Intervaly spoľahlivosti pre spoločnú strednú hodnotu - porovnanie dvoch metód

MÁRIA JANKOVÁ

majka.jankova@gmail.com

Oddelenie teoretických metód, Ústav merania, Slovenská akadémia vied, Bratislava

V tomto príspevku sa budeme zaoberať intervalovými odhadmi spoločnej strednej hodnoty v jednofaktorovom heteroskedastickom ANOVA modeli. Porovnáme dve metódy intervalového odhadu: metódu založenú na metrologickom prístupe navrhnutú Witkovským a Wimmerom v [1] a zovšeobecnené intervaly (GCI - generalized confidence intervals) navrhnuté Wangom a Iyerom v [2]. Využitím Monte Carlo simulácií skúmame frekventistické vlastnosti oboch metód.

METROGICKÁ FORMULÁCIA

V metrológii sa často stretávame s problémom stanovenia spoločnej strednej hodnoty. V praxi ide o stanovenie čo najpresnejšieho odhadu skutočnej hodnoty meranej veličiny, pričom tento odhad sa nazýva kľúčová porovnávacia referenčná hodnota (KCRV - key comparison reference value). Pre jej určenie sú k dispozícií dáta z viacerých laboratórií. Chyba pozorovaní poskytnutých každým laboratóriom pritom pozostáva z tzv. laboratórnej chyby, ktorá je pre všetky pozorovania z jedného laboratória rovnaká, a z chyby jednotlivých meraní. My budeme uvažovať situáciu s rovnomerným a normálnym rozdelením laboratórnej chyby. Formálne môžeme model zapísať ako jednofaktorový heteroskedastický ANOVA model:

$$y_{ij} = \mu + B_i + \epsilon_{ij} \qquad (1)$$

kde $j = 1, \dots k$ označuje počet zúčastnených laboratórií, $k = 1, \dots n_i$ označuje počet meraní v i-tom laboratóriu. B_i reprezentuje laboratórnu chybu a ϵ_{ij} predstavuje chybu jednotlivých meraní, o ktorej predpokladáme $\epsilon_{ij} \sim N(0, \sigma_{B,i})$.

POROVNÁVANÉ METÓDY

Označme výberový priemer $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ a výberovú disperziu $S_i^2 = \frac{1}{n_i-1}\sum(Y_{ij} - \bar{Y}_i)^2$ a ich realizácie $\bar{y}_i$ a s_i^2.

Metrologický prístup navrhnutý v [1] Witkovským a Wimmerom (metódu ďalej značíme WW). Uvažujme náhodnú premennú $\tilde{\tilde{\mu}}$ v tvare:

$$\tilde{\tilde{\mu}} = \sum_{i=1}^{k} w_i \bar{y}_i - \sum_{i=1}^{k} w_i \sqrt{\frac{s_i^2}{n_i}} T_i - \sum_{i=1}^{k} w_i B_i,$$

kde $T_i \sim t_{n_i-1}$ a w_i sú váhy volené nasledovne:

$$w_i = \frac{\left(1/\left(\sqrt{\frac{s_i^2}{n_i}}\sqrt{\frac{s_p^2}{n_i}\frac{n_i-1}{n_i-3}} + \sigma^2_{(B),i}\right)\right)}{\sum_{l=1}^{k}\left(1/\left(\sqrt{\frac{s_l^2}{n_l}}\sqrt{\frac{s_p^2}{n_l}\frac{n_l-1}{n_l-3}} + \sigma^2_{(B),l}\right)\right)},$$

kde $s_p^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2 / \sum_{i=1}^{k}(n_i - k)$.

Potom za odhad $(1-\alpha)\times 100\%$-ného intervalu spoľahlivosti pre μ berieme $(\mu_{KCRV} + q_{\alpha/2}, \mu_{KCRV} + q_{1-\alpha/2})$, kde μ_{KCRV} je stredná hodnota náhodnej premennej $\tilde{\tilde{\mu}}$ a q_β β-percentný kvantil náhodnej premennej $\tilde{\tilde{\mu}} - \mu_{KCRV}$.

