Katedra psychologie

Přednáška 2: Chyba měření a reliabilita v CTT

24. 9. 2019 | PSYn4790 | **Psychometrika: Měření v psychologii**

Katedra psychologie, Fakulta sociálních studií MU

Hynek Cígler

Obsah přednášky

1. Koncept reliability v klasické testové teorii (CTT)

- Model měření CTT.
- Koncept paralelních testů.

2. Pokročilé odhady reliability v rámci CTT.

3. Využití reliability (a validity) pro praktické závěry o výsledcích měření.

- Různé druhy chyb měření.
- Konstrukce intervalů spolehlivosti.

Klasická testová teorie (CTT)

Klasická testová teorie stojí na třech pilířích/objevech (Traub, 1997):

- Existence chyby měření I. typu (nezpůsobené ničím jiným).
- Chyba měření je náhodná veličina.
- Koncept korelace.

Spearman (1904) přišel s koeficientem proti oslabení korelace („attenuation coefficient“), chybu měření parametrizoval a umožnil vznik CTT.

- Původním účelem byl odhad korelací nezkreslených chybou měření.

Klasická testová teorie (CTT)

Důležitým impulzem byla Fergusonova komise (1932– 1940).

- Striktní požadavek aditivity (a zřetězení).
- Psychologové zřetězení nedokázali → **CTT není vědeckou teorií měření.**
- Reakcí byla Stevensova „nevědecká" „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).

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

Klasická testová teorie (CTT)

CTT model: pozorované skóre je lineární funkcí pravého skóre.

Koncept měření jako paralelních testů.

Reliabilita popisuje těsnost vztahu...

- paralelních testů
- pravého a pozorovaného skóre

Operacionalismus: Definice atributu (pravého skóru) skrze měřicí nástroj.

$$X_p = \tau_p + e_p$$

Pravý skór τ_p predikuje pozorovaný skór X_p s nějakou mírou chyby e_p.

Protože chyba nekoreluje s atributem, vztah platí přímo pro rozptyly proměnných:

$$\sigma_x^2 = \sigma_\tau^2 + \sigma_e^2$$

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.

Postup fyzikálního měření (např. délky):

- Změřím objekt n-krát a získám n měření délky označených jako d_i.
- Bodový odhad délky je průměr z těchto měření: $\mathrm{E}(d) = \frac{\sum_{i=1}^{n} d_i}{n}$
  - To $\mathrm{E}(d)$ je „expected value“ – odhad měřené hodnoty.
- Chyba tohoto měření (Standard Error /of Measurement/) je:
  - Pro jediné měření: $SE = s_d$, kde s_d je výběrová směrodatná odchylka pozorovaných hodnot d_i .
  - Pro průměr z n měření: $SE = \frac{s_d}{\sqrt{n}}$ (standardní chyba průměru!).
  - (A použijeme Studentovo t-rozložení, protože s_d je pouze pozorovaným odhadem populační σ_d.)

Paralelní testy

Na konceptu paralelních testů Spearman založil koncept reliability.

- Na reliabilitě stojí zase CTT.

Paralelní testy/měření jsou takové, pro které platí:

- A. Pravý skór je v obou 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 obou testech a pro každý subjekt stejný.
  - Důsledkem je navíc shodný rozptyl pozorovaných skórů obou testů.

Tyto předpoklady jsou v sociálních vědách příliš striktní, a proto později budeme pracovat spíše s „mírou paralelnosti“.

CTT: Paralelní testy

Úrovně paralelnosti položek (založené na faktorové analýze):

$$X_{ip} = i_i + a_i \tau_p + e_{ip}$$

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):

$$X_{ip} = i_i + a_i \tau_p + e_{ip}$$

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):

$$X_{ip} = i_i + a\tau_p + e_{ip}$$

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):

$$X_{ip} = i_i + a\tau_p + e_{ip}$$

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

- 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):

$$X_{ip} = i + a\tau_p + e_{ip}$$

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.

