MUNI FSS Katedra psychologie

Přednáška 7: Model klasické testové teorie

5. 11. 2024 | PSYn4790 | **Psychometrika: Měření v psychologii**

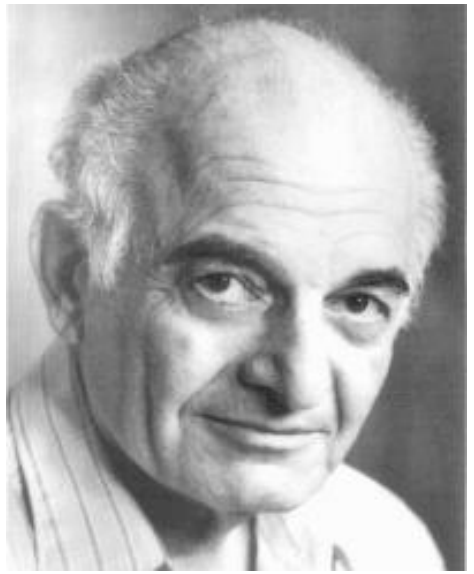
Katedra psychologie, Fakulta sociálních studií MU

Hynek Cígler









Když je psychologie mizerně placená,
musíš se stát šéfem kartelu.

caption credit: Petra Hubatka

Klasická testová teorie (CTT): Historie

Reliabilita a kontrola chyby měření.

- Spearman, Brown, Cronbach, Rulon, Guttman, Lord a Novick…

Epistemologická východiska.

- Ferguson, Campbell, Stevens, Borsboom...

Způsob tvorby skóru.

- Thurstone, Likert, Guttman…

CTT: Počátky

Předchůdci:

- Friedrich Wilhelm Bessel (1823): rozdíly mezi pozorovateli.
- Adolphe Quetelet (1842): Koncept *průměrného člověka*.
- Francis Galton (1869, 1888): Koncept *korelace*.
- Karl Pearson (1986): Výpočet korelace.

Spearman (1904): koncept reliability

- Koeficient proti oslabení korelace (attenuation formula).

$$r_{pq}^{*} = \frac{r_{pq}}{\sqrt{r_{pp'} r_{qq'}}}$$

Spearman (1910), Brown (1910): Spearman-Brown prophecy formula.

$$r_{xx'}^{*} = \frac{N r_{xx'}}{1 + (N-1) r_{xx'}}$$

CTT: Kontrola chyb a formalizace

Raný vývoj do 50. let.

- Kuder-Richardson (1937): postupy pro odhad vnitřní konzistence.
- Rulon (1939): split-half reliabilita tau-ekvivalentních testů (koeficient alfa)
- Guttman (1945): spodní hranice reliability.
- Cronbach (1951): koeficient alfa

Formalizace: Statistical Theories of Mental Test Scores.

- Lord a Novick (1968).

Zobecnění: Teorie zobecnitelnosti.

- Cronbach, Rajaratman, Gleser (1963).

Propojení s teorií latentních rysů (faktorovou analýzou) skrz kongenerický odhad reliability.

- McDonald (1970) a další.

CTT: Tvorba skóru

Intuitivní testová teorie (součet bodů; Braun a Mislevy, 2005).

Škálovací postupy.

- Hayes a Peterson (1921), Bogardus (1926), Thurstone (1928), Likert (1932), Osgood (1957).
- Guttman (40. léta)

Problematika sčítání bodů v CTT.

- CTT jako taková neobsahuje sčítání bodů.
- Většina CTT postupů však ano (vnitřní konzistence).

CTT: Tvorba skóru (součtové skóry)

Bacha na součtové skóry (ale vždyť jsou v pohodě):

- S01E01: McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0
- S01E02: Widaman, K. F., & Revelle, W. (2023). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *55*(2), 788–806. https://doi.org/10.3758/s13428-022-01849-w
- S01E03: McNeish, D. (2022). Psychometric properties of sum scores and factor scores differ even when their correlation is 0.98: A response to Widaman and Revelle. *Behavior Research Methods*, *55*(8), 4269–4290. https://doi.org/10.3758/s13428-022-02016-x
- S01E04: Widaman, K. F., & Revelle, W. (2024). Thinking About Sum Scores Yet Again, Maybe the Last Time, We Don't Know, Oh No…[1]: A Comment on McNeish (2023). *Educational and Psychological Measurement*, *84*(4), 637–659. https://doi.org/10.1177/00131644231205310

Série 2: Ale ne, součtové skóry jsou cajk (možná ne tak úplně?):

- S02E01: Sijtsma, K., Ellis, J. L., & Borsboom, D. (2024a). Recognize the Value of the Sum Score, Psychometrics' Greatest Accomplishment. *Psychometrika*, *89*(1), 84–117. https://doi.org/10.1007/s11336-024-09964-7
- S02E02: McNeish, D. (2024). Practical Implications of Sum Scores Being Psychometrics' Greatest Accomplishment. *Psychometrika*. https://doi.org/10.1007/s11336-024-09988-z
- S02E03: Mislevy, R. J. (2024). Are Sum Scores a Great Accomplishment of Psychometrics or Intuitive Test Theory? *Psychometrika*. https://doi.org/10.1007/s11336-024-10003-8
- S02E04: Sijtsma, K., Ellis, J. L., & Borsboom, D. (2024b). Rejoinder to McNeish and Mislevy: What Does Psychological Measurement Require? *Psychometrika*. https://doi.org/10.1007/s11336-024-10004-7

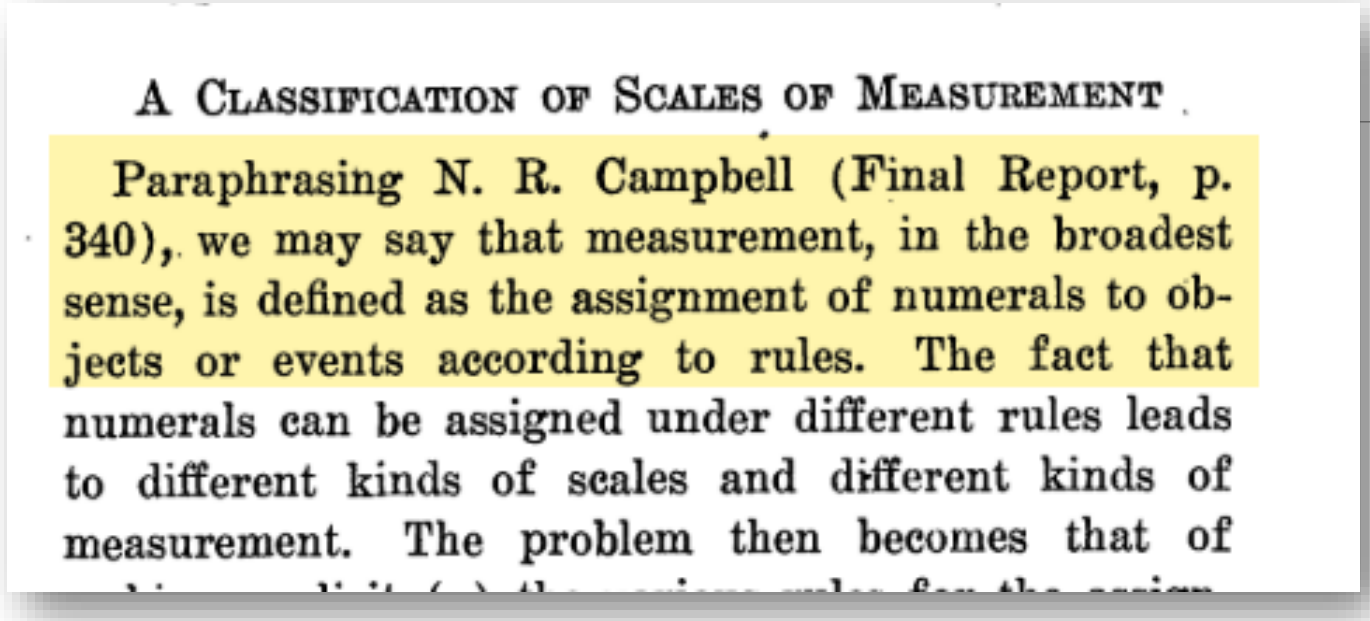
CTT: Epistemologie

Důležitým impulzem byla Fergusonova komise (1932– 1940).

- Striktní požadavek aditivity (a zřetězení).
- Psychologové zřetězení nedokázali → **měření v psychologii není vědecké.**
  - Což ale neznamená, že to není geniální nápad! ☺
- Reakcí byla Stevensova **operační teorie měření**, která rozšířila definici měření: *„...measurement, in the broadest sense, is defined as the **assignment of numerals to objects and events according to rules.**“* (Stevens, 1946, s. 677). Klíčový pojem je „**matching**“.
  - Ve skutečnosti zjednodušení konsenzu z přírodních věd: „Measurement is a method of *assigning numbers to magnitudes*“ (např. Helmholtz, 1887).
  - Klasické měření: Existuje magnituda, kterou kvantifikujeme pomocí měřicího nástroje (realismus).
  - CTT: Magnitudu „vytváříme“ s pomocí pravidla bez ohledu na povahu jevu (operacionalismus).

