

VYUŽITIE A ZNEUŽITIE CRONBACHOVEJ ALFY PRI HODNOTENÍ PSYCHODIAGNOSTICKÝCH NÁSTROJOV

MARTIN MARKO^{1,2}

Abstrakt: Na základe výsledkov testovania prebiehajúcim v psychologickej diagnostike, vznikajú závery a rozhodnutia, ktorých validita kriticky závisí od spoľahlivosti využitých posudzovacích metód. Reliabilita nástroja je z tohto dôvodu kľúčovým pojmom pri posudzovaní akéhokoľvek psychologického atribútu. Napriek množstvu alternatív, najvýznamnejším a najpoužívanejším indikátorom kvality testového skóre ostáva sprofanovaný koeficient vnútornej konzistencie - alfa (Cronbach, 1951). Postupne identifikujeme najzávažnejšie omyly a problémy, ktoré sa pri používaní tejto štatistiky opakovane objavujú. K jednotlivým bodom taktiež ponúkame riešenia, ktoré by mohli prispieť k zlepšeniu odhadu kvality skóre testovacích nástrojov a rozhodnutí, ktoré sa o ne opierajú.

Kľúčová slova: reliabilita, Cronbachova alfa, vnútorná konzistencia, psychomerika.

Reliabilita je v psychodiagnostike dôležitým zdrojom evidencie pri určovaní validity inferencií, ktoré sú založené na skóre z testov a meraní (t.j. na základe kvantitatívneho a viacpoložkového posudzovacieho nástroja, napr. dotazníka, inventára, škály, performačného testu a i.). Je to v dôsledku skutočnosti, že inštrument nemôže byť validný pokiaľ jeho meranie nie je dostatočne spoľahlivé. Či už posudzujeme schopnosť, afekt alebo osobnostnú charakteristiku, zaujíma nás určitý nepozorovaný (latentný) psychologický konštrukt. Pri snahe o diagnostiku tohto atribútu sa však nemôžeme spoliehať na používané nástroje absolútne, pretože prinášajú odhady, ktoré sú zaťažené chybou o rôznej veľkosti. Z tejto skutočnosti vyplýva potreba, aby každému nástroju, ktorý má byť použitý v praxi alebo vo výskume, bol prisúdený odhad spoľahlivosti s ktorou daný psychologický atribút posudzuje.

Tento odhad je možné vyjadriť rôznymi spôsobmi, avšak pravdepodobne žiaden indikátor kvality testového skóre nebol používaný častejšie ako Cronbachov koeficient alfa (Cronbach, 1951). Cortina (1993) na základe citačného indexu pre sociálne vedy

¹ Katedra psychológie, Filozofická fakulta UK, Gondova 2, 814 99 Bratislava

² Tento text je revidovanou verzí pôvodného článku. Seznam změn (errata) i původní verze jsou společně dostupné na <http://dx.doi.org/10.5817/TF2016-7-90>.

reportoval, že každý rok medzi 1966 až 1990 bol pôvodný Cronbachov článok (1951) citovaný približne 60 krát. Nanešťastie sa zrejme len málo štatistík zároveň spája s tak zásadnými problémami, ktoré sa týkajú problému vhodnosti použitia a interpretácie. Dôvodom, že v súvislosti s koeficientom Cronbachovej alfy (ďalej α) neustále dochádza k omylom, môže byť skutočnosť, že odborné debaty ohľadom indexov reliability sú často veľmi technické (čo platí aj pre originálny článok Cronbacha). Implikácie týchto technických debát však nie sú len záujmom technických poslucháčov, sú taktiež kriticky dôležité pre praktických pracovníkov a výskumníkov z oblasti sociálnych vied všeobecne.

