MUNI FSS
Katedra psychologie

Přednáška 6: Model klasické testové teorie

18. 10. 2021 | PSYn4790 | **Psychometrika: Měření v psychologii**

Katedra psychologie, Fakulta sociálních studií MU

Hynek Cígler





Klasická testová teorie (CTT)

Tři pilíře CTT (Traub, 1997):

- Chyby I. typu, chyba měření jako náhodná veličina, korelace.

Koeficient proti oslabení korelace (Spearman, 1904).

- Vztah reliability, chyby měření a koncept paralelních testů.
- Attenuation formula, $r_{pq}^* = \frac{r_{pq}}{\sqrt{r_{pp'} r_{qq'}}}$.

Vývoj CTT byl prakticky ukončen do 60. let: Lord a Novick (1968).

- Zlepšují se jen estimátory (koeficienty) apod.

Klasická testová teorie (CTT)

Důležitým impulzem byla Fergusonova komise (1932– 1940).

- Striktní požadavek aditivity (a zřetězení).
- Psychologové zřetězení nedokázali → **CTT není vědeckou teorií měření.**
  - 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-l

Odbočka: škálování

V 1. pol. 20. stol. se psychologie hodně zaměřovala právě na proces, jak „vznikají“ čísla při měření.

Otázka za zlatého bludišťáka: Jak z pozorování vyrobit „škálu“?

- U jednoduchých psychofyzikálních dat jasné, problém je s komplexnějšími konstrukty.

Řada různých „škál“: Hayes a Patterson (1921), Thurstone (1928), Likert (1932), Guttman (1944), Osgood (1957) a další.

Od 50. let však minimální další rozvoj.

- Etablování stávajících škál.
- Rozvoj faktorové analýzy (omezené využití u některých škál a naopak realistické vysvětlení toho, proč škály fungují).
- Pokročilejší postupy jsou komplikované a nemají všeobecné využití.
- Rozvoj měření v psychologii vedl ke standardizaci postupů.

Měření v přírodních vědách

Existuje nějaký atribut, který opakovaně měříme tím stejným nástrojem/procedurou.

Každé jedno měření má nějakou chybou, kterou neznáme.

- Jednotlivá měření se pohybují okolo skutečné hodnoty v důsledku náhodné chyby měření.

Výsledkem opakovaných měření je proto rozložení, které použijeme pro odhad skutečné hodnoty:

- **Průměr rozložení:** odhad míry atributu, $\mathrm{E}(x) = \frac{\sum_{i=1}^{N} x_i}{N}$.
  - N – počet měření; x_i – i-tá naměřená hodnota; $\mathrm{E}(x)$ – expected value (průměr, nejpravděpodobnější hodnota příštího měření.
- **Standardní chyba průměru:** odhad standardní chyby měření, $SE = \frac{s_d}{\sqrt{N}}$
  - SE – standardní chyba měření (Standard Error), s_d – výběrová směrodatná odchylka jednotlivých měření.
- Lze využít pro konstrukci *CI* atd. (za pomoci Studentova t-rozložení).

Předpoklady

Odhad průměru (standardní chyba měření) je přibližně normálně rozložený.

- Centrální limitní teorém: potřebujeme alespoň 30 měření.
- Příklady zde a zde ☺

To v psychologii není možné. Nemůžu člověka měřit 30krát tím stejným testem (vyjma jednoduchých psychofyzikálních úloh).

Kudy z toho ven? Shodná chyba měření pro všechny respondenty.

- Nikoliv *„standardní chybu průměru"* pro každého respondenta zvlášť.

Jednotlivá měření jako paralelní testy.

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

Paralelní testy

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

- CTT postup s paralelními testy a tradiční „fyzikální“ měření vedou ke stejným výsledkům.
  - To je právě Spearmanův objev.
- Test-retest, paralelní formy, shoda posuzovatelů, split-half...

Původně CTT považovala za paralelní testy pouze jejich výsledek (celkové skóre).

- Způsob konstrukce tohoto skóre je irelevantní.
- Operacionalismus: pravé skóre (a tedy měřený atribut) je definovaný měřením.

CTT tedy chápe reliabilitu jako „stabilitu“ odhadu pravého skóre napříč podmínkami (paralelním testováním).

S postupem času otázka: Jak se celkové skóre vytváří?

- **Položky jako paralelní testy.**

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):

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

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. $a_i = a$

- + 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):

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

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. $a_i = a, \mathrm{var}\big(e_{ip}\big) = \mathrm{var}(e)$

- + 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):

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

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.