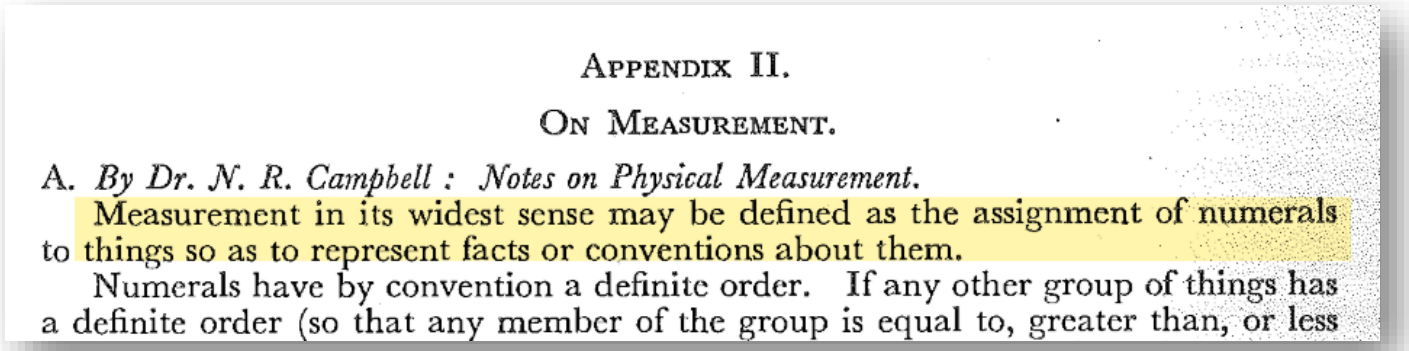
von Helmholtz, H. (1887/1971). An Epistemological Analysis of Counting and Measurement. In R. Kahl (ed.), *Selected Writings of Hermann von Helmholtz*. Wesleyan University Press.
Pro interpretaci doporučuji: Michell, J. (1993). The origins of the representational theory of measurement: Helmholtz, Hölder, and Russell. *Studies in History and Philosophy of Science Part A, 24*(2), 185–206. doi:10.1016/0039-3681(93)90045-I

Odbočka: „assignment of numerals“



A Classification of Scales of Measurement

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The problem then becomes that of



Appendix II.

On Measurement.

A. *By Dr. N. R. Campbell : Notes on Physical Measurement.*

Measurement in its widest sense may be defined as the assignment of numerals to things so as to represent facts or conventions about them.

Numerals have by convention a definite order. If any other group of things has a definite order (so that any member of the group is equal to, greater than, or less

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, *103*(2684), 677–680.

Ferguson, A., Myers, C. S., Bartlett, R. J., Banister, H., Bartlett, F. C., Brown, W., ..., & Tucker, W. S. (1940). Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *Report of the British Association for the Advancement of Science*, *2*, 331–349.

Rozdělení CTT a reprezentačního modelu

Fergusonova komise měla za následek rozdělení měření v sociálních vědách do dvou směrů.

1. Klasická testová teorie (CTT)

- Stevens (1946), Lord a Novick (1968)
- „Měření je přiřazování čísel jevům podle pravidel.“ Typicky: sečteme/zprůměrujeme body/položky.
- Nezabývá se algebraickou strukturou škály, aditivitou.
- Zaměření výhradně na celkové skóre (položky jsou jen „cestou k němu“).

2. Reprezentační model měření.

- A zejména **teorie spojitého měření** (CM; Debreu, 1960; Luce & Tukey, 1964).
- Pomocí aditivních operací vytváří algebraickou strukturu z nealgebraických dat.
  - Jinými slovy: dokáže vytvořit spojitou „míru“ v případě, že pozorujeme pouze seřazená data.
- Data musí odpovídat modelu. Využití i realistickými teoriemi (Raschův model).
  - Existuje-li latentní proměnná, která se manifestuje určitým způsobem, Raschův model bude spojitým měřením a dobře popíše data.
  - Popsal-li Raschův model dobře data, latentní proměnná může, ale nemusí existovat. Aby šlo o CM, je nutné splnit další podmínky.
  - Nepopsal-li Raschův model dobře data, latentní proměnná může, ale nemusí existovat, nicméně nepůjde o CM.

Od 70. let

CTT plně formalizovaná, malé vylepšování postupů.

CTT je zaměřená na celkové skóre, inference na úrovni položky nefunguje → impulz pro IRT.

Syntéza faktorové analýzy a CTT.

- Zaměření na kongenerické testy.
- CTT je matematicky specifickým případem FA.

Rozvoj (konfirmační) faktorové analýzy a modelů latentních proměnných (Jöreskeg).

Od r. 2000 elaborace epistemologických důsledků; „mumifikace a vykopávky“.

- Řada původních zdůvodnění zapadla, nově objevovány, přilepovány nová vysvětlení.
- Borsboom (2005).
- Kritika důsledků CTT (operacionální definice).

CTT vznikla na základech přírodních věd

Přírodní vědy: existující atribut opakovaně měříme tím samým měřicím nástrojem.

- Očekávaná hodnota atributu je průměr pozorování: $\mathrm{E}(x) = \frac{\sum_{i=1}^{N} x_i}{N}$
- Chyba měření je chybou odhadu průměru: $SE = \frac{s_x}{\sqrt{N}}$ (pokud s_x odhadujeme z dat, pak má odhad Studentova t-rozložení).

Předpoklad: Rozložení odhadu je přibližně normálně rozložené (centrální limitní teorém – potřebujeme tedy alespoň cca 30 pozorování).

V psychologii nemůžeme měřit tolikrát.

Z toho důvodu CTT zavádí **koncept paralelních testů** s několika přiměřenými předpoklady.

- CTT neřeší, jak tyto paralelní testy vznikly (tradičně ale součet položek). Problém škálování.

Ústředním konceptem v CTT je potom **reliabilita**.

Reliabilita v CTT

CTT je lineární model – všechny vztahy atributu, naměřených hodnot a chyby jsou lineární.

Základní teorém CTT je tedy lineární funkce:

$$X = \tau + e$$

Protože $\mathrm{cor}(X, e) = 0$, tak platí $\sigma_x^2 = \sigma_\tau^2 + \sigma_e^2$

- $\sigma_{a+b}^2 = \sigma_a^2 + \sigma_b^2 + 2 \cdot \mathrm{cov}(a, b)$. Pomůcka: $(a + b)^2 = a^2 + b^2 + 2ab.$

Reliabilita je pak definovaná jako rozptyl měření vysvětlený pravým skórem:

$$r_{xx'} = \frac{\sigma_\tau^2}{\sigma_x^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_e^2} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$$

Lze jednoduše dokázat, že reliabilita je zároveň korelací dvou paralelních testů:

$$r_{xx'} = \mathrm{cor}(x, x')$$

- To proto, že $\mathrm{cov}(x, x') = \sigma_\tau^2 \rightarrow \mathrm{cor}(x, x') = \frac{\mathrm{cov}(x,x')}{\sigma_x \sigma_{x'}} = \frac{\mathrm{cov}(x,x')}{\sigma_x \sigma_x} = \frac{\sigma_\tau^2}{\sigma_x^2} = r_{xx'}.$

Paralelní testy

„Dobré“ měření je takové, kdy různí lidé v různých časech dojdou různými nástroji ke stejným naměřeným hodnotám, pokud se míra samotného objektu nezměnila.

Paralelní testy/měření jsou takové, pro které platí:

- A. Pravý skór je v paralelních testech a pro každý měřený subjekt stejný
  - $T = \mathrm{E}(X) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n}$; důsledkem je shodný rozptyl pravých skórů.
- B. Chybový rozptyl je v paralelních testech a pro každý subjekt stejný.
  - Důsledkem je navíc shodný rozptyl pozorovaných skórů obou testů.

Korelace paralelních testů je pak reliabilita: $r_{xx'} = \mathrm{cor}(x, x')$.

Paralelní testy

Potíž v sociálních vědách je ale ten, že **paralelní testy neexistují**.

- Jde jen o hypotetický koncept (model).

Položky se liší...

- ... svou obtížností,
- ... těsností vztahu s univerzem,
- ... mírou náhodné chyby...

... a respondenti se rovněž napříč měřeními vyvíjejí.

Proto uvažujeme spíše o „míře paralelnosti“.

CTT: Paralelní testy

Úrovně paralelnosti položek (založené na faktorové analýze):

$$X_{ip} = i_i + a_i\tau_p + e_{ip}, \qquad e_{ip} \sim \mathrm{N}\big(0, \mathrm{var}(e_i)\big)$$

Kongenerické: Vybrané ze stejné domény. Stejná struktura rovnice pro všechny položky.

- Měří stejný rys (trs rysů), ale jiným způsobem.

Tau-ekvivalentní: Stejná lineární souvislost s měřeným atributem.

- + Shodné nestandardizované faktorové náboje („měřítko“ položky).

Paralelní: Položky měří se stejnou velikostí chyby.

- + Shodné reziduální rozptyly.

Striktně paralelní: Stejná obtížnost všech položek.

- + Shodné intercepty/průměry položek.
- U binárních položek paralelní = striktně paralelní, protože $\mathrm{var}(X_i) = P_i(1 - P_i)$.

