

Kvalita dat I.*

Klasický model měření reliability a jeho praktický aplikační význam

JAN ŘEHÁK**

Sociologický ústav AV ČR, Praha

Quality of Data I.

Classical Model of Measuring Reliability and its Practical Application

Abstract: The paper is a subject review, the purpose of which is to revitalise the concept of reliability for the sociological research. Reliability is defined by classical definition and the properties of this concept are discussed in the paper. The consequences of estimating variability and correlation are presented and the estimation formulas are given. The practical methods for estimation are also described. The paper aims at practical researchers in order to remind them of the almost forgotten and neglected notion of data quality.

Sociologický časopis, 1998, Vol. 34 (No. 1: 51-60)

Kvalita výzkumných dat se obecně posuzuje jednak z hlediska výběrových chyb, tj. z pohledu reprezentativity (vztahu datového souboru a cílové populace) na jedné straně a z hlediska chyb měření na straně druhé. Chyby měření se obvykle posuzují z hlediska reliability a z hlediska validity.

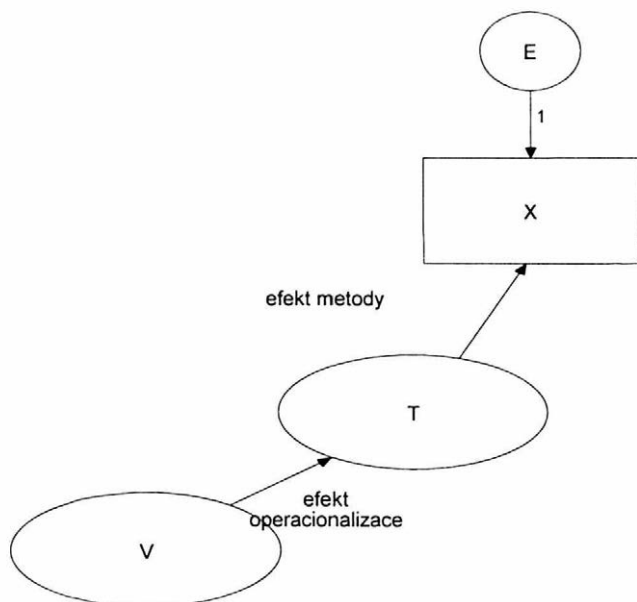
Pojem reliability se v literatuře vyskytuje často, v praktickém výzkumu a v praktickém statistickém popisu a rozhodování se však neaplikuje, přesto, že nereliabilní sociologická měření mají značně vysoký vliv na statistické výsledky a na statistickou inferenci, a tudíž i na následnou interpretaci.

Reliabilitou (spolehlivostí) měření se rozumí přesnost měření té vlastnosti, kterou ve skutečnosti měříme, *validitou* se rozumí přesnost toho, co ve skutečnosti měříme vzhledem k tomu, co chceme měřit. Reliabilita charakterizuje přesnost měření konstruktu pomocí jeho konkrétní operacionalizace, validita odpovídá přesnosti operacionalizace konceptu konstruktem. Reliabilita je proto pojmem technickým, validita substantivním. Oba tyto pojmy jsou součástí pohledu na kvalitu dat a jsou součástí klasické metodologie a teorie měření, která však byla schopna odhadovat pouze reliabilitu. Validitu v uvedeném pojetí pomocí klasických teorií měřit nelze a tento pojem byl pro praktické účely definován pouze vztahově k určenému vnějšímu kritériu; význam takového zavedení je v praxi ovšem značně omezen.

*) Tato práce vznikla v rámci grantu GA ČR číslo 403/95/0964 jako vstupní studie problému.

