$$\mathbf{X}_i = \begin{bmatrix} 1 & 0.0 & 0 \\ 1 & 0.5 & 0 \\ 1 & 1.0 & 0 \\ 1 & 1.5 & 0 \\ 1 & 2.0 & 0 \\ 1 & 2.0 & 1 \\ 1 & 2.0 & 2 \\ 1 & 2.0 & 3 \end{bmatrix}$$

pre každé i

- $\boldsymbol{\beta} = [4.5, -0.7, 0.3]'$

Simulačná štúdia A

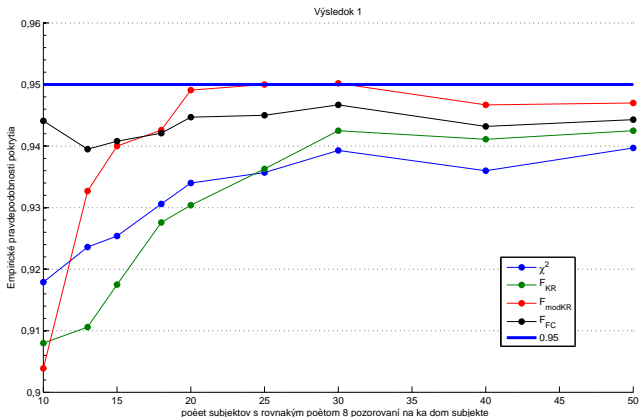
Pokus číslo 1

- matice plánu $\mathbf{Z}_i = [1, 1, 1, 1, 1, 1, 1, 1]'$ pre každé i
- η_i sú jednorozmerné náhodné efekty pochádzajúce z rozdelenia $N(0, \sigma_\eta^2)$
 - $\sigma_\eta^2 = 0.25$
- s AR(1) chybami
 - $\sigma^2 = 0.16, \rho = 0.7$
- empirické pravdepodobnosti pokrytia lineárnej kombinácie vektora pevných efektov $\mathbf{L}'\beta$ pre zvyšujúci sa počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte

Simulačná štúdia A

Pokus číslo 1 - grafické znázornenie

Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.



Simulačná štúdia A

Pokus číslo 1 - tabuľkové znázornenie

<i>počet subjektov</i>	χ^2	F_{KR}	F_{modKR}	F_{FC}
10	0.9179	0.9080	0.9039	0.9441
13	0.9236	0.9106	0.9327	0.9395
15	0.9254	0.9175	0.9400	0.9408
18	0.9306	0.9276	0.9426	0.9421
20	0.9340	0.9304	0.9491	0.9447
25	0.9357	0.9363	0.9500	0.9450
30	0.9393	0.9425	0.9502	0.9467
40	0.9360	0.9411	0.9467	0.9432
50	0.9397	0.9425	0.9470	0.9443

Tabuľka: Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.

Simulačná štúdia A

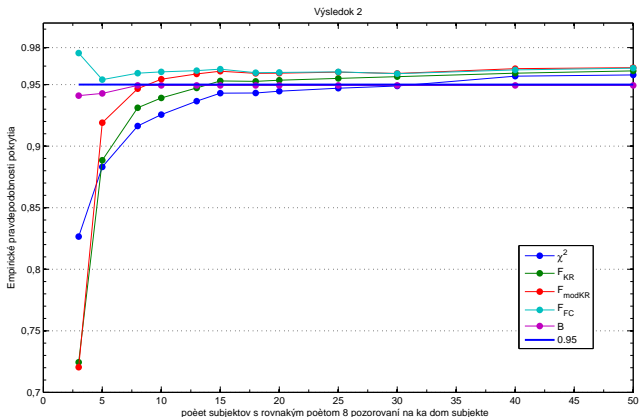
Pokus číslo 2

- matice plánu $\mathbf{Z}_i = \mathbf{X}_i$ pre každé i
- $\boldsymbol{\eta}_i$ sú 3-rozmerné náhodné vektory pochádzajúce z rozdelenia $N(\mathbf{0}, \sigma_\eta^2 \mathbf{I}_3)$, kde \mathbf{I}_3 je (3×3) -rozmerná jednotková matica
 - $\sigma_\eta^2 = 4 \cdot 10^{-14}$
- ε_i spĺňa AR(1) proces
 - $\sigma^2 = 0.16, \rho = 0.85$
- empirické pravdepodobnosti pokrytia lineárnej kombinácie vektora pevných efektov $\mathbf{L}'\boldsymbol{\beta}$ pre zvyšujúci sa počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte

Simulačná štúdia A

Pokus číslo 2 - grafické znázornenie

Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.



