

1. Wikipedia has a user friendly interface.



2. Wikipedia is usually my first resource for research.



Jak na postojové škály: Tvorba položek a jejich zpracování

SEME 2016 | 28. leden 2016 | Pedagogická fakulta MU

Hynek Cígler
Katedra psychologie a IVDMR FSS MU
hynek.cigler@mail.muni.cz

Obsah workshopu

1. Co to je škála a čím se liší od „anketních“ otázek.
2. Typy položek k měření postojů.
3. Vlastnosti škál: validita, reliabilita.
4. Tvorba a výběr položek.
5. Položková analýza.
6. Praktický workshop.

Co nezazní:

- Jde o komplexní problematiku, stihneme jen **zjednodušeně** malý **výsek**.
- Veškeré analýzy jen v MS Excel (bez využití statistického SW).

Cílem workshopu je postup tvorby škál pro výzkum, nikoliv pro individuální posouzení postojů jednotlivce.

Škála

Soubor položek, které měří stejný konstrukt.

Na rozdíl od anketní otázky **nevyhodnocujeme položky samostatně**; nezajímá nás „**pravdivostní hodnota**“ položek, ale celé škály dohromady.

Hlavní nevýhody:

- časová náročnost pro respondenta;
- náročnější vývoj, zpravidla nutnost pilotáže;
- nebezpečí reifikace (to se nicméně vztahuje i na anketní otázky).

Hlavní výhody:

- vyšší přesnost měření (reliabilita);
- škála může lépe reprezentovat to, co chceme měřit (obsahová validita);
- širší „záběr“ výsledné škály (obsáhne komplexnější konstrukt).

Škála, škálování

Budeme se zabývat postojovými škálami.

- Většina postupů ovšem platí i pro osobnostní škály, testy schopností i didaktické testy.

Občas se slovo „škála“ používá ve dvou významech:

- výsledek „škálování“ (soubor položek, test);
- samotný proces „škálování“ coby typ položky (Likertova škála atp.).

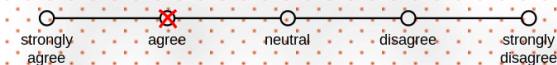
Typy položek k měření postoje

Likertova škála

- „Nakolik souhlasíte s následujícími výroky...“
- Nejjednodušší vyhodnocení a analýza.
 - Tímto se dnes budeme zabývat.
- Unipolární škála, respondent zaznamenává souhlas/frekvenci/atp. s tvrzením.
- Vyvinuto k měření osobnostních vlastností.
 - Rensis Likert.

Example Likert Scale

1. Wikipedia has a user friendly interface.



2. Wikipedia is usually my first resource for research.



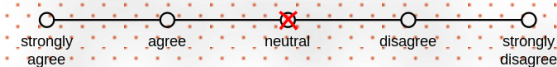
3. Wikipedia pages generally have good images.



4. Wikipedia allows users to upload pictures easily.



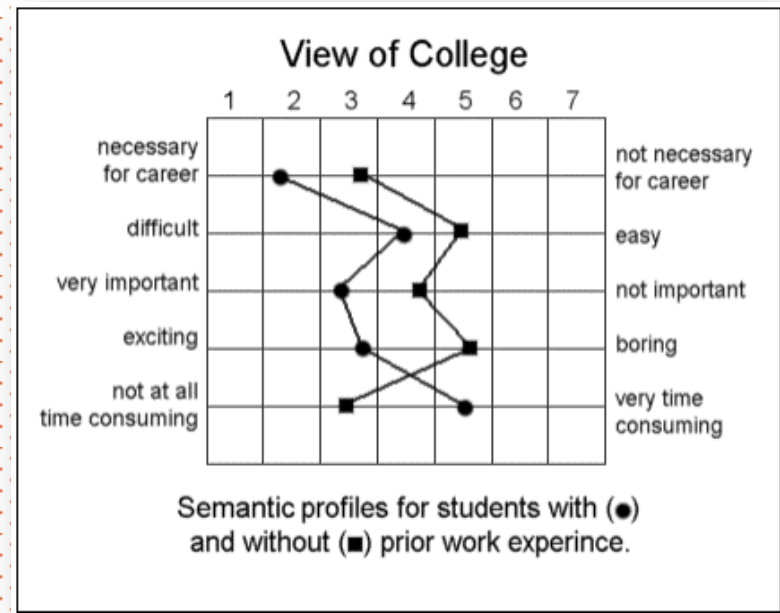
5. Wikipedia has a pleasing color scheme.