**) Veškerou korespondenci posílejte na adresu: Doc. RNDr. Jan Řehák, Sociologický ústav AV ČR, Jiřská 1, 110 00 Praha 1, tel. (02) 24 22 09 79, l. 231, fax (02) 24 22 02 78, e-mail rehak@soc.cas.cz

Situaci měření lze vyjádřit grafem:



Obr.1 Validita = efekt metody x efekt operacionalizace

V = vlastnost, kterou chceme měřit (koncept)

T = skutečná hodnota vlastnosti, kterou ve skutečnosti měříme (konstrukt)

X = měřená hodnota na stupnici operacionalizovaného konstruktů

Terminologie je v literatuře poněkud nejednotná a je věcí vkusu a vlastního rozhodnutí, který terminologický návrh uživatel přijme. *Validitou* zde budeme chápat vztah mezi V a X (v některých pracích je validitou chápán vztah V a T , který zde nazýváme *efektem operacionalizace*). *Reliabilita* v klasické teorii měření je vztahem mezi T a X , jedná se proto o reliabilitu konkrétního způsobu měření konkrétní otázky, je také označována jako *efekt metody*, vztah E a X je nespolehlivost metody měření (tj. variabilita, která vychází z jiných zdrojů než ze skutečné hodnoty).

Zatímco nutnost charakterizovat chyby výběru ve výzkumných šetřeních sociologie, veřejného mínění, marketingu, zdravotnictví, ve státní administrativě apod. je přijatou nepominutelnou zásadou, nevyhnutelnou povinností a standardem, na chyby měření, které jsou pro interpretační proces stejně nebezpečné, se zapomíná a tyto chyby se podceňují. Od osmdesátých let se tomuto aspektu kvality dat věnuje v sociologii velmi usilovná pozornost, která se v devadesátých letech velmi zesiluje [viz např. Saris a van Meurs 1990].

Cílem této práce je upozornit na pojem reliability a revitalizovat jej pro výzkumnou práci, upozornit na konsekvence, které nespolehlivá měření (tj. měření s nízkou reliability) přináší do statistických závěrů, a zároveň ukázat, jak lze některé z nich překonávat. Soustředíme se v této práci na *klasickou definici reliability* tak, jak byla vyvíjena v padesátých a šedesátých letech a podána v souhrnné teoretické a dnes již klasické, ale stále platné monografii [Lord, Novick 1968].

1. Pojem reliability

Reliabilitu definuje Lord a Novick [1968] jako koeficient determinace měření X na skutečné hodnotě T ve vztahu

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

kde E je chyba vzniklá v procesu měření, tj. při konkrétním zjišťování X .

Koeficient reliability je determinace

$$(2) \quad \text{rel}(X/T) = r^2(X, T) = \text{var}T/\text{var}X,$$

podíl variability T na variabilitě X . Uvádí se také v procentech.

Už z této definice plyne, že celková variabilita našich měření se dělí na dvě části: $\text{var}X = \text{var}T + \text{var}E$ a kvalita měření je charakterizována stupněm, do jaké se podaří eliminovat chybu, která vnáší do dat neurčitost. Část empirické variability je tedy neinterpretovatelná – zjištěná heterogenita dat zahrnuje náhodnou složku, vedoucí k interpretačnímu šumu. Reliabilita charakterizuje podíl empirické variability, tj. interpretovatelnou část variance, a to ale vzhledem k tomu, co ve skutečnosti měříme. Vyjadřuje proto efektivitu měřícího postupu, metody, kvalitu dotazníku. Uvedený podíl interpretovatelné informace reprezentuje pouze proces měření, zahrnuje pouze tu složku kvality dat, která se vztahuje k procesu individuálního zjišťování dat, tj. individuálního dotazu v rámci výzkumného šetření. Druhou složkou kvality výzkumné informace je neurčitost způsobená výběrovým postupem, tou se však zde nezabýváme ani klasická teorie měření ji ve svých modelech nezahrnuje.

Definice reliability je založena na několika předpokladech, z nichž plynou důležité praktické důsledky:

1. chyba E není závislá na hodnotě T , $r(E, T) = 0$;
2. chyba E neobsahuje systematické vychýlení, $EE = 0$ (průměr chyb při opakovaných měření je nulový);
3. při různých měřeních jsou chyby nekorelované.

Tyto předpoklady nejsou jen matematickými formalitami, ale mají velmi podstatné *aplikační* a *interpretační* důsledky. Jejich splnění předpokládá vysokou kvalifikovanost tazatelské sítě a velmi pečlivou přípravu formulací v dotazníku.