Simulačná štúdia A

Pokus číslo 2 - tabuľkové znázornenie

počet subjektov	χ^2	F_{KR}	F_{modKR}	F_{FC}	B
3	0.8265	0.7244	0.7204	0.9756	0.9410
5	0.8831	0.8885	0.9190	0.9540	0.9428
8	0.9163	0.9311	0.9466	0.9592	0.9493
10	0.9256	0.9391	0.9543	0.9603	-
13	0.9365	0.9472	0.9586	0.9613	-
15	0.9430	0.9529	0.9608	0.9625	-
18	0.9431	0.9526	0.9591	0.9596	-
20	0.9446	0.9535	0.9592	0.9598	-
25	0.9470	0.9550	0.9601	0.9603	-
30	0.9488	0.9564	0.9590	0.9588	-
40	0.9568	0.9592	0.9629	0.9619	-
50	0.9578	0.9611	0.9637	0.9633	-

Tabuľka: Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.

Simulačná štúdia A

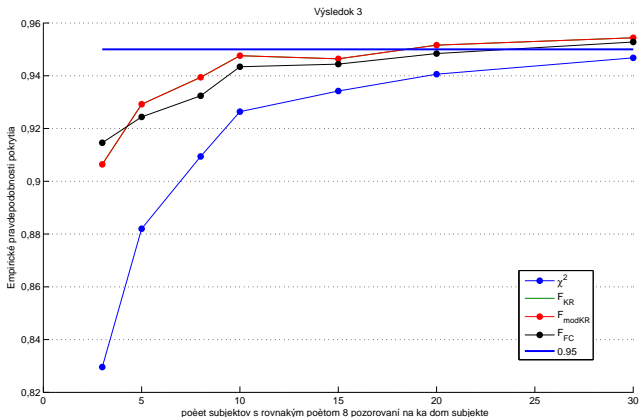
Pokus číslo 3

- matice plánu $\mathbf{Z}_i = [1, 1, 1, 1, 1, 1, 1, 1]'$ pre každé i
- η_i sú jednorozmerné náhodné efekty pochádzajúce z rozdelenia $N(0, \sigma_\eta^2)$
 - $\sigma_\eta^2 = 0.25$
- $\varepsilon_i \sim N(0, \sigma^2 \mathbf{I}_{n_i})$, kde \mathbf{I}_{n_i} je $(n_i \times n_i)$ -rozmerná jednotková matica
 - $\sigma^2 = 0.16$
- empirické pravdepodobnosti pokrytia lineárnej kombinácie vektora pevných efektov $\mathbf{L}'\beta$ pre zvyšujúci sa počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte

Simulačná štúdia A

Pokus číslo 3 - grafické znázornenie

Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.



Simulačná štúdia A

Pokus číslo 3 - tabuľkové znázornenie

<i>počet subjektov</i>	χ^2	F_{KR}	F_{modKR}	F_{FC}
3	0.8296	0.9064	0.9064	0.9146
5	0.8820	0.9292	0.9292	0.9244
8	0.9094	0.9394	0.9394	0.9324
10	0.9264	0.9476	0.9476	0.9434
15	0.9342	0.9464	0.9464	0.9444
20	0.9406	0.9516	0.9516	0.9484
30	0.9468	0.9544	0.9544	0.9528

Tabuľka: Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.

Simulačná štúdia A

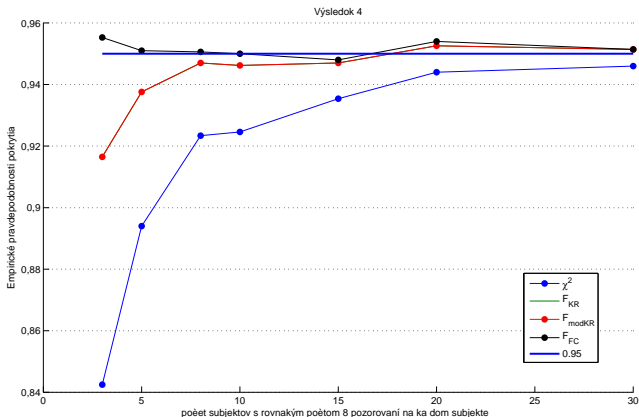
Pokus číslo 4

- matice plánu $\mathbf{Z}_i = \mathbf{X}_i$ pre každé i
- $\boldsymbol{\eta}_i$ sú 3-rozmerné náhodné vektory pochádzajúce z rozdelenia $N(\mathbf{0}, \sigma_{\eta}^2 \mathbf{I}_3)$, kde \mathbf{I}_3 je (3×3) -rozmerná jednotková matica
 - $\sigma_{\eta}^2 = 0.25$
- $\varepsilon_i \sim N(0, \sigma^2 \mathbf{I}_{n_i})$, kde \mathbf{I}_{n_i} je $(n_i \times n_i)$ -rozmerná jednotková matica
 - $\sigma^2 = 0.09$
- empirické pravdepodobnosti pokrytia lineárnej kombinácie vektora pevných efektov $\mathbf{L}'\boldsymbol{\beta}$ pre zvyšujúci sa počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte

Simulačná štúdia A

Pokus číslo 4 - grafické znázornenie

Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.



Simulačná štúdia A

Pokus číslo 4 - tabuľkové znázornenie

<i>počet subjektov</i>	χ^2	F_{KR}	F_{modKR}	F_{FC}
3	0.8425	0.9165	0.9165	0.9553
5	0.8940	0.9376	0.9376	0.9510
8	0.9234	0.9470	0.9470	0.9506
10	0.9246	0.9462	0.9462	0.9500
15	0.9354	0.9470	0.9470	0.9480
20	0.9440	0.9526	0.9526	0.9540
30	0.9460	0.9514	0.9514	0.9514

Tabuľka: Empirické pravdepodobnosti pokrytia $L'\beta$ pre rôzny počet subjektov s rovnakým počtom 8 pozorovaní na každom subjekte pre konfidenciu 0.95.

Simulačná štúdia B - Spoločné črty

- Príklad 2 - Hmotnosti prasiat
- týždenné meranie hmotnosti 16 prasiat počas 9 týždňov
- lineárny zmiešaný model

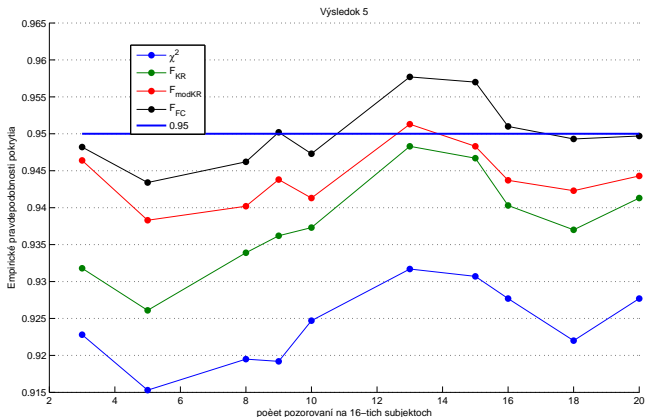
$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\eta}_i + \varepsilon_i,$$

- pričom j -ty riadok matice \mathbf{X}_i je $\mathbf{x}_{ij} = [1, j]'$ pre každé i
- $\boldsymbol{\beta} = [17.5, 6.2]'$
- $\mathbf{Z}_i = \mathbf{X}_i$
- $\boldsymbol{\eta}_i$ sú 2-rozmerné náhodné vektory pochádzajúce z rozdelenia $N(\mathbf{0}, \sigma_\eta^2 \mathbf{I}_2)$, kde \mathbf{I}_2 je (2×2) -rozmerná jednotková matica
- ε_i spĺňa AR(1) proces
 - $\sigma_\eta^2 = 0.25$
 - $\sigma_0^2 = 1.69, \rho_0 = 0.75$

Simulačná štúdia A

Pokus číslo 1 - grafické znázornenie

Empirické pravdepodobnosti pokrytia $L'\beta$ pre zvyšujúci sa počet pozorovaní na 16-tich subjektoch pre konfidenciu 0.95.



Simulačná štúdia A

Pokus číslo 1 - tabuľkové znázornenie

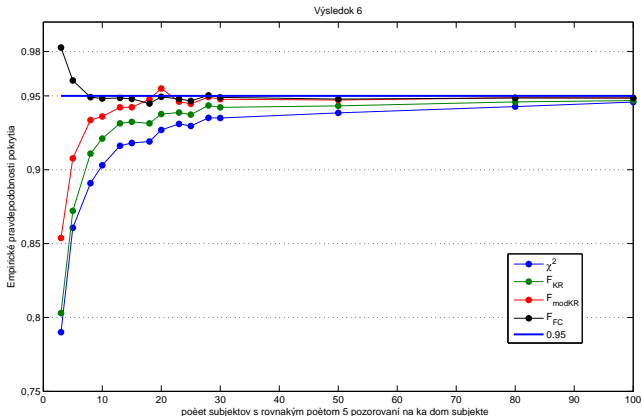
<i>počet pozorovaní</i>	χ^2	F_{KR}	F_{modKR}	F_{FC}
3	0.9228	0.9318	0.9464	0.9482
5	0.9153	0.9261	0.9383	0.9434
8	0.9195	0.9339	0.9402	0.9462
9	0.9192	0.9362	0.9438	0.9502
10	0.9247	0.9373	0.9413	0.9473
13	0.9317	0.9483	0.9513	0.9577
15	0.9307	0.9467	0.9483	0.9570
16	0.9277	0.9403	0.9437	0.9510
18	0.9220	0.9370	0.9423	0.9493
20	0.9277	0.9413	0.9443	0.9497

