

INFORMAČNÍ BULLETIN



České statistické společnosti

Ročník 25, číslo 4, prosinec 2014

HODNOCENÍ RELIABILITY ZNALOSTNÍCH A PSYCHOLOGICKÝCH TESTŮ

ESTIMATION OF RELIABILITY OF EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENTS

Patrícia Martinková, Katarína Vlčková

Adresa: Ústav informatiky AV ČR, v.v.i.,
Pod Vodárenskou věží 2, 182 07 Praha 8

E-mail: martinkova@cs.cas.cz

Abstrakt: Příspěvek se zabývá hodnocením reliability znalostních a psychologických testů. Pojednává o klasickém odhadu reliability pomocí Cronbachova alfa a jeho omezeních. Jsou zde rozebrány konkrétní situace, kdy použití Cronbachova alfa není vhodné, a pro tyto situace jsou nabídnuty obecnější nebo alternativní metody odhadu reliability.

Klíčová slova: spolehlivost, reliability, znalostní testy, psychologické testy, Cronbachovo alfa, latentní proměnné, faktorová analýza, analýza rozptylu, logistická regrese, teorie odpovědi na položku.

Abstract: This paper is devoted to estimation of reliability of educational and psychological tests. We describe the classical estimator of reliability, the so-called Cronbach's alpha, and its limitations. Situations in which Cronbach's alpha is not adequate are further elaborated. Generalized or alternative estimates are offered for these situations.

Keywords: reliability, educational tests, psychological tests, Cronbach's alpha, latent variables, factor analysis, analysis of variance, logistic regression, item response theory.

1. Úvod

O reliability bylo na stránkách tohoto časopisu pojednáno v článku [38]. Autor v něm upozornil na některá omezení klasického odhadu reliability pomocí Cronbachova alfa. Pro případ binárních položek navrhl nový odhad, tzv. logistické alfa. V této práci připomínáme některé úvahy přednesené ve zmíněném článku a dále je rozvádíme. Logistické alfa a jeho využití je tak zasazeno do kontextu dalších metod odhadu reliability v případě, kdy nejsou splněny podmínky předpokládané klasickým odhadem.

2. Reliabilita měření

Při hodnocení spolehlivosti psychologických a znalostních testů se vychází z představy, že měření Y (např. bodový výsledek v testu) je součtem skutečné hodnoty měřené vlastnosti, nepozorované (tzv. latentní) veličiny $T \sim (\mu, \sigma_T)$, např. znalosti studenta, a nezávislé náhodné chyby měření $e \sim (0, \sigma_e)$,

$$Y = T + e.$$

Přirozenou otázkou potom je, do jaké míry jsou měření Y a skutečná hodnota měřené vlastnosti T spjaty. Na tuto otázku se snaží dát odpověď teorie reliability. Reliabilita, neboli spolehlivost, je často definována pomocí korelace mezi měřením a skutečnou hodnotou, jako

$$\text{rel}(Y) = \text{corr}(Y, T)^2. \quad (1)$$

Reliabilitu lze ekvivalentně vyjádřit pomocí podílu

$$\text{rel}(Y) = \frac{\text{cov}(T + e, T)^2}{\text{var}(Y) \text{var}(T)} = \frac{\text{var}(T)}{\text{var}(Y)} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2}. \quad (2)$$

Vypovídá tedy také o tom, jaký podíl variance měření přísluší skutečné varianci měřené vlastnosti. Teoreticky jde o číslo mezi 0 a 1, spolehlivá měření, tj. měření s relativně malým podílem chybové variance, mají hodnotu reliability blízkou 1. Jelikož skutečnou hodnotu T ani chybový člen e nelze pozorovat, není možné reliabilitu odhadovat na základě jediného měření z definice (1) ani (2).

Nízká reliabilita má několik negativních důsledků. V případě znalostních testů je to mj. nepřesný odhad studentových znalostí, širší konfidenční intervaly, rozředění korelací s kritériem (což implikuje omezení validity testu), nebo potřeba vyššího počtu studentů na prokázání rozdílů mezi testovanými skupinami. Hodnocení reliability by proto mělo patřit k rutinní analýze kvality testu.

Je-li spolehlivost testu nízká (často se jako jakási mez uvažuje hodnota 0,8), lze přistoupit k opatřením, která mohou pomoci reliabilitu zvýšit. Např. k navýšení počtu položek, odebrání nevhodných položek, nebo, v případě hodnocení více hodnotiteli, k zajištění preciznějších instrukcí a tréninku, či k navýšení počtu hodnotitelů a průměrování jejich hodnocení. Tato opatření však mohou být dosti finančně a časově náročná, je proto důležité odhadovat reliabilitu správnými metodami. Špatné porozumění odhadu reliability může vést k finančně náročným úpravám nebo např. k zbytečnému odstranění důležitých položek a následnému snížení validity testu.

Jak tedy reliabilitu odhadovat? Snadno lze ukázat, že máme-li dvě stejně přesná měření téže vlastnosti, jejichž chybové složky nejsou korelované mezi sebou ani s měřenou vlastností

$$Y_1 = T + e_1, \quad Y_2 = T + e_2, \quad \text{var}(e_1) = \text{var}(e_2) = \sigma_e^2,$$

pak korelace mezi takovými měřeními je rovna jejich (shodné) reliabilitě

$$\text{corr}(Y_1, Y_2) = \frac{\text{cov}(T + e_1, T + e_2)}{\sqrt{\text{var}(Y_1) \text{var}(Y_2)}} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} = \text{rel}(Y_1) = \text{rel}(Y_2). \quad (3)$$

Tento vztah platí dokonce i pro taková měření, jejichž skutečné skóry se liší pouze o konstantu

$$T_i = T + b_i. \quad (4)$$

Takovým položkám se v testové teorii často říká *esenciálně tau ekvivalentní*. Vztah (1) už nabízí několik možností odhadu reliability:

- Stejný test můžeme probandům zadat vícekrát po sobě. Této metodě se říká **test-retest**. Narážíme zde ale na problém, že respondenti si své původní odpovědi mohou pamatovat a budou mít tendenci odpovídat stejně. Zvolíme-li delší časový úsek mezi měřeními, měřená vlastnost (např. studentova znalost) se naopak mohla změnit. Takový odhad potom vypovídá víc o stabilitě měření.
- Můžeme také připravit dvě **paralelní formy testu**, tedy zjednodušeně řečeno dvě verze testu, které si budou co nejvíce podobné – budou měřit stejnou vlastnost, jejich chybové členy ale budou nezávislé. V příkladech z matematiky lze např. vytvořit paralelní formu záměnou číselných hodnot v zadání. Pak lze hodnotit reliabilitu pomocí korelace dvou skóků. Výsledný koeficient ale kromě podílu chyby měření bude odrážet také míru, do jaké jsou použité dvě formy ekvivalentní, často se mu proto říká koeficient ekvivalence.
- Můžeme také využít skutečnosti, že test je složen z více měření. Vychází se pak z podobné představy jako u paralelních testů. Odhad je založený na korelaci mezi polovinami testu (tzv. split half koeficient) nebo na korelacích mezi jednotlivými položkami. Odhad je ještě potřeba upravit na počet položek: Předpokládáme-li, že všechny položky měří stejnou vlastnost (resp. se tyto měřené vlastnosti liší pouze o konstanty – obtížnosti/úrovně položek) a chybové členy jsou nezávislé mezi sebou a s měřenou vlastností, pak vztah mezi spolehlivostí

součtu m měření $Y_{\bullet} = \sum Y_j$ a spolehlivosti jednotlivých položek udává tzv. **Spearmanova-Brownova formule** ([27], [4], odvození lze nalézt také v článku [38]):

$$\text{rel}(Y_{\bullet}) = \frac{m \cdot \text{rel}(Y_1)}{1 + (m - 1) \cdot \text{rel}(Y_1)}. \quad (5)$$

Všimněme si, že přidáním vhodných položek lze reliabilitu přiblížit jedné. Všimněme si také, že takový odhad mj. odráží, do jaké míry jsou si položky podobné, říká se mu proto často *koeficient vnitřní konzistence*. Jedním z takových odhadů je i Cronbachovo alfa.

3. Cronbachovo alfa

K odhadu reliability se v případě znalostních testů složených z více položek, $Y_{\bullet} = Y_1 + \dots + Y_m$, nejčastěji používá tzv. Cronbachovo alfa

$$\alpha = \frac{m}{m-1} \frac{\sum \sum_{j \neq k} \text{cov}(Y_j, Y_k)}{\text{var}(Y_{\bullet})} = \frac{m}{m-1} \frac{1 - \text{tr}(\text{var}(\mathbf{Y}))}{\mathbf{1}' \text{var}(\mathbf{Y}) \mathbf{1}}.$$

To je funkcí kovarianční matice vektoru položkových měření $\mathbf{Y} = (Y_1, \dots, Y_m)'$ a je snadné jej odhadnout.

Rozeberme zde blíže případ popisovaný v článku [38], kdy jsou položkové chyby nekorelované a kdy pro ně platí vztah (4). Jedná se o situaci, kterou lze popsat smíšeným modelem analýzy rozptylu dvojného třídění: n probandů vnímáme jako náhodný vzorek z celé populace, každému je zadána stejná sada pevných m položek. Pak j -té měření (položka) se realizuje na i -tém probandovi jako Y_{ij} , které lze vyjádřit jako součet

$$Y_{ij} = T_i + b_j + e_{ij}. \quad (6)$$

Pro i -tou osobu se tedy skutečná hodnota měřené vlastnosti (např. znalost) T realizuje jako $T_i \sim (\mu, \sigma_T^2)$, konstanty b_j popisující obtížnost položek jsou shodné pro všechny osoby, a měření jsou zatížena nekorelovanými náhodnými chybami $e_{ij} \sim (0, \sigma_e^2)$. Reliabilita složeného měření $Y_{i\bullet} = \sum_{j=1}^m Y_{ij}$ je pak dle (2)

$$\text{rel}(Y_{i\bullet}) = \frac{\text{var}\left(\sum_j T_i\right)}{\text{var}\left(\sum_j Y_{ij}\right)} = \frac{m^2 \sigma_T^2}{m^2 \sigma_T^2 + m \sigma_e^2} = \frac{\sigma_T^2}{\sigma_T^2 + \frac{1}{m} \sigma_e^2},$$

což je totéž jako Cronbachovo alfa

$$\alpha = \frac{m}{m-1} \frac{\sum \sum_{j \neq k} \text{cov}(Y_{ij}, Y_{ik})}{\text{var} \left(\sum_j Y_{ij} \right)} = \frac{m}{m-1} \frac{m(m-1)\sigma_T^2}{m^2\sigma_T^2 + m\sigma_e^2} = \frac{\sigma_T^2}{\sigma_T^2 + \frac{1}{m}\sigma_e^2}.$$

V případě, kdy lze testovou situaci popsat smíšeným modelem analýzy rozptylu dvojného třídění, je tedy Cronbachovo alfa rovno reliabilitě složeného měření. Jeho použití je v takovém případě zcela legitimní. Předpoklady tohoto modelu ale mohou být v mnoha případech porušeny, je proto potřeba mít se na pozoru. Porušení vztahu (4) vede k podhodnocení reliability, korelované chyby položkových měření mohou vést i k jejímu nadhodnocení (viz [16]).

Ve zmíněném článku [38] bylo také ukázáno, že uvažujeme-li normální rozdělení $T_i \sim N(\mu, \sigma_T^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, pak průměrné čtverce MSA , MSE mají střední hodnoty

$$E MSA = m\sigma_T^2 + \sigma_e^2,$$

$$E MSE = \sigma_e^2.$$

Odhad Cronbachova alfa

$$\hat{\alpha} = \frac{m}{m-1} \frac{\sum \sum_{j \neq k} s_{jk}}{\sum \sum_{j,k} s_{jk}}, \quad \text{kde } s_{jk} = \frac{1}{n-1} \sum_{t=1}^n (Y_{tj} - \bar{Y}_{\bullet j})(Y_{tj} - \bar{Y}_{\bullet k})$$

lze tedy odvodit také pomocí průměrných čtverců, a vyjádřit jej jako funkci F statistiky běžně používané k rozhodování o hypotéze $\sigma_T^2 = 0$:

$$\hat{\alpha} = \frac{MSA - MSE}{MSA} = 1 - \frac{1}{F}. \quad (7)$$

Tato skutečnost dává statistikovi lepší porozumění Cronbachovu alfa: je blízké jedné v případě, kdy je hodnota F statistiky vysoká, tedy v případě, kdy zamítáme nulovou hypotézu o nulovosti rozptylu znalostí, a kdy znalostní test dobře rozlišuje mezi studenty. Je zde na místě upozornit na skutečnost, která je patrná již i z definice reliability, o to více pak nyní, a sice že reliability je mírou relativní a závisí také na homogenitě testované populace. Jiné výsledky proto obdržíme budeme-li např. reliability odhadovat na základě dat všech studentů, kteří skládali přijímací zkoušku, a jiný když ji budeme odhadovat pouze na základě přijatých studentů.