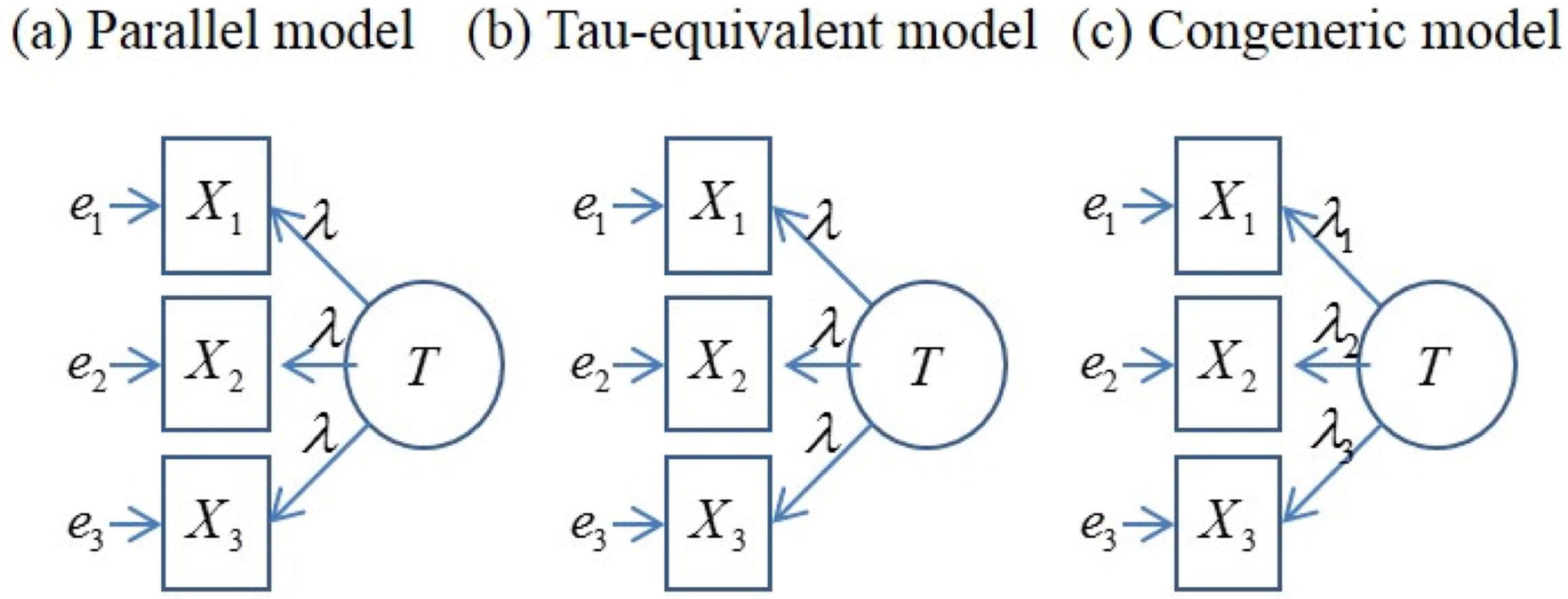
- + Shodné reziduální rozptyly.

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

- + 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

CTT: Paralelní testy



$$Var(e_1) = Var(e_2) = Var(e_3)$$

Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods, 19*(4), 651–682. https://doi.org/10.1177/1094428116656239

Reliabilita

*„The term reliability has been used in two ways in the measurement literature. First, the term has been used to refer to the **reliability coefficients of classical test theory**, defined as the **correlation between scores** on two equivalent forms of the test, presuming that taking one form has no effect on performance on the second form.*

*Second, the term has been used in a more general sense, to refer to the **consistency of scores across replications** of a testing procedure, regardless of how this consistency is estimated or reported (e.g., in terms of standard errors, reliability coefficients per se, generalizability coefficients, error/tolerance ratios, item response theory (IRT) information functions, or various indices of classification consistency).“*

Stačí to?

(AERA, 2014, s. 33)

AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

(Dvakrát) dvě pojetí reliability

Stabilita měření (operacionalismus).

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

Vysvětlený rozptyl (realismus).

- Vysvětlený rozptyl čím?
- Co považujeme za pravé skóre?

→ Klasická testová teorie.

- Dnešní přednáška.

Relativní srovnání (CTT, GT).

- Na obtížnosti položek nám nezáleží.

Absolutní srovnání (GT).

- 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

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.

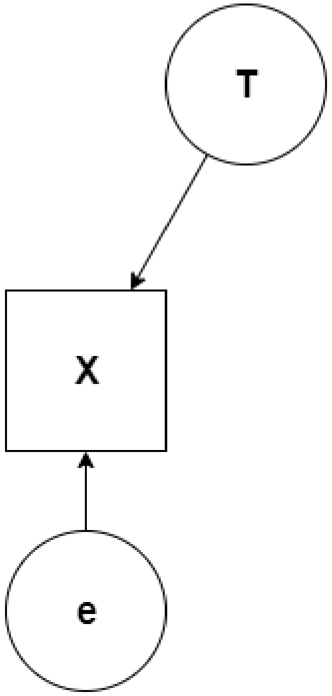
	Unidimensional		Multidimensional
	Split-Half	General	General
Parallel	Split-half parallel reliability (ρ_{SP})	Parallel reliability (ρ_P)	Multidimensional parallel reliability (ρ_{MP})
Tau-equivalent	Split-half tau-equivalent reliability (ρ_{ST})	Tau-equivalent reliability (ρ_T)	Multidimensional tau-equivalent 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)

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.