- X_{ip} – pozorované skóre osoby p na pol. i
- i_i, a_i – intercept a faktorový náboj pol. i
- τ_p – pravé skóre osoby p
- e_{ip} – náhodná chyba osoby p na pol. i (reziduum)
- $e_{ip} \sim \mathrm{N}\big(0, \mathrm{var}(e_i)\big)$ – tato chyba pochází z normálního rozložení s průměrem 0 a rozptylem $\mathrm{var}(e_i)$

Někdy též kongenerické → esenciálně tau-ekvivalentní → tau-ekvivalentní → paralelní

CTT: Paralelní testy

Úrovně paralelnosti položek (založené na faktorové analýze):

$$X_{ip} = i_i + a_i\tau_p + e_{ip}, \qquad e_{ip} \sim \mathrm{N}\big(0, \mathrm{var}(e_i)\big)$$

Kongenerické: Vybrané ze stejné domény. Stejná struktura rovnice pro všechny položky.

- Měří stejný rys (trs rysů), ale jiným způsobem.

Tau-ekvivalentní: Stejná lineární souvislost s měřeným atributem.

- + Shodné nestandardizované faktorové náboje („měřítko“ položky).

Paralelní: Položky měří se stejnou velikostí chyby.

- + Shodné reziduální rozptyly.

Striktně paralelní: Stejná obtížnost všech položek.

- + Shodné intercepty/průměry položek.
- U binárních položek paralelní = striktně paralelní, protože $\mathrm{var}(X_i) = P_i(1 - P_i)$.

Někdy též kongenerické → esenciálně tau-ekvivalentní → tau-ekvivalentní → paralelní

CTT: Paralelní testy

Úrovně paralelnosti položek (založené na faktorové analýze):

$$X_{ip} = i_i + a\tau_p + e_{ip}, \qquad e_{ip} \sim \mathrm{N}\left(0, \mathrm{var}(e_i)\right)$$

Kongenerické: Vybrané ze stejné domény. Stejná struktura rovnice pro všechny položky.

- Měří stejný rys (trs rysů), ale jiným způsobem.

Tau-ekvivalentní: Stejná lineární souvislost s měřeným atributem. $a_i = a$

- + Shodné nestandardizované faktorové náboje („měřítko“ položky).

Paralelní: Položky měří se stejnou velikostí chyby.

- + Shodné reziduální rozptyly.

Striktně paralelní: Stejná obtížnost všech položek.

- + Shodné intercepty/průměry položek.
- U binárních položek paralelní = striktně paralelní, protože $\mathrm{var}(X_i) = P_i(1 - P_i)$.

Někdy též kongenerické → esenciálně tau-ekvivalentní → tau-ekvivalentní → paralelní

CTT: Paralelní testy

Úrovně paralelnosti položek (založené na faktorové analýze):

$$X_{ip} = i_i + a\tau_p + e_{ip}, \qquad e_{ip} \sim \mathrm{N}\left(0, \mathrm{var}(e)\right)$$

Kongenerické: Vybrané ze stejné domény. Stejná struktura rovnice pro všechny položky.

- Měří stejný rys (trs rysů), ale jiným způsobem.

Tau-ekvivalentní: Stejná lineární souvislost s měřeným atributem.

- + Shodné nestandardizované faktorové náboje („měřítko“ položky).

Paralelní: Položky měří se stejnou velikostí chyby. $a_i = a, \mathrm{var}\left(e_{ip}\right) = \mathrm{var}(e)$

- + Shodné reziduální rozptyly.

Striktně paralelní: Stejná obtížnost všech položek.

- + Shodné intercepty/průměry položek.
- U binárních položek paralelní = striktně paralelní, protože $\mathrm{var}(X_i) = P_i(1 - P_i)$.

Někdy též kongenerické → esenciálně tau-ekvivalentní → tau-ekvivalentní → paralelní

CTT: Paralelní testy

Úrovně paralelnosti položek (založené na faktorové analýze):

$$X_{ip} = i + a\tau_p + e_{ip}, \qquad e_{ip} \sim \mathrm{N}\left(0, \mathrm{var}(e)\right)$$

Kongenerické: Vybrané ze stejné domény. Stejná struktura rovnice pro všechny položky.

- Měří stejný rys (trs rysů), ale jiným způsobem.

Tau-ekvivalentní: Stejná lineární souvislost s měřeným atributem.

- + Shodné nestandardizované faktorové náboje („měřítko“ položky).

Paralelní: Položky měří se stejnou velikostí chyby.

- + Shodné reziduální rozptyly.

Striktně paralelní: Stejná obtížnost všech položek. $a_i = a, \mathrm{var}\left(e_{ip}\right) = \mathrm{var}(e), i_i = i$

- + Shodné intercepty/průměry položek.
- U binárních položek paralelní = striktně paralelní, protože $\mathrm{var}(X_i) = P_i(1 - P_i)$.

Někdy též kongenerické → esenciálně tau-ekvivalentní → tau-ekvivalentní → paralelní

Paralelní testy: terminologie

původní (Lord a Novick, 1968)	alternativní (např. Cho, 2016)	náboj	chybový rozptyl	intercept
kongenerické	kongenerické	✗	✗	✗
esenciálně tau-ekvivalentní	tau-ekvivalentní	✓	✗	✗
— (později tau-ekvivalentní)	paralelní	✓	✓	✗
paralelní	striktně-paralelní	✓	✓	✓

(Dvakrát) dvě pojetí reliability

Stabilita měření (operacionalismus).

- Bez ohledu na to, jaký je „význam“ měření.
- CTT.

Vysvětlený rozptyl (realismus).

- Vysvětlený rozptyl čím?
- Co považujeme za pravé skóre?
- Může sloužit jako estimátor korelace paralelních testů, nebo mít svůj význam.
- Někdy tzv. faCTT.

→ Klasická testová teorie.

- Dnešní přednáška.

Relativní srovnání

- CTT, GT.
- Na obtížnosti položek nám nezáleží.

Absolutní srovnání.

- GT, shoda posuzovatelů v CTT.
- Položky jsou vybrané z univerza všech pol.
- Záleží, zda máme snadné či těžké položky.

→ Teorie zobecnitelnosti.

- Příští přednáška.

Dvě pojetí reliability v CTT

1. Dimension-free reliability (důraz na korelaci paralelních testů). Operacionalismus.

- Odhad vztahu (korelace) dvou paralelních měření týmž testem bez ohledu na to, co test měří.
- split-half, alfa, celková omega, *glb*

2. Model-based reliability (důraz na vysvětlený rozptyl). Realismus.

- Odhad vztahu (vysvětleného rozptylu) měřeného atributu a pozorovaného skóru.
- Rodina koeficientů omega (McDonaldova hierarchická omega).
  - „Realistická invaze do operacionalistické CTT“ ☺

Podrobně viz:

- Bentler P. M. (2009). Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability. *Psychometrika*, *74*(1), 137–143. doi:10.1007/s11336-008-9100-1
- Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods*, *19*(4), 651–682. doi:10.1177/1094428116656239

Dvě pojetí reliability v CTT

Reliabilita jako korelace:

$$r_{xx'} = \mathrm{cor}(x, x')$$

Ryzí logický pozitivismus.

Reliabilita vyjadřuje **stabilitu** odhadu pravého skóre nehledě na to, co toto pravé skóre je nebo jak vzniklo.

Reliabilita jako vysvětlený rozptyl:

$$r_{xx'} = \frac{\sigma_\tau^2}{\sigma_x^2}$$

Toto už není čistá CTT – ušpiněno modelem s latentními proměnnými.

Něco vysvětlený rozptyl muselo „způsobit“.

- Jak velkou roli to „něco“ má na vznik pozorovaného skóre?

Za dodržení smysluplných podmínek jsou obě otázky ekvivalentní.

- Lokální nezávislost (jednodimenzionalita) položek/konstruktu.
- Unikátní rozptyl je plně nesystematický.

Systematický přístup k reliabilitě

Table 3. Names of Reliability Coefficients Currently Used in the Literature.

	Unidimensional		Multidimensional
	Split-Half	General	General
Parallel	Spearman–Brown formula	Standardized alpha	(Not yet published)
Tau-equivalent	Flanagan–Rulon formula Flanagan formula Rulon formula Guttman's λ_4	Cronbach's alpha Coefficient alpha Guttman's λ_3 Hoyt method KR-20	Stratified alpha
Congeneric	Raju (1970) coefficient Angoff–Feldt coefficient Angoff coefficient	Composite reliability Construct reliability Congeneric reliability Omega Unidimensional omega Raju (1977) coefficient Classical congeneric reliability coefficient	Omega Omega total McDonald's omega Multidimensional omega

Cho (2016)

Systematický přístup k reliabilitě

Table 4. Names and Notations of Reliability Coefficients Suggested in This Study.

	Unidimensional		Multidimensional
	Split-Half	General	General
Parallel	Split-half parallel reliability (ρ_{SP})	Parallel reliability (ρ_P)	Multidimensional parallel reliability (ρ_{MP})
Tau-equivalent	Split-half tau-equivalent reliability (ρ_{ST})	Tau-equivalent reliability (ρ_T)	Multidimensional tau-equivalent reliability (ρ_{MT})
Congeneric	Split-half congeneric reliability (ρ_{SC})	Congeneric reliability (ρ_C)	Bifactor model Bifactor reliability (ρ_{BF}) Second-order factor model Second-order factor reliability (ρ_{SOF}) Correlated factors model Correlated factors reliability (ρ_{CF})

Cho (2016)

Ad hoc notace v tomto příkladu: Pozorovaná rychlost v konkrétního člověka se od jeho „pravé rychlosti“ $\mathrm{E}(v)$ liší o e_v. Směrodatná odchylka e_v je standardní chybou měření rychlosti, σ_v.

Odbočka: Kde se bere chyba měření?