Cronbachovo alfa je populárním odhadem a funkce pro jeho výpočet je obsažena ve většině statistických balíků. Jeho rutinní použití nebo vyžadování však není vhodné. Odporuje také doporučením obsaženým v publikaci [28],

kde se zdůrazňuje, že neexistuje jediný optimální odhad, naopak vědec musí pečlivě zvážit, kterou metodu odhadu reliability zvolí. V následujících sekcích se podíváme blíže na situace, kdy použití Cronbachova alfa vhodné není, a nabídneme alternativní metody odhadu reliability.

4. Příklad více latentních proměnných

Jak bylo zmíněno, Cronbachovo alfa předpokládá, že všechny položky měří jedinou latentní proměnnou T , resp. že pro ně platí vztah (4). Tento předpoklad lze považovat za splněný např. u úloh testujících algebraické schopnosti, jako je schopnost sčítat a odečítat do 100. Naproti tomu u slovních úloh, které testují jak znalost elementární algebry, tak schopnost číst a orientovat se v textu, je správnější uvažovat dvě latentní proměnné. V psychologii je častý ještě vyšší počet latentních proměnných. V takových případech porušení předpokladu (4), a nejsou-li současně porušeny další předpoklady, bude Cronbachovo alfa podhodnocovat skutečnou reliability (viz [16]).

Jedním možným přístupem k odhadu reliability v případě více latentních proměnných je využití faktorové analýzy (viz také [1]). Zde pojednáme o odhadu ω_t navrženém v článku [14] McDonalde. Bližší popis tohoto odhadu a jeho porovnání s jinými odhady lze nalézt např. v článku [22].

Mějme m -položkové měření $\mathbf{Y} = (Y_1, \dots, Y_m)'$ se střední hodnotou $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)'$ a rozptylem $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_m^2)'$. Potom označme standardizované skóry

$$Z_j = \frac{Y_j - \mu_j}{\sigma_j}, \quad j = 1, \dots, m.$$

Zavedením standardizovaných skórů se eliminuje vliv jednotek, přičemž se korelační matice vzhledem k původním skórum nezmění. Předpokládejme, že m -položkové měření $\mathbf{Z} = (Z_1, \dots, Z_m)'$ lze popsat pomocí malého počtu $k < m$ faktorů

$$Z_j = c_j g + a_{j1} f_1 + \dots + a_{jk} f_k + e_j, \quad j = 1, \dots, m. \quad (8)$$

g zde představuje jeden obecný faktor (např. studentovo obecné nadání), $f_1, \dots, f_k$ jsou faktory společné pro některé, ale ne všechny položky. O faktorech předpokládáme, že jsou vzájemně nekorelované a nekorelované s chybovým vektorem $\mathbf{e}$. Bez újmy na obecnosti také můžeme předpokládat, že faktory mají střední hodnotu rovnou nule a rozptyl roven jedné. Vektor konstant $\mathbf{c} = (c_1, \dots, c_m)'$ a matice konstant $\mathbf{A} = (a_{ij})_{i,j=1}^m$ se nazývají faktorové zátěže (factor loadings) a jsou odhadovány z dat. V níže uvedené proceduře omega je jako výchozí způsob odhadování faktorových zátěží využívána

metoda minimálních reziduí, v níž jsou faktory extrahovány tak, aby byl minimalizován součet čtverců reziduálních korelačních koeficientů, které neleží na diagonále.

O chybových členech e_j se předpokládá, že jsou vzájemně nekorelované a že mají rozptyly u_j^2 . Číslu u_j^2 se říká unicity a vyjadřuje podíl variance j -té položky, která není vysvětlitelná společnými faktory g a f . Doplnkem unicity je tzv. komunalita $h_j^2 = 1 - u_j^2$. Jde o podíl variance j -tého měření, která je vysvětlitelná společnými faktory. Komunality mají tvar

$$h_j^2 = c_j^2 + \sum_{i=1}^n a_{ij}^2. \quad (9)$$

McDonald pak definuje odhad reliability ω_t jako celkový podíl variance, která je vysvětlitelná společnými faktory g a f

$$\omega_t = \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1} + \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}}{V_z} = 1 - \frac{\sum_{j=1}^m (1 - h_j^2)}{V_z} = 1 - \frac{\sum_{j=1}^m u_j^2}{V_z}, \quad (10)$$

kde $V_z = \mathbf{1}'\text{var}(\mathbf{Z})\mathbf{1}$ je rozptyl součtu měření $\text{var}(\mathbf{Z}_\bullet)$.

Do jaké míry měří test pouze jediný latentní rys popisuje tzv. hierarchické omega (viz také [15]). To je definováno jako podíl variance příslušející pouze obecnému faktoru g

$$\omega_h = \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1}}{V_z}. \quad (11)$$

K prezentaci těchto metod budeme používat data z článku [30]. Jedná se o devět skóre ze souboru testů, které byly vyplněny 710 respondenty. Data jsou reprezentována korelační maticí 9×9 . K výpočtu jsou zvoleny dva faktory schopností, představující plynulost slov a verbální schopnosti. Vhodnost této volby je podrobněji odůvodněna v článku [31]. McDonaldovy odhady získáme v R [17] použitím balíku `psych` [21] a zadáním příkazů

```
library(psych)
data(Tucker)
summary(omega(Tucker,nfactors=2))
```

Hierarchické omega ω_h nabývá hodnoty

$$\omega_h \doteq 0,54, \quad (12)$$

tj. přibližně 54 % variability měření přísluší obecnému faktoru. McDonaldův odhad reliability ω_t má hodnotu

$$\omega_t \doteq 0,9. \quad (13)$$

V tomto případě bychom použitím Cronbachova alfa získali o něco nižší odhad reliability 0,86. Podhodnocení skutečné reliability by mohlo v některých případech vést k mylnému závěru o nedostatečné reliabilitě testu a jeho nákladné revizi.

5. Příklad korelovaných chyb

Jedním z dalších předpokladů modelu (6) je nekorelovanost chyb mezi položkami. Zatímco nesplnění požadavku (4) vede výhradně k podhodnocení odhadu reliability, korelovanost chyb položek může vést k podhodnocení i nadhodnocení reliability (viz např. [18] nebo [7]).

Korelované chyby položek si lze snadno představit v případě, kdy se část otázek vztahuje k tématu uvozujícímu textu. Takové skupiny položek jsou běžné v testech studijní připravenosti nebo např. testech cizího jazyka, jako je TOEFL. V článku [8] jsou popsány také další možné příčiny narušení předpokladu nekorelovaných chyb, které mohou vést k nadhodnocení reliability: např. výrazně omezený čas na vyřešení testu (tzv. rychlostní test), nebo skutečnost, že předcházející položky ovlivňují položky následující.

V člancích [18] a [7] je navrženo pro odhad reliability v takovém případě využít strukturního modelování. Pro případy, kdy se jedná o binární nebo ordinální položky, je v článku [9] navržen odhad reliability pomocí nelineárního strukturního modelování.

6. Příklad více zdrojů chyb

Model, který stojí za Cronbachovým alfa také předpokládá pouze jediný typ chyb – náhodnou chybu měření. V praxi je situace často mnohem složitější. Studenty např. může hodnotit více hodnotitelů, studenti mohou test skládat na různých místech s odlišnými podmínkami apod. Dostáváme se tak často k složitějším modelům analýzy rozptylu, mnohdy hierarchickým a s neúplným designem. Kromě hodnocení reliability coby obecné opakovatelnosti měření s ohledem na rozptyly všech náhodných efektů může situace přinést také další otázky, např.

- Je dostatečná shoda mezi hodnotiteli? Nebo: Nakolik se reliabilita zvýší, budou-li místo jediného hodnotitele hodnotit všechny studenty hodnotitelé dva nebo tři a budeme-li uvažovat jejich průměr?

- Hodnotí hodnotitelé z jedné školy stejně jako z druhé?

V tomto obecnějším pojetí se už teorii, která se zabývá odhadem opakovatelnosti měření říká **Teorie zobecnitelnosti**, pro popis statistických metod viz např. [3]. Při odhadu reliability se vychází ze vztahu (2). Komponenty rozptylu lze v méně komplikovaných situacích odhadnout pomocí průměrných čtverců z příslušného ANOVA modelu, podobně jako ve vzorci (7) v případě Cronbachova alfa. Ve složitějších případech lze využít také bayesovské odhady a další metody. V R lze pro odhad rozptylů ve smíšeném modelu analýzy rozptylu využít např. funkce `lmer` z knihovny `lme4` [2].

Tuto metodu budeme demonstrovat na datech z Rakouského průzkumu národních vzdělávacích standardů pro osmý ročník v psaní v německém jazyce. K dispozici máme odpovědi 135 studentů, kteří odpověděli na 5 úkolů. Každý student byl hodnocen jedním až sedmi hodnotiteli z množiny 16 hodnotitelů, každý hodnotitel hodnotil vždy všechny položky u daného studenta. Za každý úkol mohli studenti obdržet 0 až 3 body, studenti byli hodnoceni na základě součtu položek.

V tomto případě tedy budou výsledné odpovědi záviset na položce, hodnotiteli, schopnosti studenta a jejich interakcích. Z důvodu nedostatku dat nebudeme zahrnovat trojnou interakci. Předpokládáme tedy mj., že pokud hodnotitelé straní některým studentům, činí tak konzistentně napříč položkami. Formálně tento model zapíšeme následujícím způsobem:

$$Y_{ijk} = A_i + B_j + C_k + D_{ij} + E_{ik} + F_{jk} + e_{ijk}, \quad \begin{aligned} i &= 1, \dots, n, \\ j &= 1, \dots, m, \\ k &= 1, \dots, l, \end{aligned} \quad (14)$$

kde A_i značí náhodný efekt schopnosti studenta s rozptylem σ_s^2 , B_j značí náhodný efekt položky s rozptylem σ_p^2 a C_k značí náhodný efekt hodnotitele s rozptylem σ_h^2 . Proměnné D_{ij} , E_{ik} , F_{jk} , jsou náhodné interakce nezávisle proměnných s rozptyly σ_{sp}^2 , σ_{sh}^2 , σ_{ph}^2 . Proměnná e_{ijk} značí náhodnou chybu měření s rozptylem σ_e^2 . Jelikož se soustředíme na celkové výsledky studentů nezávisle od výkonu ostatních, zajímáme se o chybu náhodných efektů ve tvaru

$$\begin{aligned} \Delta_{i.k} &= Y_{i.k} - mA_i = \sum_{j=1}^m (Y_{ijk} - A_i) \\ &= \sum_{j=1}^m B_j + mC_k + \sum_{j=1}^m D_{ij} + mE_{ik} + \sum_{j=1}^m F_{jk} + \sum_{j=1}^m e_{ijk}. \end{aligned} \quad (15)$$

Rozptyl této chyby bude mít tvar

$$\sigma_{\Delta_{i..k}}^2 = m\sigma_p^2 + m^2\sigma_h^2 + m\sigma_{sp}^2 + m^2\sigma_{sh}^2 + m\sigma_{ph}^2 + m\sigma_e^2. \quad (16)$$

Reliabilitu součtu položek vyjádříme jako poměr

$$\text{rel}(Y_{i..k}) = \frac{m^2\sigma_s^2}{m^2\sigma_s^2 + \sigma_{\Delta_{i..k}}^2}. \quad (17)$$

V našem případě tedy bude odhad reliability součtu položek roven 0,702.