Zovšeobecnené intervaly navrhnuté v [2] Wangom a Iyerom (ďalej metódu značíme WI). Za dolnú a hornú hranicu $(1-\alpha)\times 100$ percentného intervalu spoľahlivosti berieme $q_{\alpha/2}$ a $q_{1-\alpha/2}$ kvantily náhodnej premennej R_μ, kde náhodná premenná R_μ má tvar:

$$R_\mu = \frac{\sum_{i=1}^{k}(\bar{y}_i - B_i)n_i W_i / [(n_i-1)s_i^2]}{\sum_{i=1}^{k} n_i W_i / [(n_i-1)s_i^2]} - Z\sqrt{\frac{1}{\sum_{i=1}^{k} n_i W_i / [(n_i-1)s_i^2]}}$$

kde $W_i \sim \chi^2_{n_i-1}$ a $Z \sim N(0,1)$.

METODOLÓGIA

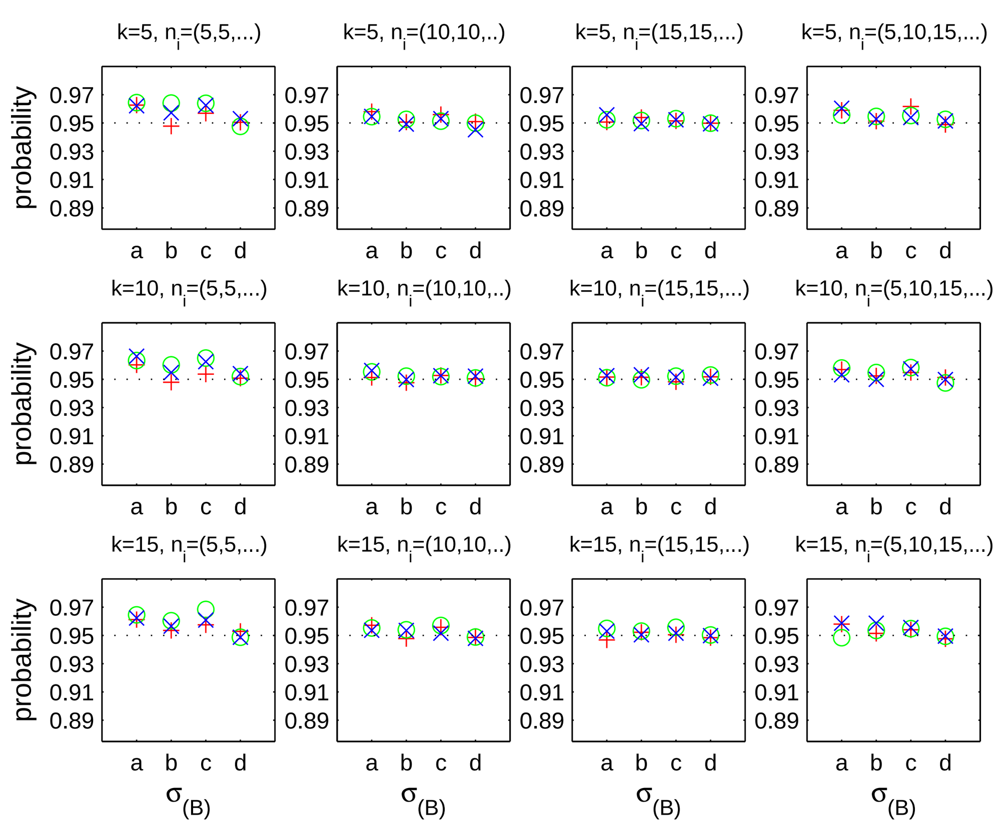
Budeme uvažovať viacero situácií s rozličnými parametrami modelu (1) - nazvime ich dizajny. Pre každý dizajn budeme generovať 10000 krát sériu dát. Pre každú sériu generovaných dát konštruujeme podľa vyššie navrhnutých metód $(1-\alpha)\times 100\%$ intervaly spoľahlivosti pre μ. Metódy porovnávame na základe dvoch frekventistických vlastností, percentuálnej úspešnosti pokrytia skutočnej hodnoty μ a na základe relatívnej priemernej dĺžky nasimulovaných intervalov.