Typy položek k měření postoje

Osgoodův sémantický diferenciál

- „Vyberte mezi následujícími výroky na škále..“
- Komplikace při designu, analýze i interpretaci.
- Podobné Likertově škále, ale škála je bipolární – respondent volí míru, v jaké se přiklání k některému z antonym.
- Přímo vyvinuto k měření postojů.
 - Charles E. Osgood



Typy položek k měření postoje

Q-sort (Q-třídění)

- „Seřadte následující výroky...“
„Roztřídte pojmy podle významnosti...“
- Někdy též **Q-metodologie**.
 - Q jako questions oproti R jako respondents.
- Komplikovanější analýza a jiný účel (Q-factor analysis), těžší interpretace.
- Užitečné např. pro pilotáž.
- Respondent třídí výroky podle souhlasu; omezení v počtu jednotlivých kategorií.
- Vyvinuto k zachycení subjektivity (intra- i interindividuální).
 - William Stephenson.

Least Important -2	-1	0	+1	Most Important +2

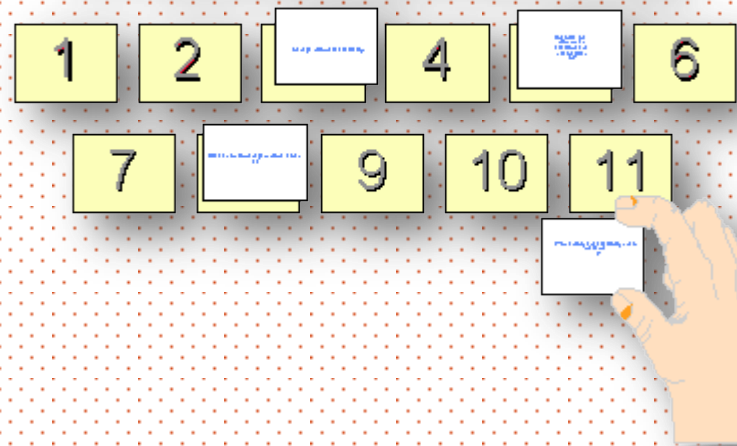
Typy položek k měření postoje

Thurstonova škála

- „Seřadte následující výroky...“
- Komplikace při designu, analýze i interpretaci.
- Podobné Q-sortu, ale každému „kroku“ přiřazujeme pouze jediné „tvrzení“.
- První metoda vyvinutá k měření postojů.
 - Louis Leon Thurstone

1 = least favorable to the concept

11 = most favorable to the concept



Struktura Likertovy položky

1. Instrukce (není nezbytné)

- „U každého z následujících výroků označte míru, nakolik s ním souhlasíte.“
- „U každé z následujících činností vyberte, jak často ji provádíte.“
- Slouží k sjednocení „odpovědního stylu“ respondentů, poskytuje další nutné instrukce k odpovídání.

2. Podnět (stimulus)

- „Měla by vláda uzavřít státní hranici?“
- „Jak často si vyhledáváte informace o migrační krizi?“
- Vlastní „otázka“. Krátké tvrzení, které vzbudí „reakci“ na základě znalosti instrukce.

3. Odpověď složená z více „distraktorů“ (response)

- „souhlasím – spíše souhlasím – spíše nesouhlasím – nesouhlasím“
- „téměř nikdy – občas – často – skoro pořád“
- Vlastní „odpověď“ na daný podnět.
- Sama o sobě nás nezajímá, důležité je, jak funguje ve shodě s ostatními položkami.

Podnět Likertovy položky

Krátký, jednoduchý, jednoznačný, rychlý na pochopení.

- „Lidé jsou obecně dobří.“

Pozor na záporny, zejména dvojité.

- „Nemyslím si, že...“; „Lidé by neměli nedávat najevo...“

Měl by vzbudit rychlou, bezprostřední reakci.

- Nezajímá nás tak úplně pravdivostní hodnota možných odpovědí, neměly by proto spouštět racionální rozhodovací procesy, ale naopak intuitivní pocit souhlasu (frekvence atp.).

Neměl by „konstruovat“ odpověď.