Pretože α je pre psychologickú diagnostiku veľmi dôležitou a zároveň často nesprávne chápanou štatistikou, vyvstáva silná potreba prešetriť najdôležitejšie problémy jej porozumenia a uplatnenia. Cieľom tohto článku je preto identifikovať a objasniť ťažkosti, ktoré sa týkajú jednak bežného poznania, no taktiež uviesť niektoré praktické riešenia, ktoré by mohli prispieť k zlepšeniu odhadu reliability testových skóre v psychologickú diagnostiku. Pre tieto účely sme sa snažili sumarizovať najhlavnejšie body diskusie o koeficiente α v poslednom období a doplniť ho o niekoľko vlastných simulácií. Nakoľko je cieľom článku uľahčiť porozumenie tejto štatistiky, snažili sme sa vyhnúť nadbytočnému formalizmu.

Problémy interpretácie koeficientu α

Alfa nevyjadruje homogenitu škály. Jedným z najväčších nedorozumení medzi teoretikmi a praktikmi je zamieňanie pojmov interná konzistencia a homogenita. Vnútoraná konzistencia odkazuje na vzájomné vzťahy (korelácie) položiek nejakého nástroja, pričom homogenita sa spája s unidimenzionalitou (Schmitt, 1996) Vnútoraná konzistencia a homogenita sú však čiastočne navzájom prepojené. Už Cronbach (1951) uviedol, že koeficient α podceňuje reliabilitu pokiaľ korelácie medzi položkami nie sú homogénne.

Pre ukážku uvádzame dve korelačné matice šiestich položiek, ktorých konfigurácia bola nastavená tak aby korelačné koeficienty jednej matice boli nehomogénne a druhej homogénne (tabuľka 1; vytvorené podľa Schmitt, 1996). Pre každú z matíc sme vypočítali hodnoty koeficientu α , Revelleovej β (Revelle, 1979) a vykonali konfirmačnú faktorovú analýzu (CFA). Ako uvádzajú Zimbard a kol. (2005), formálna definícia koeficientu α a β je veľmi podobná, obe je možné vyjadriť výrazom $k^2 \bar{\sigma}_{ij} / \text{Var}(X)$. Avšak, zatiaľ čo $\bar{\sigma}_{ij}$ pri výpočte α vyjadruje priemernú medzipoložkovú kovarianciu, v prípade koeficientu β ide o priemernú kovarianciu medzi položkami v dvoch častiach testu, ktoré boli rozdelené tak, aby medzi nimi bola minimálna priemerná kovariancia (z tohto dôvodu je $\alpha \geq \beta$).

Napriek nápadnej odlišnosti týchto matíc, výsledky Cronbachovej α boli zhodné. Alfa neprinesla žiadnu informáciu o rozdieloch v danej štruktúre korelačných koeficientov. V prípade koeficientu β , je už rozdiel nápadný. I keď sa hodnoty Cronbachovej α nelíšili, výsledky CFA demonštrujú tento kritický bod jednoznačne. Zatiaľ čo matica

A nevyhovovala unidimenzionálnemu riešeniu, maticu B bolo možné opísať modelom s jedným faktorom.

Tabuľka 1. Vnútoraná konzistencia a CFA matíc (podľa Schmitt, 1996)

A) Nehomogénne položky							B) Homogénne položky						
Položka	1	2	3	4	5	6	Položka	1	2	3	4	5	6
1	–						1	–					
2	.8	–					2	.5	–				
3	.8	.8	–				3	.5	.5	–			
4	.3	.3	.3	–			4	.5	.5	.5	–		
5	.3	.3	.3	.8	–		5	.5	.5	.5	.5	–	
6	.3	.3	.3	.8	.8	–	6	.5	.5	.5	.5	.5	–
Konzistencia: $\alpha = .86$, $\beta = .53$; CFA: $\chi^2(9) = 1913$, $p < .01$, $CFI = .59$, $TLI = .31$, $RMSEA = .46$, $SRMR = .24$							Konzistencia: $\alpha = .86$, $\beta = .83$; CFA: $\chi^2(9) = 6.26$, $p = .71$, $CFI = 1$, $TLI = 1$, $RMSEA = .00$, $SRMR = .008$						