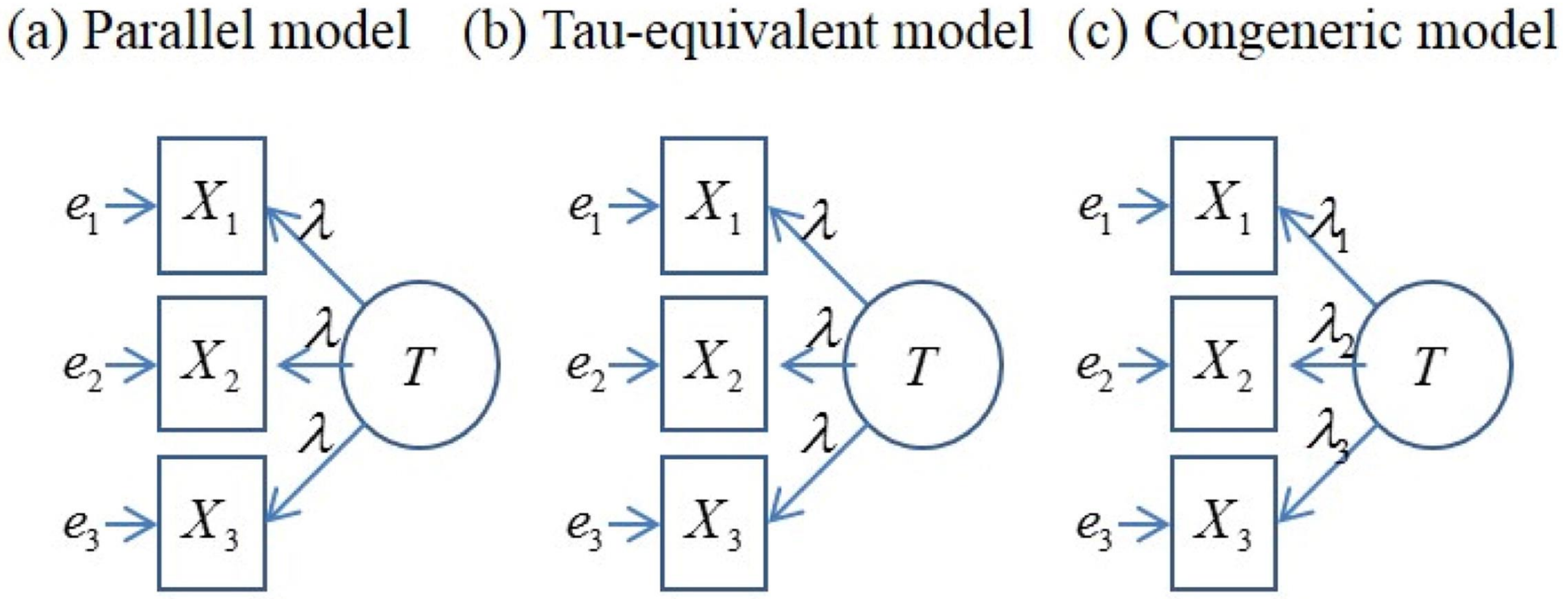
- Shodné reziduální rozptyly.

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

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

Reliabilita $r_{xx'}$ testu x je definovaná jako vysvětlený rozptyl pozorovaného skóre pravým skóre:

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

- Úpravy platí, protože dosazujeme podle vzorce $\sigma_x^2 = \sigma_\tau^2 + \sigma_e^2$.

Relativní část pozorovaného rozptylu způsobený rozptylem měřeného atributu.

CTT: **Vysvětlený rozptyl** pozorovaného skóre pravým skóre.

Reliabilita: Spearman

Spearman odvodil, že pokud reliabilita $\boldsymbol{r_{xx'}}$ je podíl rozptylu měření vysvětlený atributem („**vysvětlený rozptyl**“) a tedy...

... odmocnina z reliability $\sqrt{\boldsymbol{r_{xx'}}}$ je korelace atributu a měření (**korelace pravého a pozorovaného skóre**), pak...