- „Také si myslíte, že...“

Implikuje formát odpovědi.

- Míra souhlasu, důležitosti, pravdivosti, vyjádření frekvence...
- Inspirace: <https://www.clemson.edu/centers-institutes/tourism/documents/sample-scales.pdf>

Odpověď na Likertovu položku

JE VHODNÉ DODRŽET

Stejný směr a počet stupňů napříč položkami jedné škály.

- Distraktory musí být ordinálně seřazené.
- „Vzdálenosti“ by měly být přibližně podobné.

Volby odpovědí by měly být přibližně normálně rozdělené.

- Nepříliš extrémní distraktory („nikdy“, „zcela určitě“).

Krátké, jednoduché distraktory.

JE NUTNÉ ROZHODNOUT

Počet možných odpovědí.

- Vyšší počet zdánlivě více informace, odpovědi se ale začínají „překrývat“, nejsou odlišitelné.

Střední hodnota?

- Někdy vhodné, jindy nikoliv.

Rozdílná valence?

- Komplikovanější otázka.

Možnost zvolit „nevím“?

- Pokud vůbec, zpravidla ne doprostřed škály, protože „tak na půl, je mi to jedno“ je významově odlišné od „nevím, nemám dost informací“.

Odpověď na Likertovu položku

Stejný typ odpovědí v rámci jedné škály (nelze sčítat hrušky a jablka).

- Bud' souhlas, nebo frekvence, nebo...

V rámci jedné škály by měly mít všechny položky stejný počet distraktorů (souvisí s požadavkem shodného rozptylu, viz dále).

- Lze částečně řešit standardizací odpovědí např. na z-skóry.

Formát otázky/odpovědi implikuje měřený konstrukt.

- „Často chodíte...“ → „souhlasím, spíše souhlasím...“ („pocit častosti“)
vs.
„Jak často chodíte...“ → „každý den, alespoň jednou týdně...“ („odhad frekvence“)

Zvážit použití „kotev“ a jejich popis (zpravidla výhodné).

- Konkrétní pojmenování jednotlivých distraktorů.
- Pozor na „VAŠ“ (vizuální analogové škály), procentové škály atp.
 - V těchto případech kotvy chybí a je posílen odpověďový styl respondentů na úkor konstruktů.

Zkreslení odpovědí

Tzv. „**response bias**“ – ovlivnění odpovědi na položku něčím jiným, než měřeným kostruktem.

Souvisí s „**response style**“ – určitým stylem odpovídání konkrétního respondenta v konkrétní situaci, který může zneplatnit výsledky.

- Může být jak záměrné, tak i nezáměrné.

Typické zdroje zkreslení:

- Nahodilé odpovědi a záměrné zneplatnění výsledků.
- Sociální žádoucnost.
- Tendence k (ne)souhlasu.
- Simulování a sebeznevýhodňování.
- Tendence k extrémním nebo průměrným odpovědím.
- Hádání (u postojové škály asi nehrozí 😊).

Kvalitní distraktory a podněty zpravidla redukují, existují i specifické postupy.

Latentní proměnná

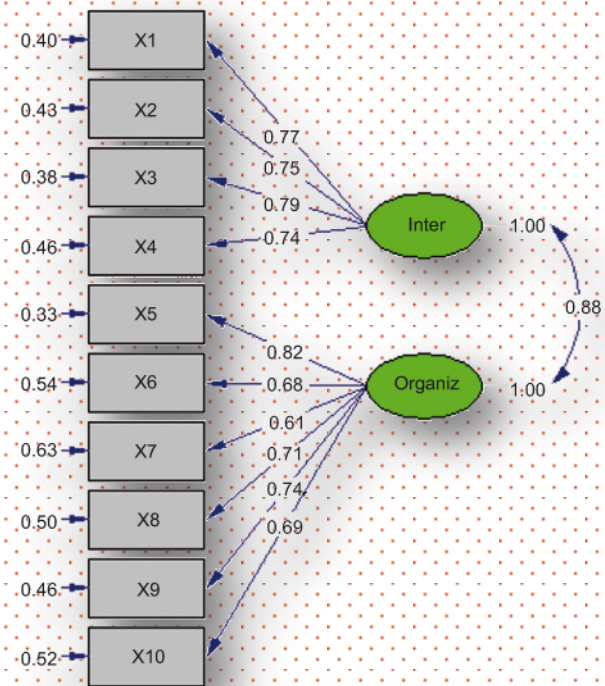
Předpoklad škálování:

- Odpovědi na položky „způsobuje“ přímo nepozorovatelná latentní proměnná.
 - Např. celkově *protiuprchlícký postoj* „způsobuje“ *nesouhlas* s otázkami týkajícími se různých aspektů přijímání uprchlíků.