Skutočnosť, že α je funkciou vzájomnej prepojenosti (a počtu) položiek nesie dôležitú implikáciu pre interpretáciu merania – koeficient α neprináša žiadnu informáciu o tom, či nástroj posudzuje jednoliaty konštrukt. Samotná α preto nestačí k tomu, aby mohlo byť testové skóre správne interpretované. Cortina (1993) odporučil aby bola α reportovaná spolu s hodnotou variability korelácií medzi položkami. Veľká variabilita môže podľa neho indikovať multidimenzionalitu nástroja alebo veľkú chybu odhadu týchto korelácií. Za vhodné pokladáme doplniť štatistiku α koeficientom Revelleovej β . Koeficient β vyjadruje najhoršie rozdelenie testu, resp. minimalizovanú kovarianciu medzi dvoma časťami testu (Zinbarg et al., 2005). Rozdiel medzi α a β je preto možné využiť pre indikáciu miery heterogenity medzi položkami.

Maximalizácia koeficientu alfa nie je vždy žiaduca. Ako poznamenal Streiner (2003, str. 217), „hlavnou tendenciou klasickej testovej teórie je, aby škály mali vysokú mieru internej konzistencie podľa Cronbachovho koeficientu alfa“. Okrem vyššie poskytnutých argumentov, uvedieme dve situácie, ktoré od citovanej tendencie odrádzajú.

Existujú škály, ktoré sa skladajú z položiek, ktoré nie sú (resp. ani by nemali byť) navzájom korelované. Je rozdiel, ak viaceré indikátorov (položiek) závisí od latentného konštraktu (reflektívny model), alebo ak tieto indikátory daný konštrukt spôsobujú (formatívny model). Príkladom formatívneho modelu môže byť škála stresu, v ktorej majú participanti vyjadriť frekvenciu výskytu rôznych typov stresujúcich zážitkov. V tomto prípade by sme nepredpokladali, že jednotlivé položky (stresové situácie) by spolu mali nutne korelovať. Napriek tomu, že koeficient α by pri danom nástroji mohol byť nízky, jeho položky by bolo možné sčítať pre vyjadrenie množstva stresujúcich

zážitkov v danom období. Podobný formatívny test môže byť dosť reliabilný, avšak reliabilitu je nutné odhadovať odlišným spôsobom.

Cattel (1978) upozornil na ďalší problém spojený s veľmi vysokými hodnotami koeficientu α . Vysoké interkorelácie medzi položkami značila, že položky sú navzájom redundantné a daný konštrukt je posudzovaný príliš špecificky. V takomto prípade je vysoká vnútorná konzistencia nástroja na úkor nízkej obsahovej validity, t.j. miery s akou daný nástroj postihuje všetky aspekty daného konštrukt. Ak koeficient α dosahuje vysoké hodnoty už pri menšom množstve položiek, značí to, že dané položky opakovane posudzujú užšiu časť konštrukt. Z tohto dôvodu môže byť tendencia maximalizovať vnútornú konzistenciu nežiaduca. Druhý aspekt zvyšovania vnútornej konzistencie sa týka priamej závislosti koeficientu α od počtu položiek nástroja. Pridávaním počtu položiek je preto možné zvýšiť α nad tradične odporúčané hodnoty (napr. 0.7 až 0.8) aj v keď ich vzájomné korelácie sú relatívne nízke.