Tabuľka: Empirické pravdepodobnosti pokrytia $\mathbf{L}'\beta$ pre rôzny počet pozorovaní na 16-tich subjektoch pre konfidenciu 0.95.

Simulačná štúdia A

Pokus číslo 2 - grafické znázornenie

Empirické pravdepodobnosti pokrytia $L'\beta$ pre zvyšujúci sa počet subjektov s rovnakým počtom 5 pozorovaní na každom subjekte pre konfidenciu 0.95.



Simulačná štúdia A

Pokus číslo 2 - tabuľkové znázornenie

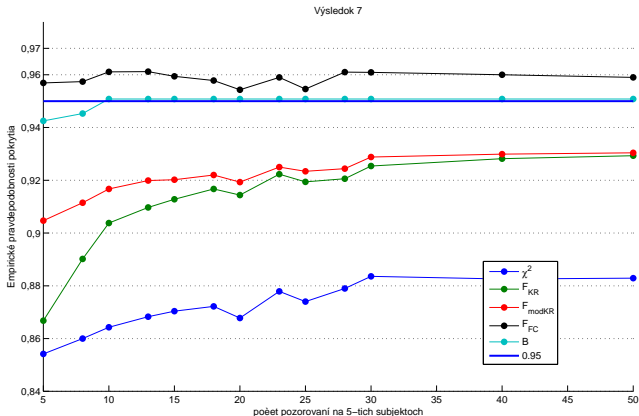
<i>počet pozorovaní</i>	χ^2	F_{KR}	F_{modKR}	F_{FC}
3	0.9228	0.9318	0.9464	0.9482
5	0.9153	0.9261	0.9383	0.9434
8	0.9195	0.9339	0.9402	0.9462
9	0.9192	0.9362	0.9438	0.9502
10	0.9247	0.9373	0.9413	0.9473
13	0.9317	0.9483	0.9513	0.9577
15	0.9307	0.9467	0.9483	0.9570
16	0.9277	0.9403	0.9437	0.9510
18	0.9220	0.9370	0.9423	0.9493
20	0.9277	0.9413	0.9443	0.9497

Tabuľka: Empirické pravdepodobnosti pokrytia $\mathbf{L}'\beta$ pre rôzny počet subjektov s rovnakým počtom 5 pozorovaní na každom subjekte pre konfidenciu 0.95.

Simulačná štúdia A

Pokus číslo 3 - grafické znázornenie

Empirické pravdepodobnosti pokrytia $L'\beta$ pre pre zvyšujúci sa počet pozorovaní na 5-tich subjektoch pre konfidenciu 0.95.



Simulačná štúdia A

Pokus číslo 3 - tabuľkové znázornenie

<i>počet pozorovaní</i>	χ^2	F_{KR}	F_{modKR}	F_{FC}	B
5	0.8542	0.8668	0.9047	0.9569	0.9425
8	0.8600	0.8902	0.9115	0.9574	0.9453
10	0.8643	0.9038	0.9167	0.9611	0.9508
13	0.8683	0.9097	0.9199	0.9612	-
15	0.8704	0.9128	0.9202	0.9594	-
18	0.8722	0.9167	0.9220	0.9578	-
20	0.8678	0.9144	0.9193	0.9543	-
23	0.8779	0.9223	0.9250	0.9590	-
25	0.8740	0.9194	0.9234	0.9546	-
28	0.8790	0.9206	0.9244	0.9610	-
30	0.8836	0.9254	0.9288	0.9609	-
40	0.8826	0.9282	0.9299	0.9600	-
50	0.8829	0.9293	0.9304	0.9590	-

Tabuľka: Empirické pravdepodobnosti pokrytia $L'\beta$ pre pre rôzny počet pozorovaní na 5-tich subjektoch pre konfidenciu 0.95.

Ďakujem za pozornosť