SIMULAČNÁ ŠTÚDIA

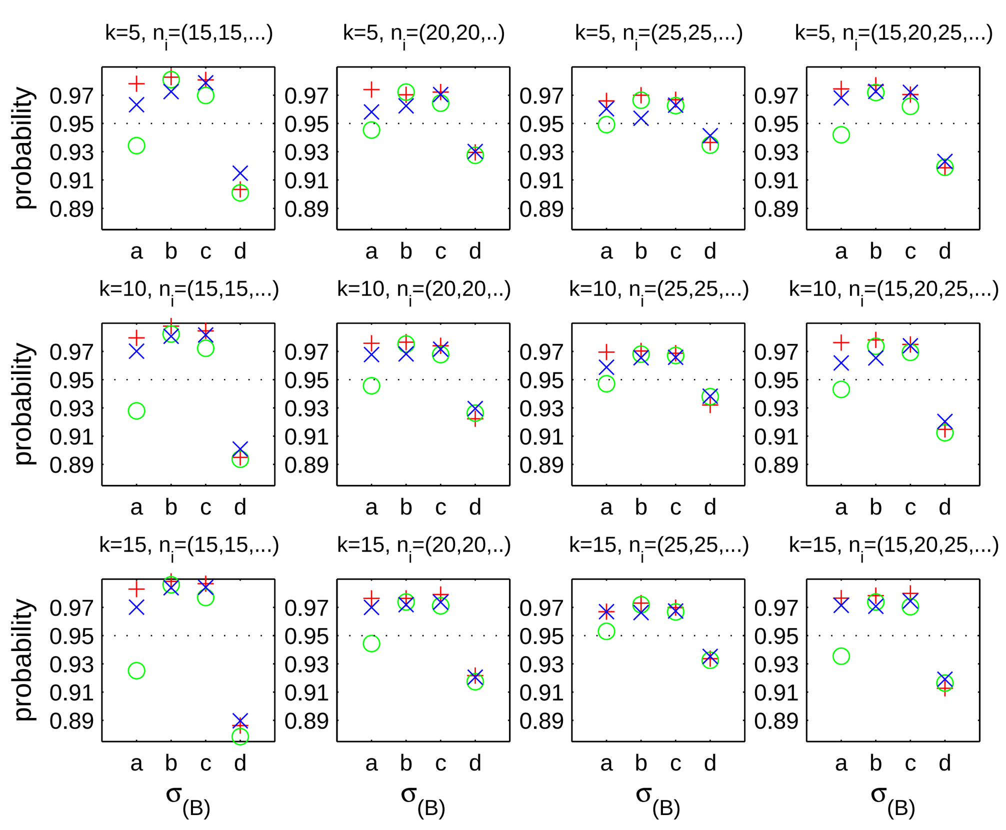
Zvolíme bez ujmy na všeobecnosti $\mu = 0$. Testujeme na hladine významnosti $\alpha = 0.05$. Rôzne dizajny vyberáme podľa [1]. Počet zúčastnených laboratórií je $5, 10$ alebo 15, t.j. $k \in \{5, 10, 15\}$. Počet pozorovaní v i-tom laboratóriu je $n_i = 5$, $n_i = 10$, $n_i = 15$ alebo $n_i = \{5, 10, 15, 5, 10, 15, 5, 10, 15, 5, 10, 5, 5, 10, 15\}, i = 1, \dots k$. Vzory pre $\sigma_{B,i}$ sú označené nasledovne: $\sigma_{B,i} = 1$ značené a, $\sigma_{B,i} = 5$ značené b, $\sigma_{B,i} \in \{1,2,3,4,5,1,2,3,4,5,1,2,3,4,5\}, i = 1, \dots k$ značené c, a $\sigma_{B,i} = 0$ značené d. Pre vybrané vzory $\sigma_{A,i}$ je zvolené nasledovné značenie: $\sigma_{A,i} = 1$ značené $+$, $\sigma_{A,i} = 5$ značené o a $\sigma_{A,i} \in \{1,2,3,4,5,1,2,3,4,5,1,2,3,4,5\}, i = 1, \dots k$ značené $\times$.