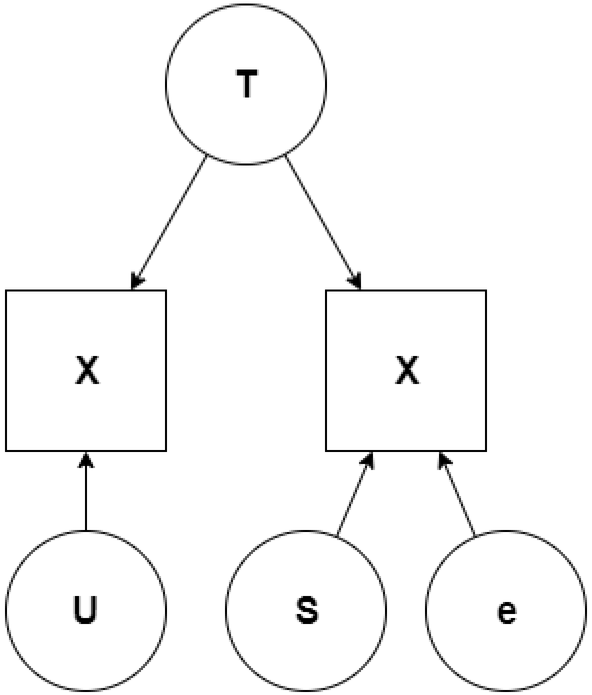
Unikátní rozptyl U ale lze rozdělit na:

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

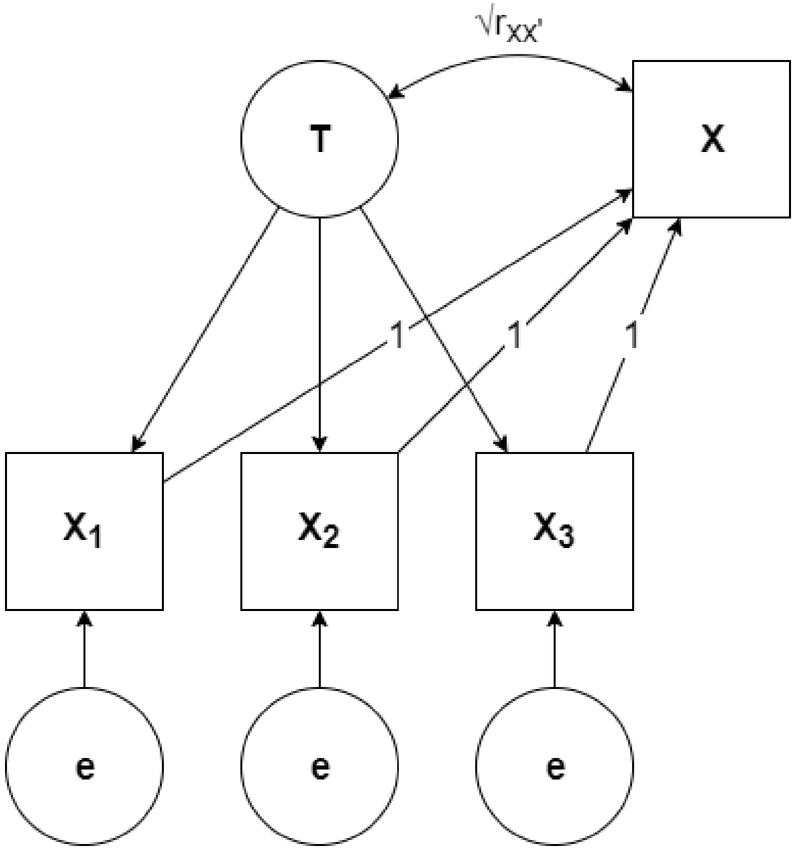
Zatímco S přispívá ke korelaci paralelních testů, chyba e nikoli.

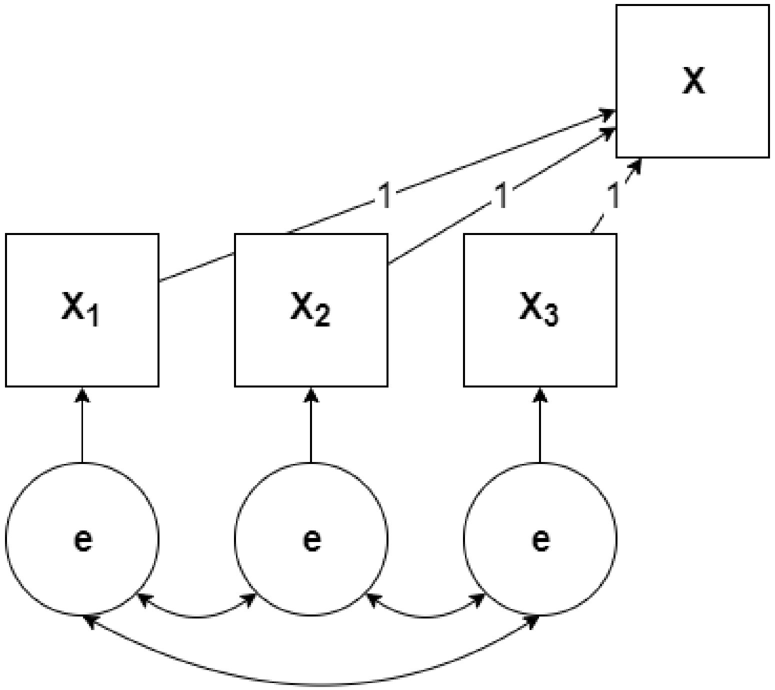
Tyto složky ale nelze oddělit při jediné administraci testu a S je považován celý za chybu.

- Proto v longitudinálních SEM modelech korelovaná rezidua v čase.