Měříte průměrnou rychlost běhu na **1 km** s výsledkem **4 minuty** (tedy rychlostí $15\ {}^{km}/_{h} = 4{,}17\ {}^{m}/_{s}$. Jaký je postup výpočtu chyby měření σ_v této rychlosti v?

Postup:

- Vzdálenost jsme měřili s chybou e_d a čas s chybou e_t. Protože $v = \frac{s}{t}$, platí $e_v = \frac{e_s}{e_t}$.
- Výsledné rozložení není definované, ale bude přibližně normální a $\sigma_v = \frac{s}{t}\sqrt{\left(\frac{\sigma_s^2}{s^2} + \frac{\sigma_t^2}{t^2}\right)}$ – čím menší chyby v poměru k naměřeným hodnotám, tím menší chyba.
- Pokud měříme s malou chybou, řekněme $\sigma_d = 0{,}5m$ a $\sigma_t = 0{,}1s$, pak $\boldsymbol{\sigma_v = 0{,}0027\ {}^{m}/_{s}}$.

Řekněme, že „měříte“ podíl správných odpovědí na 30 položek v testu. „Naměříte“ 20/30.

- Zdánlivě těch 30 i 20 máte „změřené“ přesně.
- **V čem se obě situace liší?**

Je rozdíl v průměrné rychlosti jednoho pokusu a běžné (průměrné) průměrné rychlosti (tedy schopnosti běhat) kilometrové tratě.

Odbočka: Kde se bere chyba měření?

V psychologii nás nezajímá chyba měření konkrétního výkonu...

- (Ta je zcela zanedbatelná.)

... ale ***průměrná hodnota výkonu*** napříč paralelními situacemi (CTT) nebo ***úroveň latentní schopnosti*** (realismus), která výkon způsobuje.

- Ta je výrazně větší, protože pozorovaný výkon napříč situacemi značně kolísá.

Nechceme vědět, jak rychle člověk běžel, ale jak rychle běhá.

- „běžel“ = X_{pi} (manifestní proměnná)
- „běhá“ = $\mathrm{E}(X_{pi}) = \tau_p$ (pravé skóre / latentní proměnná).

Odbočka: Kde se bere chyba měření?

Reliabilita příkladu s během tedy bude $r_{vv'} = \mathrm{cor}(v, v') = 1 - \frac{\sigma_e^2}{\sigma_v^2}$.

- Čím více bude kolísat rychlost běhání napříč pokusy (σ_e^2), tím nižší reliabilita bude.
- Čím více se lidé liší ve schopnosti běhat a tedy i pozorované výkony (σ_v^2), tím vyšší reliabilita bude.
- Pozn.: V tomto CTT designu neodlišíme zdroje chyby (kolísání výkonu, nepřesnost měření délky a času). To bude řešit teorie zobecnitelnosti (GT).

Úkol: Za jakých okolností se může lišit „model based“ a „dimension-free“ reliabilita?

Chceme zjistit schopnost rychle uběhnout 1 km. Na výkon mají vliv ale i další proměnné.

- Měříme třeba rychlost na 500 metrech.
- Kvalita obuvi, náročnost terénu.
- Motivace, výše odměny, únava z předchozího závodu...

Pro přesný odhad model-based chyby potřebujeme s těmito dalšími proměnnými **manipulovat**.

Reliabilita… jakého skóre?

CTT není závislá na způsobu vzniku pozorovaného skóre.

- Funguje stejně, jde-li o součet položek, naměřený čas, či cokoli jiného.

Způsob konstrukce skóre je zcela arbitrární.

Postupy založené na položkách (alfa, omega) a další „spodní hranice reliability“ ale při výpočtu předpokládají konkrétní způsob vzniku skóru.

- Typicky jde o odhady reliability „součtu položek“.
- Koeficienty lze ale snadno upravit do podoby reliability nějakého váženého součtu a podobně.

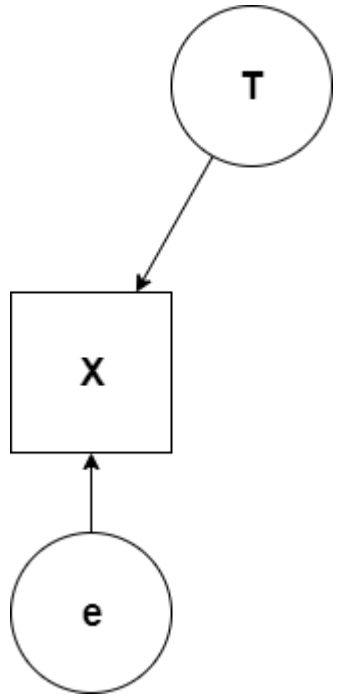
Ve většině přednášky předpokládáme, že skóre vzniká jako součet položek.

- **A tedy mluvíme o *„reliabilitě součtu položek“*.**

Spodní hranice reliability

Lower-bound of reliability.

Zpravidla předpokládáme, že unikátní rozptyl položek je chyba (e).



Spodní hranice reliability

Lower-bound of reliability.

Zpravidla předpokládáme, že unikátní rozptyl položek je chyba (e).

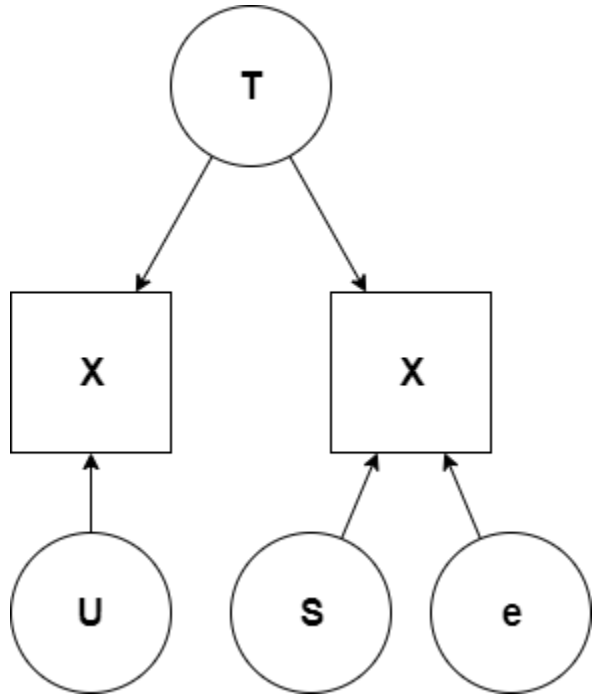
Unikátní rozptyl U ale lze rozdělit na:

- specifický S (systematický pro daného člověka)
- chybový e (náhodný)

Zatímco S přispívá ke korelaci paralelních testů, chyba e nikoli.

Tyto složky ale nelze oddělit při jediné administraci testu a S je považován celý za chybu.

- Proto v longitudinálních SEM modelech korelovaná rezidua v čase.



Spodní hranice reliability

Brněnský test domácích mazlíčků:

- Máte rádi psy? (ne = 0, ano = 1)
- Máte rádi kočky? (ne = 0, ano = 1).

Celkové skóre 0–2.

- Korelace položek r = 0,6.
- Obliba koček a psů je plně stabilní lidskou charakteristikou.
  - Buď je milujete, nebo nenávidíte.

Jaká bude reliabilita dotazníku?

Vnitřní konzistence:

$$r_{SB} = \frac{2 \cdot 0{,}6}{1 + 0{,}6} = 0{,}75$$

- Spodní hranice reliability.
- Rozptyl vysvětlený „láskou k mazlíčkům“.

Test-retest:

$$r_{xx'} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_e^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + 0} = 1$$

- Korelace paralelních testů.
- Na obě otázky odpovíte vždy stejně.

Koeficienty založené na paralelních testech

Split-half přístupy

Alfa

Split-half

Reliabilita jako stabilita.

Problémy se split-half:

- Nelze ověřit předpoklady paralelnosti.
- Test je zkrácený na polovinu.
- Existuje velké množství rozdělení testu na dvě poloviny.
  - Různá rozdělení → různé odhady.
    - Tohle byl jeden z Cronbachových motivů pro alfu (která je průměrem split-half reliabilit).

Split-half

SPEARMANŮV-BROWNŮV PŘÍSTUP

Spearmanův-Brownův věštecký vzorec:

- $$r^*_{xx'} = \frac{N r_{xx'}}{1 + (N-1) r_{xx'}}$$

- N – změna délky testu, v případě split-half N=2:

$$r^*_{xx'} = \frac{2 r_{xx'}}{1 + r_{xx'}}$$

Předpoklad: **paralelní poloviny**.

- Při nedodržení příliš „optimistický“, může nadhodnocovat nebo podhodnocovat.

GUTTMANOVA λ_4

Guttman (1945) publikoval $\lambda_{1–6}$:

$$\lambda_4 = \frac{4\sigma^2_{pq}}{\sigma^2_x}$$

- σ^2_{pq} – kovariance polovin testu
- $\sigma^2_x = \sigma^2_p + \sigma^2_q + 2\sigma^2_{pq}$ – rozptyl celého testu.

$\lambda_4 = \alpha$ (ve dvoupoložkovém testu)

- **tau-ekvivalentní poloviny** (jinak podhodnocuje)
- Proto se dnes λ_4 používá jako ρ_{glb} – split-half maximalizovaná pomocí nejlepšího možného rozdělení.
- „Příliš dobré rozdělení“→ na malých vzorcích nadhodnocuje.
- Rozdílná délka testů: hrubé podhodnocení.