Výjimečně:

- Jednotlivé odpovědi (symptomy) „konstruují“ latentní proměnnou.
 - Např. *setkání se s negativním chováním sester* má v důsledku vliv na vytvoření *negativního postoje k nemocnici*.
- Situace se zásadně epistemologicky liší a má důsledky pro odlišnou práci s položkami.

Je nezbytné vědět, která ze situací platí (u postojů zpravidla první).



$\chi^2=45,51$, $df=34$, $p=0,08964$, $RMSEA=0,043$

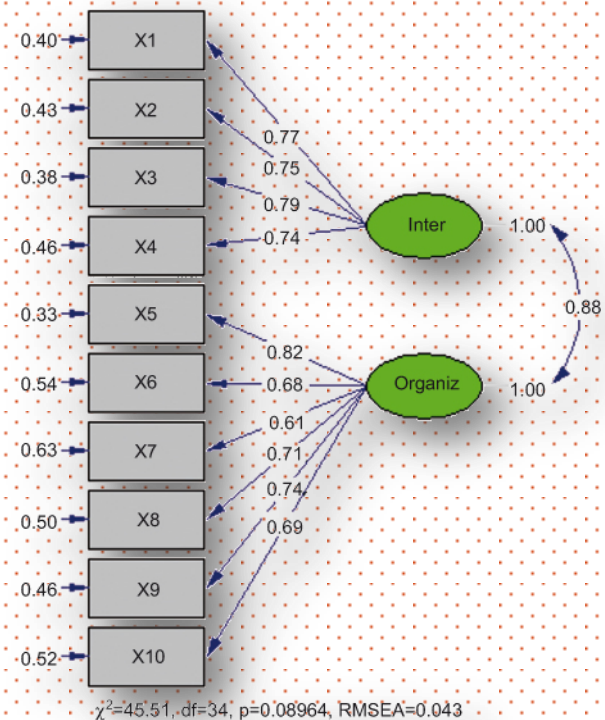
Latentní proměnná

Celkový skór je zpravidla součet nebo průměr odpovědí na jednotlivé položky.

- Model tzv. klasické testové teorie (CTT). Existují alternativní modely (zejm. teorie odpovědi na položku, IRT).

$$X = \tau + e$$

- Vzorec platí pro celkový skór, metaforicky i pro položku.
- X = pozorovaný skór: součet položek. Manifestní proměnná.
- τ = pravý skór: skutečná hodnota respondenta na škále (nelze přímo měřit). Latentní proměnná.
- e = chyba měření: míra odlišnosti pravého a pozorovaného skóre.



Základní vlastnosti škály

RELIABILITA

Jak přesně dokážeme z pozorovaného skóre usuzovat na pravé skóre?

Jak spolehlivě/přesně měříme?

Relativní nepřítomnost chyby měření.

Základní způsoby ověření:

- vnitřní konzistence
- test-retest, paralelní formy
- shoda posuzovatelů

VALIDITA

Jak moc je hypotetický pravý skór shodný s tím, co jsme chtěli měřit?

Co měříme?

Podmínkou validity je reliabilita

- Když neměříme přesně, nemá smysl přemýšlet o tom, co měříme.

My se budeme zabývat jen tzv. „obsahovou validitou“.

Tvorba položek: Základní pojmy

Položky by měly plně a rovnoměrně pokrývat **obsahové univerzum**, aniž by se vzájemně příliš překrývaly svým významem.

- Měly by být **jednodimenzionální**, měřit právě jeden konstrukt.
- Všechny by měly mít stejný vztah s pravým skóre, neměly by tvořit dvojčky či trojičky podobných položek (**tau ekvivalentní**).
- **Psychometrický paradox**: Příliš „úzký“ výběr položek vede ke zvýšení reliability (položky jsou si velmi podobné), ale klesá validita (nepostihujeme všechny možné významy konstruktů) – měří sice přesně, ale jen úzký výsek problematiky.