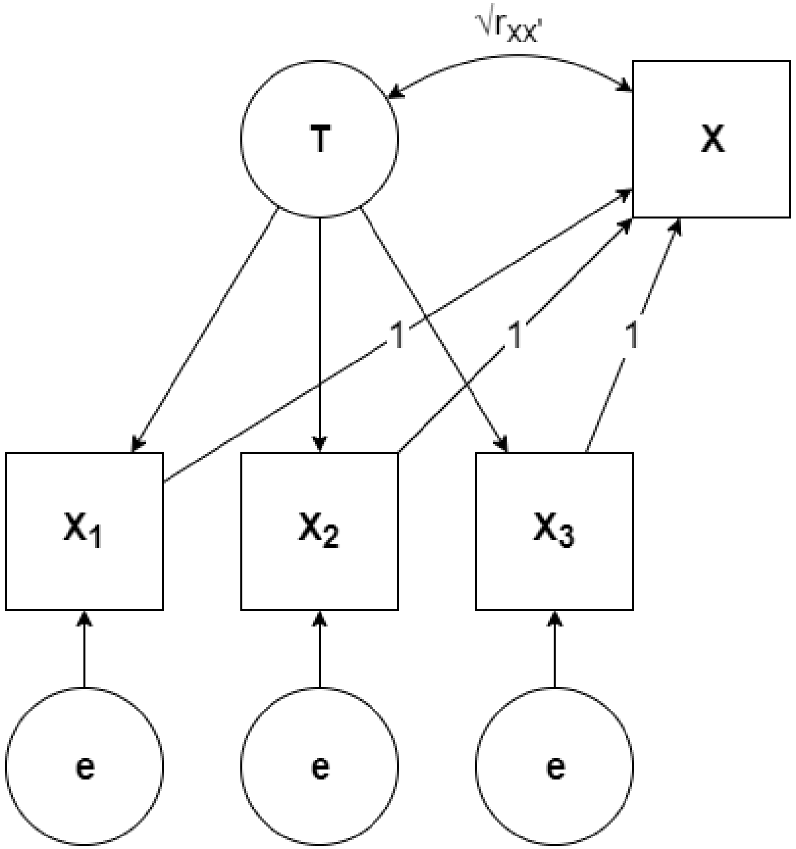
Formativní vs. reflektivní model

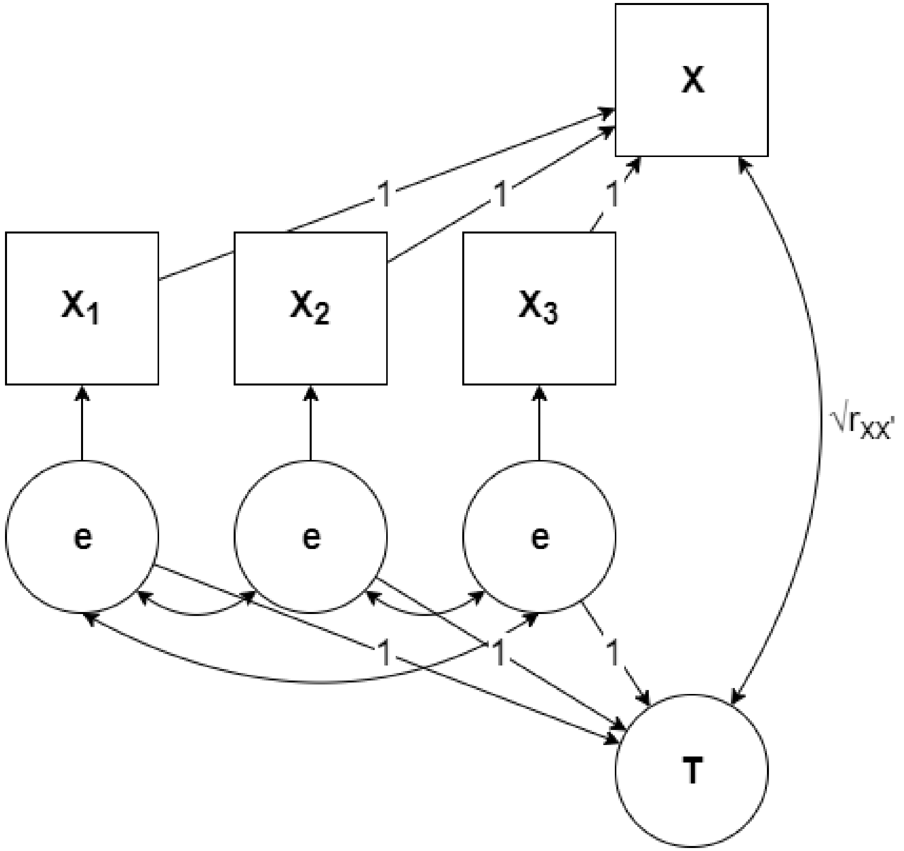




(Takto konstruovaný model se někdy označuje jako Raykovovo omega.)

Formativní vs. reflektivní model





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:

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

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

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

Předpoklad: **paralelní poloviny**.

- Při nedodržení příliš „optimistický“, může nadhodnocova 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 je λ_4 dnes chápána jako maximalizovaná split-half pomocí nejlepšího možného rozdělení.

