

Kvalita dat III.:

Empirické výsledky měření reliability pro vybrané míry a stupnice

JAN ŘEHÁK**

Sociologický ústav AV ČR, Praha

IRENA BÁRTOVÁ

JANA HAMANOVÁ

Fakulta sociálních věd UK, Praha

Quality of Data III.

Empirical Results of Reliability Measurement for Selected Variables and Scales

Abstract: The 'test-retest' study for assessing reliability has been carried out for various question formats and scale lengths as well as for various substantive topics. The results selected and displayed in the paper are relevant for sociological surveys. The selection of particular items here has been motivated both by our attempt to help relevant substantive projects that have been regularly carried out, as well as by continual discussions about the likes and dislikes of researchers that are not, however, methodologically supported.

Sociologický časopis, 1998, Vol.34 (No. 3: 363-372)

Cílem statí je prezentovat výsledky metodologického experimentu retestování postojových položek a využití výsledků klasického modelu teorie měření tak, jak byl model shrnut v monografii Lord, Novick [1968] a stručně uveden také v přehledové statí Řehák [1998a]. Chceme nabídnout některé metodologické závěry o reliability položek, jejichž formát se v praxi výzkumu běžně používá a zároveň přitom také ilustrovat postupy teorie měření.

Studium kvality sociologických dat se liší podle typu otázek a přístupů – rozdílný charakter informace umožňuje i různé pohledy na kvalitu a různé metody jejího hodnocení. Měření subjektivních vlastností a postojů (v této statí pojednáváme pouze o číselných proměnných) se provádí několika způsoby a ohodnocení kvality závisí na konkrétním způsobu dotazu a typu otázky. Jsou to a) přímá otázka, b) indikační proměnné pro latentní vlastnost, c) kompozice pojmu – součtové indexy, d) model pojmu – konstrukce proměnné pomocí funkcí a operací. Zaměřujeme se zde na přímé otázky a položky baterií.

V této statí vycházíme z dat dvou akcí „Výzkum spokojenosti a hodnotových preferencí“ a „Etnocentrismus“ projektu GA ČR číslo 403/95/0964, v jehož rámci tato práce také vznikla a z řady tří výzkumů společnosti SC&C pro Českou televizi „Senátní volby 1996“, organizovaných panelovým designem.

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

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

1. Reliabilita dotazu jednotlivých otázek

Studium reliability pomocí metod 'test-retest' naráží na nebezpečí změny zkoumané subjektivní vlastnosti. Proto je metoda vhodná především pro studium měření postojů, neboť jednou ze základních charakteristik postojů, na rozdíl od mínění nebo názoru, je jejich vysoká stabilita. Je také vhodná pro dotazy na objektivní informace, případně i jiné vlastnosti, jejichž neměnnost je prokázána, resp. předpokládána. V této stati vybíráme několik typů otázek, které se liší formou i různou rolí v analýze dat a stupněm neurčitosti a subjektivnosti pro respondenta.

a) subjektivní sebezařazování: dotaz na politickou pozici (sedmibodová škála)

Běžnou otázkou ve výzkumu politických postojů a veřejného mínění je sebezařazování na politické levoprávní škále. Otázka byla v experimentu MTMM95 položena dvakrát (v obou jeho částech) v listopadu a opakovaně v prosinci 1995, a byla také položena třikrát v panelu výzkumu senátních voleb: (1) den před prvním kolem senátních voleb, (2) den před druhým kolem senátních voleb a (3) tři týdny po volbách. Zjištěné korelace a tudíž odhady reliability se pohybují okolo hodnoty 0,75, a tedy 25 % variability odpovědí na politickou škálu (stupnice 1-7) je chybová variance. Z empirických měření tedy plyne, že variabilita odpovědí je ze 75 % určena skutečnou hodnotou a 25 % zbývá na nespolehlivost měření. Výsledky experimentu jsou shrnuty v tabulce 1.