Obsahové univerzum (universe): veškeré možné významy, kterých může nabývat měřená proměnná.

- Kvalitou vybraných položek zajišťujeme tzv. výběrovou, vlastní obsahovou validitu (měřím to, co měřit chci).
- Obsahuje „všechny možné myslitelné položky“.

Položková banka: seznam skutečně vytvořených, potenciálně použitelných položek, ze kterých sestavíme škálu.

Finální set položek je dobré volit až na základě **pilotáže**.

Fasetový model tvorby položek

Identifikuji všechny možné „fasety“ měřeného konstruktů.

- Proces se někdy označuje jako **dekompozice obsahového univerza**.
- Můžu si na základě teoretických předpokladů zvolit „váhu“ jednotlivých faset.
- Mohou se fasety překrývat či „křížit“?
- Fasety mohou pocházet z jedné nebo více domén.

Vytvořím položky pro každou fasetu tak, aby jejich počet reprezentoval váhu jednotlivých faset.

- Je lepší vymyslet více položek, a až poté jejich počet redukovat.

Ve všech fázích je důležitá diskuze s dalšími odborníky na dané téma.

Jednotlivé fasety mohou ve výsledku představovat různé dimenze/faktory postoje, může jít ale i o různé „aspekty“ toho samého.

Fasetový model: příklad

Postoje k přijímání uprchlíků.

Doména 1:

obecné postojové složky

- 1A: kognitivní
- 1B: afektivní
- 1C: behaviorální

Doména 2: specifika situace

- 2A: soucit s uprchlíky
- 2B: strach o vlastní bezpečí
- 2C: mezinárodní závazky
- 2D: informovanost o problematice
- ...

Příklady položek:

- 1A×2C – Odmítnutí uprchlických kvót poškodí pozici ČR v rámci EU.
- 1C×2B – Bude-li ČR přijímat migranty, budou muset občané vzít bezpečnost do vlastních rukou.
- 1B×2B – Kvůli migrantům mám vážné obavy o bezpečnost své rodiny.
- 1A×2D – V médiích o uprchlících slyšíte... (téměř vůbec, alespoň jednou denně...).

Položková analýza

Sada postupů, které vedou k výběru nejlépe fungujících položek.

- Zaměříme se na analýzy v rámci klasické testové teorie.
- Seznam postupů není tak docela ohraničen.

Zaměřuje se typicky na:

- Deskriptivy položek.
 - M, SD, obtížnost, rozložení (zešíkmení).
- Deskriptivy škály
 - M, SD, rozložení
 - Vhodnější nahradit např. faktorovou analýzou, ale k tomu potřebujeme statistický SW.
- Diskriminační účinnost položek.
 - Korigovaná korelace položky se škálou, Cronbachovo alfa po odstranění položky, ULI index.
- Odhad reliability
 - Cronbachovo alfa...
- Korelační matici
 - Průměr, medián korelací.
 - Měla by být „vyvážená“, žádná položka by neměla vybočovat.

Položková analýza

Sada postupů, které vedou k výběru nejlépe fungujících položek.

- Zaměříme se na analýzy v rámci klasické testové teorie.
- Seznam postupů není ohraničený.

Typicky se pokračuje:

- Analýzou korelační matice
 - Tohle je dobré udělat i v Excelu jako základní kontrola vzájemných vztahů položek.
 - Průměr, medián korelací.
 - Měla by být „vyvážená“, žádná položka by neměla vybočovat.
- Faktorová analýza
 - Analýza počtu faktorů (mnoho postupů).
 - Identifikace faktorů a jejich souvislosti s položkami.

Deskriptivy položek

Směrodatná odchylka

- Položky musí variovat mezi respondenty (mít vysokou SD).
- „Kde není rozptyl, není nic dalšího.“

Průměr, popularita (obtížnost)

- Položky by optimálně měly být pokud možno průměrně obtížné.
- Příliš „lehké“ či „obtížné“ položky špatně diferencují.
- Popularita: $p = M/\max$ (průměrný/maximální skór položky).

Zešikmení

- Souvisí logicky s průměrem, zpravidla není potřeba analyzovat separátně.

Frekvence distraktorů

- Distraktory by měly být normálně rozdělené a všechny využité.

Deskriptivy škály

Škála by měla být normálně rozdělená, výrazné zešikmení může být problém.