Platnost prvního předpokladu je silně závislá na znění otázek a neutralitě v chování tazatele a v jeho celkové důvěryhodnosti, korelovanost chyby a skutečné hodnoty může být způsobena maskováním skutečné hodnoty (snížování příjmů, stylizace při výzkumu hodnot).

Druhý předpoklad je přirozeným požadavkem pro chybu, která se chová náhodně. Je velmi závislý na neutrálním chování tazatele, na jeho zdrženlivosti v projevování reakcí na odpovědi, na jeho celkové disciplinovanosti. Tazatel nesmí nijak působit na

odpovědi. Proto se také nesmí s respondentem znát, aby odpovědi nepředjímaly znalost tazatelových názorů či nebyly zkresleny obavou z nediskrétnosti.

Třetí předpoklad je pro praxi výzkumu základní a vyjadřuje nutnost nezávislého měření na různých jednotkách šetření. Sem patří např. zásada dobrého proškolení tazatelské sítě, neboť tuto chybu často do dat vnášejí nekvalifikovaní tazatelé ať už svým celkovým chováním a vystupováním, nebo chováním při samotném rozhovoru. Patří sem také zásada, že tazatel *nesmí* provádět rozhovor se svými známými a příbuznými a *nesmí* provádět více rozhovorů v jedné rodině. Vlivem nekvalifikovaných tazatelských sítí se vnáší do dat skupinková korelovanost chyb měření. Základem tohoto předpokladu je opět neutralita a korektní vystupování tazatele.

Uvedené poznámky ukazují racionální základy pro nutné chování tazatelské sítě a to, že tyto zásady nejsou jen umělými komplikacemi při sběru dat. Porušení uvedených předpokladů má za následek automatickou nepřesnost závěrů analýzy, která na nich závisí.

Pro praktické zjišťování reliability je důležitý pojem *paralelních měření*. Dvě měření jsou paralelní, platí-li:

$$\begin{aligned} X &= T + E \\ X' &= T + E' \end{aligned}$$

přičemž E a E' jsou nekorelované a jejich rozptyly jsou stejné. V datové praxi tak musí mít dvě proměnné z paralelních měření stejné průměry a stejné rozptyly.

To znamená, že obě proměnné měří totéž. Paralelními měřeními mohou být nezávislá opakování stejného měření nebo položení obsahově ekvivalentní otázky či stejně obtížné položky v testu. Tento pojem je důležitý pro metodologii odhadu reliability. Platí důležitý vztah:

$$(3) \quad \text{rel}(X/T) = r^2(X, T) = r(X, X').$$

Korelace mezi dvěma paralelními měřeními téže vlastnosti je zároveň čtvercem korelace (tj. koeficientem determinace) mezi skutečnou a měřenou hodnotou dané veličiny. To znamená, že pokud se nám podaří zkonstruovat paralelní měření téhož, můžeme pomocí jejich korelace odhadnout reliability.

2. Vlastnosti a důsledky reliability

Spolehlivost/nespolehlivost měření (dotazu) má některé praktické konsekvence, které mohou mít vliv na empirickou analýzu i na interpretaci, a proto je zde uvádíme. Všechny další vlastnosti platí v případě, že přijmeme klasickou definici reliability tak, jak byla uvedena výše.

1. Průměrná hodnota měření X v souboru dat se rovná přibližně průměru skutečných hodnot T , na průměry proto spolehlivost nemá vliv ve smyslu očekávání (důsledek předpokladu, že očekávaná hodnota E je nulová).

$$(4) \quad \text{průměr } X = \text{průměr } T + \text{průměr } E$$

a protože očekáváme průměr E přibližně nulový, neliší se průměr X od průměru T podstatně. Sbližování obou se zrychluje s počtem pozorování. Platí ovšem, že čím vyšší je spolehlivost, a tím nižší chyba, tím rychleji se průměry X a T sbližují. Tato „přibližnost“ roste s počtem pozorování.