Faktory skresľujúce odhad koeficientu α

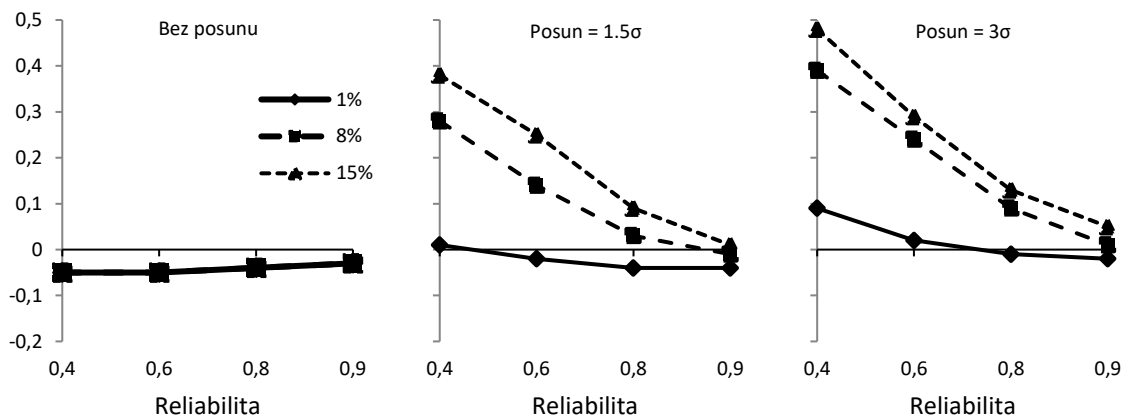
Narušenie α extrémnymi pozorovaniami

Ako naznačili Lind a Zumbo (1993), koeficient α spadá do štatistík využívajúcich metódu najmenších štvorcov a preto je náchylná ku neprimeranému skresleniu extrémnymi dátovými bodmi (tzv. „outliers“, inak nazývané aj ako „vzdialené pozorovania“, „extrémne dátové hodnoty“ alebo „kontaminanty“). Za extrémne sa považujú také pozorovania, ktoré možno na základe určitého kritéria považovať za vzdialené od prevažnej časti dát. Matematicky sa miera extrémnych pozorovaní vyjadruje pomerom výberu pozorovaní z kontaminovanej distribúcie s odlišnými parametrami k výberu z normálnej distribúcie. Ak pozorovanie pochádza z distribúcie, ktorá sa od cieľovej distribúcie líši parametrom variability, hovoríme o symetrickom extrémnom pozorovaní. V prípade, že extrémne pozorovanie pochádza z distribúcie, ktorá sa líši v miere centrálnej tendencie (napr. priemernej hodnote), ide o asymetrickú kontamináciu.

Extrémne pozorovania môžu mať vo všeobecnosti veľký vplyv na menej robustné odhady modelov a ich parametrov (napr. Wilcox, 2005). Napriek tomu, že ich dopad na korelácie, regresné parametre, t testy a F testy je relatívne známy, len málo štúdií sa venovalo ich efektu na odhad koeficientu α . Liu a Zumbo (2007) zistili, že veľkosť výberu a štandardná odchýlka symetrickej kontaminácie neovplyvňuje ani skreslenie ani replikovateľnosť koeficientu α . Prítomnosť asymetrických odľahlých pozorovaní však viedla k značným preceneným odhadom koeficientu α a horšej replikovateľnosti výsledku. Čím bola proporcia kontaminácie vyššia, tým väčšie bolo aj pozitívne skreslenie koeficientu α , to však platilo menej ak populačná reliabilita bola vyššia. Výsledky ukazujú, že populačná hodnota koeficientu $\alpha = .40$ môže byť v dôsledku extrémnych pozorovaní výrazne skreslená (napr. v dôsledku 15 % extrémnych pozorovaní s posunom 3 smerodajných odchýlok, môže byť uvedená hodnota skreslená až na $\alpha = .90$). V ďalšej štúdiu Liu, Wu

a Zumbo (2010) tieto zistenia generalizovali aj na binárne a ordinálne posudzovacie škály (napr. Likertové škály, ktoré sa v psychológii často využívajú).

Skreslenie α



Obrázok 1: Vplyv teoretickej reliability (.4 až .9), množstva (1% až 15%) a veľkosti (posun priemeru 0 až 3 σ) extrémnych hodnôt na odhad reliability pomocou koeficientu α (grafy boli vytvorené na základe výsledkov v Liu, Wu, & Zumbo, 2010).