Založeno na jediné korelaci → relativně nepřesný odhad reliability.

Split-half: Nestejné poloviny

Spearmanův-Brownův i Guttmanův přísup předpokládá stejně dlouhé poloviny testu.

Odvozeno z SB-vzorce (při stejné délce by poloviny byly paralelní):

- Horstova (1951)[1]: $r_H = \frac{r_{12}\sqrt{r_{12}^2+4\pi_1\pi_2(1-r_{12}^2)}-r_{12}^2}{2\pi_1\pi_2(1-r_{12}^2)}$, kde π_1 a π_2 jsou délky polovin testu.

Odvozeno z Guttmanovy λ_4 (při stejné délce by poloviny byly tau-ekvivalentní):

- Raju (1977): $\beta = \frac{\sigma_{12}}{\pi_1\pi_2\sigma_x^2}$
- Délku polovin lze odhadnout na základě jejich rozptylu jako $\pi_1 = \frac{\sigma_1^2+\sigma_{12}}{\sigma_x^2}$, $\pi_2 = \frac{\sigma_2^2+\sigma_{12}}{\sigma_x^2}$, což lze dosadit:
- **Angoffův-Feldtův koeficient** (1953, 1975): $r_{AF} = \frac{4\sigma_{12}}{\sigma_x^2-\frac{(\sigma_1^2-\sigma_2^2)^2}{\sigma_x^2}}$

Stručný přehled: Cígler, H., & Chvojka, E. (2022, February 16). Reliability estimation in tests composed of two items only: Admissible and Plausible reliability ranges. [Unpublished preprint]. https://doi.org/10.31234/osf.io/9w738

[1] Horst (1951) má chybu ve vzorci 2, pro korektbní vzorec viz např. Warrense (2016).

Položky jako paralelní testy
Cronbachovo alfa (Guttmanova λ_3)

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

- σ_i^2 – rozptyl položky *i*, $\sum_{i=1}^{k} \sigma_i^2$ je diagonála var-kovar matice (unikátní rozptyl položek = chyba)
- σ_x^2 – rozptyl celého testu, tedy suma var-kovar matice (sdílený rozptyl položek)
- k – počet položek (ne celý unikátní rozptyl je chybou, proto korekce $\frac{k}{k-1}$, aby reliabilita mohla být 1)
- V případě binárních položek je výsledek shodný s výpočetně jednodušším KR-20.

Předpoklady:

- Tau-ekvivalentní položky (při nedodržení je korekce $\frac{k}{k-1}$ nedostatečná → podhodnocení reliability).
- Jednodimenzionalita (nadhodnocení i podhodnocení dle typu).
- Alfa není ukazatelem jednodimenzionality (viz např. Marko, 2016).

Výhody: Přesný odhad (ve srovnání se split-half), jednoduchý/jednoznačný postup, tradice.

Varianty koeficientu alfa

Standardizované alfa.

- Pro výpočet použita korelační matice → reliabilita součtu standardizovaných položek.
- Použitelné v případě položek s rozdílnou odpověďovou škálou, tedy i pozorovaným rozptylem a výrazným narušením předpokladu tau-ekvivalence.

Ordinální alfa (Zumbo, Gadermann, Zeisser, 2007)

- Alfa spočítané nad maticí polychorických korelací.
- Zcela jiný význam, není použitelné pro běžnou praxi.
- Není srovnatelné s jinými odhady reliability (viz např. Chalmers, 2017).

Stratifikované Cronbachovo alfa

Nejjednodušší odhad reliability součtu subtestů – Cronbach (1965):

$$\alpha_{strat} = 1 - \frac{\sum_{i=1}^{k}\left[\omega_i^2 \sigma_i^2 (1 - r_{ii'})\right]}{\sigma_Z^2}$$

- ω_i „váha“ testu *i*
- σ_i^2 rozptyl testu *i*
- $r_{ii'}$ reliabilita testu *i*
- Pro výpočet stačí kovarianční matice a alfy subtestů.

Předpokladem je nejen tau-ekvivalence položek v testech, ale i tau-ekvivalence testů.

- A nekorelované chyby měření testů.

Např.: *„Jaká bude test-retest korelace celkového IQ skóre, pokud jsou obě měření paralelní?“*

Alpha: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha

Série článků po roce 2009, zejména:

- Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. Psychometrika, 74(1), 107–120. doi
- Bentler, P. M. (2009). Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability. *Psychometrika*, *74*(1), 137–143. doi
- Sijtsma, K., & Pfadt, J. M. (2021). Part II: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha: Discussing Lower Bounds and Correlated Errors. Psychometrika, 86(4), 843–860. https://doi.org/10.1007/s11336-021-09789-8
- V češtině: Marko, M. (2016). Využitie a zneužitie Cronbachovej alfy pri hodnotení psychodiagnostických nástrojov. *Testfórum*, *5*(7). https://doi.org/10.5817/TF2016-7-90

Ve stručnosti:

- Alfa není odhadem reliability. Alfa je spodní hranicí reliability. Výhody, nevýhody.
- Koeficient alfa má své užití, které ale není odhad reliability. „Kontrola kvality".
- Máme lepší estimátory reliability, které ale mohou nadhodnocovat.
- Koeficient alfa trpí, nebo naopak netrpí (a.) předpokladem nekorelovaných reziduí, (b.) jednodimenzionality, (c.) tau-ekvivalence.
- Potíže s epistemologickými východisky, různými cíli a způsoby využití.

Alpha: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha

Moje naprosto nezávazné doporučení:

Kdy ano:

- Ukazatel kvality diagnostických nástrojů.
- Odhad standardní chyby měření diagnostických nástrojů.
- Odhad reliability v případě dlouhých, přiměřeně normálně rozdělených testů s vícebodovými položkami.
- → **spodní hranice reliability**. Plní kontrolu kvality.

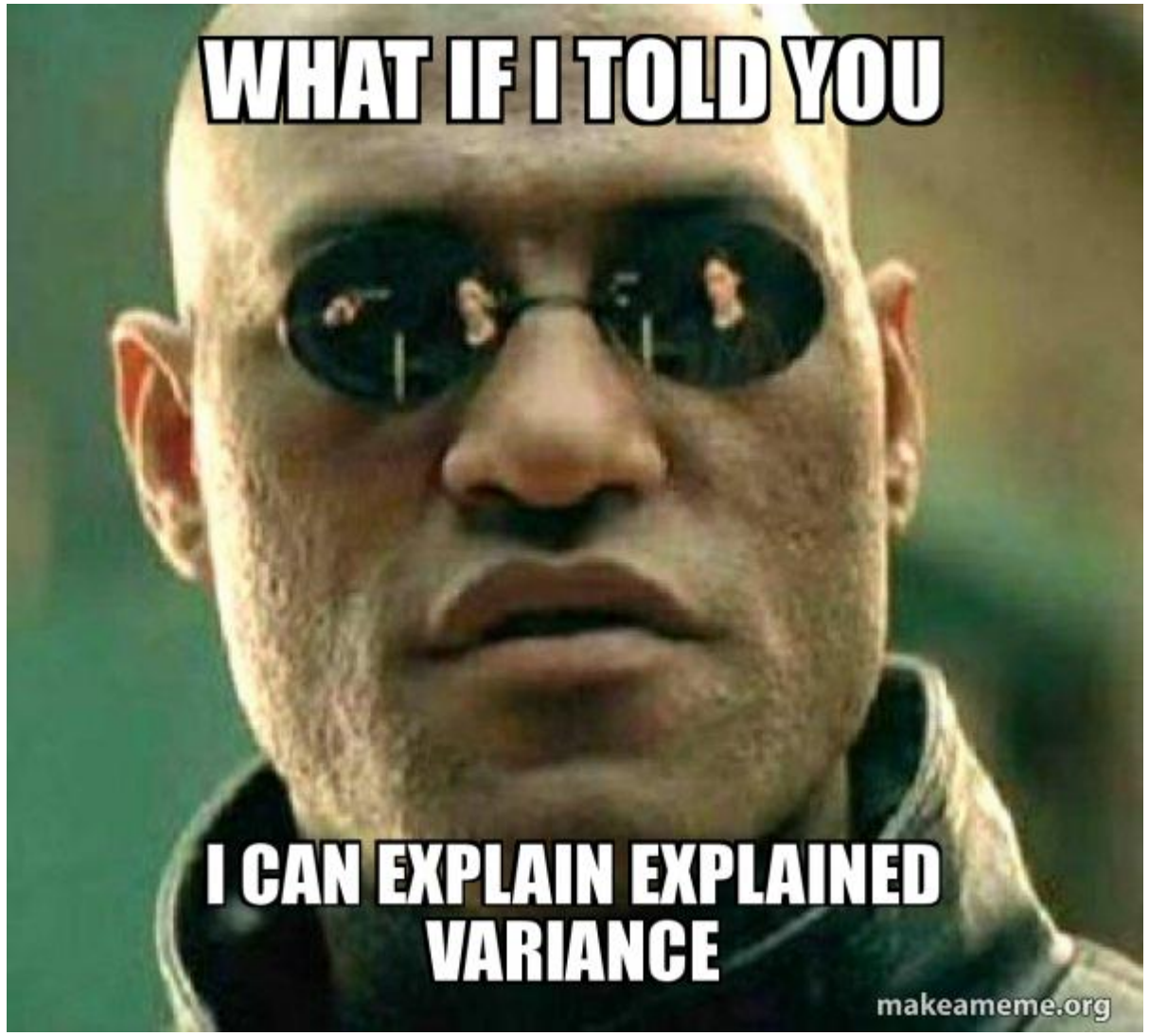
Kdy ne:

- Odhad reliability výzkumných nástrojů.
- Korekce na nereliabilitu.
- Odhad reliability kontrolních proměnných v multivariační regresi.
- Podhodnocení reliability → nadhodnocení případných korekcí.