Další otázkou je, jaká by byla reliability testu, kdyby se výsledek řídil průměrem hodnocení více (např. dvou nebo tří) hodnotitelů. Označme chybu takových průměrných hodnocení

$$\bar{\Delta}_{i..} = \frac{1}{l} \sum_{k=1}^l Y_{i..k} - mA_i \quad (18)$$

$$\begin{aligned} &= \frac{1}{l} \sum_{k=1}^l \left(\sum_{j=1}^m B_j + mC_k + \sum_{j=1}^m D_{ij} + mE_{ik} + \sum_{j=1}^m F_{jk} + \sum_{j=1}^m e_{ijk} \right) \\ &= \sum_{j=1}^m B_j + \frac{m}{l} \sum_{k=1}^l C_k + \sum_{j=1}^m D_{ij} + \frac{m}{l} \sum_{k=1}^l E_{ik} + \\ &\quad + \frac{1}{l} \sum_{k=1}^l \sum_{j=1}^m F_{jk} + \frac{1}{l} \sum_{k=1}^l \sum_{j=1}^m e_{ijk}. \end{aligned} \quad (19)$$

Rozptyl $\bar{\Delta}_{i..}$ bude tedy roven

$$\sigma_{\bar{\Delta}_{i..}}^2 = m\sigma_p^2 + \frac{m^2}{l}\sigma_h^2 + m\sigma_{sp}^2 + \frac{m^2}{l}\sigma_{sh}^2 + \frac{m}{l}\sigma_{ph}^2 + \frac{m}{l}\sigma_e^2 \quad (20)$$

a reliability vyjádříme jako

$$\text{rel}(\bar{Y}_{i..}) = \frac{m^2\sigma_s^2}{m^2\sigma_s^2 + \sigma_{\bar{\Delta}_{i..}}^2}. \quad (21)$$

Pokud bychom tedy uvažovali, že každého studenta hodnotí právě dva náhodní hodnotitelé a výsledek se řídí jejich průměrem, pak reliability takového hodnocení by stoupla na 0,805. V případě tří hodnotitelů by reliability průměrného součtu byla dokonce 0,847.

```
library(lme4)
library(sirt)
data(data.ratings1)
# data.ratings1<-
# read.csv2("http://www.cs.cas.cz/martinkova/data.ratings1.csv")
m=5 # počet položek
length(levels(data.ratings1$rater)) # počet hodnotitelů
[1] 7

# dlouhý formát dat
data.long=reshape(data.ratings1,direction="long",varying
  =list(3:7),v.names = "ratings", timevar = "items")

# fitování modelu
(l1<-lmer(ratings~(1|items)+(1|rater)+(1|idstud)+(1|items:rater)
  +(1|idstud:rater)+(1|idstud:items),data=data.long))

# extrakce odhadů rozptylu
matrixSD<-as.matrix(lme4::VarCorr(l1))
Var<-data.frame(c(matrixSD,attr(matrixSD,"sc")^2))
names(Var)=c("idstud:items","idstud:rater","idstud",
  "items:rater","rater","items","Residual")

# funkce pro výpočet reliability průměru l hodnotitelů
reliabilita<-function(m,l,Var){
  delta_i=m*Var["items"]+m^2/l*Var["rater"]+m*Var["idstud:items"]+
  m^2/l*Var["idstud:rater"]+m/l*Var["items:rater"]+m/l*Var["Residual"]
  rel=m^2*Var["idstud"]/(m^2*Var["idstud"]+delta_i)
  return(as.numeric(rel))
}

reliabilita(m,1,Var)
[1] 0.7017009
reliabilita(m,2,Var)
[1] 0.805332
reliabilita(m,3,Var)
[1] 0.8470301
```

7. Příklad nesplnění normality, případ binárních položek

Odhad Cronbachova alfa vychází z představy normálního rozdělení jednotlivých měření. Jeho použití tedy může být nevhodné, je-li tento předpoklad porušen. Robustní verzi Cronbachova alfa se zabývá např. článek [5].

V případě znalostních testů je častým jevem použití binárních položek. Do jaké míry je použití Cronbachova alfa adekvátní zde? V tomto případě už nelze měření vnímat jako součet dvou nezávislých veličin – skutečné hodnoty a náhodné chyby. Místo toho je obvyklé modelovat pravděpodobnost správného zodpovězení na položku např. pomocí logistické regrese. Nejjednodušší, tzv. Raschův model předpokládá, že pravděpodobnost správné odpovědi i -tého studenta na j -tou otázku je

$$P_{ij}(X_{ij} = 1|T_i, b_j) = \frac{e^{T_i - b_j}}{1 + e^{T_i - b_j}}, \quad (22)$$

kde T_i je latentní vlastnost studenta a b_j je parametr obtížnosti položky.

Reliabilitu v takovém případě nelze definovat pomocí vztahu (2). Lze ji definovat obecněji pomocí podmíněných rozptylů a podmíněných středních hodnot (viz např. Marcoulides a Raykov, 2011):

$$\text{rel}(Y) = \frac{\text{var}(E(Y|T))}{\text{var}(E(Y|T)) + E(\text{var}(Y|T))} = \frac{\text{var}(E(Y|T))}{\text{var}(Y)}. \quad (23)$$

Výsledné integrály (viz např. [11]) často nelze vyjádřit explicitně, lze je však odhadnout numericky.

Definice (23) již není totožná s definicí pomocí paralelních forem, jak tomu bylo ve vztahu (3), rozdíly jsou však v praktických situacích zanedbatelné [10]. Také Spearmanova-Brownova formule (5) platí v Raschově modelu pouze přibližně (viz [12]).

Inspirován vztahem (7) navrhl Zvára [38] pro případ binárních položek jako alternativu ke Cronbachovu alfa použití tzv. logistického alfa

$$\alpha_{\log} = 1 - \frac{n - 1}{X^2}, \quad (24)$$

kde X^2 je statistika používaná v logistické regresi pro testování modelu bez efektu studenta. Logistické alfa může přinést jisté vylepšení odhadu, obzvláště pro případ většího počtu položek a relativně malého vzorku studentů (viz [11]), kdy nepodhodnocuje reliabilitu natolik jako Cronbachovo alfa. Je ale výpočetně složitější a stejně jako Cronbachovo alfa neřeší případy s více latentními proměnnými, korelovanými chybami položek nebo více zdroji chyb, zmíněné v tomto článku.

8. Závěr

Reliabilita znalostního testu je důležitým indikátorem jeho kvality a coby taková by měla být rutinně odhadována. Cronbachovo alfa se hodí jako odhad reliability v případě, kdy náhodná chyba je jediným zdrojem chyby měření, položky popisují jedinou latentní proměnnou, položky jsou nekorelované a lze předpokládat, že položkové skóry se řídí přibližně normálním rozdělením. V článku byly představeny možnosti odhadu reliability v případech, kdy tyto předpoklady splněny nejsou.

Zdrojový kód k použitým analýzám lze nalézt na stránce
<http://www.cs.cas.cz/martinkova/reliabilita.R>

Poděkování

Autorky děkují recenzentům za jejich podnětné připomínky, které pomohly k významnému zlepšení článku.

Literatura

- [1] Jiří Anděl. *Matematická statistika*. SNTL, 1985.
- [2] Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann. *lme4: Linear mixed-effects models using Eigen and S4*, 2014.
- [3] Robert L. Brennan. *Generalizability Theory (Statistics for Social and Behavioral Sciences)*. Springer, 2001 edition, September 2001.
- [4] William Brown. Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3: 296–322, 1910.
- [5] AAndreas Christmann, Stefan Van Aelst. Robust Estimation of Cronbach's Alpha. *J. Multivar. Anal.*, 97(7): 1660–1674, August 2006.
- [6] John Fox, Zhenghua Nie, Jarrett Byrnes, Michael Culbertson, Saikat DebRoy, Michael Friendly, Richard Jones, Adam Kramer, Georges Monette Monette, and R-Core. *Structural Equation Models*, 2013.
- [7] Samuel B. Green, L. Scott. Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, 7: 251–270, 2000.
- [8] Samuel B. Green, Yanyun Yang. Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1): 121–135, March 2009.
- [9] Samuel B. Green, Yanyun Yang. Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, 74(1): 155–167, March 2009.

- [10] Seonghoon Kim. A note on the reliability coefficients for item response model-based ability estimates. *Psychometrika*, 77(1): 153–162, January 2012.
- [11] Patrícia Martinková, Karel Zvára. Reliability in Rasch model. *Kybernetika*, 43(3): 315–326, 2007.
- [12] Patrícia Martinková, Karel Zvára. Reliability of composite dichotomous measurements. *European Journal for Biomedical Informatics*, 6(2): 14–23, 2010.
- [13] A. E. Maxwell. The effect of correlated errors on estimates of reliability coefficients. *Educational and Psychological Measurement*, 28(3): 803–811, 1968.
- [14] Roderick P. McDonald. Generalizability in factorable domains: “domain validity and generalizability”. *Educational and Psychological Measurement*, 38(1): 75–79, April 1978.
- [15] Roderick P. McDonald. *Test theory: A unified treatment*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 1999.
- [16] Melvin R. Novick, Charles Lewis. Coefficient alpha and the reliability of composite measurements. 32(1): 1–13, 1967.
- [17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. ISBN 3-900051-07-0.
- [18] Tenko Raykov. Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied psychological measurement*, 22(4): 375–385, 1998.
- [19] Tenko Raykov, George A. Marcoulides. *Introduction to Psychometric Theory*. New York: Taylor & Francis, 2011.
- [20] William Revelle. Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14: 57–74, 1979.
- [21] William Revelle. *Procedures for Psychological, Psychometric, and Personality Research*, 2014.
- [22] William Revelle, Richard E. Zinbarg. Coefficients alpha, beta, omega and the glb: comments on sijtsma. *Psychometrika*, 74(1): 145–154, 2009.
- [23] Alexander Robitzsch. *sirt: Supplementary Item Response Theory Models*, 2014.
- [24] Yves Rosseel, Daniel Oberski, Jarrett Byrnes, Leonard Vanbrabant, Victoria Savalei, Ed Merkle, Michael Hallquist, Mijke Rhemtulla, Myrsini Katsikatsou, and Mariska Barendse. *lavaan: Latent Variable Analysis*, 2014.
- [25] Richard J. Shavelson, Noreen M. Iftah. Generalizability theory: 1973–1980. *British Journal of Mathematical and Statistical Psychology*, 34: 133–166, 1981.
- [26] Patrick E. Shrout, Joseph L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2): 420–428, 1979.

- [27] Charles E. Spearman Correlation calculated from faulty data. *British Journal of Psychology*, 3: 271–295, 1910.
- [28] *Standardy pro pedagogické a psychologické testování*. Praha: Testcentrum, 2001.
- [29] Louis L. Thurstone, Thelma G. Thurstone. Factorial studies of intelligence. *Psychometric Monographs*, 2(94): 133–166, 1941.
- [30] Ledyard R. Tucker. An inter-battery method of factor analysis. *Psychometrika*, 23(2): 111–136, June 1958.
- [31] Ledyard R. Tucker, Charles Lewis. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1): 1–10, March 1973.
- [32] Noreen M. Webb, Richard J. Shavelson. Generalizability theory: Overview. 2005.
- [33] Qingyu Yang, Nailong Zhang, Yili Hong. Reliability analysis of repairable systems with dependent component failures under partially perfect repair. *Reliability, IEEE Transactions on*, 62(2): 490–498, June 2013.
- [34] Donald W. Zimmerman. Test reliability and the Kuder-Richardson formulas: Derivation from probability theory. *Educational and Psychological Measurement*, 32: 939–954, 1972.
- [35] Richard E. Zinbarg, Iftah Yovel, William Revelle. Estimating omega for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement*, 31(2): 135–157, 2007.
- [36] Richard E. Zinbarg, Iftah Yovel, William Revelle, Wen Li. Cronbach's alpha, Revelle's beta, McDonald's omega: Their relations with each and two alternative conceptualizations of reliability. *Psychometrika*, 70: 123–133, 2005.
- [37] Richard E. Zinbarg, Iftah Yovel, William Revelle, Roderick McDonald. Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for omega. *Applied Psychological Measurement*, 31(2): 135–157, 2006.
- [38] Karel Zvára. Měření reliability aneb Bacha na Cronbacha. *Informační bulletin České statistické společnosti*, 13(2): 13–20, 2002.

NA ČEM ZÁVISÍ SPOLEHLIVOST DOTAZNÍKU?

WHAT DOES THE RELIABILITY OF QUESTIONNAIRE DEPEND ON?

Zdeněk Půlpán

Adresa: Na Brně 1952/39, 500 09 Hradec Králové 9

E-mail: Zdenek.Pulpan@klikni.cz

Abstrakt: Spolehlivost Re dotazníku závisí na populaci, pro kterou je určena, a na neurčitosti skórovací procedury (popsané dobře nebo špatně v manuálu dotazníku). V práci je ukázán vliv neurčitosti skórování na odhad spolehlivosti dotazníku. Jako nejspolehlivější se jeví dotazník aplikovaný v populaci, jejíž úroveň není ani příliš nízká ale ani ne příliš vysoká; měřeno ukazatelem úrovně populace p , což je pravděpodobnost správné odpovědi, by mělo být přibližně pro každou položku $p \sim 0,5$. Odhad spolehlivosti dotazníku předpokládá splnění i řady podmínek pro konstrukci dotazníku. Je to například jistá konzistence položek, znalost obtížnosti položek a jejich rozdělení v pořadí zadávání, dostatečný počet dotazníkových položek a jejich vzájemná korelace.

Klíčová slova: Spolehlivost dotazníku, interpretace statistických metod, teorie dotazníku, chyba měření.

Abstract: The reliability Re of the questionnaire depends on the population for which the questionnaire was addressed, and on the uncertainties of scoring procedures (described well or poorly in the manual of the questionnaire under consideration). The highest reliability is achieved when the questionnaire is applied in the population, whose level due to the questionnaire is not too low nor too high; it is evident that larger value Re is achieved if level of population (level of population is probability of the correct answer) is $p \sim 0.5$. Estimating the reliability of the questionnaire is assumed to satisfy a number of additional conditions for the construction of the questionnaire. Reliability of the questionnaire depends on the population for which it is intended and on the uncertainty scoring procedures (described well or poorly in the manual of the questionnaire). This work shows the influence of uncertainty of scoring procedures to estimate the reliability of the questionnaire. It is, for example, a certain consistency of items, their difficulty and breakdown in order entry, the number of questionnaire items and their correlation.

Keywords: Reliability of Questionnaire, Interpretation of Statistical Methods, Theory of Questionnaire, Error of Measurement.