Tabulka 1. Odhad reliability metodou 'test-retest' pro dotaz na levo-pravou politickou orientaci
(sedmibodová škála, 1 = extrémní levice, 7 = extrémní pravice)

A) Experiment 1: MTMM 95

	odhad	n
Soubor A	0,80	392
Soubor B	0,74	394
průměr z obou částí	0,77	

B) Experiment 2: MTMM 96

	odhady					
vlny:	1-2	1-3	2-3	průměr	n	rozpětí
soubor	0,77	0,70	0,76	0,74	603	0,06
pohlaví						
muži	0,78	0,75	0,79	0,77	306	0,03
ženy	0,75	0,65	0,72	0,71	294	0,10
vzdělání						
základní	0,71	0,81	0,81	0,78	53	0,10
bez maturity	0,78	0,72	0,74	0,75	180	0,06
maturita	0,73	0,62	0,71	0,69	270	0,11
VŠ	0,84	0,77	0,83	0,81	96	0,07

vlny:	odhady			průměr	n	rozpětí
	1-2	1-3	2-3			
<i>subjektivní sociální skupina</i>						
dolní	0,81	0,82	0,84	0,82	34	0,03
nižší střední	0,84	0,74	0,82	0,80	141	0,09
střední	0,71	0,61	0,67	0,66	258	0,11
vyšší střední	0,78	0,83	0,90	0,84	38	0,12
horní	0,63	0,87	0,71	0,74	6	0,24
odmítl rozdělení	0,80	0,68	0,78	0,75	62	0,12
neví	0,74	0,72	0,84	0,77	29	0,12
odmítl odpověď	0,89	0,82	0,87	0,86	17	0,07
<i>region</i>						
Praha	0,70	0,39	0,50	0,53	51	0,31
středočeský	0,75	0,71	0,79	0,75	77	0,07
jihočeský	0,76	0,74	0,79	0,76	59	0,05
západočeský	0,53	0,56	0,79	0,63	36	0,27
severočeský	0,85	0,79	0,81	0,82	54	0,06
východočeský	0,74	0,62	0,81	0,72	91	0,18
jihomoravský	0,76	0,81	0,72	0,76	115	0,10
severomoravský	0,88	0,79	0,82	0,83	113	0,09
<i>zaměstnání</i>						
dělník	0,78	0,74	0,78	0,77	97	0,04
zaměstnanec státní správy	0,85	0,78	0,78	0,80	69	0,08
ostatní zaměstnanec	0,75	0,64	0,74	0,71	200	0,11
rolník	0,94	1,00	0,94	0,96	6	0,06
podnikatel, živnostník	0,70	0,68	0,73	0,70	84	0,04
důchodce	0,83	0,72	0,71	0,76	45	0,12
žena v domácnosti	0,57	0,78	0,86	0,74	10	0,29
student	0,61	0,84	0,55	0,67	16	0,29
nezaměstnaný	0,88	0,84	0,94	0,88	11	0,10
ostatní	0,82	0,69	0,84	0,78	30	0,15
<i>věk</i>						
do 29 let	0,71	0,68	0,71	0,70	163	0,03
do 44 let	0,74	0,66	0,78	0,73	225	0,12
do 59 let	0,83	0,77	0,79	0,79	155	0,06
60 let a více	0,86	0,72	0,71	0,77	57	0,15

Poznámky: a) n pro jednotlivé dvojice vln u druhého experimentu se může lišit o jedno nebo dvě pozorování vzhledem k chybějícím údajům – uveden je medián počtu pozorování;
b) u experimentu 1 nebyly výpočty provedeny pro skupiny vzhledem k malému výběrovému rozsahu.

Z této tabulky je vidět, že a) spolehlivost otázky je do značné míry *invariantní* a kolísá především u malých skupin (přirozený jev pro malá *n*); b) korelace mezi časově posloupnými měřeními naznačují, že v čase nastává mírný posun, tj. že reliability a stabilita jsou časově závislé veličiny (koeficient mezi první a třetí vlnou je zpravidla menší než koeficienty sousedních vln. Spolehlivost se tedy snižuje s odstupem doby.