Koeficienty založené na vysvětleném rozptylu

omega

FSD



Model-based reliabilita: **omega**

Rodina koeficientů; Betlerova, Raykovova, ... a zejm. **McDonaldova omega**.

Obecný vzorec (Bollen, 1980; Raykov, 2001):

$$\omega = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2 + \sum_{i=1}^{n} \sigma_{e;i}^2 + 2\sum_{i<j} \sigma_{ij}^2} = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2}{\sigma_x^2}$$

- λ_i = faktorový náboj položky i
- σ_ψ^2 = rozptyl faktoru, σ_x^2 = celkový pozorovaný rozptyl
- $\sigma_{e;i}^2$ = reziduální rozptyl položky i
- σ_{ij}^2 = kovariance položek i, j

Bez předpokladu tau-ekvivalence (rozdílné faktorové náboje jsou zohledněny).

Model-based reliabilita: **omega**

Rodina koeficientů; Betlerova, Raykovova, ... a zejm. **McDonaldova omega**.

Obecný vzorec (Bollen, 1980; Raykov, 2001):

$$\omega = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2 + \sum_{i=1}^{n} \sigma_{e;i}^2 + 2\sum_{i<j} \sigma_{ij}^2} = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2}{\sigma_x^2}$$

- λ_i = faktorový náboj položky i
- σ_ψ^2 = rozptyl faktoru, σ_x^2 = celkový pozorovaný rozptyl
- $\sigma_{e;i}^2$ = reziduální rozptyl položky i (náhodný chybový rozptyl)
- σ_{ij}^2 = kovariance položek i, j (systematický chybový rozptyl)

- vysvětlený rozptyl
- chybový rozptyl
- celkový rozptyl

Bez předpokladu tau-ekvivalence (rozdílné faktorové náboje jsou zohledněny).

Model-based reliabilita: **omega**

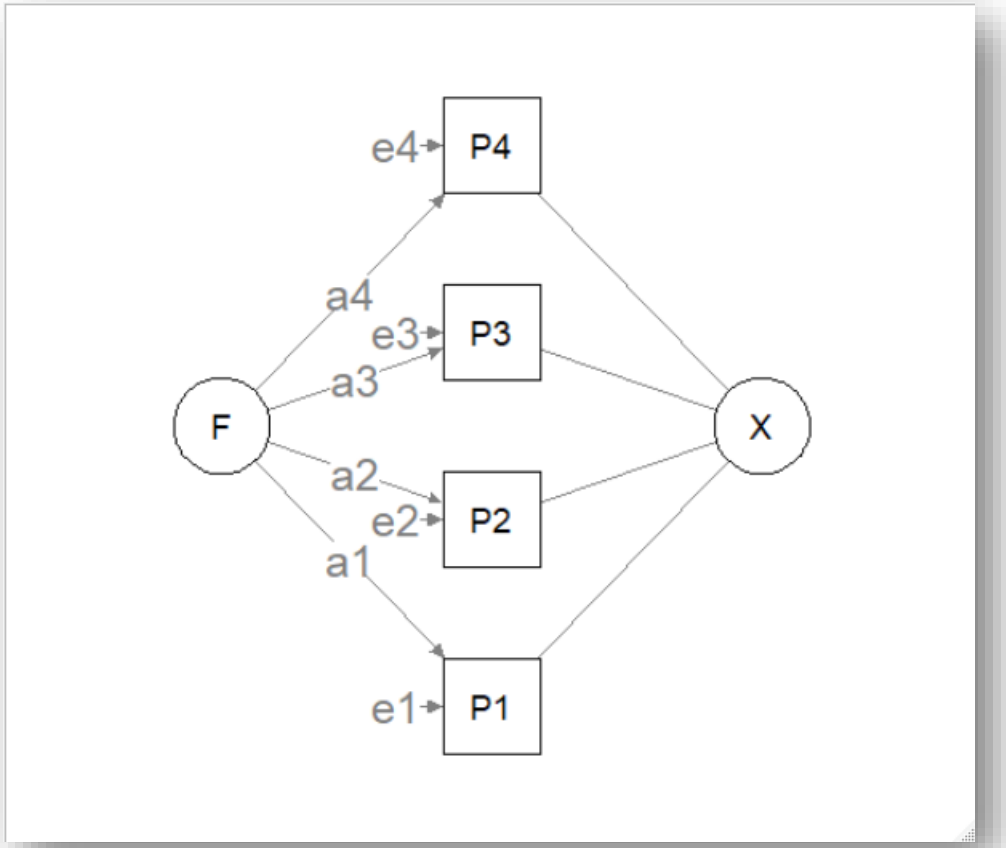
Použití koeficientu omega nás nutí zamyslet se, co je pravým skóre.

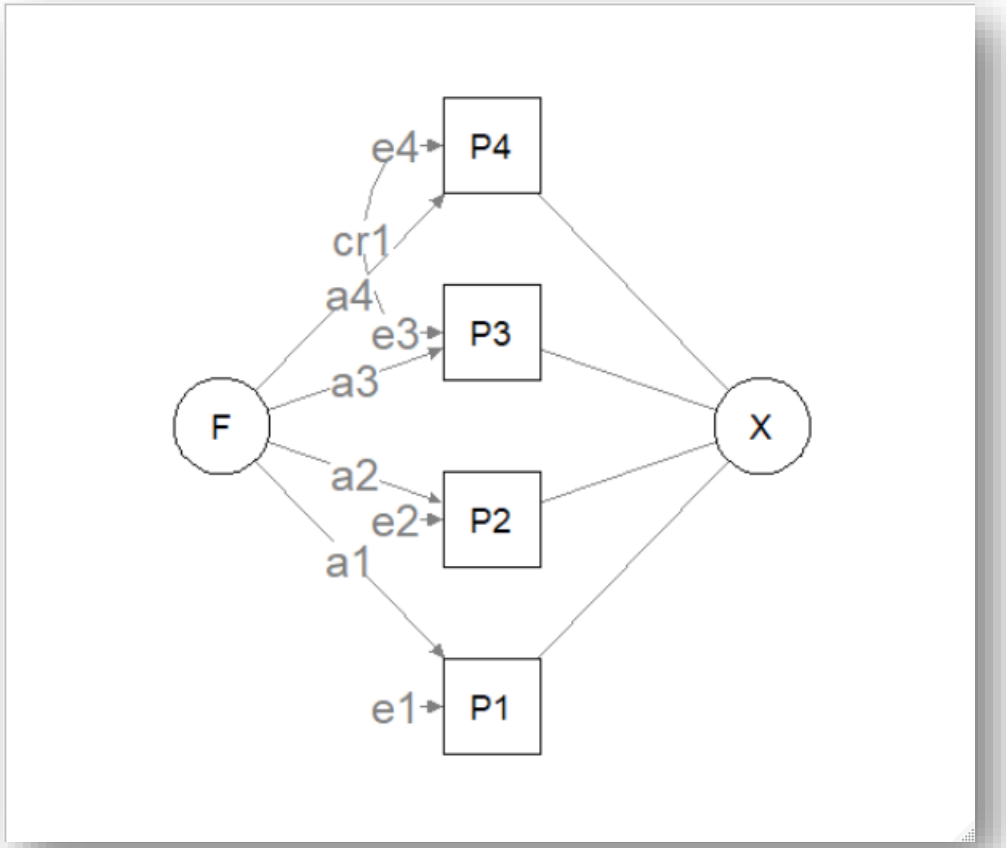
Co je to, co chceme měřit?

A také je nutné uvažovat nad modelem měření.

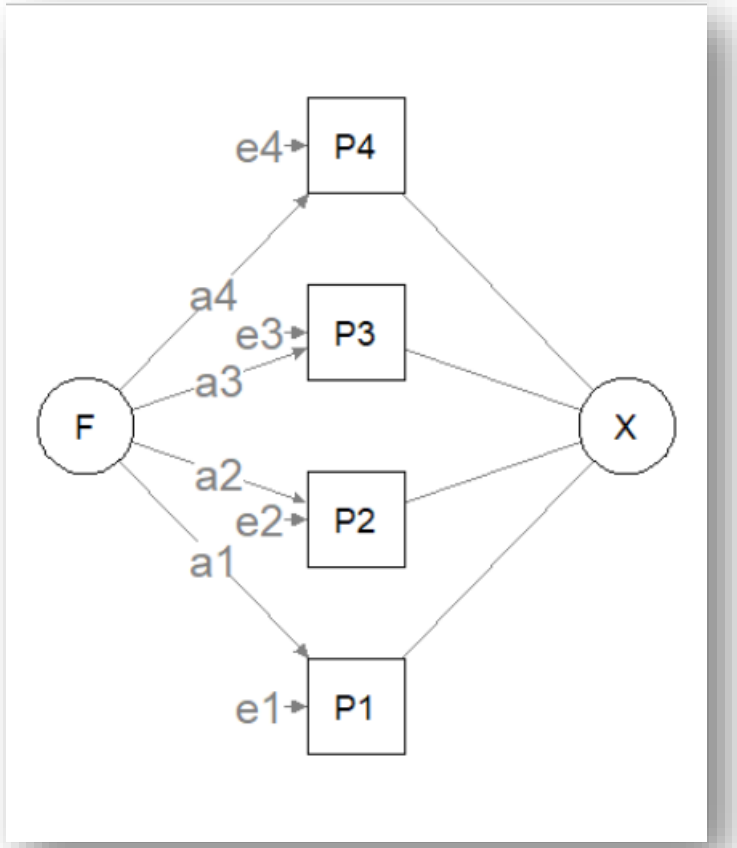
- V případě ordinální CFA Greenova-Yangova (2009) „prahová“ korekce.
  - Green, S. B., & Yang, Y. (2009). Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. Psychometrika, 74(1), 155–167. https://doi.org/10.1007/s11336-008-9099-3