Nejdůležitější charakteristikou dotazníku (testu) je po validitě jeho spolehlivost (označuje se také jako reliabilita). „*Spolehlivý je takový test, který při opětném použití u týchž zkoušených přináší v podstatě tytéž výsledky.*“ [8] V pracích [1], [2], [3], [4], [6] nebo [7] je ukázáno, jak lze spolehlivost dotazníku statistickými prostředky odhadnout. Zde poukážeme na některé vlastnosti nejčastěji používaných ukazatelů spolehlivosti, které je třeba respektovat již při konstrukci dotazníku a pak při jejich užití a interpretaci z nich odvozených hodnot.

Spolehlivost statisticky definujeme pomocí testového výsledku X a chybového faktoru ε . Výsledkem testování je nejčastěji tzv. hrubý testový skór, který je součtem skórů jednotlivých položek. O chybovém faktoru se předpokládá, že je určen buď osobnostními charakteristikami respondenta (např. únavou) nebo charakteristikami majícími svůj původ v dotazníkových položkách nebo ve způsobu skórování. Pro statistickou definici spolehlivosti si představujeme testový výsledek X vyjádřený ve tvaru součtu

$$X = T + \varepsilon, \quad (1)$$

kde T je skutečná, přímému měření nedostupná ideální hodnota testového výkonu. Všechny tři veličiny X , T a ε chápeme jako náhodné; u veličin T a ε předpokládáme ještě jejich nekorelovanost. Spolehlivost dotazníku Re je pak definována vztahem

$$Re = \frac{DT}{DX} = 1 - \frac{D\varepsilon}{DX}, \quad 0 < Re < 1, \quad (2)$$

kde $D\varepsilon > 0$ označuje rozptyl chybového faktoru, $DT > 0$ rozptyl ideálního, $DX > 0$ rozptyl registrovaného testového výsledku (nenulovost rozptylů je přijatelným předpokladem pro reálnou situaci, zatíženou neurčitostí).

Vezmeme-li v úvahu pouze skórovací proceduru, můžeme ukázat jak ovlivňuje spolehlivost dotazníku. Předpokládejme proto, že

$$\varepsilon = \varepsilon_1 + \varepsilon_2, \quad (3)$$

kde ε_1 je chybový faktor, ovlivněný respondentem a ε_2 chybový faktor, ovlivněný metodikou skórování (jiné faktory neuvažujeme). Je přirozené předpokládat nezávislost náhodných veličin ε_1 a ε_2 . Proto můžeme psát pro rozptyly

$$D\varepsilon = D\varepsilon_1 + D\varepsilon_2. \quad (4)$$

Dosadíme-li tento vztah do (2), máme

$$Re = 1 - \frac{D\varepsilon}{DX} = 1 - \frac{D\varepsilon_1}{DX} - \frac{D\varepsilon_2}{DX}. \quad (5)$$

Když by se nám například podařilo odhadnout $\Delta_2 Re = \frac{D\varepsilon_2}{DX}$, získali bychom informaci o tom, jak se sníží spolehlivost dotazníku vlivem chyby skórování.

Předpokládejme, že používáme skórování jen typu 0 – 1 (tj. např. za správně zodpovězenou položku určujeme skór 1, nesprávně zodpovězenou skór 0) na intervalu $\langle 0; 1 \rangle$ latentní bazické spojité intervalové stupnici určitého atributu (znalosti... , schopnosti...). Rozlišujeme pro ilustraci pouze dva možné druhy neurčitosti skórování A a B , které jsou popsány dvěma druhy spojitých náhodných veličin s hustotami pravděpodobnosti jejichž grafy jsou na obr. 1 a obr. 2. V případě A – viz obr. 1 (volíme pak ε_2^A – s horním indexem A) se jedná o dvě variantní hustoty (jedna pro případ, že se volí skór 1, druhá pro volbu skóru 0) definované vztahy:

$$\begin{aligned} \text{a) volba skóru 1: } f_1(x) &= \begin{cases} 2x & \text{když } 0 \leq x \leq 1, \\ 0 & \text{jinde;} \end{cases} \\ \text{b) volba skóru 0: } f_2(x) &= \begin{cases} 2 - 2x & \text{když } 0 \leq x \leq 1, \\ 0 & \text{jinde.} \end{cases} \end{aligned} \quad (6A)$$

Podobně i ve druhém případě B máme dvě hustoty náhodných veličin ve tvaru:

$$\begin{aligned} \text{a) pro volbu skóru 1: } f_1(x) &= \begin{cases} 8x - 4 & 0,5 \leq x \leq 1, \\ 0 & \text{jinde;} \end{cases} \\ \text{b) pro volbu skóru 0: } f_2(x) &= \begin{cases} 4 - 8x & 0 \leq x \leq 0,5, \\ 0 & \text{jinde.} \end{cases} \end{aligned} \quad (6B)$$

Obě náhodné veličiny z a) i b) vztahů (6A) mají stejné rozptyly. Rozptyl obou náhodných veličin je roven $\frac{1}{18} \sim 0,0556$. V dotazníku s n nezávislými chybami položek je pak celkový rozptyl $D\varepsilon_2^A$ roven

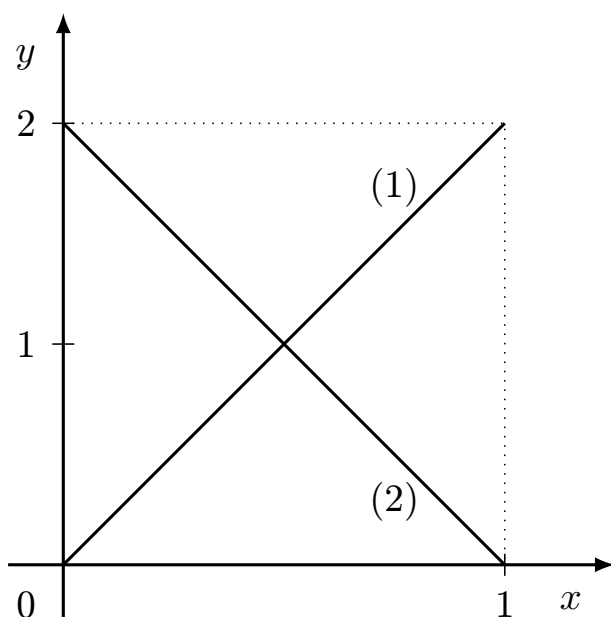
$$D\varepsilon_2^A = \frac{n}{18} \sim 0,0556 \cdot n. \quad (7)$$

Rozptyl v případě a) i b) vztahů (6B) je opět stejný a zde roven $\frac{1}{72} \sim 0,0139$. Proto za uvedených podmínek rozhodování je (volíme v tomto případě značení ε_2^B – s horním indexem B)

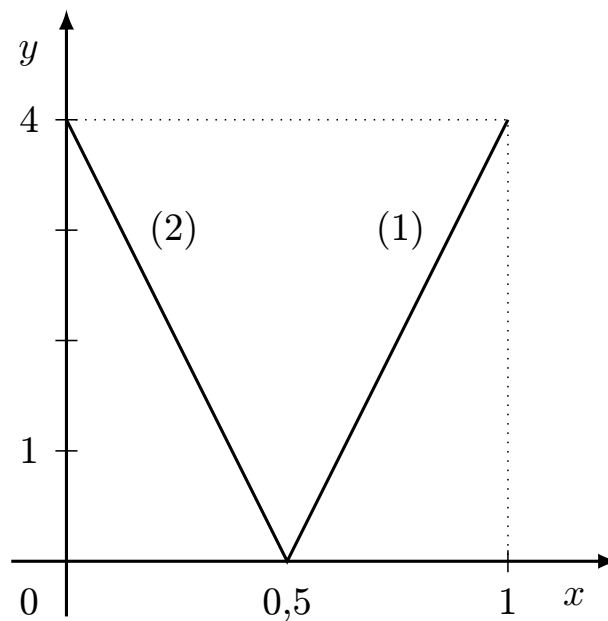
$$D\varepsilon_2^B = \frac{1}{72} \cdot n \sim 0,0139 \cdot n. \quad (8)$$

Z výrazů (7) a (8) můžeme posoudit velikost změny spolehlivosti dotazníku v důsledku změny neurčitosti odhadu dotazníkového skóru podílem, který již nezávisí na DX a n :

$$\frac{\Delta_2 Re_A}{\Delta_2 Re_B} = \frac{D\varepsilon_2^A}{D\varepsilon_2^B} = \frac{\frac{n}{18}}{\frac{n}{72}} = 4. \quad (9)$$



Obrázek 1



Obrázek 2

Grafy hustot rozdělení (6A) a (6B).

Na obrázku 1 je vyjádřena předpokládaná neurčitost při diagnostikování skóre 0 (označeno (2)) a 1 (označeno (1)) v prvním případě A modelových hustot (6A) a na obrázku 2 totéž ve druhém případě B modelových hustot (6B). Ve druhém případě se zřejmě jedná o přesnější a tudíž i spolehlivější diagnózu.

Volba skórovací procedury je vždy i volbou neurčitosti skórování vzhledem k dotazníkovým položkám. Je-li naše rozhodování obtížnější (viz obr. 1), je zatíženo větší neurčitostí volby skóru. Zde je to vyjádřeno čtyřikrát větší hodnotou $\Delta_2 Re_A$ vzhledem k $\Delta_2 Re_B$.

Uvažujme nyní zjednodušenou situaci, kdy zadáváme dotazník v souboru respondentů přibližně stejné úrovně, kterou charakterizuje parametr $p \in \langle 0; 1 \rangle$. Předpokládejme přitom ještě, že dotazník je složen z položek *stejně obtížnosti* (z této představy vychází obecně psychology a pedagogy formulovaný předpoklad tzv. “vnitřní konzistence” dotazníku; pojem stejné obtížnosti položek je však možné vztahovat jak k dotazníku při určité a známé homogenitě výkonů členů populace, tak také k populaci při konzistenci položek dotazníku). Pak můžeme psát (když $n > 1$)

$$\begin{aligned}
 X &= \sum_{i=1}^n X_i; \quad EX_i = p; \quad DX_i = p(1 - p); \\
 DX &= \sum_{i=1}^n DX_i + 2 \sum_{i < j} \varrho_{ij} \sqrt{DX_i} \sqrt{DX_j} = np(1 - p)[1 + (n - 1) \cdot \bar{\varrho}], \\
 \bar{\varrho} &= \frac{2}{n(n-1)} \sum_{i < j} \varrho_{ij}; \quad 0 \leq p \leq 1,
 \end{aligned}$$

kde X_i označuje skóre i -té položky ($X_i \in \{0, 1\}$), EX_i , resp. DX_i , střední hodnotu, resp. rozptyl skóre i -té položky a ϱ_{ij} korelaci mezi skóre i -té a j -té položky.

Ze vztahu (2) pro ukazatel spolehlivosti dotazníku pak dosazením za DX dostaneme za přijatelných předpokladů $0 < p < 1$, $n > 1$ a $1 + (n - 1)\bar{\varrho} > 0$ vztah pro Re ve tvaru

$$Re = 1 - \frac{D\varepsilon}{np(1 - p)[1 + (n - 1)\bar{\varrho}]}. \quad (10)$$

Z předchozího vztahu (10) vidíme, že spolehlivost dotazníku (odhadovaná prostřednictvím ukazatele (2)) se zvyšuje se zvětšováním počtu dotazníkových položek n a průměrné korelace mezi položkami. Na druhé straně je ale také spolehlivost závislá na úrovni populace (odhadované pomocí p), ve které je dotazník aplikován. Nelze tedy chápat ukazatel Re a všechny další ukazatele spolehlivosti, vycházející ze vztahu (2) (jako je např. vzorec Spearman-Brownův (12) nebo Cronbachovo alfa (16)) jako charakteristiku pouze dotazníku. Ze vztahu (10) plyne, že větší hodnoty Re odhadneme v populaci průměrné úrovně ($p \sim 0,5$) než v populaci, kde $p \rightarrow 0$ nebo $p \rightarrow 1$. (Připomeňme ještě, že paradoxně by byl vysoce spolehlivý dotazník s mnohonásobně opakovanou stejnou položkou. A to je v rozporu se smyslem každého šetření pomocí dotazníku.)

Je ještě jedna zajímavá interpretace spolehlivosti Re . Když bychom měli dvě ekvivalentní verze (a), (b) určitého dotazníku, pak bychom mohli předpokládat, že platí:

$$\begin{aligned} X^a &= T + \varepsilon^a; \quad X^b = T + \varepsilon^b; \quad EX^a = ET + E\varepsilon^a = ET + E\varepsilon^b = ET; \\ D\varepsilon^a &= D\varepsilon^b = D\varepsilon, \quad DX^a = DX^b = DT + D\varepsilon^a = DT + D\varepsilon^b = DT + D\varepsilon. \end{aligned}$$

Zřejmě je pak i oprávněný předpoklad nezávislosti chybové složky ε na výsledku T .