2. Rozptyl hodnoty X v datovém souboru je větší, než je rozptyl skutečných hodnot T .

Platí:

$$(5) \quad \text{var } X = \text{var } T + \text{var } E.$$

Důsledek pro analýzu dat a obzvláště pro interpretaci je podstatný ve výše uvedeném smyslu: v explikační analýze je naším úkolem vysvětlit variabilitu a odpovědět na otázku, co je příčinou empirické rozmanitosti. Vzorec udává, že jen část variability (proporčně rovná reliability) má substantivní smysl, zatímco zbytek je způsoben chybami a jeho vysvětlování znamená uvádění artefaktů. Ze vzorce plyne, že skutečnou variabilitu, kterou chceme podchytit a vysvětlit, můžeme odhadnout jako:

$$(6) \quad \text{var } T = \text{rel}(X/T) * \text{var } X.$$

3. Jednotlivé hodnoty T_i v souboru dat lze podle klasické teorie měření odhadnout podle regresního vztahu T na X takto:

$$(7) \quad \text{odhad } T = R(T/X) = \text{rel}(X/T) * X + (1 - \text{rel}(X/T)) * \text{průměr } X.$$

a obdobnou rovnici lze také napsat pro odhad chyby:

$$(8) \quad \text{odhad } E = R(E/X) = (1 - \text{rel}(X/T)) * (X - \text{průměr } X).$$

Odhadem skutečné hodnoty je vážený průměr měřené hodnoty (s vahou reliability) a průměru všech hodnot (s komplementární vahou). Postup odpovídá logice redukováného skutečného rozptylu, menší skutečný rozptyl znamená, že odhady individuálních hodnot jsou sevřenější kolem průměru měřené veličiny než měřené hodnoty samy (ty jsou navíc rozptýleny chybou E).

4. Reliabilita má vliv na výpočet konfidenčních intervalů (intervalů spolehlivosti) pro skutečné hodnoty a na testování hypotéz o nich. Místo rutinně zjišťovaného intervalu spolehlivosti, který zahrnuje i chybovou část variance, je možné použít vzorce s odhadem skutečných směrodatných chyb:

$$(9) \quad \text{průměr } T = \text{průměr } X \pm (z/\text{sqrt}(n)) * s_X * \text{sqrt}(\text{rel}(X/T)),$$

kde s_X je směrodatná odchylka hodnot X a z je skór spolehlivosti pro zvolenou úroveň.¹

5. Korelace dvou proměnných v nespolehlivých měřeních má značnou deflační tendenci: Pro dvě proměnné X a Y s koeficienty spolehlivosti $\text{rel}(X/T)$ a $\text{rel}(Y/U)$ resp. platí, že empirická korelace se součinnově redukuje podle *redukční formule (attenuace)*:

$$(10) \quad r(X, Y) = \sqrt{\text{rel}(X/T)} * r(T, U) * \sqrt{\text{rel}(Y/U)}$$

Pro analýzu je zajímavá korelace mezi skutečnými hodnotami, proto:

$$(11) \quad r(T, U) = r(X, Y) / \sqrt{\text{rel}(X/T) * \text{rel}(Y/U)}$$

Pokud je reliabilita nízká, je deflace korelačního koeficientu značná a *chyba měření maskuje skutečné vztahy*. Proto nám v analýze mnoho zajímavých korelátů může uniknout. Z této rovnice je vidět, že i nízké koeficienty získané ve výzkumu mohou ve skutečnosti vycházet ze zajímavých vztahů, jejichž skutečná korelovanost může být skryta za chybou měření. Chyba způsobená nereliabilitou měření proto vytváří mlhu

¹⁾ Termín spolehlivost se používá ve dvojím významu: a) ve smyslu reliability, jak byl definován výše pro indikaci přesnosti měření, b) ve smyslu konfidenčního intervalu pro vyjádření přesnosti statisticky odhadnutých populačních, resp. modelových parametrů (jako je např. průměr). Na tuto dvojznačnost české terminologie je třeba dát v praxi pozor.

kolem skutečných vztahů a ohrožuje naše korelační a asociační závěry a ověřování apriorních korelačních, resp. asociačních hypotéz.