„Příliš dobré rozdělení“→ na malých vzorcích nadhodnocuje.

Pokud je kovariance větší než kterýkoli z rozptylů: hrubé podhodnocení.

Založeno na jediné korelaci → 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): $r_{AF} = \frac{4\sigma_{12}}{\sigma_x^2-\frac{(\sigma_1^2-\sigma_2^2)^2}{\sigma_x^2}}$

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

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.

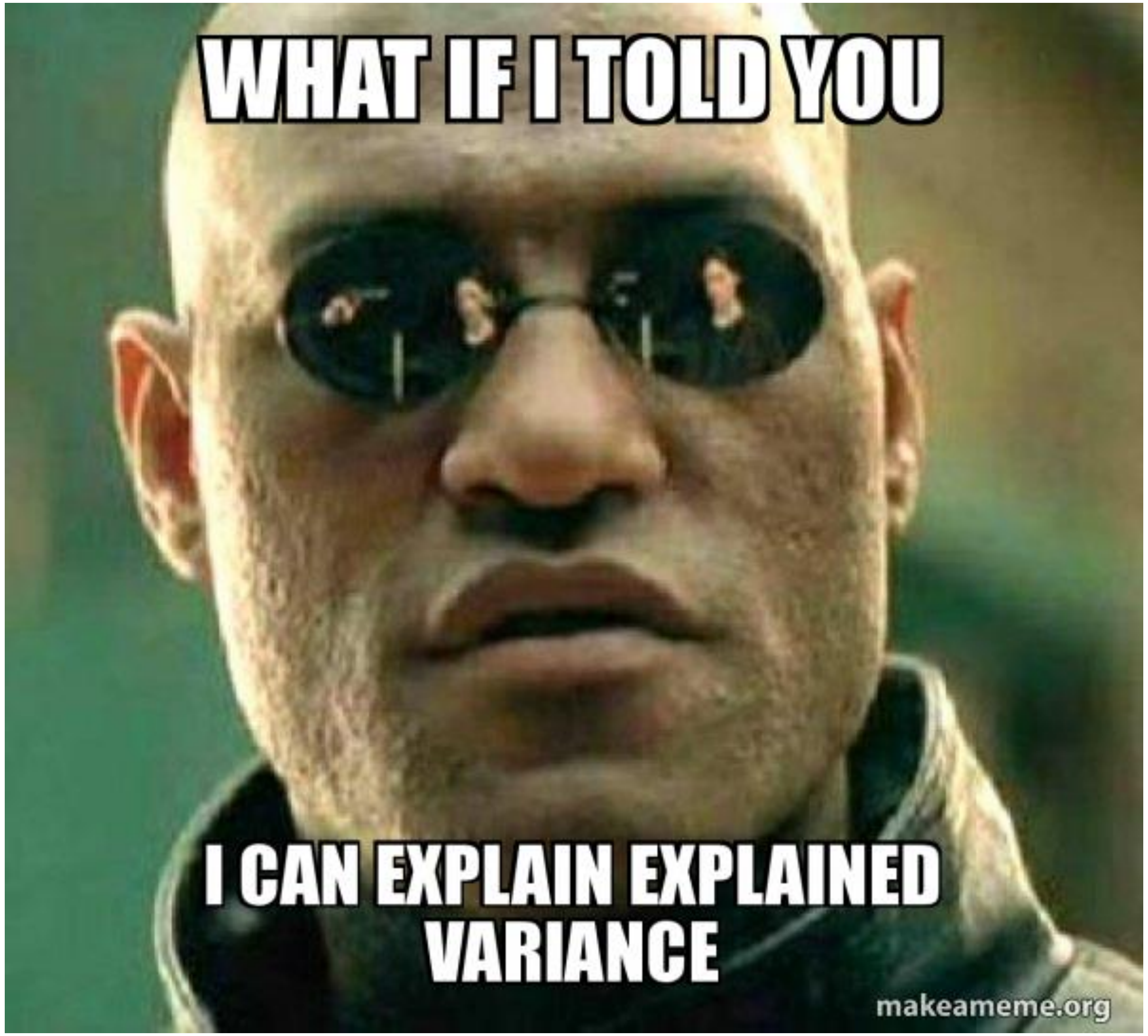
- A nekorelované chyby měření testů.

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

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):

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

- 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 přímo započítány).