VÝSLEDKY

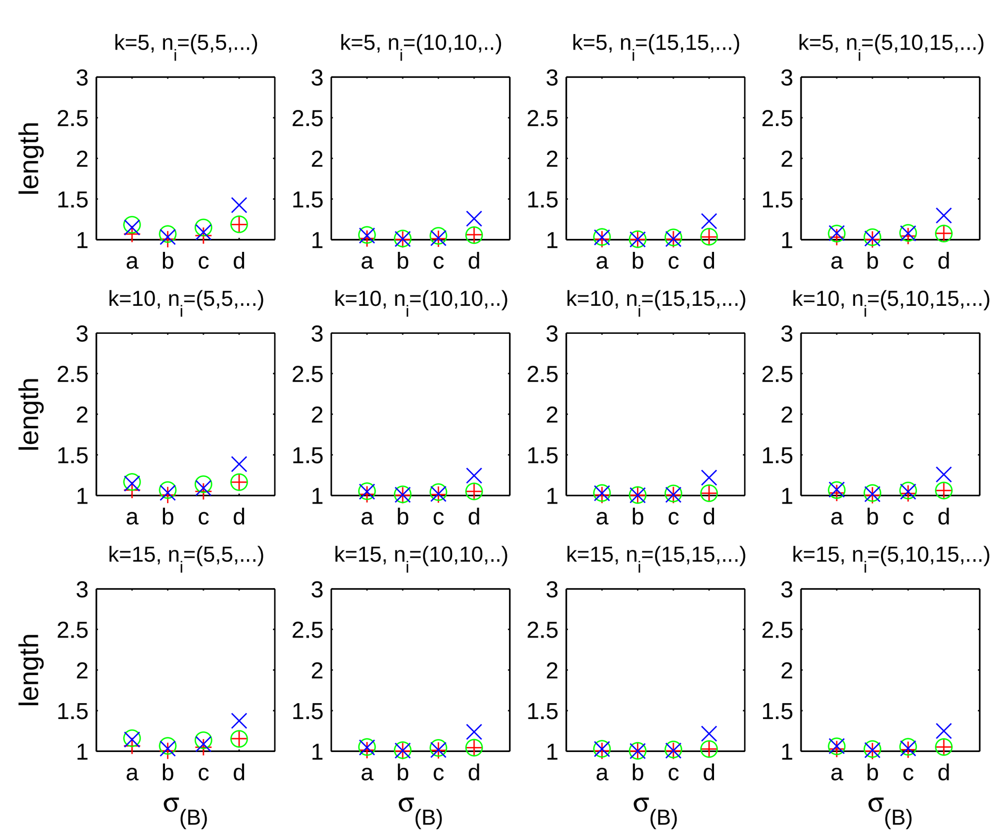
Z obrázkov pre model predpokladajúci normálne rozdelenie B_i vidíme, že pre zvolené dizajny dáva lepšie empirické pravdepodobnosti pokrytia a dĺžky intervalov WW metóda. Poznamenajme, že toto platí pre malý rozsah pozorovaní, t.j. max. počet pozorovaní v laboratóriu je 15.