Při nezávislých realizacích je pak korelační koeficient pozorovatelných hodnot těchto dvou verzí roven

$$\varrho(X^a, X^b) = \frac{DT}{DX} = 1 - \frac{D\varepsilon}{DX} = Re. \quad (11)$$

Pro realizovatelný odhad míry spolehlivosti dotazníku se ukázala jako schůdná metoda, předpokládající existenci k danému dotazníku ještě aspoň jednoho dotazníku ekvivalentního s dotazníkem daným (co do výsledků v rovnocenné populaci respondentů). To však je často téměř nemožné. Obchází se

to ale tím, že se daný dotazník rozdělí na dva (jeden dotazník bude např. s lichými pořadovými čísly položek, druhý se sudými). Je-li pak r_{12} výběrový korelační koeficient mezi výsledky obou polovin dotazníku, odhaduje se spolehlivost původního dotazníku následujícím postupem užitím vztahu (2) a (11):

$$Re = \frac{D(2T)}{D(X^a + X^b)} = \frac{4DT}{2DX(1 + r_{12})} \sim \frac{2r_{12}}{1 + r_{12}}. \quad (12)$$

A to je známý vztah Spearman-Brownův. (Předpokládá se, že výběrová korelace $r_{12} > 0$, protože obě poloviny dotazníku měří totéž; v případě, že $r_{12} \leq 0$, je třeba prověřit hlavně validitu dotazníku.)

Za předpokladu, že chybové složky dvou komplementárních částí dotazníku jsou nekorelované ($\text{cov}(\varepsilon^a, \varepsilon^b) = 0$), můžeme pro rozptyl rozdílů výsledků obou částí psát

$$\begin{aligned} D(X^a - X^b) &= D((T + \varepsilon^a) - (T + \varepsilon^b)) = D(\varepsilon^a - \varepsilon^b) = \\ &= D\varepsilon^a + D\varepsilon^b = D(\varepsilon^a + \varepsilon^b) = D\varepsilon. \end{aligned} \quad (13)$$

Užitím předchozího vztahu (13) a definičního vztahu (2) můžeme odhad spolehlivosti dotazníku vyjádřit pak také ve tvaru

$$Re \sim 1 - \frac{s_{X^a - X^b}^2}{s_X^2}, \quad (14)$$

kde $s_{X^a - X^b}^2$, s_X^2 jsou výběrové rozptyly rozdílů testových skóre obou částí, resp. celého dotazníku. Dostaneme tak vztah Rulonův.

Spolehlivost Re můžeme odhadovat i tak, že ve vztahu (2) vyjádříme $D\varepsilon$ pomocí (13) a za podmínky, že $X = X^a + X^b$:

$$\begin{aligned} Re &= \frac{DX - D\varepsilon}{DX} = \frac{D(X^a + X^b) - D(X^a - X^b)}{DX} = \frac{4 \cdot \text{cov}(X^a, X^b)}{DX} = \\ &= 2 \cdot \frac{DX - DX^a - DX^b}{DX} = 2 \cdot \left(1 - \frac{DX^a + DX^b}{DX}\right). \end{aligned}$$

$$\text{Proto je } Re \sim 2 \left(1 - \frac{s_{X^a}^2 + s_{X^b}^2}{s_X^2}\right). \quad (15)$$

Vztah (15) je vztahem Guttmanovým.

Odhad reliability jako ukazatele vnitřní konzistence dotazníku se nejčastěji realizuje pomocí Cronbachova koeficientu alfa. Tento ukazatel je považovaný za jeden z nejužívanějších koeficientů vnitřní konzistence

$$Re \sim \frac{n}{n-1} \left(1 - \frac{\sum_1^n s_i^2}{s^2}\right) = \alpha, \quad (15)$$

kde s_i^2 je odhad rozptylu i -té položky a s^2 je odhad rozptylu celého dotazníku. [2] Vychází z předpokladu, že všechny položky, měřící týž atribut, by měly mít mezi sebou kladné a dostatečně vysoké vzájemné korelace.

Na základě toho, že dotazník je složen z různých ekvivalentních podčástí, odvodili Kuder a Richardson velké množství ukazatelů spolehlivosti. Nejznámější jsou ukazatelé, které se označují KR20, KR21, ... (o tom podrobněji v [1], [3], [6], [7] a [9]). Koeficient KR20 je vlastně koeficient (16), aplikovaný na dotazník s dichotomickými položkami (byl publikován již v roce 1937, Cronbach až v roce 1951 vytvořil jeho zobecnění a teorii vnitřní konzistence dotazníku). Užívá se hlavně v případě tzv. výkonových zkoušek, kdy se používají dotazníky s položkami uspořádanými vzestupně vzhledem k jejich obtížnosti (měřené například jako $1 - p$). Každý zkoušený by měl vyřešit všechny položky až do určité položky ve stanoveném pořadí.

O spolehlivosti hovoříme i v případě, kdy neměříme, ale posuzujeme. Pak spolehlivost je mírou shody posuzovatelů (počítá se Kendallovo tau nebo koeficienty kappa). [9] Modernější pojetí užívá hlubších psychologicko-statistických pohledů na dotazníkové výsledky (v psychologii se tato teorie nazývá *teorie zobecnitelnosti*). Jde o specifické modely analýzy rozptylu na data, vzniklá dotazníkovým šetřením za jistých podmínek. [9], [11]

Závěr

Spolehlivost dotazníku, odhadovaná prostřednictvím výrazu (2), je číslo mezi 0 a 1, vyjadřující do jaké míry se můžeme o výsledek měření opírat při jeho interpretaci (jaká je stabilita výsledného skóru při případném opakování). Čím je větší, tím spolehlivější jsou výsledky měření. Spolehlivost dotazníku ale také závisí na úrovni populace, pro kterou je dotazník určen a na neurčitosti skórovací procedury (popsané dobře či špatně v manuálu uvažovaného dotazníku). Vyšší spolehlivost má dotazník, aplikovaný v populaci, jejíž úroveň vzhledem k dotazníku není příliš nízká nebo také příliš vysoká. Různí ukazatelé spolehlivosti se vypočítávají na základě zjednodušujících předpokladů o stavbě dotazníku jako je například jistá homogenita, konzistence, rovnocennost či závislost výsledků jistých podčástí dotazníku.

To je třeba vždy respektovat při interpretaci ukazatelů spolehlivosti. Platí však, že dotazník nemůže být užitečným nástrojem, když není *validní*. Validita je míra toho, že dotazník měří to, co měřit má. Proto reliabilita jako míra zatížení skóre chybami (nebo konzistence opakované procedury měření) má smysl jen tam, kde víme, co měříme. Spoléhání na korelační koeficienty je sice užitečné, ale jen do určité míry. Čím vyšší je vnitřní konzistence, tím více skóre položek mohou svědčit (ze vzájemných korelací) o tomtéž (mají

vysoké a kladné vzájemné korelace), reliabilita se zvyšuje, ale snižuje se validita dotazníku (tj. vypovídací hodnota o sledovaném atributu). Atributy mají více dimenzí, mají-li se zachytit v dotazníkových reakcích, musí být obsaženy v různých dotazníkových položkách (které pak mohou mít různé vzájemné korelace), proto úpravou dotazníku ke zvýšení validity se často současně snižuje jeho spolehlivost (označuje se to jako tzv. psychometrický paradox). Proto je nutné nejprve zjistit, jak probíhá odpovídání na položky a kde jsou možné zdroje chyb při interpretaci skóre. Zde se účinně může uplatnit logistická regrese, ale hlavně zdravý rozum, viz [2], [3], [9] a [10]. Jiný než „klasický“ pohled na problematiku dotazníku podává [12].

Literatura

- [1] Půlpán, Z.: *K problematice zpracování empirických šetření v humanitních vědách*, 1. vyd. Academia, Praha 2004. 180 s. ISBN 80-200-1221-4.
- [2] Zvára, K.: Měření reliability aneb bacha na Cronbacha, *Informační bulletin České statistické společnosti*, 13(2): 13–20, 2002. ISSN 1210-8022. doi: 10.5300/IB
- [3] Lienert, G. A., Raatz, U.: *Testaufbau und Testanalyse*, 6. vyd. Beltz, 1998. 442 s. ISBN 978-3-621-27424-1.
- [4] Cronbach, L. J.: Coefficient alpha and the internal structure, *Psychometrika*, 16(3): 297–334, 1951. ISSN 0033-3123. doi: 10.1007/BF02310555
- [5] Richardson, M. W.: Notes on the rational of item analysis, *Psychometrika*, 1(1): 69–76, 1936. ISSN 0033-3123. doi: 10.1007/BF02287926
- [6] Lord, F. M.: *Applications of Item Response Theory to Practical Testing Problems*, Hillsdale, NJ, Erlbaum 1980. 274 s. ISBN 978-0898590067.
- [7] Richardson, M. W., Kuder, G. F.: The calculation of test reliability coefficients based upon the method of rational equivalence, *Journal of Educational Psychology*, 30(9): 1939, 681–687. ISSN 0022-0663. doi: 10.1037/h0054933
- [8] Mužić, V.: *Testy vědomostí*, 1. vyd. SPN – pedagogické nakladatelství, Praha, 1971. 130 s. Přeložil Vladimír Togner.
- [9] Urbánek, T. a kol.: *Psychometrika (Měření v psychologii)*, Portál, Praha 2011.
- [10] Guilford, J. P.: *Podstawowe metody statystyczne w psychologii i pedagogice*, PWN, Warszawa 1964.
- [11] Hendl, J.: *Přehled statistických metod*, Portál, Praha 2009.
- [12] Půlpán, Z.: *Odhad informace z dat vágní povahy*, Academia, Praha 2012.

SIMULTÁNNÍ KONFIDENČNÍ ELIPSOIDY

SIMULTANEOUS CONFIDENCE ELLIPSOIDS

Lubomír Kubáček

Adresa: Přírodovědecká fakulta Univerzity Palackého v Olomouci,
17. listopadu 12, 771 46 Olomouc

E-mail: kubacek.lubos@gmail.com

Abstrakt: Konfidenční elipsoid, jako míra spolehlivosti určení vícerozměrného parametru, není vhodný v případě, že se zajímáme o několik subvektorů tohoto parametru současně. Jestliže subvektory jsou jednorozměrné, infomaci získáme pomocí Schefféovy nerovnosti. Problémem je, jestliže subvektory jsou více než jednorozměrné.

Klíčová slova: konfidenční elipsoid, simultánní konfidenční elipsoidy.

Abstract: The confidence ellipsoid, as a measure of reliability of estimator of a multivariate parameter, is not appropriate if we are interested simultaneously in several subvectors of the parameter. If these subvectors are one-dimensional information can be gained from the Scheffé's inequality. Problems arise in the situation of non one-dimensional subvectors.

Keywords: confidence ellipsoid, simultaneous confidence ellipsoids.

1. Úvod

Při vytyčování přímé linie (např. u rychlostních koridorů) se přímost kontroluje pomocí několika bodů na linii. Standardní postup – testování hypotézy, že daných n bodů leží na přímce – vychází z konfidenčního elipsoidu v $2n$ -rozměrném prostoru a neumožňuje v případě zamítnutí hypotézy zjistit, který bod se významně odchyluje. Pro inženýra je mnohem výhodnější postup, který přímo indikuje vadný bod. Teoretická východiska takového postupu uvádíme v článku. Více podobných – potřebami praxe motivovaných – postupů najde čtenář v [1].

Uvažujme regresní model

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

kde $\mathbf{Y}$ je n -rozměrný náhodný vektor normálně rozdělený, $\mathbf{X}$ je známá $n \times k$ matice s hodnotí $r(\mathbf{X}) = k < n$ a $\boldsymbol{\Sigma}$ je pozitivně definitní kovarianční matice náhodného vektoru $\mathbf{Y}$; k -rozměrný vektor $\boldsymbol{\beta}$ je neznámý parametr.

Konfidenční elipsoid $\mathcal{E}_{\boldsymbol{\beta}}$ s hladinou confidence $1 - \alpha$ pro parametr $\boldsymbol{\beta}$ je (viz [3])

$$\mathcal{E}_\beta = \{\mathbf{u} : (\mathbf{u} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{u} - \hat{\boldsymbol{\beta}}) \leq \chi_k^2(0; 1 - \alpha)\},$$

kde $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{Y}$ je nejlepší nestranný lineární odhad parametru $\boldsymbol{\beta}$ a $\chi_k^2(0; 1 - \alpha)$ je $(1 - \alpha)$ -kvantil centrálního chí-kvadrát rozdělení s k stupni volnosti.

Jestliže $\boldsymbol{\Sigma} = \sigma^2 \mathbf{V}$, kde $\mathbf{V}$ je známá pozitivně definitní matice a σ^2 je neznámý skalár, potom

$$\mathcal{E}_\beta = \{\mathbf{u} : (\mathbf{u} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} (\mathbf{u} - \hat{\boldsymbol{\beta}}) \leq \hat{\sigma}^2 k F_{k, n-k}(0; 1 - \alpha)\},$$

kde

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y}, \quad \hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})}{n - k}$$

a $F_{k, n-k}$ je $(1 - \alpha)$ -kvantil centrálního Fisherova-Snedecorova rozdělení s k a $n - k$ stupni volnosti.