K odhadu reliability byl použit pro srovnání také jednoduchý model faktorové analýzy, který předpokládá, že jedna latentní (skutečná) hodnota ovlivňuje všechny tři měření stejně. Pomocí

konfirmační faktorové analýzy by bylo možné takový model (shodnost faktorových zátěží) testovat. Používáme však explorační model, aby se projevila variabilita odhadů. Reliabilita v tomto přístupu odpovídá čtvrtci faktorového koeficientu. Pro tři dotazy v jejich pořadí dostáváme (extrakce metodou maximalizace Cronbachova koeficientu alfa) postupně hodnoty reliabilit: 0,74, 0,82, 0,69 (průměr 0,75). Odhady modelu se pohybují v podobných intervalech jako přímé párové korelace. Model však znamená jisté a značné zjednodušení vztahu – vychází z předpokladu, že skutečná měřená hodnota je stejná při všech třech měřeních; tedy jde o předpoklad značně nereálný, který nedovoluje žádnou variabilitu skutečné hodnoty v čase (a to ani systematické posunutí ani přirozenou náhodnou změnu). Přesto zmiňujeme tento alternativní přístup pro srovnání. V souhrnném pohledu se oba přístupy vzájemně potvrzují. Rozvinutím tohoto přístupu je quasi-simplexový model [viz např. Řehák 1998b], který však pro tři opakování také nepovažujeme za reálně aplikovatelný a vhodný.

Výsledek odhadu párovými korelacemi můžeme považovat za metodologicky příznivý, jak relativně vysoké procento odhadnuté spolehlivosti, tak invariance měřícího prostředku dávají základ pro bezpečnou aplikaci této často používané otázky. Výsledek ovšem platí pouze pro sedmibodový formát škály, ve kterém zde byla otázka na politickou orientaci prověřována. Poznamenejme, že v empirických korelacích této proměnné očekáváme přibližnou redukci korelačního koeficientu s jinými proměnnými poměrem 0,866 (odmocnina z 0,75).

b) subjektivní sebezařazování: vrstvy

Z otázek subjektivně-sebezařazovacího typu byla v experimentech použita také otázka na subjektivní sebezařazení do společenské vrstvy tak, jak se používá ve výzkumech stratifikace, chudoby, trhu práce apod.; pětibodová ordinální škála dotazu na zařazení se k dolní, nižší střední, střední, vyšší střední a horní vrstvě společnosti vyhovuje více než 80 % populace; odmítnutí škály nebo neschopnost zařadit se bylo postupně ve třech dotazech: 18,9, 16,3, 15,4. Korelovanost (a tedy formální odhad spolehlivosti) dichotomií „zařadil se/nezařadil se“ je (postupně 1.-2. dotaz, 1.-3. dotaz, 2.-3. dotaz): 0,662, 0,617, 0,673 (tento odhad je však silně deflatován vzhledem k vysoké asymetrii četností dichotomických proměnných a je zde uveden pouze ilustrativně; jak známo, koeficient korelace není vhodnou mírou vztahu pro takovéto asymetrické situace).

Odhad reliability těch, kteří se do škály zařadili je vysoký (postupně 1.-2. dotaz, 1.-3. dotaz, 2.-3. dotaz): 0,883, 0,849, 0,857 a v průměru 0,86. Domníváme se, že tento výsledek je pro výzkumy sociální stratifikace velmi důležitý, neboť potvrzuje důležitost proměnné: sebezařazení je stabilní a tedy uvědomělé. Znamená také, že uvedených pět stimulů ordinální škály je významově přijímáno. Metodologická skepse, která je u proměnných obdobného typu vždy na místě, ztrácí z tohoto hlediska na relevanci: představa lidí o svém společenském zařazení je silně stabilní. V důsledku toho také očekáváme silné korelace této proměnné a to ať už v roli závislé, nebo v roli nezávislé proměnné, redukce korelací způsobená nespolehlivostí je pro tuto proměnnou maskována relativně málo (koeficient redukce je 0,927).