... reliabilita $r_{xx'}$ je korelace dvou paralelních testů.

Attenuation formula (korekce proti oslabení):

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

- Kde r_{pq}^* je odhad korelace pravých skórů p, q, r_{pq} je pozorovaná korelace testů p a q a $r_{pp'}$, $r_{qq'}$ jsou jejich reliability.

Protože korelace pravých skórů $r_{pq}^* \leq 1$, lze odhadnout maximální možnou pozorovanou korelaci 2 testů jako:

$$r_{pq} \leq \sqrt{r_{pp'} r_{qq'}}$$

- **Korelace nemůže být vyšší než odmocnina součinu reliabilit!**

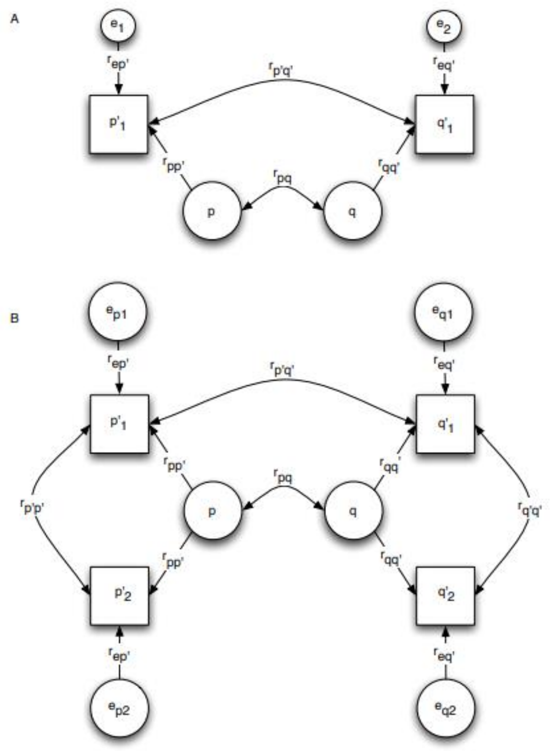


Fig. 7.1 Spearman's model of attenuation and reliability. Panel A: The true relationship between p and q is attenuated by the error in p' and q'. Panel B: the correlation between the latent variable p and the observed variable p' may be estimated from the correlation of p' with a parallel test.

Pozor, notace na diagramu neodpovídá notaci jinde.

https://www.personality-project.org/r/book/Chapter7.pdf

Reliabilita

Doporučuji: 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

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

Dvě klíčová pojetí reliability jako vnitřní konzistence podle toho, co považují za σ_τ^2:

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

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

2. Model based reliability (důraz na vysvětlený rozptyl)

- Odhad vztahu (vysvětleného rozptylu) měřeného atributu a pozorovaného skóru.
- Rodina koeficientů omega (McDonaldova omega, hierarchická omega, celková omega).

Reliabilita

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)

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), tradice.

Variainty 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).

Split-half

SPEARMAN-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.

Předpoklad: paralelní položky.

- Při nedodržení příliš „optimistický“, může nadhodnocovat.
- Založeno na jediné korelaci → nepřesný odhad.

GUTTMANOVA λ_4

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

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

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

Dnes zpravidla jako koeficient GLB (greatest-lower bound of reliability).

- Rozdělení testu tak, aby λ_4 bylo maximální.
- „Příliš dobré rozdělení“→ na malých vzorcích nadhodnocuje.