Problémem je určení konfidenčních elipsoidů $\mathcal{E}_{\beta_i}$, kde $i = 1, \dots, m$, kde $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_m)'$, přičemž $\dim \boldsymbol{\beta}_i \geq 2$, $i = 1, \dots, m$, pro které platí

$$P\{\forall \{i = 1, \dots, m\} \boldsymbol{\beta}_i \in \mathcal{E}_{\beta_i}\} \geq 1 - \alpha.$$

Elipsoidy s touto vlastností nazýváme simultánní konfidenční elipsoidy.

2. Řešení

Nechť $\mathcal{M}(\mathbf{A}) = \{\mathbf{A}_{k,l} \mathbf{u} : \mathbf{u} \in R^l\}$ je sloupcový podprostor matice $\mathbf{A}$ v k -rozměrném lineárním vektorovém prostoru R^k . Potom projekční matice v Euklidově normě na $\mathcal{M}(\mathbf{A})$ je $\mathbf{P}_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$, kde „ $-$ “ označuje pseudoinverzi (podrobněji v [2]).

Protože $\mathbf{P}_A$ je symetrická a idempotentní platí $\mathbf{P}_A = \mathbf{J}\mathbf{J}'$, kde $\mathbf{J}$ je $k \times r(\mathbf{A})$ matice. Zřejmě platí $\mathcal{M}(\mathbf{A}) = \mathcal{M}(\mathbf{J})$ a $\mathbf{J}'\mathbf{J} = \mathbf{I}_{r(\mathbf{A})}$ (identická matice rozměru $r(\mathbf{A})$).

Nechť $\mathbf{K}'$ je $[k - r(\mathbf{A})] \times k$ matice s vlastností $\mathbf{K}'\mathbf{J} = \mathbf{0}$, $r(\mathbf{K}) = k - r(\mathbf{A})$ a $\mathbf{K}'\mathbf{K} = \mathbf{I}_{k-r(\mathbf{A})}$. Matice $(\mathbf{J}, \mathbf{K})$ je tedy ortogonální a regulární.

Nechť $\mathbf{C}$ je pozitivně definitní, symetrická $k \times k$ matice. Potom průmět elipsoidu

$$\mathcal{E} = \{\mathbf{x} : \mathbf{x}'\mathbf{C}\mathbf{x} \leq c^2\}$$

na sloupcový podprostor matice $\mathbf{P}_A$

je

$$\mathbf{P}_A \mathcal{E} = \{\mathbf{P}_A \mathbf{x} : \mathbf{x}'\mathbf{C}\mathbf{x} \leq c^2\}.$$

Vzhledem k regularitě matice $(\mathbf{J}, \mathbf{K})$ platí:

$$\begin{aligned} \mathbf{x}'\mathbf{C}\mathbf{x} &= \mathbf{x}'(\mathbf{J}, \mathbf{K}) \left[\begin{pmatrix} \mathbf{J}' \\ \mathbf{K}' \end{pmatrix} \mathbf{C}^{-1}(\mathbf{J}, \mathbf{K}) \right]^{-1} \begin{pmatrix} \mathbf{J}' \\ \mathbf{K}' \end{pmatrix} \mathbf{x} \\ &= \mathbf{x}'(\mathbf{J}, \mathbf{K}) \begin{pmatrix} \mathbf{J}'\mathbf{C}^{-1}\mathbf{J} & \mathbf{J}'\mathbf{C}^{-1}\mathbf{K} \\ \mathbf{K}'\mathbf{C}^{-1}\mathbf{J} & \mathbf{K}'\mathbf{C}^{-1}\mathbf{K} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{J}' \\ \mathbf{K}' \end{pmatrix} \mathbf{x} \\ &= \mathbf{x}'(\mathbf{J}, \mathbf{K}) \begin{pmatrix} \boxed{11} & \boxed{12} \\ \boxed{21} & \boxed{22} \end{pmatrix} \begin{pmatrix} \mathbf{J}' \\ \mathbf{K}' \end{pmatrix} \mathbf{x}, \end{aligned}$$

kde

$$\begin{aligned} \boxed{11} &= (\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1} + (\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{C}^{-1}\mathbf{K}[\mathbf{K}'\mathbf{C}^{-1}\mathbf{K} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1} \\ &\quad \times \mathbf{J}'\mathbf{C}^{-1}\mathbf{K}]^{-1}\mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1} \\ \boxed{12} &= -(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{C}^{-1}\mathbf{K}[\mathbf{K}'\mathbf{C}^{-1}\mathbf{K} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{C}^{-1}\mathbf{K}]^{-1}, \\ \boxed{21} &= -[\mathbf{K}'\mathbf{C}^{-1}\mathbf{K} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{C}^{-1}\mathbf{K}]^{-1}\mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}, \\ \boxed{22} &= [\mathbf{K}'\mathbf{C}^{-1}\mathbf{K} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{C}^{-1}\mathbf{K}]^{-1}. \end{aligned}$$

Nechť $\mathbf{M}_A = \mathbf{I} - \mathbf{P}_A$. Vzhledem k definici matice $(\mathbf{J}, \mathbf{K})$ matici $[\mathbf{K}'\mathbf{C}^{-1}\mathbf{K} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{C}^{-1}\mathbf{K}]^{-1}$ lze vyjádřit jako $[\mathbf{K}'(\mathbf{M}_A\mathbf{C}\mathbf{M}_A)^+\mathbf{K}]^{-1}$, kde „+“ označuje Mooreovu-Penroseovu pseudoinverzi (podrobněji viz [2]).

Označme $\mathbf{u} = \mathbf{J}'\mathbf{x}$, $\mathbf{v} = \mathbf{K}'\mathbf{x}$. Potom $\mathbf{P}_A\mathbf{x} = \mathbf{J}\mathbf{u}$ a

$$\begin{aligned} \mathbf{P}_A\mathcal{E} &= \left\{ \mathbf{J}\mathbf{u} : \mathbf{u}'(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u} + [\mathbf{v}' - \mathbf{u}'(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{C}^{-1}\mathbf{K}] \right. \\ &\quad \times [\mathbf{K}'(\mathbf{M}_A\mathbf{C}\mathbf{M}_A)^+\mathbf{K}]^{-1}[\mathbf{v} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u}] \leq c^2 \left. \right\}. \end{aligned}$$

Evidentně platí

$$\left\{ \mathbf{J}\mathbf{u} : \mathbf{x}'\mathbf{C}\mathbf{x} \leq c^2 \right\} \subset \left\{ \mathbf{J}\mathbf{u} : \mathbf{u}'(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u} \leq c^2 \right\}.$$

Na druhé straně pro každé $\mathbf{u}$ s vlastností $\mathbf{u}'(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u} \leq c^2$ existuje vektor $\mathbf{x}$ takový, že $\mathbf{x}'\mathbf{C}\mathbf{x} \leq c^2$. Plyne to z následující úvahy.

Vektor $\mathbf{x}$, pro který platí $\mathbf{u} = \mathbf{J}'\mathbf{x}$, lze vyjádřit jako $\mathbf{x} = \mathbf{x}_p + \mathbf{K}\mathbf{y}$, kde $\mathbf{y} \in R^{k-r(A)} = \mathcal{M}(\mathbf{K}')$ je libovolný a $\mathbf{x}_p$ je partikulárním řešením rovnice $\mathbf{J}'\mathbf{x} = \mathbf{u}$. Pro toto $\mathbf{x}$ platí $\mathbf{u} = \mathbf{J}'\mathbf{x} = \mathbf{J}'\mathbf{x}_p$ ($\mathbf{J}'\mathbf{K} = \mathbf{0}$) a $\mathbf{v} = \mathbf{K}'\mathbf{x}_p + \mathbf{y}$ ($\mathbf{K}'\mathbf{K} = \mathbf{I}_{k-r(A)}$), tedy

$$\mathbf{v} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u} = \mathbf{K}'\mathbf{x}_p + \mathbf{y} - \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u}.$$

Jestliže $\mathbf{y} = -\mathbf{K}'\mathbf{x}_p + \mathbf{K}'\mathbf{C}^{-1}\mathbf{J}(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u}$, potom $\mathbf{x}'\mathbf{C}\mathbf{x} = \mathbf{u}'(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u}$, tzn.

$$\forall \{\mathbf{u} : \mathbf{u}'(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})^{-1}\mathbf{u} \leq c^2\} \exists \{\mathbf{x} : \mathbf{x}'\mathbf{C}\mathbf{x} \leq c^2\}.$$

Tím jsme dokázali tvrzení.

Věta 1. Nechť $\mathbf{C}$ je $k \times k$ pozitivně definitní matice a $\mathbf{A}$ je libovolná $k \times l$ matice. Potom průmět elipsoidu $\mathcal{E} = \{\mathbf{x} : \mathbf{x}'\mathbf{C}\mathbf{x} \leq c^2\}$ na podprostor $\mathcal{M}(\mathbf{A})$ je

$$\mathbf{P}_A\mathcal{E} = \{\mathbf{J}\mathbf{u} : \mathbf{u}'(\mathbf{J}'\mathbf{C}^{-1}\mathbf{J})\mathbf{u} \leq c^2\},$$

kde $\mathbf{P}_A = \mathbf{J}\mathbf{J}'$, $\mathbf{J}'\mathbf{J} = \mathbf{I}_{r(A)}$, $\mathbf{P}_A\mathbf{x} = \mathbf{J}\mathbf{u}$. (Viz též [1], kde jsou řešeny i jiné analogické problémy.)

Zvolme za matici $\mathbf{A}$ libovolný k -rozměrný vektor $\mathbf{h}$ a nechť $\mathbf{x} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$. Potom $\mathbf{J} = \mathbf{h}/\sqrt{\mathbf{h}'\mathbf{h}}$, $\mathbf{J}'\mathbf{x} = \mathbf{h}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)/\sqrt{\mathbf{h}'\mathbf{h}}$ a $\mathcal{E} = \{\boldsymbol{\beta} - \boldsymbol{\beta}_0 : (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{C}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \leq c^2\}$,

$$\begin{aligned} \mathbf{P}_h\mathcal{E} &= \left\{ \mathbf{P}_h\mathbf{x} : \mathbf{x}'\mathbf{C}\mathbf{x} \leq c^2 \right\} = \left\{ \mathbf{J}\mathbf{u} : \mathbf{u}' \left(\frac{\mathbf{h}'}{\sqrt{\mathbf{h}'\mathbf{h}}} \mathbf{C}^{-1} \frac{\mathbf{h}}{\sqrt{\mathbf{h}'\mathbf{h}}} \right)^{-1} \mathbf{u} \leq c^2 \right\} \\ &= \left\{ \frac{\mathbf{h}}{\sqrt{\mathbf{h}'\mathbf{h}}} \frac{\mathbf{h}'}{\sqrt{\mathbf{h}'\mathbf{h}}} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) : (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \frac{\mathbf{h}}{\sqrt{\mathbf{h}'\mathbf{h}}} \left(\frac{\mathbf{h}'}{\sqrt{\mathbf{h}'\mathbf{h}}} \mathbf{C}^{-1} \frac{\mathbf{h}}{\sqrt{\mathbf{h}'\mathbf{h}}} \right)^{-1} \right. \\ &\quad \left. \times \frac{\mathbf{h}'}{\sqrt{\mathbf{h}'\mathbf{h}}} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \leq c^2 \right\} \\ &= \left\{ \frac{\mathbf{h}\mathbf{h}'}{\mathbf{h}'\mathbf{h}} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) : (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \mathbf{h}\mathbf{h}' (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \leq c^2 \mathbf{h}'\mathbf{C}^{-1}\mathbf{h} \right\} \\ &= \left\{ \frac{\mathbf{h}\mathbf{h}'}{\mathbf{h}'\mathbf{h}} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) : |\mathbf{h}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)| \leq c\sqrt{\mathbf{h}'\mathbf{C}^{-1}\mathbf{h}} \right\}. \end{aligned}$$

Jestliže $\boldsymbol{\beta}_0 = \hat{\boldsymbol{\beta}}$, $c^2 = \chi_k^2(0; 1 - \alpha)$ a $\mathbf{C} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$, je $\mathcal{E}$ konfidenční elipsoid pro parametr $\boldsymbol{\beta}$ a platí

$$\begin{aligned} P\{\boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \in \mathcal{E}\} &= P\{\forall \{\mathbf{h} \in R^k\} \mathbf{P}_h\mathcal{E} \ni \mathbf{P}_h(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\} \\ &= P\{\forall \{\mathbf{h} \in R^k\} : |\mathbf{h}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})| \leq \sqrt{\chi_k^2(0; 1 - \alpha)} \sqrt{\mathbf{h}'\mathbf{C}^{-1}\mathbf{h}}\} = 1 - \alpha, \end{aligned}$$

důkaz druhé rovnosti je obsahem Schefféovy věty (podrobněji ve [4]).