c) preferenční dvoupólové stupnice

Častým a výzkumníky oblíbeným formátem otázky je preferenční dotaz, porovnání dvou protilehlých vlastností, jevů, předmětů, položek nebo dotaz typu sémantického diferencálu: respondents na dvoupólové škále vyjadřuje stupeň své preference mezi dvěma póly. Dotaz se provádí většinou s grafickou podporou předložené škály. V našem experimentu se provedený test-retest týkal baterie otázek zkoumajících, jakou představu mají lidé o

vlastnostech senátora: *Jaké vlastnosti má mít Váš senátor? Označte, čemu dáváte přednost!*. Pět položek s pětibodovou škálou reprezentovalo dvojice pólů: (1) chová se nezávisle-prosazuje zájmy své strany, (2) zkušený politik-nezávislý odborník, (3) přizpůsobivý-zásadový, (4) klade důraz na místní problémy-klade důraz na celostátní problémy, (5) prosazuje rovnost všech lidí-prosazuje svobodu pro každého jednotlivce. Baterie byla položena respondentům třikrát v rámci projektu „Senátní volby 96“ (SC&C pro ČT), a máme tedy tři koeficienty korelace, které odhadují reliabilitu. V tabulce uvádíme výsledky. Je vidět na první pohled, že z hlediska přesnosti měření jednotlivce jsou odhady reliability podstatně nižší a je otázka, zda jsou ještě pro empirickou práci přijatelné. Je tedy otázka, zda je tato forma otázky vhodná. Nízké koeficienty jsou invariantní mezi skupinami (rozsáhlou tabulku zde neuvádíme) a žádné závislosti spolehlivosti na sociálních skupinách nebyly odhaleny; pro tento formát otázky se neprosadila žádná z hypotéz typu „s vyšším vzděláním očekáváme zřetelněji formulovaný názor a zvýšenou schopnost jej vyjádřit“, které jsme při přípravě experimentu formulovali.

V tabulce 2 je vidět, že odhady pomocí faktorové analýzy jsou, stejně jako u levo-pravé škály, srovnatelné s přímými korelacemi.

Tabulka 2. Test-retest pro baterii preferenčních otázek
Jaké vlastnosti má mít Váš senátor? (pětibodová škála)

A) reliabilita odhadnutá z přímých korelací

		1.-2. dotaz	1.-3. dotaz	2.-3. dotaz	průměr	rozpětí
nezávislý	straník	0,452	0,437	0,556	0,482	0,119
politik	odborník	0,387	0,355	0,42	0,387	0,065
přizpůsobivý	zásadový	0,388	0,444	0,398	0,41	0,056
lokální	celospolečenský	0,478	0,475	0,514	0,489	0,039
rovnost	svoboda	0,482	0,459	0,613	0,518	0,154
<i>průměr</i>		0,437	0,434	0,5	0,457	0,066
<i>rozpětí</i>		0,095	0,12	0,215	0,131	0,258

B) reliabilita odhadnutá z modelu faktorové analýzy (čtverce faktorových skóre)
(extrakce metodou maximalizace Cronbachova koeficientu alfa)

		první dotaz	druhý dotaz	třetí dotaz	průměr	rozpětí
nezávislý	straník	0,425	0,486	0,637	0,516	0,212
politik	odborník	0,384	0,393	0,448	0,408	0,064
přizpůsobivý	zásadový	0,496	0,3	0,529	0,442	0,229
lokální	celospolečenský	0,503	0,464	0,555	0,507	0,091
rovnost	svoboda	0,403	0,588	0,635	0,542	0,232
<i>průměr</i>		0,442	0,446	0,561	0,483	0,119
<i>rozpětí</i>		0,119	0,288	0,189	0,134	0,337

Zdroj: *Senátní volby 1996*, SC&C pro Českou televizi.