Neměl by nastat efekt stropu či podlahy.

Diskriminační účinnost pol.

Korigovaná korelace položky se škálou

- Korelace položky s testem po odečtení skóre položky (korelace položky se součtem ostatních položek).
 - Nekorigovaná korelace je zkreslená v případě menšího počtu položek, resp. položek s vysokým rozptylem, které hodně přispívají k celkovému skóre testu.
- „Jak moc souvisí vysoké skóre v položce s vysokým skóre v testu?“
- Chceme co nejvyšší.

Cronbachovo alfa po odstranění položky

- Typický výstup např. z SPSS.
- Jak se změní reliabilita, když odstraníme položku?
- Chceme co nejnižší.

ULI index

- Rozdíl průměrných skóre v položce nejlepší a nejhorší třetiny (čtvrtiny, poloviny...) respondentů. Chceme co nejvyšší.
- Spíše zastaralý index z doby, kdy se korelace musely počítat ručně.
- Bývá zvykem „standardizovat“ (rozdíl popularit)

Reliabilita

$$r_{xx'} = \frac{\sigma_t^2}{\sigma_x^2} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$$

- σ_t^2 je rozptyl pravého skóre, σ_x^2 pozorovaného a σ_e^2 chyby měření.

Reliabilita je:

- Podíl rozptylu měřeného rysu na celkovém rozptylu.
- Virtuální „korelace metody se sebou samou“.
- $\sqrt{r_{xx'}} = r_{xt}$ – odmocnina z reliability je korelace naměřených skóre se skutečnou hodnotou měřeného pravého skóre

Reliabilita by měla být nad 0,7 (cca třetina chyby), lépe nad 0,8.

- Dobrá metoda má reliabilitu nad 0,9.
- Co je dobré, záleží vždy na účelu škály (výzkum, posouzení osob...).

Reliabilita

Nejpoužívanějším ukazatelem reliability je Cronbachovo alfa:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_T^2} \right)$$

- k – počet položek
- $\sum_{i=1}^k \sigma_i^2$ – součet rozptylů položek
- σ_T^2 – rozptyl celkového skóre

Předpoklady pro použití Cronbachova alfa (zjednodušeně):

- Všechny položky měří měřený rys shodným způsobem a ve stejné míře.
- Jednodmenzionalita.
- Pokud předpoklady nejsou splněny (což je vždy), alfa do určité míry podhodnotí odhad reliability.
 - Cronbachovo alfa je tedy jedním z ukazatelů minimální reliability testu.

Reliabilita

V rychlosti (a u určitých druhů testů je to žádoucí) lze nahradit tzv. split-half reliabilitou.