Jestliže se vrátíme k maticím $\mathbf{A}$, zřejmě platí

$$P\{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \in \mathcal{E}\} = P\{\forall \{\mathbf{A}\} \mathbf{P}_A(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \in \mathbf{P}_A\mathcal{E}\}.$$

V dalším rozdělíme k -rozměrný parameter $\boldsymbol{\beta}$ na m subvektorů $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m$, přičemž $\dim \boldsymbol{\beta}_i = k_i \geq 2$.

Jestliže za matici $\mathbf{A}_i$ zvolíme matici

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{0}_{\sum_{j=1}^{i-1} k_j, \sum_{j=1}^{i-1} k_j}, & \mathbf{0}_{\sum_{j=1}^{i-1} k_j, k_i}, & \mathbf{0}_{\sum_{j=1}^{i-1} k_j, \sum_{j=i+1}^m k_j} \\ \mathbf{0}_{k_i, \sum_{j=1}^{i-1} k_j}, & \mathbf{I}_{k_i}, & \mathbf{0}_{k_i, \sum_{j=i+1}^m k_j} \\ \mathbf{0}_{\sum_{j=i+1}^m k_j, \sum_{j=1}^{i-1} k_j}, & \mathbf{0}_{\sum_{j=i+1}^m k_j, k_i}, & \mathbf{0}_{\sum_{j=i+1}^m k_j, \sum_{j=i+1}^m k_j} \end{pmatrix},$$

dostáváme důsledek.

Důsledek.

$$P\left\{\forall\{i = 1, \dots, m\} \mathbf{P}_{A_i}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \in \mathbf{P}_{A_i} \mathcal{E}\right\} \geq P\left\{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \in \mathcal{E}\right\} = 1 - \alpha,$$

kde

$$\mathbf{P}_{A_i} \mathcal{E} = \left\{ \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i \\ \mathbf{0} \end{pmatrix} : (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)' (\mathbf{C}^{ii})^{-1} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i) \leq \chi_k^2(0; 1 - \alpha) \right\}$$

a

$$\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{C}^{11}, & \mathbf{C}^{12}, & \dots, & \mathbf{C}^{1m} \\ \mathbf{C}^{21}, & \mathbf{C}^{22}, & \dots, & \mathbf{C}^{2m} \\ \dots & \dots & \dots & \dots \\ \mathbf{C}^{m1}, & \mathbf{C}^{m2}, & \dots, & \mathbf{C}^{mm} \end{pmatrix}.$$

Elipsoidy

$$\mathcal{E}_{\beta_i} = \{\boldsymbol{\beta}_i : (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)' (\mathbf{C}^{ii})^{-1} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i) \leq \chi_k^2(0; 1 - \alpha)\}, \quad i = 1, \dots, m,$$

jsou simultánní elipsoidy pro subvektory $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m$ parametru $\boldsymbol{\beta}$.

Jestliže zopakujeme předchozí úvahy pro model $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{V})$, potom dospějeme k následujícímu tvrzení.

Věta 2. V modelu $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{V})$ jsou simultánní elipsoidy pro subvektory $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m$, dány vztahem

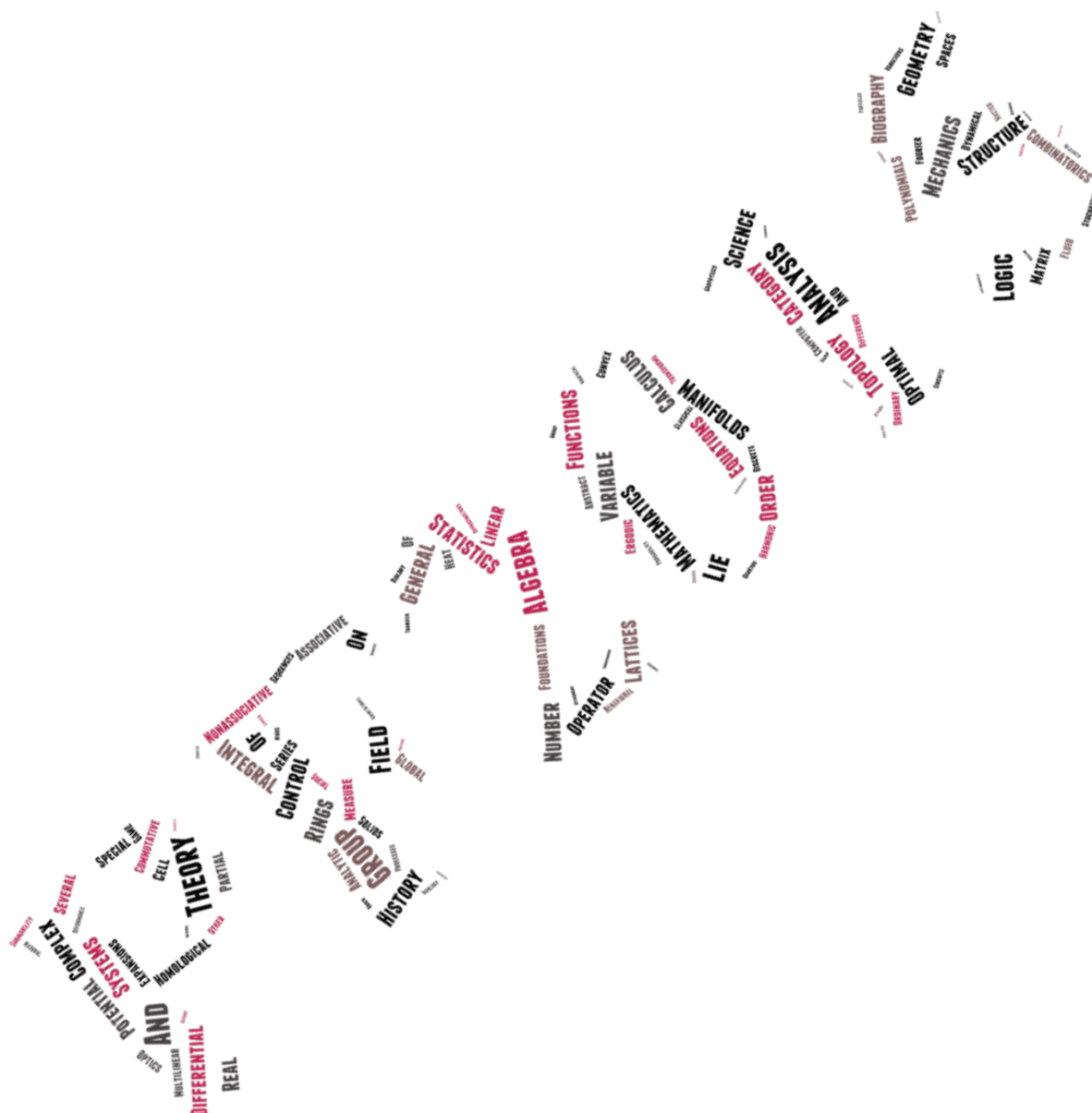
$$\begin{aligned} \mathcal{E}_{\beta_i} = \left\{ \boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i : (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)' (\mathbf{C}_0^{ii})^{-1} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i) \right. \\ \left. \leq \widehat{\sigma}^2 k F_{k, n-k}(0; 1 - \alpha) \right\}, \quad i = 1, \dots, m, \end{aligned}$$

kde

$$\mathbf{C}_0^{-1} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} = \begin{pmatrix} \mathbf{C}_0^{11}, & \mathbf{C}_0^{12}, & \dots, & \mathbf{C}_0^{1m} \\ \mathbf{C}_0^{21}, & \mathbf{C}_0^{22}, & \dots, & \mathbf{C}_0^{2m} \\ \dots & \dots & \dots & \dots \\ \mathbf{C}_0^{m1}, & \mathbf{C}_0^{m2}, & \dots, & \mathbf{C}_0^{mm} \end{pmatrix}.$$

Literatura

- [1] Kubáček, L. (2013): *Statistical Theory of Geodetic Networks*. Praha–Zdiby, Výzkumný ústav geodetický, topografický a kartografický, 2013.
- [2] Rao, C. R., Mitra, S. K. (1971): *Generalized Inverse of Matrices and Its Applications*. New York, J. Wiley, 1971
- [3] Rao, C. R. (1973): *Linear Statistical Inference and Its Applications* (2nd edition). New York, J. Wiley, 1973
- [4] Scheffé, H. (1967): *The Analysis of Variance* (5th printing). New York–London–Sydney, J. Wiley, 1967



ZASEDÁNÍ PŘEDSTAVITELŮ NÁRODNÍCH STATISTICKÝCH SPOLEČNOSTÍ V6 V PRAZE

AN INTERNATIONAL MEETING OF THE REPRESENTATIVES OF STATISTICAL SOCIETIES OF V6 IN PRAGUE

Hana Řezanková, Marek Malý

E-mail: rezanka@vse.cz, mmaly@szu.cz

Ve dnech 9. – 10. října 2014 se v Praze konalo jubilejní desáté setkání představitelů národních statistických společností z šesti zemí střední Evropy (V6). V této skupině byly do letošního roku kromě České republiky zastoupeny Maďarsko, Rakousko, Rumunsko, Slovensko a Slovinsko. Letos se nezúčastnila delegace z Rumunska, naopak byla zastoupena Polská statistická společnost, jejíž představitelé byli na zasedání přizváni.

Oficiální jednání probíhalo 10. října v budově Českého statistického úřadu. Ze zahraničních hostů se ho zúčastnili Dr. Istvan Szabo a Dr. Éva Laczka, místopředseda a generální tajemnice Maďarské statistické společnosti, prof. Grażyna Trzpiot, místopředsedkyně Polské statistické společnosti, Mag. Margit Epler, místopředsedkyně Rakouské statistické společnosti, RNDr. Peter Mach, místopředseda Slovenské statistické a demografické společnosti a Dr. Andrej Blejec, předseda Statistické společnosti Slovinska. Jako host byla přizvána paní docentka Marie Bohatá, dřívější předsedkyně ČSÚ (1999–2003) a zástupkyně generálního ředitele Eurostatu (2004–2014). Za ČSÚ byl přítomen pan vrchní ředitel sekce makroekonomických statistik, Ing. Jaroslav Sixta, který byl v minulé funkčním období členem výboru ČStS. Současný výbor ČStS zastupovali prof. Antoch, prof. Dohnal, RNDr. Malý, prof. Pícek, prof. Řezanková a Ing. Vozár.

Na úvod předsedkyně ČStS prof. Řezanková přivítala všechny přítomné a požádala pana vrchního ředitele Sixtu, aby zahájil zasedání. Po jeho úvodním slovu a přivítání účastníků jménem vedení Českého statistického úřadu vystoupila paní docentka Bohatá s příspěvkem „Second Round of Peer Reviews in the European Statistical System“, po němž následovala bohatá diskuze. Druhou zvanou přednášku přednesl profesor Antoch, který seznámil účastníky s úskalími algoritmu pro ověřování podpisů pro návrh kandidátů v prezidentských volbách v ČR. Po krátké přestávce jednání pokračovalo informacemi delegací z jednotlivých zemí o aktivitách národních statistických společností uskutečněných od minulého setkání v loňském roce v Lublani a také o aktivitách plánovaných na další období. Na závěr zasedání Margit Epler informovala delegáty statistických společností o návrhu stanoviska rady

ESGAB (European Statistical Governance Advisory Board) proti porušování prvního principu dokumentu European Statistics Code of Practice o nezávislosti statistických úřadů a požadavku na profesní způsobilost jeho top managementu, které bylo připraveno jako podpora prohlášení Německé statistické společnosti (DStG). Vyzvala delegáty statistických společností, aby podpořili prohlášení DStG a stanovisko rady ESGAB. Tento návrh byl jednohlasně přijat (prohlášení společností V6 bylo formulováno v následujících dnech po jednání prostřednictvím e-mailu a postupně podepsáno zástupci jednotlivých společností). Po skončení oficiálního jednání byla pořízena společná fotografie účastníků před budovou Českého statistického úřadu. Zasedání představitelů národních statistických společností bylo zakončeno společným obědem.

Poděkování patří pracovníkům kanceláře předsedkyně Českého statistického úřadu, jmenovitě panu řediteli Sidorovi a paní Kašparové, kteří pro účastníky zajistili zasedací místnost s občerstvením a poskytli materiály vydané na ČSÚ. Díky vedoucímu oddělení statistiky cestovního ruchu Ing. Vančurovi zahraniční hosté obdrželi z CzechTourismu informační materiály a dárek. Společnou fotografii z jednání pořídil Ing. Růžicka z oddělení mezinárodní spolupráce. Výbor České statistické společnosti si váží pomoci a vstřícnosti všech jmenovaných a děkuje za významné přispění při organizování setkání V6 svému kolegovi Ing. Vozárovi.

Před oficiálním jednáním se účastníci setkali již ve čtvrtek odpoledne, kdy navštívili Klementinum, jehož Astronomická věž poskytla nádherný pohled na sluncem ozářenou Prahu, a vedli neformální diskuse.

Schůzky představitelů národních statistických společností přispívají k vzájemné informovanosti a užší spolupráci mezi společnostmi. Hlavní výbor Polské statistické společnosti podpořil návrh společností skupiny V6 na zapojení polských statistiků a přejmenování skupiny na V7. Termín a místo příštího zasedání zatím nebyly stanoveny. Podle pravidelného pořadí se jednání mělo konat v Rumunsku, ale momentálně není zřejmé, jaké bude další zapojení Rumunské statistické společnosti do společných aktivit.