Nižší než padesátiprocentní reliabilitu, tj. více než polovinu variability způsobenou vztahem respondenta a metody vyjádřeným schopností přesně a spolehlivě odpovědět považujeme již za problematické. Hodnoty kolem 0,45 jsou jistě hraniční pro to, aby otázka a její daný formát byly vůbec jednotlivě používány. Výpovědní schopnost dat je při takto nízké reliabilitě zpochybnitelná: *buď je chybná metoda, nebo respondent nemá dostateč-*

ně vykrystalizovaný názor. V každém případě se korelace takovéto proměnné redukuje koeficientem 0,78 a veškeré explorační analýzy, založené na dekompozici variability jsou proto nepřesné a neostře. Nízká reliabilita zamlžuje vztahy. Poznamenejme, že použité otázky byly koncipovány tak, aby každá z nich měla samostatnou analytickou roli, každá vlastnost má svůj samostatný význam. V praxi by bylo velmi obtížné každou z uvedených dimenzí nějakým způsobem dublovat a indikovat ji několika dotazy. Uvedené závěry platí pro každou položku v její samostatné analytické aplikaci. Nicméně i v těchto případech lze reliabilitu v analýze dat zvýšit, a to v případě, že jednotlivé položky se spojují do obecnějších faktorů a syntetických pohledů – v lineární kombinaci příbuzných a vzájemně korelovaných položek (např. při používání odhadu hodnot latentních faktorů ve faktorové analýze) se reliabilita syntetického indexu zvyšuje tím, že se spojuje vzájemně se kontrolující a doplňující informace z položek do jedné proměnné.

d) dotaz na objektivní údaje o čistém příjmu a na subjektivní hodnocení příjmu

Příjem a plat v sociologických šetřeních působí problémy, neboť respondenti vypovídají neochotně a velmi často odmítají informaci o příjmu sdělit. Informace o příjmu jsou však pro mnoho sociologických i příbuzných analýz podstatné, a to jak z důvodů odhadů výše příjmu v populaci a ve skupinách obyvatelstva, tak jako vysvětlující nebo zprostředkující proměnná v modelech sociologických teorií. Proto jsme do experimentu MTMM 95 zahrnuli tyto dotazy v několika otázkách a dotazu byl ponechán zcela přirozený průběh (nebyly používány žádné podpůrné metody, jak respondenty přesvědčit o tom, aby tuto informaci sdělili).

Uváděná výše platu (respondent měl sdělit přímou hodnotu, nikoliv zařadit svůj plat do kategorií) prokázala v části A experimentu MTMM 95 spolehlivost 0,927 a v části B 0,945, (průměr 0,936) tedy velmi vysokou (a nečekanou) hodnotu. Výsledky jsou však poznamenány tím, že jen 26 % respondentů na tuto otázku odpovědělo (v části A i v části B). Výsledek je zajímavý: ti, co odpovídají, odpovídají spolehlivě, avšak jakékoliv zobecnění výsledku není možné, protože odhad populačního průměru je vychýlený a nepoužitelný. Navíc je poměr odmítnutí tak vysoký, že lze vyslovit hypotézu (bohužel vzhledem k malým četnostem obtížně z uvedených dat testovatelnou) o tom, že vychýlení se projevuje v různých vrstvách, třídách a skupinách různě. Jelikož tato proměnná je v mnoha výzkumech kritická (ať už v roli závislé nebo nezávislé proměnné), uvedená čísla zdůrazňují velmi závažnou metodologickou úlohu – redukovat počet odmítnutí a nalézt cestu, jak odhadnout příjem tak, aby výsledky byly použitelné. Ve výzkumu MTMM 96 testování neprobíhalo, plat však sdělilo pouze 37 % a asi 50 % přímo odmítlo na otázku odpovědět. (Poznamenejme, že uvedená procenta jsou počítána z celého souboru, nikoliv z části, která má vlastní příjem). Z těchto faktů plyne, že přímý dotaz na příjem (často některými uživateli vyžadovaný) nemá výzkumný smysl a je nutné se dotazovat zařazováním do kategorií nebo vybudovat speciální metodologii přímý dotaz podporující.