Z uvedených dôvodov je nutné hodnoty konzistencie psychodiagnostických nástrojov odhadovať na dátach s presnou špecifikáciou. Prípadnú kontamináciu pozorovaných distribúcií je možné zmierniť klasickými spôsobmi detekcie extrémnych pozorovaní a aberantných vektorov odpovedí (napr. tzv. person-fit štatistiky, *lz*). Ako menej invazívnu alternatívu je možné využiť robustné metódy odhadu parametrov s využitím metódy bootstrap alebo orezania (napr. winsorizovanú korelačnú maticu; Wilcox, 2005).

Narušenie α ordinalitou škál

Väčšina psychodiagnostických nástrojov obsahuje položky ordinálneho charakteru. Použitie koeficientu α v prípade takýchto (diskrétnych) škál je problematické z dôvodu narušenia predpokladu intervalovej úrovne merania. V predchádzajúcom výskume autori Gelin, Beasley a Zumbo (2003) demonštrovali, že použitie Likertových škál môže značne redukovať veľkosť odhadu koeficientu α v prípade, ak má škála rozsah menší ako päť bodov (t.j. ak položkám daného nástroja testovaná osoba prideluje hodnoty v rozsahu menšom ako 1 až 5). Pri väčšom počte bodov na Likertových škálach odhad reliability aproximuje populačnú reliabilitu. Táto problematika je veľmi dôležitá pretože odhad reliability psychodiagnostických metód a didaktických testov, ktoré využívajú Likertové (alebo dokonca binárne) hodnotenie, môže byť do značnej miery skreslený.

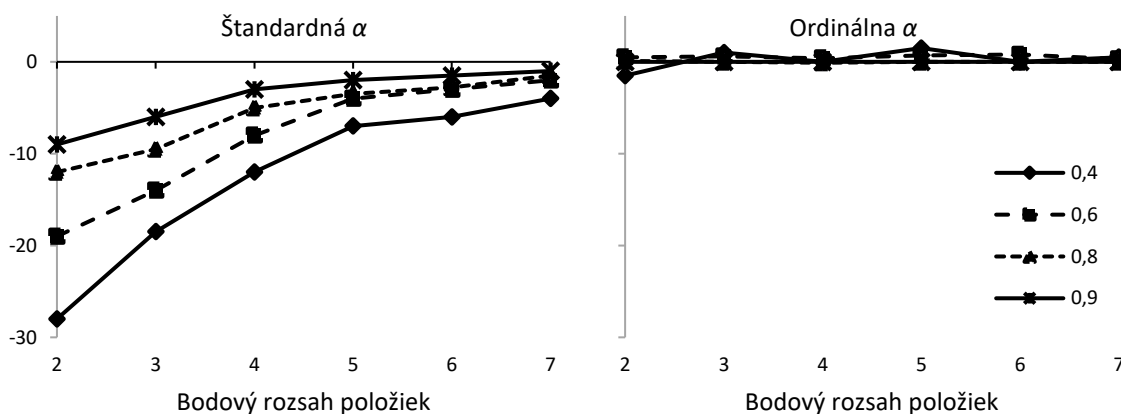
Odhad koeficientu α sa zvyčajne odvodzuje z Pearsonovej korelačnej matice, ktorá predpokladá spojitosť premenných, resp. intervalovú úroveň merania. Uvádzané skreslenie preto nastáva v prípade, že je tento predpoklad narušený. Z tohto dôvodu je v prípade ordinálnych dát optimálne určovať mieru asociácie medzi položkami pomocou

polychorických korelačných koeficientov. Tie sú vhodné pre odhady lineárnych vzťahov aj za podmienok, že premenné sú indikované na ordinálnej úrovni merania (Flora & Curran, 2011).