Tento fakt je tak závažný, že je ilustrován v tabulkách A a B pomocí některých vybraných situací:

Tabulka A. Hodnoty empiricky zjištěného korelačního koeficientu při vybraných kombinacích koeficientů reliability

			$r(T/U)$								
odmocnina			hodnoty empiricky zjištěné korelace při snížené spolehlivosti								
rel(X/T)	rel(X/T)	součinu	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	1
0,4	0,4	0,4	0,12	0,16	0,20	0,24	0,28	0,32	0,36	0,38	0,40
0,5	0,5	0,5	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,48	0,50
0,6	0,6	0,6	0,18	0,24	0,30	0,36	0,42	0,48	0,54	0,57	0,60
0,7	0,7	0,7	0,21	0,28	0,35	0,42	0,49	0,56	0,63	0,67	0,70
0,8	0,8	0,8	0,24	0,32	0,40	0,48	0,56	0,64	0,72	0,76	0,80
0,9	0,9	0,9	0,27	0,36	0,45	0,54	0,63	0,72	0,81	0,86	0,90
0,95	0,95	0,95	0,29	0,38	0,48	0,57	0,67	0,76	0,86	0,90	0,95
0,4	0,95	0,616	0,18	0,25	0,31	0,37	0,43	0,49	0,55	0,59	0,62
0,5	0,9	0,671	0,20	0,27	0,34	0,40	0,47	0,54	0,60	0,64	0,67
0,6	0,8	0,693	0,21	0,28	0,35	0,42	0,48	0,55	0,62	0,66	0,69
0,7	0,7	0,700	0,21	0,28	0,35	0,42	0,49	0,56	0,63	0,67	0,70
0,8	0,9	0,849	0,25	0,34	0,42	0,51	0,59	0,68	0,76	0,81	0,85
0,9	0,8	0,849	0,25	0,34	0,42	0,51	0,59	0,68	0,76	0,81	0,85
0,95	0,7	0,815	0,24	0,33	0,41	0,49	0,57	0,65	0,73	0,77	0,82

Tabulka B. Hodnoty přepočteného skutečného korelačního koeficientu při vybraných kombinacích koeficientů reliability

rel(X/T)	rel(Y/U)	odmocnina součinu	$r(X,Y)$ odhad hodnoty skutečné korelace při snížené spolehlivosti							
			0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95
0,4	0,4	0,4	0,75	1,00	-	-	-	-	-	-
0,5	0,5	0,5	0,60	0,80	1,00	-	-	-	-	-
0,6	0,6	0,6	0,50	0,67	0,83	1,00	-	-	-	-
0,7	0,7	0,7	0,43	0,57	0,71	0,86	1,00	-	-	-
0,8	0,8	0,8	0,38	0,50	0,63	0,75	0,88	1,00	-	-
0,9	0,9	0,9	0,33	0,44	0,56	0,67	0,78	0,89	1,00	-
0,95	0,95	0,95	0,32	0,42	0,53	0,63	0,74	0,84	0,95	1,00
0,4	0,95	0,616	0,49	0,65	0,81	0,97	-	-	-	-
0,5	0,9	0,671	0,45	0,60	0,75	0,89	-	-	-	-
0,6	0,8	0,693	0,43	0,58	0,72	0,87	-	-	-	-
0,7	0,7	0,700	0,43	0,57	0,71	0,86	-	-	-	-
0,8	0,9	0,849	0,35	0,47	0,59	0,71	0,82	0,94	-	-
0,9	0,8	0,849	0,35	0,47	0,59	0,71	0,82	0,94	-	-
0,95	0,7	0,815	0,37	0,49	0,61	0,74	0,86	0,98	-	-

Poznámka: Některé kombinace označené „-“ nejsou v praxi možné.

Uvedené vlastnosti mohou být v praxi využity buď jako teoretická znalost, která je vodítkem pro *interpretace* běžně používaných statistických výsledků, nebo pro *přepočty*, pokud jsou známy koeficienty reliability pro jednotlivá měření.