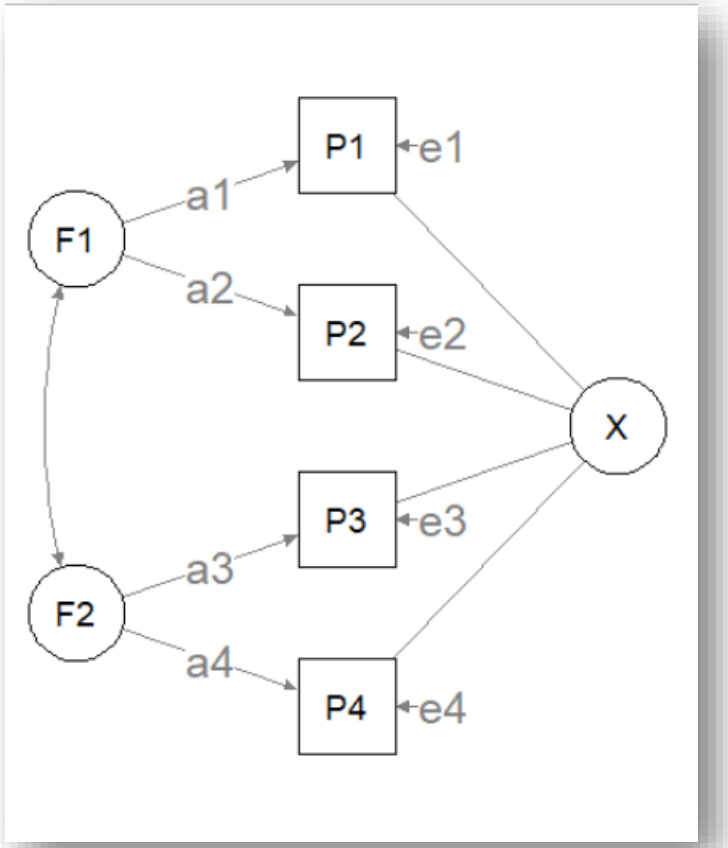
Omega: Multidimenzionalita

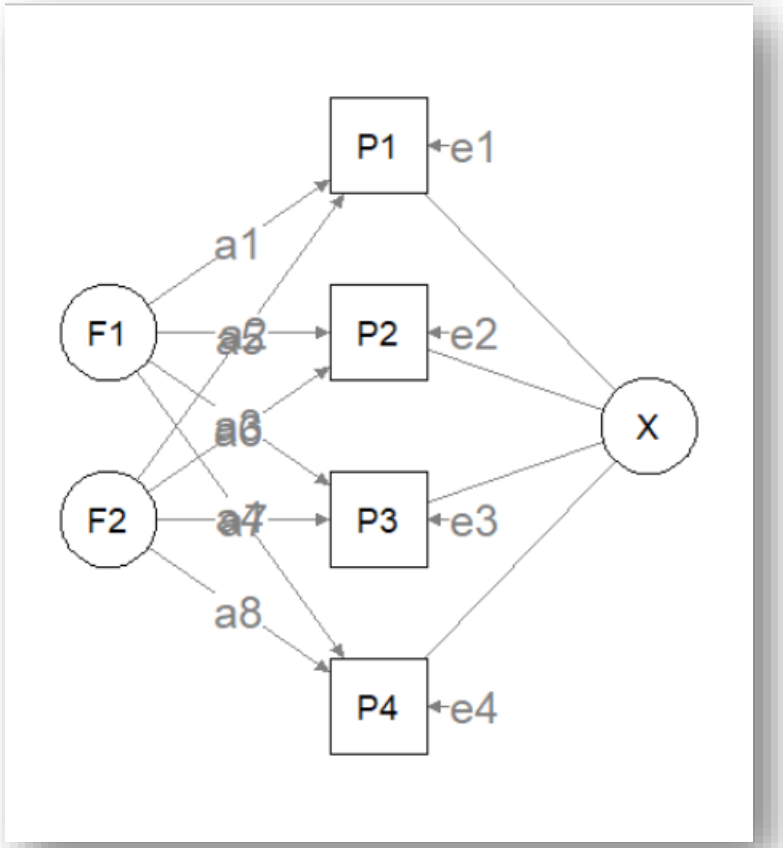




Omega: Multidimenzionalita







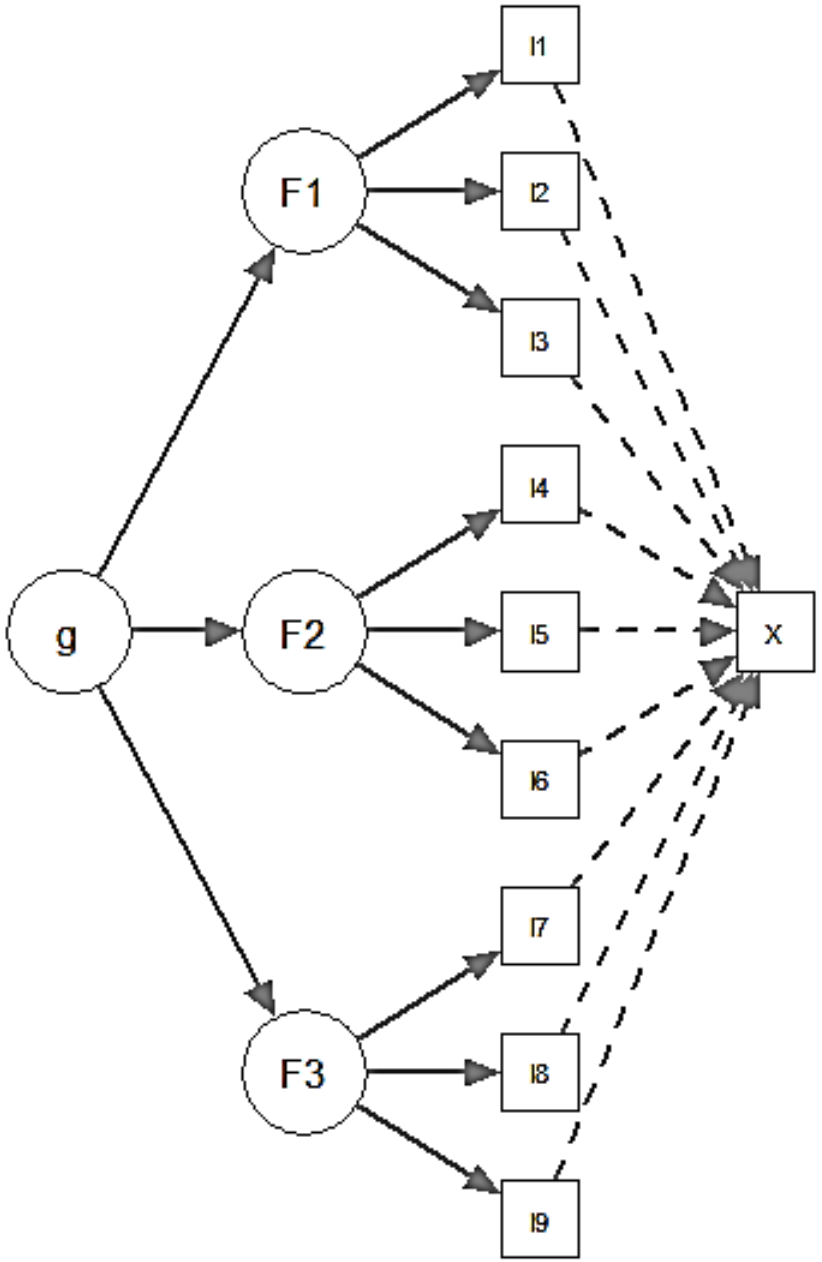
Omega: Multidimenzionalita

Hierarchická omega (omega hierarchical):

- Rozptyl součtu položek vysvětlený daným faktorem.
- V případě faktoru druhého řádu (g) jsou specifické rozptyly faktorů prvního řádu považovány za chybu.
- **Model based reliabilita:** velmi záleží na definici modelu.

Celková omega (omega total):

- Rozptyl součtu položek vysvětlený všemi faktory prvního řádu.
- Odhad test-retest reliability součtu položek, pokud se míra žádného z atributů nezmění.
- **Odhad korelace paralelních testů** při dodržení přiměřených předpokladů



Přehled dalších (FA) koeficientů

Revellova β (1978): Nejnižší podíl rozptylu, který lze vysvětlit jediným faktorem.

- Odhad nejhorší možné split-half reliability.
- $\beta = \frac{k^2 \bar{\sigma}_{ij}}{\sigma_x^2}$, kde $\bar{\sigma}_{ij}$ je průměrná kovariance napříč dvěma nejhůře rozdělenými polovinami testu.