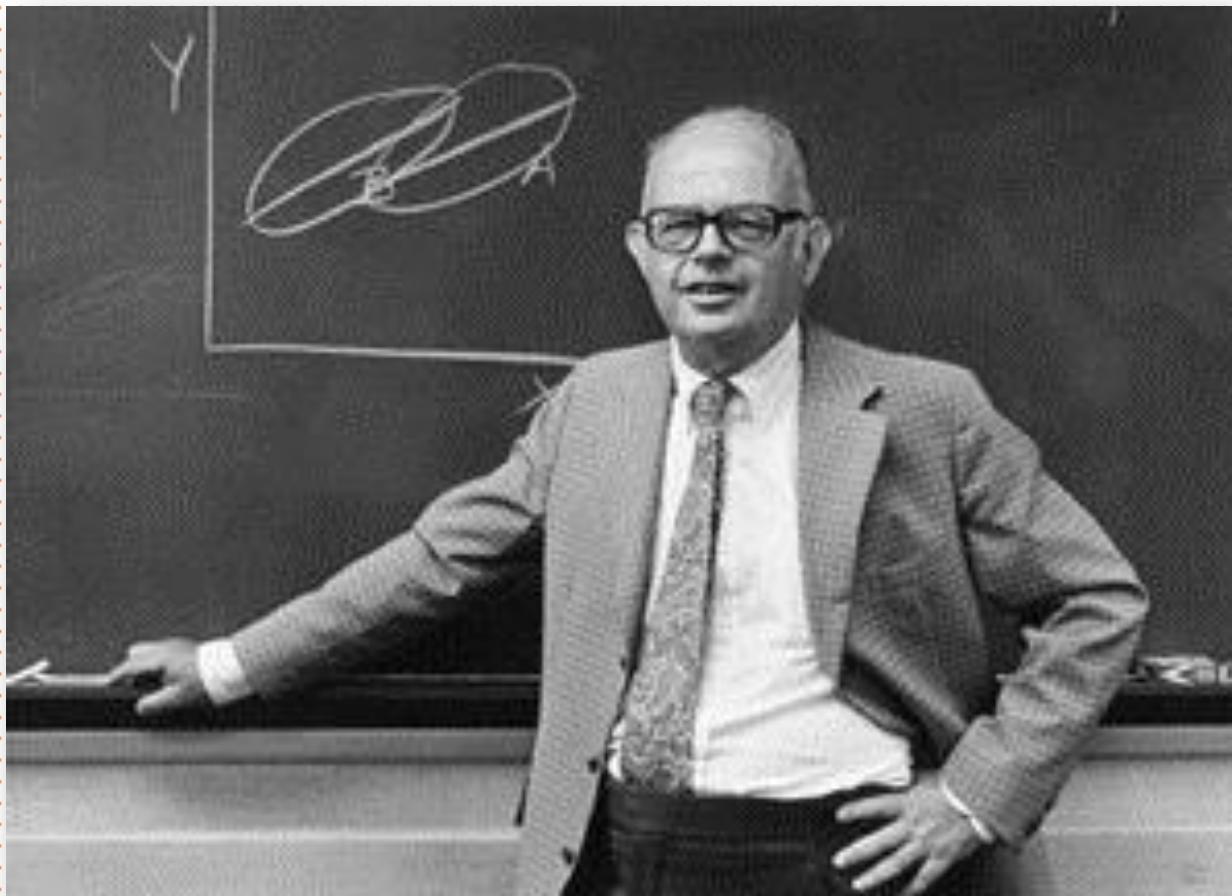
Korelace dvou polovin položek korigovaná na délku testu.

- Typicky se použijí liché a sudé položky či náhodný výběr.

Korekce se nazývá „Spearmanův-Brownův věštecký vzorec“.

- $$r_{xx'} = \frac{mr'_{xx'}}{1+(m-1)r'_{xx'}} = \frac{2r'_{xx'}}{1+r'_{xx'}}$$
- m je poměr celého a zkráceného testu (v tomto případě tedy 2) a $r'_{xx'}$ je nekorigovaná korelace půlek testů.
- Korekce je nutná, protože nás nezajímá reliabilita půlek testu, ale celého, který je delší, a má tedy vyšší reliability než každá z půlek sama o sobě.

Split-half vychází zpravidla vyšší než Cronbachovo alfa a reliabilitu mírně nadhodnocuje (nesplněné předpoklady věštění).



Praktický workshop

Postoje k uprchlíkům

Základní fasety:

Očekávané faktory postojů k uprchlíkům a uprchlické krizi.

- Obavy z uprchlíků, strach z krize.
- Soucit s uprchlíky, obecné pocity sounáležitosti.
- Postoj k pomoci uprchlíkům.
- Důvěra ve schopnost vlády vyřešit situaci (obecně, bez ohledu na valenci ostatních faktorů)
- Informovanost a vyhledávání informací o situaci (bez ohledu na povahu zdroje).

Nějaké další? Jsou některé navíc?

- Co např. vztah k EU?

Zadání:

- Rozdělte se do skupin a vyberte si jeden z faktorů.
- Proveďte dekompozici obsahového univerza, vytvořte fasetový model.
- Zvolte formát položek.
- Vytvořte položkovou banku a vyberte set adekvátního množství položek.

Bude následovat:

- Vygenerování dat jedné (či několika) vybraných dimenzí.
- Položková analýza (Excel).

Závěrem

Prezentace a Excel bude k dispozici na adrese:

<http://is.muni.cz/www/175803/pedf2016/>

Doporučená literatura:

- Urbánek, T., Denglerová, D., & Širůček, J. (2011). *Psychometrika: Měření v psychologii*. Praha: Portál.
- Jakékoliv učebnice o didaktických testech (některé i v knihovně na PedF).
- Literatura k předmětu zkoumání 😊

Kontakty, dotazy, konzultace:

- hynek.cigler@mail.muni.cz

Díky za pozornost!