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?“*

Omega

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

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

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

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

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

Omega

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

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

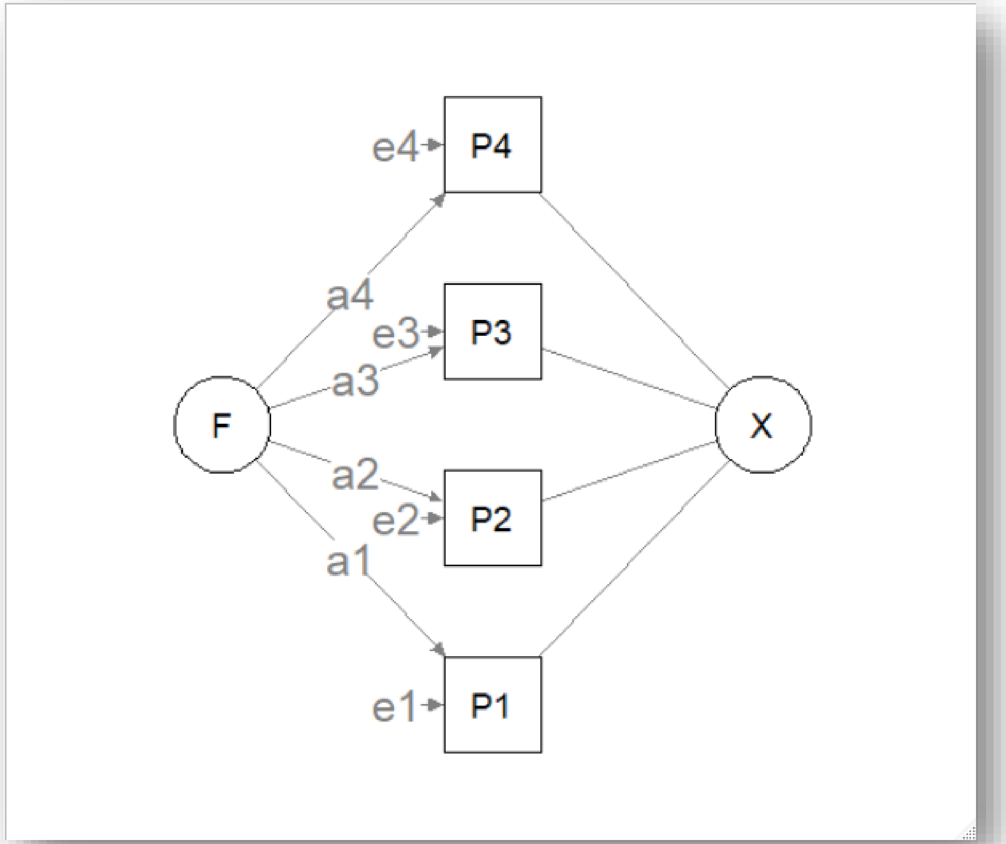
$$\omega = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 \sigma_\psi^2 + \sum_{i=1}^{n} \sigma_{e;i}^2 + 2\sum_{i<j} \sigma_{ij}^2}$$