Subjektivní hodnocení svého platu a své finanční situace vykazují odhady reliability standardně nad 0,60. U těchto otázek se dramatický pokles poměru odpovědí nevyškytl.

Tabulka 3. Reliabilita otázek na plat a finanční situaci

	n	Odhad reliability
škála 5 bodů adekvátnost platu	388	0,674
škála 5 bodů porovnání finanční situace s průměrnou rodinou	420	0,675
škála 9 bodů finanční situace – spokojenost	423	0,659
škála 9 bodů finanční situace – vzdálenost od ideálu	422	0,642

O něco nižší reliability devítibodových škál oproti pětibodovým u obsahově příbuzných otázek můžeme pouze konstatovat, výsledek nelze z těchto dat zobecnit. Hypotézu však podporuje také výsledek testu pro dotaz na spokojenost (viz dále).

e) otázky na spokojenost

Otázky na spokojenost mají v empirickém výzkumu vysokou frekvenci využití: používají se při studiu kvality života a také při praktických aplikacích výzkumu trhu a v personalistice. Ve výzkumu se používají různé typy škál i různé typy dotazu. Pro experiment 'test-retest' byly použity přímé otázky s devítibodovou stupnicí.

Tabulka 4. Odhad reliability metodou test-retest, devítibodová škála spokojenosti

	n	Odhad reliability
spokojenost s politickým vývojem	417	0,681
spokojenost s prací	394	0,578
spokojenost s osobním životem	422	0,607
celková spokojenost v životě	421	0,531

Časté otázky na spokojenost v různých kontextech života nevykazují tedy vysokou reliability, nespolehlivost až 47 % je pro empirické výzkumy používající tento formát dotazu nepřijemným výsledkem. Položky spokojenosti mají samostatnou analytickou roli, a proto u nich v takovém kontextu většinou nemůžeme reliability zvýšit vytvářením sumačních indexů z paralelních, resp. alespoň obsahově příbuzných položek, indikujících jeden společný latentní faktor. Sumační indexy spokojenosti mají jiný charakter, součet znamená kompozici pojmu ze složek, které nemusí být (a často také nejsou) korelované. Korelace položek v případě takto dekomponovaných pojmů nemá jen charakter obsahového překrytí, a tudíž indikační vlastnosti společné latence, ale věcnou, kontextovou, případně i kauzální interpretaci. Konstrukce položek obsahovou dekompozicí pojmu odpovídá tedy jinému účelu než zvýšení reliability opakovaným dotazováním na jednu vlastnost jejími různými indikátory (důsledky). Proto je nutno věnovat metodologii takových dotazů dále zvýšenou pozornost a hledat jiné prostředky pro dosažení vyšší spolehlivosti dat, pokud je vyžadována (jiný formát, resp. model dotazu, nepřímé otázky, indikační baterie).

2. Reliabilita škál souhlasu

Nízké reliability jednotlivých položek obsahově příbuzných baterií jsou v praxi běžné. V důsledku aplikace Brownových vzorců se používají pro měření postojoyvých vlastností syntetické součtové indexy, vytvořené jako sumy, resp. průměry stejně orientovaných položek. Položení dotazu na jednu vlastnost (postoj) v několika vzájemně se zastupujících položkách vede k tomu, že se reliability jejich součtu podstatně zvyšuje. Odhady reliability mohou sloužit jako základ aplikace Brownovy formule pro hledání dostatečné délky (po-

čtu položek) v součtovém indexu. Experimenty obsahovaly různé dotazové kombinace obsahu a formy. Zde prezentujeme položky se čtyřbodovými a pětibodovými stupnicemi, jejichž hodnoty byly pojmenovány a v pojmenovaném tvaru předloženy respondentům na kartě.

a) položky souhlasu s výroky na čtyřbodové škále

Baterie výroků na ekonomické, sociální, národnostní a politické názorové orientace zahrnovala směs různých obsahově od sebe odlišných položek. V tabulce 5 shrnujeme odhady reliabilit pro celou populaci.