6. Spolehlivost součtu (průměru) proměnných. Další důležitou vlastností reliability je možnost jejího podstatného zlepšení v případě opakovaného *dotazu* na totéž pomocí paralelních měření a spojení těchto údajů do součtových indexů (obdoba nezávislého opakování měření v přírodních vědách). Právě zde se prosazuje relevance pojmu paralelních měření. Náhodná chyba se při spojení více měření podstatně eliminuje.

Jsou-li X_1 a X_2 dvě paralelní měření a obě mají stejnou spolehlivost $rel(X)$, platí pro spolehlivost součtového indexu $Y = X_1 + X_2$ (nebo stejně i pro spolehlivost průměrového indexu $Y = (X_1 + X_2)/2$) známá *Spearman-Brownova formule*:

$$(12) \quad rel(Y) = 2 * rel(X) / (1 + rel(X))$$

Obecněji platí pro součet Y , resp. průměr Y , více (K) položek

$$(13) \quad rel(Y) = K * rel(X) / (1 + (K - 1) * rel(X))$$

kde

$$Y = \sum X_i$$

Vzorec ukazuje, jak se zvýší spolehlivost při kombinování více paralelních položek do jedné míry. V tabulce C ilustrujeme vztah číselně.

Tabulka C. Spolehlivost sumačních indexů

Počet položek		Spolehlivost jedné položky				
1	0,2	0,4	0,6	0,8	0,9	0,95
2	0,333	0,571	0,75	0,889	0,947	0,974
3	0,429	0,667	0,818	0,923	0,964	0,983
4	0,5	0,727	0,857	0,941	0,973	0,987
5	0,556	0,769	0,882	0,952	0,978	0,99
6	0,6	0,8	0,9	0,96	0,982	0,991
7	0,636	0,824	0,913	0,966	0,984	0,993
8	0,667	0,842	0,923	0,97	0,986	0,993
9	0,692	0,857	0,931	0,973	0,988	0,994
10	0,714	0,87	0,938	0,976	0,989	0,995
12	0,75	0,889	0,947	0,98	0,991	0,996
14	0,778	0,903	0,955	0,982	0,992	0,996
16	0,8	0,914	0,96	0,985	0,993	0,997
18	0,818	0,923	0,964	0,986	0,994	0,997
20	0,833	0,93	0,968	0,988	0,994	0,997
30	0,882	0,952	0,978	0,992	0,996	0,998
50	0,926	0,971	0,987	0,995	0,998	0,999
100	0,962	0,985	0,993	0,998	0,999	0,999

Výsledek je opět aplikačně nesmírně důležitý. Ukazuje, jak se v měření projevuje spojení informace z dvojnásobného nebo vícenásobného opakování dotazu na tutéž vlastnost pomocí *baterie položek*. Rychlost zvyšování spolehlivosti s rostoucím počtem položek je podstatná pro plánování položkových baterií, ze kterých se indikačně a sumačně měří postojové a jiné latentní proměnné. Ekonomie času a neschopnost či neochota respon-

denta odpovídat na dlouhá interview je tu pochopitelně v konfliktu s přesností, která s délkou baterie roste. Zvyšování spolehlivosti při položkovém opakování dotazu na totéž provádíme paralelu a analogii k opakování pokusu. Alespoň přibližný odhad reliability umožňuje volit vhodnou velikost měřící baterie zajišťující apriori žádanou přesnost. Princip je obdobný: opakovaná měření vedou k vzájemné kompenzaci náhodných nezávislých chyb, které v průměru postupně konvergují k nule. Tabulka C charakterizuje konvergenci přesnosti. (Teoreticky by se pro enormně velké baterie spolehlivost blížila k jedné bez ohledu na nízkou nenulovou spolehlivost jedné položky.)

Obecnějším výsledkem je *odhad dolní hranice spolehlivosti pro součtovou proměnnou Y*, jejíž složky nemusí být paralelními měřeními:

$$(14) \quad \text{rel}(Y) \geq \alpha = \frac{K}{K-1} \left(1 - \frac{\sum \text{var } X_k}{\text{var } Y}\right)$$

která je známa jako *Cronbachovo alfa*.