Z tohto dôvodu Zumbo, Gadermann a Zeisser (2007) predstavili spôsob estimácie koeficientu α za pomoci polychorickkej korelačnej matice – ide o tzv. *ordinálnu alfu*. V uvedenej simulačnej štúdií autori porovnávali odhady ordinálnej α a štandardnej α za rôznych podmienok. Výsledkom bolo zistenie, že klasický odhad koeficientu α bol drasticky ovplyvnený Likertovým typom odpovede. Tento záver spresnila ďalšia štúdia (Gadermann, Guhn, & Zumbo, 2012), kde autori modelovali množstvo atenuácie odhadu Cronbachovej α (atenuácia bola definovaná v percentách, ako pomer rozdielu odhadnutej hodnoty α a simulovanej reliability ku simulovanej reliability). Výsledky preukázali, že znižujúcim sa rozsahom škály položky (od 7-stupňovej až po binárnu) sa zvyšovala atenuácia odhadu štandardnej α , pričom tento efekt bol tým silnejší, čím menšia bola teoretická reliability súboru položiek. V prípade ordinálnej α sa efekt atenuácie nepreukázal. Porovnanie Cronbachovej a ordinálnej α ilustruje obrázok 2.

Uvedené zistenia upozorňujú na nesprávne použitie Cronbachovej α , ktoré v prípade ordinálnych položiek s nižším počtom bodov, môže viesť k vážnejšiemu preceneniu chyby merania (t.j. podceneniu spoľahlivosti) a aj celkovej diskvalifikácii posudzovacích nástrojov.

Atenuácia α (%)



Obrázok 2: Porovnanie vplyvu teoretickej reliability (.40 až .90) a bodového rozsahu položiek (2 až 7) na odhad reliability pomocou štandardného a ordinálneho koeficientu α (grafy boli vytvorené na základe výsledkov v Gadermann, Guhn, & Zumbo, 2012).

Narušenie α heterogenitou položiek

Ako ukázali už Lord a Novick (1968), na to aby bol koeficient α neskreslený, test musí spĺňať predpoklad tau-ekvivalentencie (variabilita položiek vysvetliteľná „true“ skóre je medzi položkami rovnaká). Absolútne homogénna škála je len teoretická idealizácia. Bežná prax nedisponuje nástrojmi, ktoré by posudzovali cieľový konštrukt úplne homogénne. Častejšie sa stáva, že položky testu sú zoskupené do skupín s viac či menej odlišnými obsahovými doménami. Z tohto dôvodu je možné vyhnúť sa skresleným odhadom tak, že jednotlivé položky sa rozdelia do stratifikovaných vrstiev. Každá stratifikovaná vrstva obsahuje skupinu takých položiek, ktoré sú unidimenzionálne (resp. homogénne). Reliabilita kompozitného skóre môže byť potom vypočítaná na základe Feldtovho a Brennanovho vzorca (1989) pomocou reliability (r_i) a variancie (σ_i^2) každej vrstvy položiek a variancie kompozitného skóre (σ_c^2):

$$\alpha_{STRAT} = 1 - \frac{\sum \sigma_i^2 (1 - r_i)}{\sigma_c^2}.$$

Pre porovnanie použitia klasického a stratifikovaného koeficientu α sme nasimulovali 1000 odpovedí pre deväť nehomogénnych položiek fiktívnej škály (intervalový charakter), ktoré tvorili tri vrstvy. Simulácia bola realizovaná pomocou jazyka R (R Core Team, 2013) v programe RStudio (RStudio Team, 2015). V rámci každej vrstvy položky navzájom korelovali na úrovni $r = .50$, korelácia medzi položkami z odlišných vrstiev bola nastavená na $r = .20$. Odhad pre tieto dáta bol $\alpha = .75$, respektíve $\alpha_{STRAT} = .85$. Odhad štandardnej chyby merania určitého psychologického atribútu (napr. inteligencie) by bol v prípade použitia nestratifikovaného koeficientu α vyšší o približne 25 %. Takýto rozdiel už môže mať vecný význam pri interpretácii daného skóre v procese psychodiagnostiky.