Tabulka 5. Odhad reliability metodou test-retest, různé položky, čtyřbodová škála souhlasu

	n	Odhad reliability
celkový rozvoj ekonomiky	419	0,475
práva menšin i na úkor práv většiny	419	0,311
nezávislí odborníci	415	0,397
vedoucí představitelé – vždy uvádět pravdu	417	0,538
horní hranice příjmu	419	0,584
stát nepřipustí nadbytek, pokud chudoba	416	0,550
rozvoj ekonomiky versus strádání občanů	414	0,471
silní a schopní řídí	414	0,433
menšina může oponovat, ale ne protivit se většině	417	0,289
někteří jsou způsobilější – tradice a rodinný původ	411	0,485
vláda odpovídá za práva menšin	416	0,439

Reliabilita souhlasu s výroky je zde velmi nízká. Analytická práce s jednotlivými položkami je proto bezesporu velmi zpochybněna – samy o sobě by se tedy neměly v analýze používat, neboť statistická výpověď z nich je velmi neurčitá. 50 % až 60 % a více variability je chybových a hledat její vysvětlení nemá smysl – naopak je nebezpečné analýzu provádět: výsledky reflektují nahodilou strukturu dat.

b) pětibodové škály souhlasu

Tabulka 6. Odhad reliability pro souhlas s výroky o vztahu k svému národu pětibodová škála souhlasu

	n	Odhad reliability
hrdost na národní historii	421	0,618
negace českého způsobu života	418	0,442
ostatní země se mohou od Čechů učit	422	0,590
Češi mají více důvodů k hrdosti než ostatní	418	0,597

Porovnání výsledků pětibodové škály se čtyřbodovou se nabízí hypotéza o tom, že absence středového bodu čtyřbodové škály snižuje spolehlivost odpovědi a vnáší do dat zbytečnou variabilitu přechodů mezi hodnotami 2 a 3 pro osoby, které by se chtěly zařadit do středu.

Výsledky indikují výraznou preferenci pětibodové škály před čtyřbodovou.

Závěry

Reliabilita zjišťovaná na instrumentech různého formátu a pro různá substantivní témata se ukazuje jako velmi variabilní a závislá na formátu a na kontextových okolnostech. Metoda „test-retest“ odhaduje pomocí svého modelu spolehlivost, resp. nespolehlivost dat, současně však v praxi reflektuje komplexní vlastnost stability odpovědí. Ta zahrnuje reliabilitu instrumentu, pevnost/stabilitu postoje či názoru a soustředěnost respondenta, resp. rušivé vlivy při realizaci dotazu. Tyto složky nelze od sebe metodou separovat. Výsledky však (ať už v nich dominuje kterákoliv uvedená složka) v každém případě vypovídají o kvalitě dat, o jejich upotřebitelnosti pro přímé odhady i pro jejich vstup do komparačních, korelačních, regresních, strukturních, klastrových apod. analýz. Nízká variabilita jednak maskuje a zamlžuje reálné vztahy tím, že získané empirické korelace a komparační statistiky příslušné vztahy nevykazují a samozřejmě i znehodnocuje konfidenční intervaly, neboť klasické vzorce pro směrodatnou chybu dodatečnou variabilitu neobsahují. Chyba z nespolehlivosti se přičítá v analýze dat k chybě výběru.

Některé výsledky experimentu mohou u některých sociologů podpořit jistou metodologickou skepsi k empirické sociologii. Nedomníváme se však, že je to nutný závěr. Výsledky jsou především varováním proti metodologickému diletantismu a proti podceňování přípravy empirických šetření z důvodů nadměrné subjektivní přesvědčenosti o kvalitě „svých“ otázek a o tom, jak tyto otázky „dobře pracují“, poskytují-li nám výsledky, které očekáváme a které se hodí pro argumentování svých teorií. Tyto výsledky tedy by měly být spíše tlakem na zvýšení důrazu v přípravě kvalitního datového základu (byť náročné a nákladné), v žádném případě odmítnutí některých měřicích baterií a formátů otázek nemůže být použito jako argument proti empirické sociologii.