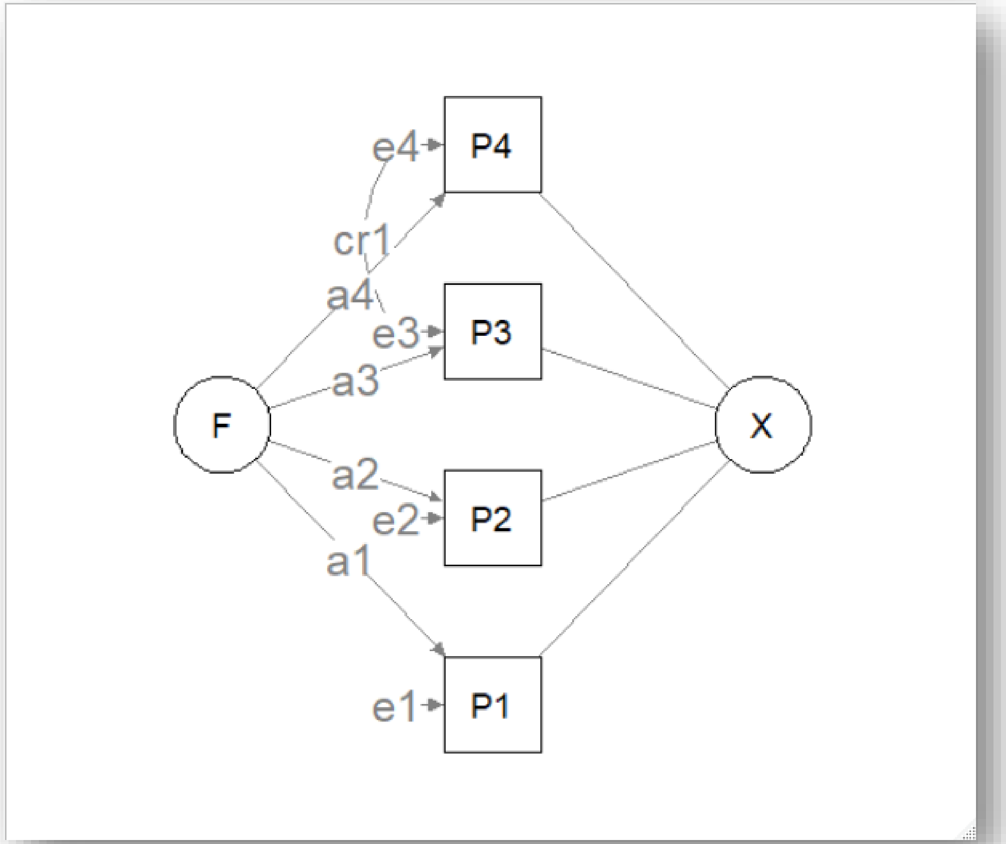
- vysvětlený rozptyl
- chybový rozptyl

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

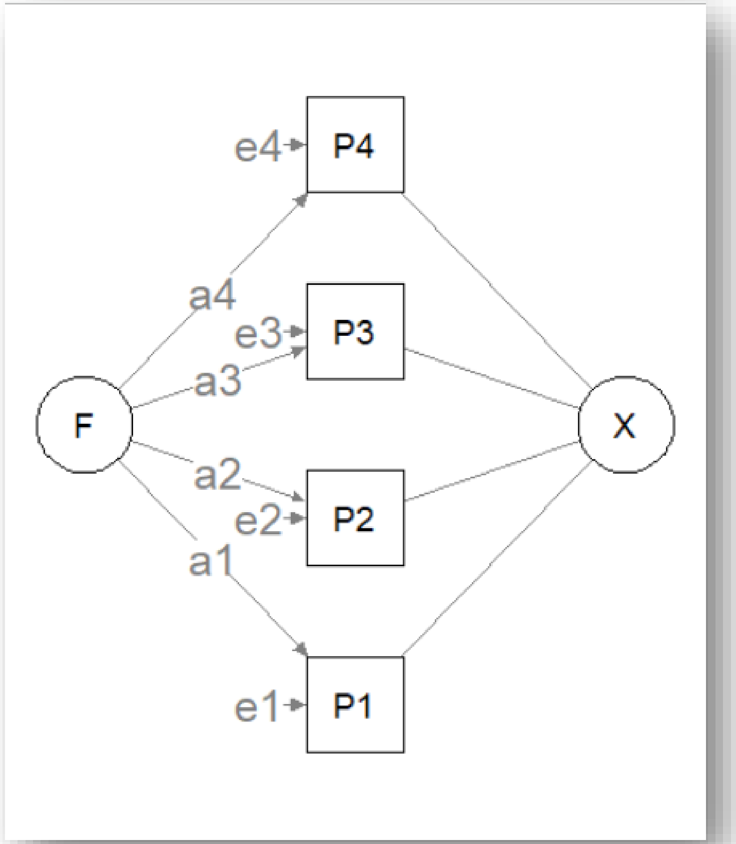
Bez předpokladu tau-ekvivalence (rozdílné faktorové náboje jsou zohledněny).