Lze je vyjádřit také jako

$$(15) \quad \alpha = \frac{K * C / V}{1 - (K - 1)(C / V)}$$

kde C je průměrná kovariance mezi všemi položkami a V je průměrný rozptyl všech položek.

Pro standardizované položky platí vzorec, v němž R je průměr všech korelačních koeficientů mezi položkami:

$$(16) \quad \alpha = \frac{K * R}{1 - (K - 1)(R)}$$

Cronbachovo *alfa* můžeme také interpretovat jako koeficient korelace mezi daným součtovým indexem a mezi kterýmkoliv jiným testem, jenž má stejný počet položek ze stejného univerza položek a který měří totéž.

Pro studium součtových indexů existuje mnoho různých prostředků, jako jsou Guttmanovy odhady dolní hranice spolehlivosti, Kuder-Richardsonovy vzorce (speciální tvary Spearman-Brownovy formule pro dichotomické položky) a další. Teorie měření pokrývá velmi mnoho reálných a praktických situací.

3. Způsoby odhadu reliability

Pro odhad reliability potřebujeme vždy opakovaná měření téhož. Odhadové postupy jsou založeny na vzorcích či odkazech uvedených výše a na pojmech korelace a variance, tudíž nutně předpokládá zjišťování variability a souběhu variabilit jednotlivých měření (tedy variabilitu odpovědí jednoho respondenta v rámci baterie položek měřících stejnou vlastnost). Přitom opakování musí zajistit vlastnosti paralelních měření, především nezávislost chyb.

Z klasické teorie měření plynou tři známé metody odhadu reliability:

A) test-retest;

B) interní metody odhadu v baterii položek;

C) paralelní měření.

Přes různé nedostatky těchto metod a teoretickou kritiku jsou tyto základní a dobře použitelné postupy v praxi přijímány a využívány.

A) *Test-retest*: znamená opakování dotazu, tj. stejné otázky po jisté době. V praxi se používá časový posun tří až šesti neděl, ovšem časování replikace závisí na problému tak, aby byl zajištěn předpoklad nezávislosti chyb obou dotazů a potlačen efekt zapamatování, a zároveň ale nenastává systematická změna v postoji. Odhadem spolehlivosti je korelační koeficient obou měření.

B) *Interní metody odhadu* v baterii položek vycházejí z předpokladu, že baterie otázek (úloh) splňuje podmínky paralelních měření, a proto lze využít teoretických výsledků pro měření „vedle sebe“ v jednom dotazníku. Zde se vyskytuje několik postupů a používá se několika měř, které jednak odhadují spolehlivost položky a jednak spolehlivost výsledného součtového indexu. Pro baterii položek je možné, především pomocí faktorové analýzy, zkontrolovat, zda položky rovnoměrně sytí společný faktor měřené vlastnosti. Pro takovou skupinu položek můžeme použít:

a) *odhad Cronbachova alfa* pro index podle výše uvedených vzorců; postup je doplněn měřeními, jako je korelace položky a indexu po jejím vynechání a spolehlivost takto redukováného indexu, odhad spolehlivosti pro jednu položku, příp. další míry,

b) „*split half*“ – rozdělení množiny položek na dvě stejně velké (náhodné) části a výpočet korelace mezi nimi; obě části se považují za paralelní měření, proto je jejich korelační koeficient odhadem reliability každé z obou částí a koeficient reliability pro celý index se spočte podle příslušné Spearman-Brownovy formule. (Jiný odhad poskytuje Guttmanova formule pro „split half“, která je obecnější, nepředpokládá stejné reliability obou částí ani stejné variance.)

c) *paralelní měření* – v dotazníku se ptáme na jednu věc dvakrát pomocí obsahově ekvivalentních otázek stejného formátu; za předpokladu, že jde o paralelní měření, můžeme vzájemný korelační koeficient interpretovat jako odhad reliability. Tento způsob má především nedostatek v obtížně zjistitelné ekvivalenci obou znění dotazu.