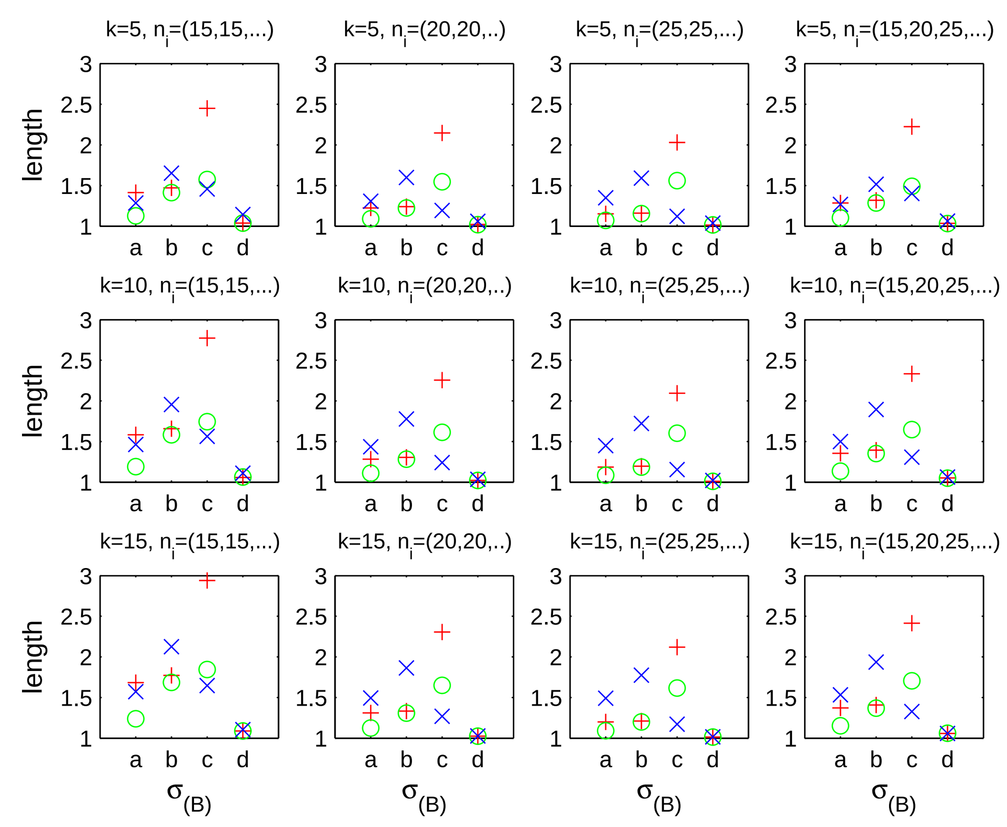
Obr 1.: Empirické pravdepodobnosti pokrytia pre $(1-\alpha)\times 100\%$ intervaly spoľahlivosti získané WW metódou pre $\alpha = 0.05$.$B_i \sim N(0, \sigma^2_{B,i})$.



Obr 2.: Empirické pravdepodobnosti pokrytia pre $(1-\alpha)\times 100\%$ intervaly spoľahlivosti získané WI metódou, $\alpha = 0.05$.$B_i \sim N(0, \sigma^2_{B,i})$.



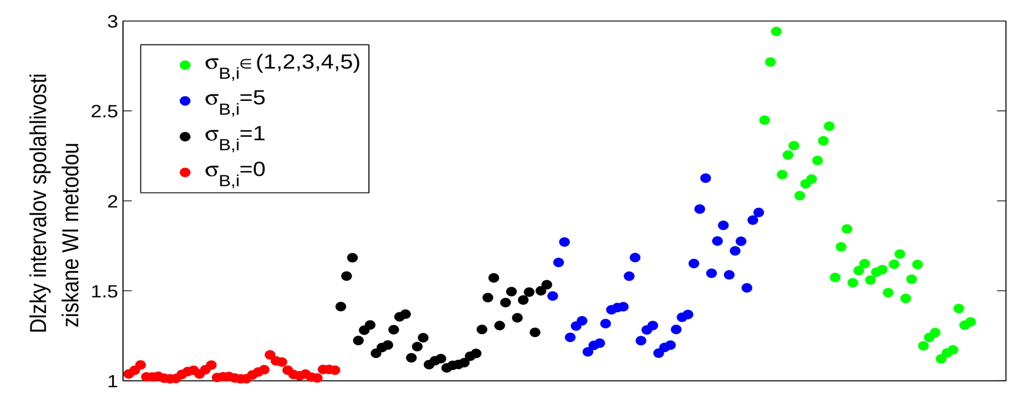
Obr 3.: Relatívne dĺžky $(1-\alpha)\times 100\%$ intervalov spoľahlivosti získaných WW metódou v porovnaní s exaktnými intervalmi spoľahlivosti skonštruovanými so znalosťou všetkých parametrov modelu.$B_i \sim N(0, \sigma^2_{B,i})$.