Revellova omega: celková omega (ω_{tot}) odhadnutá s pomocí EFA.

- S větším počtem faktorů (typicky tři) a Schmid-Leiman transformací.

Bentlerův koeficient *glb* (Greatest Lower-Bound of reliability, 1980):

- Dimension-free vnitřní konzistence.
- Princip: odhad ω_{tot} pro tolik faktorů, kolik jich nevede k negativnímu reziduálnímu rozptylu žádné z položek.
- $\rho_{glb} = 1 - \max \frac{1'\Psi 1}{1'\Sigma 1}$, s poslední pozitivně semi-definitní maticí $(\Sigma - \Psi)$ (kde Σ je pozorovaná matice, Ψ reziduální matice a 1 je jednotková matice.

SW implementace

Pozor: omega v JASPu a JAMOVI je vhodným ukazatelem jen tehdy, pokud jednodimenzionální model sedí na data.

Balíček `psych` v R (funkce `splitHalf`, `omega`, `glb.fa`).

- Pozor: funkce `omega` defaultně využívá korelační, nikoliv kovarianční matici (`covar=FALSE`).

Funkce `semTools::compRelSEM` odhadne reliabilitu `lavaan` modelu. Vhodnější než psych balíček (lepší estimátory).

- Vhodné i pro ordinální data – Greenova-Yangova (2009, vzorec 21) korekce.
- Možnost exploračního řešení s pomocí funkce `lavaan::efa`.

$$\rho_{X\tilde{X}} = \frac{\sum_{j=1}^{J}\sum_{j'=1}^{J}[\sum_{c=1}^{C-1}\sum_{c'=1}^{C-1}\Phi_2(\tau_{V_jc},\tau_{V_{j'}c'},\sum_{k=1}^{K}\sum_{k'=1}^{K}\lambda_{V_j^*F_k}\lambda_{V_{j'}^*F_{k'}}\rho_{F_kF_{k'}})-(\sum_{c=1}^{C-1}\Phi_1(\tau_{V_jc}))(\sum_{c=1}^{C-1}\Phi_1(\tau_{V_{j'}c}))]}{\sum_{j=1}^{J}\sum_{j'=1}^{J}[\sum_{c=1}^{C-1}\sum_{c'=1}^{C-1}\Phi_2(\tau_{V_jc},\tau_{V_{j'}c'},\rho_{V_j^*V_{j'}^*})-(\sum_{c=1}^{C-1}\Phi_1(\tau_{V_jc}))(\sum_{c=1}^{C-1}\Phi_1(\tau_{V_{j'}c}))]}. \quad (21)$$

Určitost (determinace) faktorových skórů Factor score determinacy

Koeficienty omega jsou odhadem reliability součtu položek (všechny položky mají váhu 1).

Občas pracujeme s odhady faktorových skórů (lineární kombinací položek).

- Vážený průměr všech položek; váha je spočítaná na základě f. nábojů a reziduálních rozptylů.
- $C = \Sigma_y \Lambda_y^T (\Lambda_y \Sigma_y \Lambda_y^T + \Theta_y)^{-1}$ maticový vzorec výpočtu, není podstatné.

Výhody: Vyšší reliabilita (váhy položek jsou optimálně zvolené).

Nevýhody: Sample dependency (zvláště u malých vzorků nepřesný odhad parametrů FA modelu).

Factor score determinacy (FSD) = podíl rozptylu odhadu faktorového skóre vysvětlený faktorem.

Reliabilita rozdílu

Jak reliabilní je používání rozdílu mezi dvěma testy?

- Například VIQ a PIQ ve WAIS-III?

$$r_{x-y} = \frac{\sigma_x^2 r_{xx'} + \sigma_y^2 r_{yy'} - 2 r_{xy} \sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2 - 2 r_{xy} \sigma_x \sigma_y},$$

- kde σ_x^2 a σ_y^2 jsou rozptyly obou testů, $r_{xx'}$ a $r_{yy'}$ jejich reliability a r_{xy} je jejich korelace.
- jmenovatel je roven rozptylu výsledných rozdílů.

Pokud $\sigma_x^2 = \sigma_y^2 = \sigma_{xy}^2$ (v případě standardizovaných testů), pak:

- $r_{x-y} = \sigma_{xy}^2 \frac{r_{xx'} + r_{yy'} - 2 r_{xy}}{2 - 2 r_{xy}}$

Reliabilita rozdílu

Standardní chybu (SE) rozdílu lze spočítat s pomocí SD a SE vpravo, nebo prostřednictvím vzorce.

Toto je důvod, proč je problematická interpretace rozdílu vysoce korelovaných subtestů.

- $r_{xx'}$, $r_{yy'}$ – reliability testů x a y
- r_{xy} – korelace testů x a y
- **r_{x-y} – reliabilita rozdílu**
- SD_{x-y} – SD rozdílu
- SE_{x-y} – standardní chyba rozdílu
- $CI_{95\%}$ – šířka 95% intervalu spolehlivosti

$r_{xx'}$	$r_{yy'}$	r_{xy}	r_{x-y}	SD_{x-y}	SE_{x-y}	$CI_{95\%}$
0,7	0,8	0	**0,75**	21,2	10,6	20,8
0,7	0,8	0,2	**0,69**	19,0	10,6	20,8
0,7	0,8	0,4	**0,58**	16,4	10,6	20,8
0,7	0,8	0,6	**0,38**	13,4	10,6	20,8
0,7	0,7	0,6	**0,25**	13,4	11,6	22,8
0,9	0,9	0,8	**0,50**	9,5	6,7	13,1
0,9	0,9	0,45	**0,82**	15,7	6,7	13,1
0,6	0,6	0,5	**0,20**	15,0	13,4	26,3
0,7	0,7	0,65	**0,14**	12,5	11,6	22,8

Kompozitní reliabilita obecně

Srovnání reliability rozdílu a kompozitní reliability (stratifikovaná Cronbachova alfa).

Je evidentní, že korelace testů má opačný vliv na výslednou reliabilitu. S rostoucí korelací:

- reliabilita rozdílu klesá;
- kompozitní reliabilita roste.

Příčinou je rozdílné nasčítání chypového rozptylu podle „součtového“ vzorce

$$\mathrm{var}(A \pm B) = \mathrm{var}(A) + \mathrm{var}(B) \pm 2\mathrm{cov}(A, B)$$

- Pomůcka: $(a \pm b)^2 = a^2 + b^2 \pm 2ab$
- Chyba se vždy sčítá, zatímco pravé skóry se sčítají nebo odčítají.

$r_{xx'}$	$r_{yy'}$	r_{xy}	r_{x-y}	r_{x+y}
0,7	0,8	0	**0,75**	**0,75**
0,7	0,8	0,2	**0,69**	**0,79**
0,7	0,8	0,4	**0,58**	**0,82**
0,7	0,8	0,6	**0,38**	**0,84**
0,7	0,7	0,6	**0,25**	**0,81**
0,9	0,9	0,8	**0,50**	**0,94**
0,9	0,9	0,45	**0,82**	**0,93**
0,6	0,6	0,5	**0,20**	**0,73**
0,7	0,7	0,65	**0,14**	**0,82**

Otázky na závěr

Reliabilita čeho?

- Odhadu latentní proměnné? Stability pozorovaného skóre?
- Součtu položek, odhadu faktorového skóre?

Stabilita skóre napříč (jakými?) podmínkami?

Reliabilita není jedna.

- Záleží na epistemologických východiscích i účelu měření.

Moje osobní doporučení

Alfa je tradiční „deskriptivní“ ukazatel s jednoznačným výpočtem. Je dobré jej uvádět.

- Ale jde o podhodnocenou spodní hranici reliability.
- Z hlediska model-based reliability může nadhodnocovat i podhodnocovat.

Omega koeficienty nejsou vhodné, pokud faktorový model nedobře popisuje data.

- Výjimkou je omega jednofaktorového modelu, které je vždy lepším estimátorem než alfa.

V případě nejasné faktorové struktury lze využít některý z *glb* koeficientů.

- V případě velkého vzorku λ_4, v případě menšího (ale stále dostatečného) Bentlerovo ρ_{glb}.

V případě jasné faktorové struktury je vhodnější omega koeficient. Lze si vybrat:

- Celková omega: Odhad dimension-free reliability jako uvažované stability skóru.
- Hierarchická omega: Odhad model-based reliability jako spolehlivosti usuzování na míru latentního rysu.

Moje osobní doporučení

Nepoužívejte dvoupoložkové testy! ☺

- Pokud je už použijete, ideální je Angoff-Feldtův koeficient, SB ale poslouží rovněž.

Je potřeba vyvážit „jednoduchost“ postupu vs. jeho „vhodnost“ pro dané řešení.

- Potíže s omega koeficienty tkví v tom, že existuje mnoho postupů výpočtu s rozdílnými výsledky.
- Je jednoduché se do toho zamotat. **Pokud vůbec netušíte, alfa (téměř vždy) poslouží!**

„Nebezpečné“ situace, kdy je dobré se zamyslet:

- Velmi krátké testy (do pěti položek?).
- Výrazně komplikovaná faktorová struktura...
- ... a zejména korelované chyby měření (reziduální kovariance).
- Výrazné porušení předpokladu tau-ekvivalence.
- Dvoupoložkové testy o nestejné délce.