Záver

Koeficient Cronbachova α je zrejme najpoužívanejšou štatistikou odhadu reliability kvantitatívnych nástrojov v psychodiagnostike. Prvým cieľom predkladaného textu bolo zvýšiť porozumenie tomuto indexu a upozorniť na najbežnejšie omyly, ktoré sa týkajú interpretácie jeho významu a veľkosti. Ako druhé sme identifikovali tri najhlavnejšie faktory, ktoré spôsobujú značné skreslenie koeficientu α a poskytli alternatívy, ktoré zabezpečia jeho realistickejšie odhady. V prípade podozrenia na prítomnosť extrémnych pozorovaní odporúčame prešetriť distribučné charakteristiky jednotlivých premenných (resp. položiek). V prípade substantívneho narušenia šikmosti, špicatosti alebo výskytu extrémnych pozorovaní je vhodné pre výpočet α využiť robustný ekvivalent pearsonovej korelačnej matice. Obzvlášť vhodná môže byť (jednostranná alebo obojstranná) winsorizácia, alebo detekcia mnohorozmerne atypických odpovedí. Ordinálny charakter škál, ktorý je typický pre absolútnu väčšinu psychodiagnostických nástrojov, je možné zohľadniť využitím ordinálnej verzie α , ktorá vychádza z tetrachorickej alebo polychorickej korelačnej matice. Vzhľadom na výsledky uvádzaných simulačných štúdií pokladáme index ordinálnej α , v prípade odhadu vnútornej konzistencie škál s binárnymi (resp. ordinálnymi) položkami, za nadradený. Na záver, skreslenie α v dôsledku

diferencovanej štruktúry koštruktú je možné riešiť prostredníctvom stratifikovaného koeficientu α . Za predpokladu homogenity a konzistencie položiek v rámci jednotlivých dimenzií komplexnejšieho koštruktú prináša tento index lepší odhad teoretickej reliability meracieho nástroja. Z uvedeného však vyplýva, že pred výpočtom α je nevyhnutné overiť dimenzionalitu posudzovaných nástrojov. Odporúčania v predkladanom texte je možné využiť k zlepšeniu hodnotenia psychodiagnostických metód a psychodiagnostických rozhodnutí, ktoré na nich stavajú.

Marko, M. (2016): (Mis)application of Cronbach's alpha in the evaluation of diagnostic instruments

Abstract: *The validity of inferences and decisions formed on the basis of testing in psychological diagnostics is critically dependent on reliability of utilized assessment tools. Therefore, whenever a psychological attribute is to be scored, reliability of the diagnostic means becomes crucial. Despite many alternatives, Cronbach's alpha (1951), as profaned internal consistency statistic, remains the most prominent and most widely used indicator of test score quality. Because of this importance, our article is aimed to identify the most serious misconceptions and misapplications which repeatedly occur when using this statistics. For each particular point of this issue, we suggest a simple solution which may improve the estimation of the score quality of tests and therefore enhance decisions that are drawn from them.*

Keywords: *reliability; Cronbach's Alpha; internal consistency; psychometrics*

Referencie

- Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Boston, MA: Springer US.
- Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications, 78(1), 98–104
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.: pp. 105-146). New York: Macmillan.
- Flora, D. B., & Curran, P. J. (2011). An evaluation of alternative methods of estimation for

confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491.

Gelin, M. N., Beasley, T. M., & Zumbo, B. D. (2003). What is the impact on scale reliability and exploratory factor analysis of a Pearson correlation matrix when some respondents are not able to follow the rating scale? Paper presented at the annual meeting of the American Educational Research Association (AERA) in Chicago, Illinois.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating Ordinal Reliability for Likert-Type and Ordinal Item Response Data: A Conceptual, Empirical, and Practical Guide. *Practical Assessment, Research & Evaluation*, 17(3).

Lind, J. C., & Zumbo, B. D. (1993). The continuity principle in psychological research: An introduction to robust statistics. *Canadian Psychology*, 34, 407-412

Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item responses. *Educational & Psychological Measurement*, 70, 5-21.

Liu, Y., & Zumbo, B. D. (2007). The Impact of Outliers on Cronbach's Coefficient Alpha. *Educational and Psychological Measurement*, 67(4), 620–634.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14(1), 57–74.

RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80(3), 217–222.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.

Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, 53, 33-49.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β and McDonald's ω_H : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21–29.