Model-based reliabilita: **omega**

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

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

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

- vysvětlený rozptyl
- chybový rozptyl
- celkový rozptyl

- 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

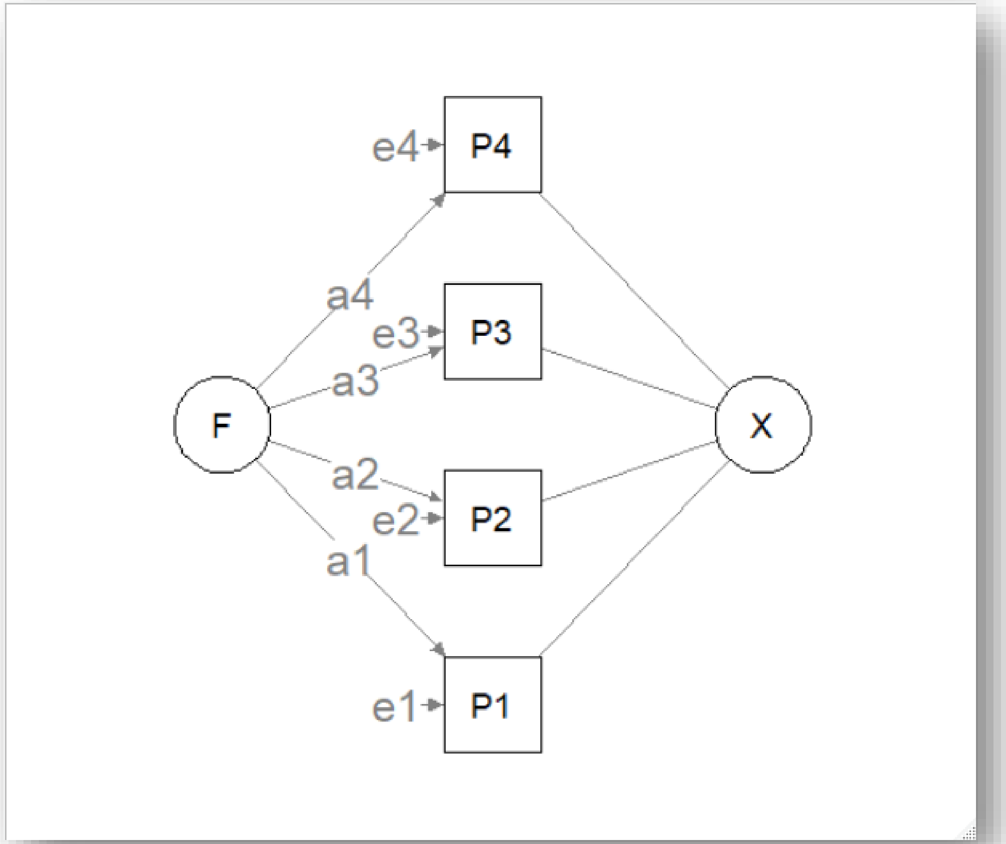
Bez předpokladu tau-ekvivalence (rozdílné faktorové náboje jsou zohledněny).