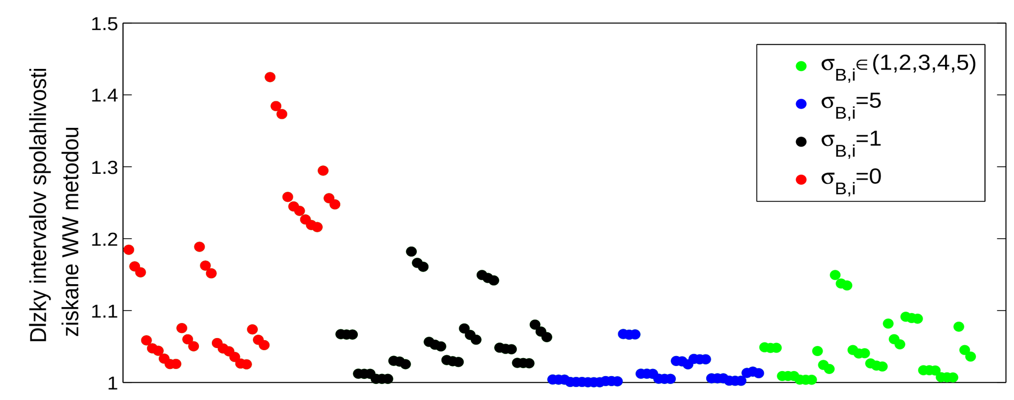
Obr 4.: Relatívne dĺžky $(1-\alpha)\times 100\%$ intervalov spoľahlivosti získaných WI metódou v porovnaní s exaktnými intervalmi spoľahlivosti skonštruovanými so znalosťou všetkých parametrov modelu.$B_i \sim N(0, \sigma^2_{B,i})$.

Pre obidve porovnávané vlastnosti intervalov pozorujeme nasledovné závislosti. Empirické pravdepodobnosti pokrytia sa blížia optimálnej hodnote $(1-\alpha)\times 100\%$ s rastúcim počtom pozorovaní v jednotlivých laboratóriách pre obidve metódy, pričom rozdiely medzi metódami sa zmenšujú.

Ďalej je pozorovaná závislosť dĺžky intervalov spoľahlivosti od parametrov $\sigma_{A,i}$ a $\sigma_{B,i}$,výraznejšia však od $\sigma_{B,i}$, rôzna pre každú z metód. Závislosti od $\sigma_{B,i}$ sú zobrazené v nižšieuvedených grafoch.



Obr 5.: Relatívne dĺžky intervalov spoľahlivosti získaných WI metódou v závislosti od jednotlivých dizajnov zoradených podľa $\sigma_{B,i}$.



Obr 6.: Relatívne dĺžky intervalov spoľahlivosti získaných WW metódou v závislosti od jednotlivých dizajnov zoradených podľa $\sigma_{B,i}$.

Priemerná dĺžka intervalov spoľahlivosti získaná WW metódou je najkratšia pre dizajny so $\sigma_{B,i} = 5$ a najdlhšia pre $\sigma_{B,i} = 0$. Pre WI metódu sú najkratšie intervaly pre dizajny so $\sigma_{B,i} = 0$ a najdlhšie pre $\sigma_{B,i} \in 1, 2, 3, 4, 5$.

Záverom treba spomenúť, že jednotlivé dizajny sme uvažovali aj s predpokladom, že $B_i \sim U(-\sqrt{3}\sigma_{B,i}, \sqrt{3}\sigma_{B,i})$. Intervaly pre ten istý dizajn s rôznymi predpokladmi na rozdelenie B_i dávajú veľmi podobné pokrytia - maximálny rozdiel v pokrytí pre WW metódu oproti situácií s $B_i \sim N(0, \sigma^2_{B,i})$ je $r_{max,WW} = 0.79\%$, pre WI metódu $r_{max,WI} = 0.98\%$. Rovnako pozorujeme minimálne rozdiely v relatívnej dĺžke intervalov, pre WW metódu je $l_{max,WW} = 0.017$ a pre WW je $l_{max,WI} = 0.14$.

Poďakovanie. Práca vznikla vďaka podpore grantov VEGA 1/0077/09, VEGA 2/0019/10.

Literatúra.

[1] WITKOVSKÝ,V.,WIMMER,G.: *Confidence Interval for Common Mean in Interlaboratory Comparisons with Systematic Laboratory Biases*. MEASUREMENT SCIENCE REVIEW,Vol.7, Section 1, No. 6, 2007

[2] WANG,C.M.,Iyer, H.K.: *A generalized confidence interval for a measurand in the presence of type-A and type-B uncertainties*. Measurement, Vol. 39, No. 9, pp. 856-863, 2006