Každá z uvedených metod má své nedostatky. Je proto vhodné je podle možnosti kombinovat a vzájemně kontrolovat.

Metody studia reliability byly později rozvinuty do komplexnějších modelů umožněných modelováním kovariančních struktur, a jejichž smyslem je zachytit jak reliability, tak pojmy příbuzné. Jejich schémata vyžadují samostatné pojednání. Uvedeme zde jen názvy modelů, pod nimiž nalezneme tyto přístupy v literatuře: A) Quasi Simplex Model, B) Model opakovaných násobných měření (RMM – Repeated Multimethod Model), C) MTMM design (multitrait-multimethod).

Závěr

Zjišťování reliability metodou „*test-retest*“ podle klasické teorie měření není obtížné a je založeno na známých a jednoduchých vzorcích a postupech. Procedury dostupné ve statistických programech (např. procedura Reliability v *SPSS Base*) umožňují aplikaci metody „*split-half*“ a postup prověřování reliability podstatně zjednodušují. Obzvláště u sumačních indexů, jejichž smyslem je právě zvýšit spolehlivost měření a redukovat chybovou varianci, je možné postup použít pro zjištění nutné, resp. postačující délky (počtu položek) indexu, a tak zajistit požadovanou spolehlivost a přitom nepřehlcovat

dotazník zbytečně pokládány otázkami. Při předkládání nového měřicího prostředku uživateli je ovšem dokumentace reliability nutnou podmínkou takové nabídky.

Základním limitem přístupu však je nemožnost zjišťovat validitu, a proto celý postup vychází z výzkumníkovy přesvědčení, že měřicí položky a indexy jsou validní a vyjadřují požadovanou vlastnost. Pro souběžnou kontrolu obou charakteristik kvality dat je zapotřebí speciálního dotazového uspořádání a větší množství informací. Přístup, který takovou úlohu řeší, se nazývá MTMM design (multi-trait-multi-method). Rozšíření postupu test-retest bylo navrženo v pracích Heise [1969], Wiley a Wiley [1970] a byl použit v Alwin [1989] pod názvem Quasi Simplex Model. Joreskog [1971] navrhl model později použitý Sarisem [1982] a nazývaný „repeated multimethod approach“. Tyto postupy vyžadují specifické designy sběru dat a v běžné praxi nejsou plně realizovatelné. Vzhledem k závažnosti pojmu kvality dat jsou však organizovány speciální projekty, jejichž účelem je zajistit nutné výzkumné uspořádání a získat informaci pro naplnění odpovídajícího modelu.

JAN ŘEHÁK je vědeckým pracovníkem Sociologického ústavu AV ČR, kde se zabývá metodologií a analýzou dat.

Literatura

- Alwin, D. F. 1989. „Problems in the data estimation and interpretation of reliability of survey data.“ *Quality and Quantity* 23: 409-442.
- Heise, D. R. 1969. „Separating reliability and stability in test-retest correlation.“ *American Sociological Review* 34: 93-101.
- Joreskog, K. G. 1971. „Statistical analysis of sets of congeneric tests.“ *Psychometrika* 36: 109-133.
- Lord, F. M., M. R. Novick 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley Publ. Comp.
- Saris, W. E. 1982. „Different questions, different variables.“ In *A second generation of multivariate analysis, Vol. 2. Measurement and Evaluation*, ed. by D. C. Fornell. New York: Praeger.
- Saris, W. E., A. van Meurs (eds.) 1990. *Evaluation of measurement instruments by meta-analysis of multitrait multimethod studies*. Amsterdam: North-Holland.
- Wiley, David E., James A. Wiley 1970. „The estimation of measurement error in panel data.“ *American Sociological Review* 35: 112-117.

Summary

The paper provides a review of the classical theory of reliability and its practical consequences. The concept of reliability and its properties are introduced and discussed from the practical point of view. Properties of the reliability model are dealt with in the context of the survey research and the rules following from the model are given. Various practical measures for checking the reliability are summarised. The methods of reliability estimation in the research context are also described for methodological applications.