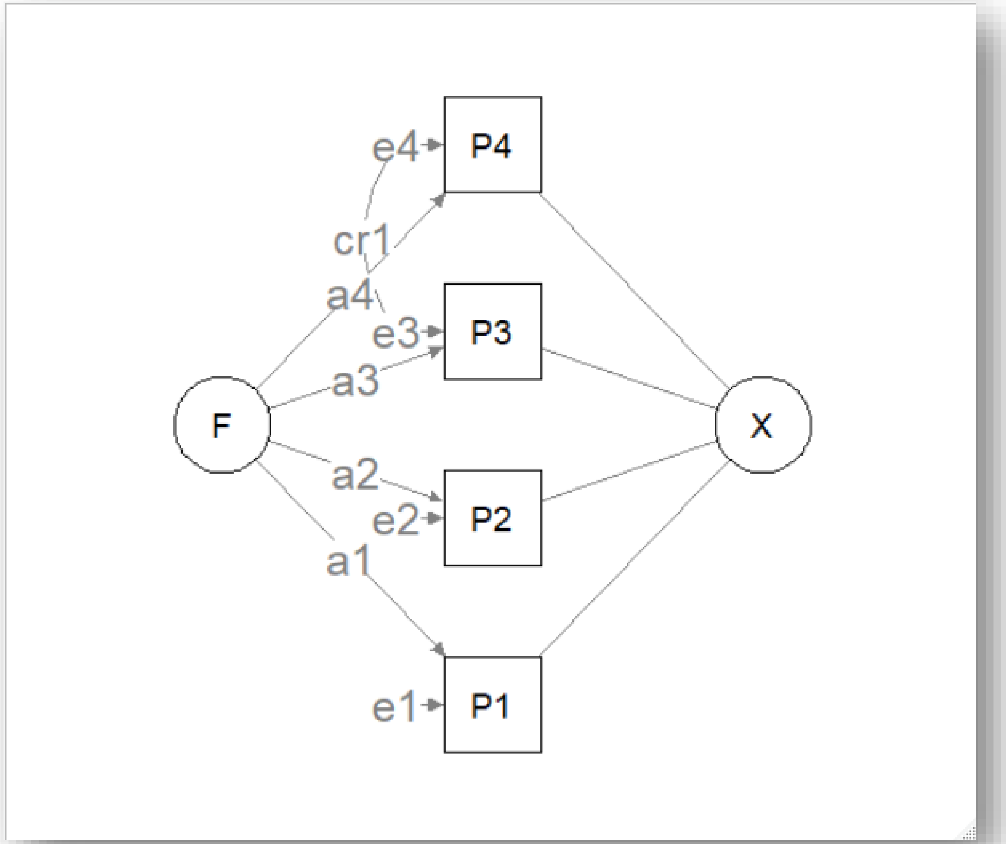
Model-based reliabilita: **omega**

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

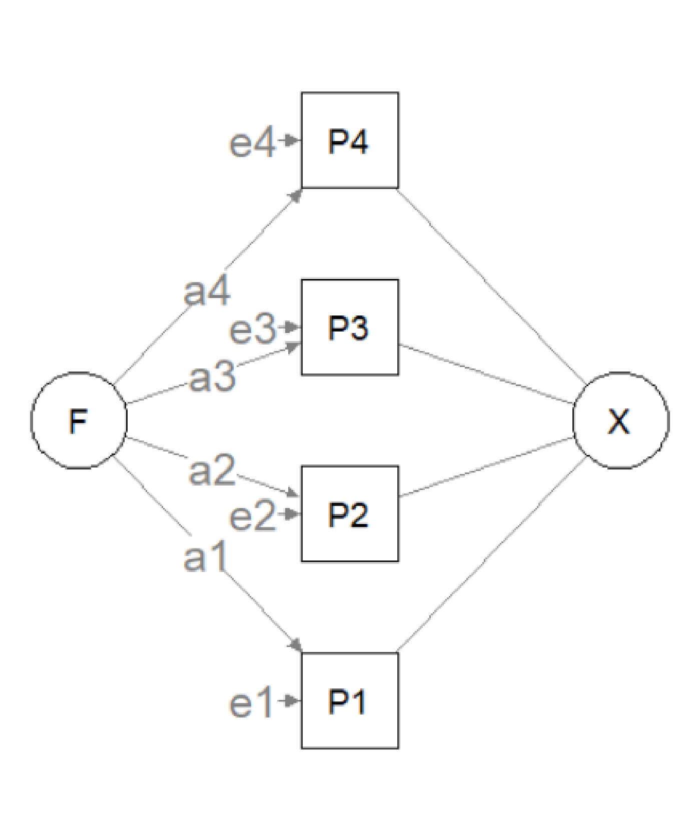
Co je to, co chceme měřit?

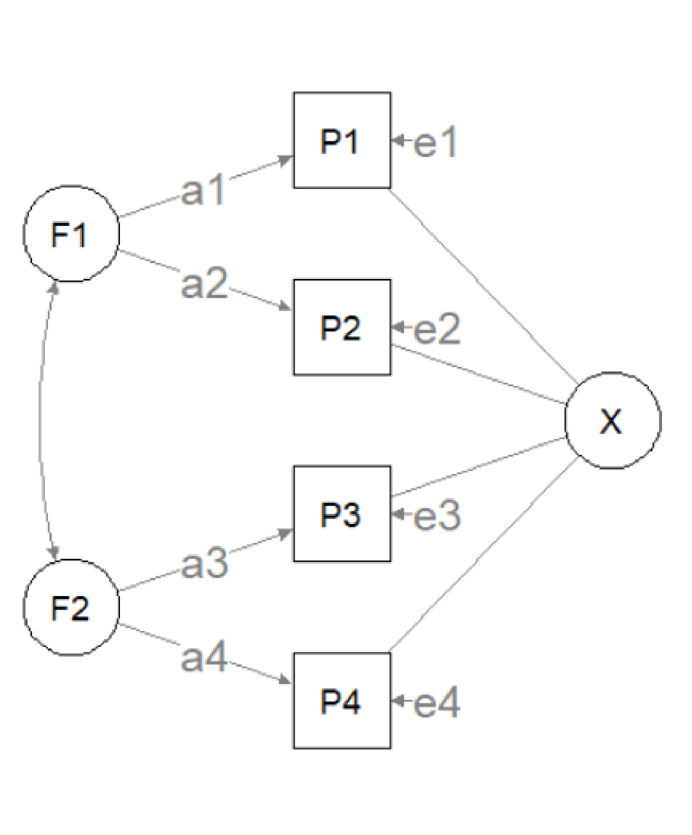
Omega: Multidimenzionalita

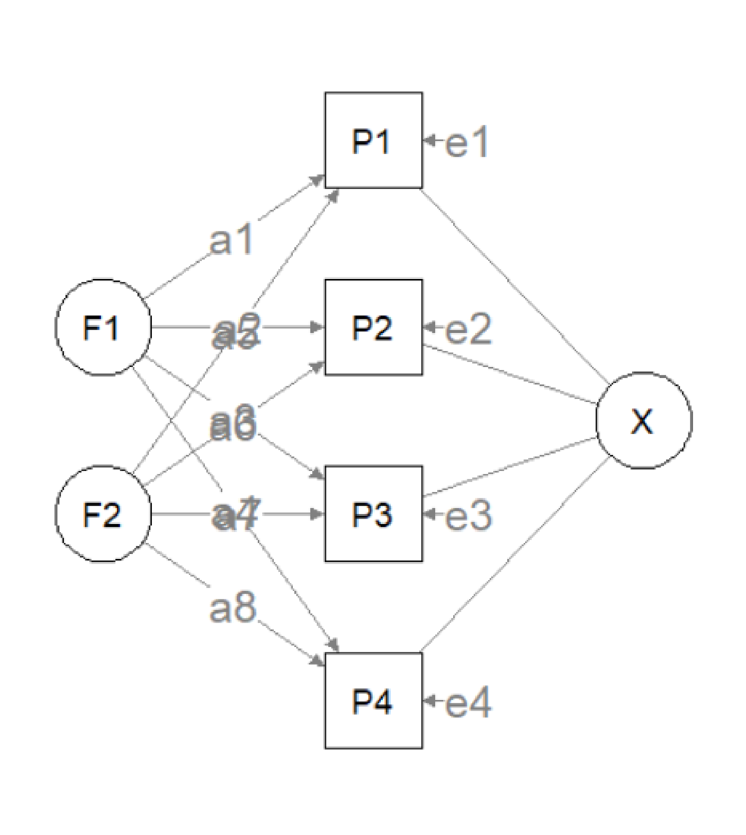




Omega: Multidimenzionalita







Omega: Multidimenzionalita

Hierarchická omega (omega hierarchical):