Metoda „test-retest“ byla použita pro různé formáty a délky škál i pro různé tematické okruhy. Výsledky, které byly vybrány a prezentovány v článku, považujeme za důležité pro empirickou sociologii. Výběr položek byl motivován snahou pomoci pravidelně zkoumaným tématům a projektům i nekončícími diskusemi o vhodnosti délky škál, založenými ovšem na subjektivních preferencích a argumentech jednotlivých výzkumníků. Cílem studie je ovšem přispět konkrétními výsledky pro postupnou metodologickou generalizaci.

Stručný souhrn závěrů pro praktickou práci a přípravu dotazníků:

- a) sedmibodová škála levo-pravé politické orientace vykazuje vysokou spolehlivost kolem průměru 0,77, a to s přirozenou, ale nikterak vysokou variabilitou mezi skupinami a kroky experimentu;
- b) sebezařazení na subjektivní škále sociální třídy založené na pětibodové stupnici je vysoce reliabilní (0,86) s více než 80 % odpovědí – výsledek je příspěvkem k diskusím o nespolehlivosti této proměnné; výsledek také přináší jasnou evidenci proti kritice smysluplnosti tohoto pojmu a jeho operacionalizace;
- c) velmi nízkou reliabilitu (v průměru 0,46) vykazují bipolární stupnice preferencí – zde byly aplikovány k očekávaným vlastnostem senátora;
- d) při přímém dotazu na částku osobního měsíčního příjmu (bez jakékoliv podpory jinými otázkami) byl zjištěn zajímavý fakt: pouze 27 % respondentů odpovědělo na tuto otázku – ti, co odpověděli, však poskytli informaci s velmi vysokou reliabilitou – 0,94;

e) otázky na subjektivní ohodnocení příjmu byly odpovídány bez problémů s velmi malým procentem odmítnutí; spolehlivost pětibodových stupnic je 0,67, devítibodových 0,65;

f) spolehlivost satisfakčních otázek na devítibodové stupnici byla v rozpětí od 0,53 do 0,68;

g) často užívané otázky na souhlas s výroky, které se používají pro měření postojů, byly testovány se čtyřbodovou a pětibodovou stupnicí:

4-bodová stupnice ukázala naprosto nevyhovující spolehlivost s hodnotami v rozpětí mezi 0,29 a 0,58;

5-bodová stupnice má spolehlivost vyšší, a to mezi 0,44 a 0,62;

přesto, že položky souhlasu s výroky jsou používány převážně k vytváření součtových indexů, resp. faktorových skóre, jejichž spolehlivost se značně zvyšuje s počtem položek, výsledek varuje před používáním jednotlivých otázek (resp. i „krátkých“ indexů), a to především v korelacích – takové využití vyžaduje vysoký stupeň opatrnosti (pokud se k němu vůbec rozhodneme);

výsledky jednoznačně vypovídají ve prospěch pětibodové stupnice ve srovnání se čtyřbodovou z hlediska spolehlivosti.

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

IRENA BÁRTOVÁ absolvovala Matematicko-fyzikální fakultu UK v Praze v oboru pravděpodobnost a matematická statistika. Zabývá se matematickými a matematicko-statistickými aplikacemi v hydrometeorologii, územním plánování a v sociologii. Přednáší statistiku na katedře sociologie Fakulty sociálních věd UK v Praze.

JANA HAMANOVÁ absolvovala Fakultu sociálních věd UK v Praze, obor sociologie. Pracuje v sociologické skupině SC&C, s.r.o. Praha jako vedoucí výzkumných projektů a konzultantka.

Literatura

Lord, F. M., M. R. Novick 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley Publ. Comp.

Řehák, J. 1998a. „Kvalita dat I. Klasický model měření reliability a jeho praktický aplikační význam.“ *Sociologický časopis* 34: 51-60.

Řehák, J. 1998b. „Kvalita dat II. Přístupy ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur.“ *Sociologický časopis* 34: 195-204.