Hana Řezanková a Marek Malý



KOMENTÁŘ K SHEWHARTOVÝM REGULAČNÍM DIAGRAMŮM

SHEWHART CONTROL CHARTS: NOTES AND COMMENTS

Eva Jarošová

Adresa: Katedra logistiky a řízení kvality, Škoda Auto Vysoká škola,
Mladá Boleslav

E-mail: jarosova@is.savs.cz

Ráda bych zareagovala na článek Petra Klímka z červnového čísla Informačního bulletinu. Řízení kvality je zde omezeno na popis Shewhartových regulačních diagramů a z úvodu druhé kapitoly by čtenář mohl nabýt dojmu, že buď lze použít Shewhartovy diagramy, nebo je třeba přejít k vícerozměrným diagramům či dokonce k řešení stochastických diferenciálních rovnic. Chtěla bych doplnit některé možnosti postupu v případě nesplněných předpokladů,

především však musím oponovat tvrzení, že předpoklady je třeba ověřit před vlastní konstrukcí Shewhartova regulačního diagramu.

Bráno doslova by to znamenalo, že diagram je možné použít jen tehdy, jsou-li předpoklady splněny, to znamená jen v případě, kdy je proces pod statistickou kontrolou (úroveň a variabilita jsou konstantní, působí pouze náhodné vlivy inherentní procesu). Splnění předpokladů však není podmínkou použití Shewhartova diagramu, ale zárukou toho, že diagram bude mít očekávané vlastnosti, tedy např. že riziko falešného signálu bude 0,0027. Dostat proces pod kontrolu je jedním z cílů tohoto diagramu. Teprve po odstranění zvláštních příčin, které jsou signalizovány body mimo regulační meze, může mít ověřování předpokladů smysl.

Důsledky nesplnění předpokladů při regulaci procesu se ve vědeckých člancích spíše zveličují; v praxi se platnost předpokladů začne řešit především v případě, kdy dochází k častějším falešným signálům, tj. situacím, kdy se výskyt bodu mimo meze nepodařilo vysvětlit. Příčinou může být např. jiný tvar rozdělení nebo pozitivní autokorelace, která se může vyskytnout i u stabilního procesu, často je však důsledkem proměnlivé střední hodnoty. Výrobní procesy a možnosti jejich monitorování se od r. 1924 značně změnily a v současnosti se uvažuje osm různých modelů chování procesu v čase, viz ČSN ISO 21747:2010. Rozlišuje se přitom okamžité a výsledné rozdělení hodnot znaku. Uvedené předpoklady jsou splněny jen u jednoho z nich.

Statistická regulace procesu (nikoli řízení procesů, jak uvádí autor) probíhá ve dvou etapách. Zvláštní příčiny, které vedou k větším změnám úrovně procesu, působí hlavně v první etapě. Je třeba je identifikovat, a pokud je to možné, zásahem do procesu zabránit jejich opakování a dostat proces „pod kontrolu“. Po identifikaci odstranitelných zvláštních příčin se příslušné podskupiny vypustí a regulační meze se revidují. V druhé etapě SPC jde přinejmenším o udržení současného stavu nebo také o jeho zlepšení, doporučují se proto citlivější typy diagramů, tedy CUSUM nebo EWMA.

Není-li proces ani po odstranění zvláštních příčin pod kontrolou, tj. dochází-li stále k překračování regulačních mezí nebo existují-li náznaky systematických nebo náhodných změn střední hodnoty, je třeba provést důkladnější analýzu procesu spočívající v kontrole předpokladů. Častějším falešným signálům je třeba při regulaci zabránit, neboť snaha posunout střední hodnotu procesu „správným“ směrem může vést k tzv. přeregulování, jehož důsledkem je zvýšení variability procesu. Za tímto účelem byly pro procesy, u nichž nelze splnění uvedených předpokladů z různých důvodů dosáhnout, navrženy další metody. Omezíme-li se na diagramy pro průměry nebo pro individuální hodnoty, můžeme zvolit jiný model rozdělení a určit pravděpodobnostní meze pro zvolené riziko falešného signálu, využít transformace, sestrojít modifiko-

vaný diagram či diagram s rozšířenými mezemi, případně korigovat odhad směrodatné odchylky při zjištěné pozitivní autokorelaci. V modifikovaném regulačním diagramu jsou nové meze odvozeny od mezí určených specifikací, v diagramu s rozšířenými mezemi se vzdálenost mezí zvětší o předpokládaný rozsah změn střední hodnoty. V posledním případě se využívá hodnota výběrové autokorelační funkce pro zpoždění 1. Kromě těchto diagramů je také možné pracovat s rezidui regresního modelu v případě neodstranitelného trendu nebo s rezidui ARIMA modelu v případě autokorelace.

Spíše výjimečně lze vybrané diagramy nalézt v některém ze známých statistických softwarových produktů (např. regresní diagram nebo ARIMA diagram ve Statgraphicsu), obvykle si uživatel musí regulační meze vypočítat sám a zadat je programu pro konstrukci regulačního diagramu.

S regulací procesu je těsně spjata analýza způsobilosti, jejímž cílem je kvantifikovat schopnost procesu plnit „požadavky zákazníka“, tj. udržet hodnoty sledovaného znaku uvnitř tolerančního pole. Předpoklady je třeba ověřovat i v rámci této analýzy, v závislosti na modelu chování procesu se vybírá způsob výpočtu ukazatele způsobilosti či výkonnosti. O analýze způsobilosti se autor nezmiňuje, ačkoli ji program zjevně provádí, viz ukazatele C_p a C_{pk} na obr. 4 a 7.

Z popisu listu *Quality Control* vyplývá, že program *XLStatistics* je možné využít k ilustraci Shewhartových regulačních diagramů, tj. k zakreslení centrální přímky, regulačních mezí a bodů představujících výběrové charakteristiky, tedy k úspoře kroků, které by bylo třeba vykonat v samotném Excelu. Není už jasné, zda obsahuje něco navíc, tedy zda umožňuje základní analýzu diagramu, která spočívá nejen v kontrole bodů mimo regulační meze, ale také v testech zvláštních seskupení, viz norma ČSN ISO 8258. Není uvedeno, zda program obsahuje nějaké metody ověření předpokladů, jako test normality či pravděpodobnostní graf, případně některou z metod testování autokorelace.

Pro smysluplnou realizaci SPC je nutná nejen diagnostika procesu, ale i možnost úpravy regulačních mezí v případech, kdy je Shewhartův diagram nevhodný. Autor v závěru konstatuje, že *XLStatistics* bohatě pokrývá potřebu běžných uživatelů. Uvážíme-li, že „nevyhovujících“ procesů je v praxi většina, platí to i pro uživatele v oblasti kontroly kvality?

POZVÁNKA NA KONFERENCI PROBASTAT 2015 INVITATION TO PROBASTAT 2015 CONFERENCE

Viktor Witkovský

E-mail: witkovsky@savba.sk

The International Program Committee and the Local Organizing Committee in collaboration with the Institute of Measurement Science of the Slovak Academy of Sciences, Faculty of Mathematics, Physics and Informatics of the Comenius University, and Mathematical Institute of the Slovak Academy of Sciences

**Invite you to participate in the
PROBASTAT 2015**
the 7th International Conference on Probability and
Statistics
June 28 – July 3, 2015, Smolenice Castle, Slovakia



This is the next in a long-running series of PROBASTAT conferences with the purpose of stimulating exchanges of ideas and research in all aspects of statistics. The history of the PROBASTAT conference goes way back for approximately 40 years. From that time, researchers have met and enjoyed the friendly and relaxed academic atmosphere in the beautiful environment of Smolenice Castle. We hope that the seventh international PROBASTAT will continue this tradition.

PROBASTAT 2015 will include invited and contributed talks and posters in the following research areas:

- Regression and Mixed Models;
- Optimal Experimental Design;
- Dynamical Statistical Models;
- Innovative Statistical Practices.

A special issue of Statistical Papers (Springer, IF 0.683) is planned for the conference proceedings; the papers will undergo the usual peer review process required by the journal.

More detailed information (including information on the conference fee, payment options, manuscript submissions, and available accommodation/prices in Smolenice Castle) is available at the conference web page

<http://www.um.sav.sk/en/upcoming-conferences/probastat-2015.html>

Contact Address

PROBASTAT 2015
Institute of Measurement Science
Slovak Academy of Sciences
Dúbravská cesta 9
84104 Bratislava, Slovakia

E-mail: witkovsky@savba.sk
Tel.: +421 905 223191
Fax.: +421 2 5477594

MIKUKLÁŠ 2014

REPORT ON MIKUKLÁŠ 2014 MEETING

Jaromír Antoch, Gejza Dohnal

V úterý 16. prosince 2014 se v respiriu MFF UK v Karlíně uskutečnil již tradiční pátý seminář ČStS obecně přezdívaný Mikukláš. Dvacet pět účastníků z celé ČR s velkým zájmem vyslechlo devět zajímavých přednášek. Mezi nimi největší pozornost a zájem vyvolala přednáška Ing. J. Vopravila z ČSÚ a 1. LF UK na téma *Odhady nelegální ekonomiky v České republice*, soustředěná na drogový trh, sexbysnys a nelegální alkohol. Zaujaly též přednášky kolegů z TU v Liberci, věnované aplikacím statistiky při analýze textilních materiálů. Místo Mikukláše se vzhledem k datu spíše vánočnímu zúčastnila pouze jeho trochu zarostlá kopie, která ale nezapomněla přinést punčošku plnou nejenom dobrot, ale i nedostatkového černého uhlí. Na závěr příjemně stráveného dne si účastníci při svíčkách zazpívali a mohli tak alespoň na chvíli zapomenout na klasický shon a stres konce roku. Doufáme, že i příští akce naší společnosti se ponesou ve stejném duchu. Vše nejlepší a příjemně prožité svátky přejí všem členům společnosti organizátoři.

POZVÁNKA NA ČLENSKOU SCHŮZI ČSTS 2015

INVITATION TO ČSTS MEETING 2015

Redakce

Výbor České statistické společnosti zve všechny členy na členskou schůzi, která se bude konat ve **čtvrtek 29. ledna 2015 od 13 hodin v zasedací místnosti Českého statistického úřadu**, Na padesátém 81, Praha 10, a následujícím programem:

- Zahájení
- Zpráva o činnosti ČStS v roce 2014
- Zpráva o hospodaření ČStS v roce 2014 a návrh rozpočtu na rok 2015
- Volby předsedy, ostatních členů výboru a revizora
- Odborná přednáška (doc. Valenta: Úvod do analýzy přežívání)
- Vyhlášení výsledků voleb
- Různé

Těšíme se na Vaši účast,

výbor České statistické společnosti

Obsah

Vědecké a odborné statě

Patrícia Martinková, Katarína Vlčková

Hodnocení reliability znalostních a psychologických testů 1

Zdeněk Půlpán

Na čem závisí spolehlivost dotazníku? 16

Lubomír Kubáček

Simultánní konfidenční elipsoidy 24

Zprávy a informace

Hana Řezanková, Marek Malý

Zasedání představitelů národních statistických společností V6 v Praze 30

Eva Jarošová

Komentář k Shewhartovým regulačním diagramům 32

Pozvánky na akce

Viktor Witkovský

Pozvánka na konferenci ProbaStat 2015 35

Jaromír Antoch, Gejza Dohnal

Mikukláš 2014 36

Redakce

Pozvánka na členskou schůzi ČStS 2015 36

Informační bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214.

The Information Bulletin of the Czech Statistical Society is published quarterly.
The contributions in bulletin are published in English, Czech and Slovak languages.

Předsedkyně společnosti: prof. Ing. Hana ŘEZANKOVÁ, CSc., KSTP FIS VŠE v Praze, nám. W. Churchilla 4, 130 67 Praha 3, e-mail: hana.rezankova@vse.cz.

Redakce: prof. RNDr. Gejza DOHNAL, CSc. (šéfredaktor), prof. RNDr. Jaromír ANTOCH, CSc., prof. Ing. Václav ČERMÁK, DrSc., doc. Ing. Jozef CHAJDIK, CSc., doc. RNDr. Zdeněk KARPÍŠEK, CSc., RNDr. Marek MALÝ, CSc., doc. RNDr. Jiří MICHÁLEK, CSc., prof. Ing. Jiří MILITKÝ, CSc., doc. Ing. Josef TVRDÍK, CSc., Mgr. Ondřej VENCÁLEK, Ph.D.

Redaktor časopisu: Mgr. Ondřej VENCÁLEK, Ph.D., ondrej.vencalek@upol.cz.
Informace pro autory jsou na stránkách společnosti, <http://www.statspol.cz/>.

DOI: 10.5300/IB, <http://dx.doi.org/10.5300/IB>
ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)

Toto číslo bylo vytištěno s laskavou podporou Českého statistického úřadu.