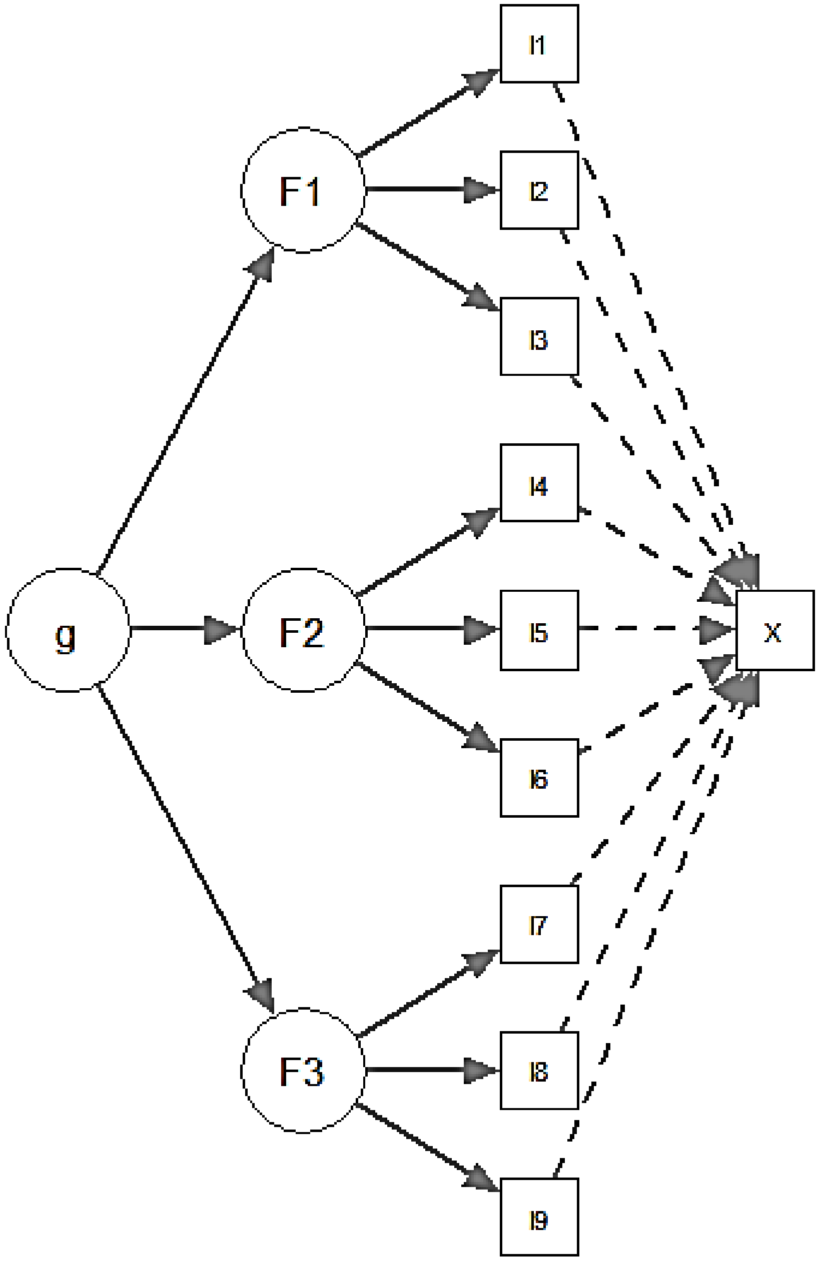
- Rozptyl součtu položek vysvětlený daným faktorem.
- V případě faktoru druhého řádu (g) jsou specifické rozptyly faktorů prvního řádu považovány za chybu.
- **Model based reliabilita:** velmi záleží na definici modelu.

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.

Explorační omega (Revelle):

- Celková omega spočítaná na základě EFA.
- omega funkce v psych balíčku v R.



Přehled dalších (FA) koeficientů

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

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 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 dobrým 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.

Funkce `reliability` v `semTools` balíčku odhadne reliabilitu `lavaan` modelu. Vhodnější než psych balíček (lepší estimátory).

- Pro hierarchické modely (faktory vyššího řádu) funkce `reliabilityL2`.
- Vhodné i pro ordinální data – Greenova-Yangova (2009, vzorec 21) korekce.
- Možnost exploračního řešení s pomocí funkce `semTools::efaUnrotate`.

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

Určitost faktorových skórů

Factor score determinacy.

Koeficienty omega pracují se součtem položek (všechny položky mají váhu 1).

Občas pracujeme s odhady faktorových skórů.

- 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?

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

- 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:

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

Reliabilita rozdílu

$r_{xx'}$	$r_{yy'}$	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

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

Kompozitní reliabilita

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?

Pravého 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 extrahovaná s využitím jediného faktoru, které je vždy lepší 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

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.