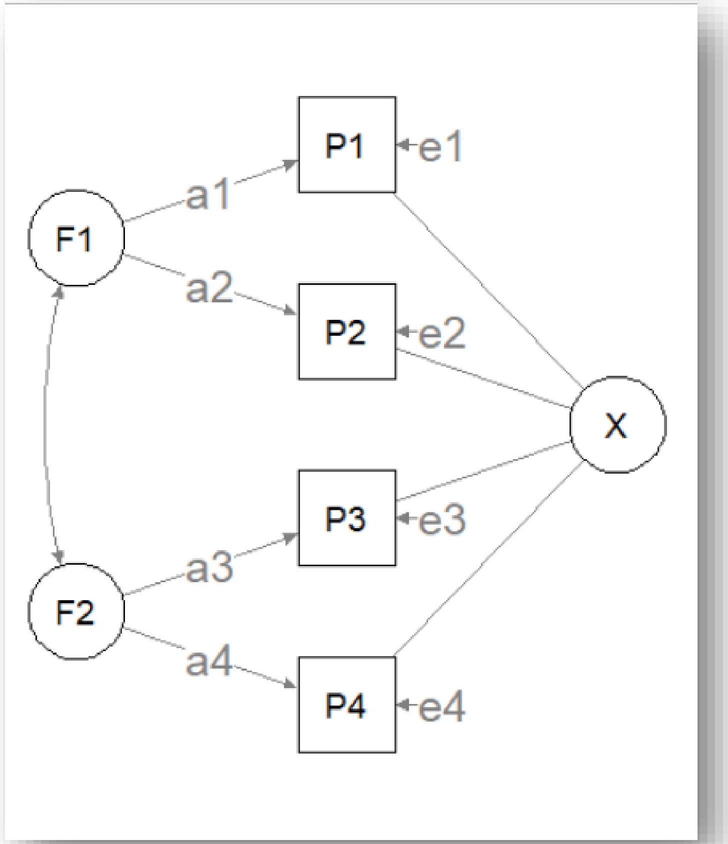
Omega: Multidimenzionalita 1

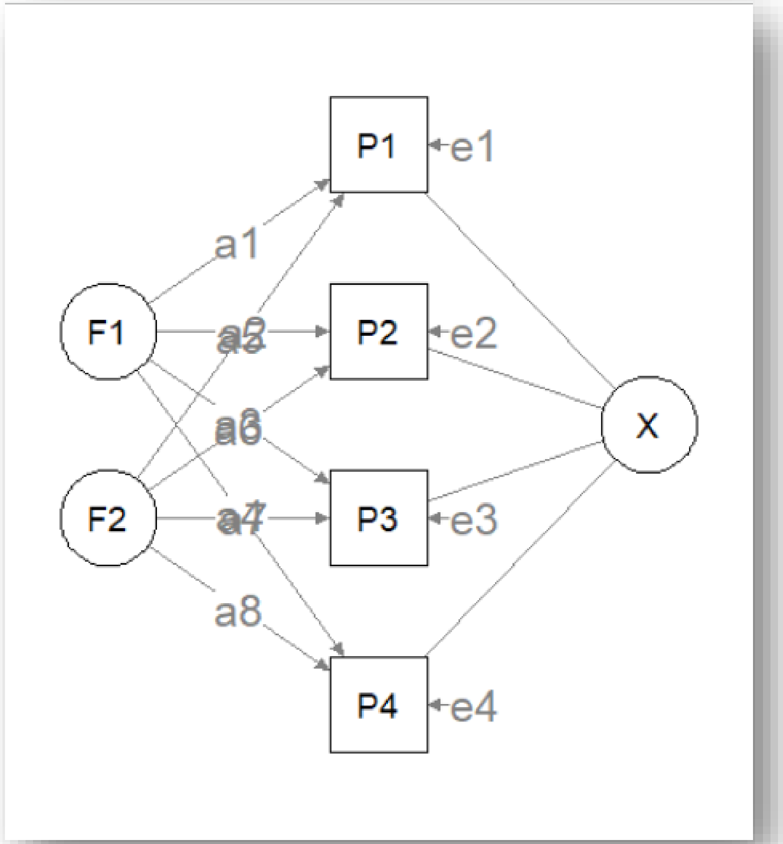




Omega: Multidimenzionalita 2







Omega: Multidimenzionalita 3

Hierarchická omega (omega hierarchical):

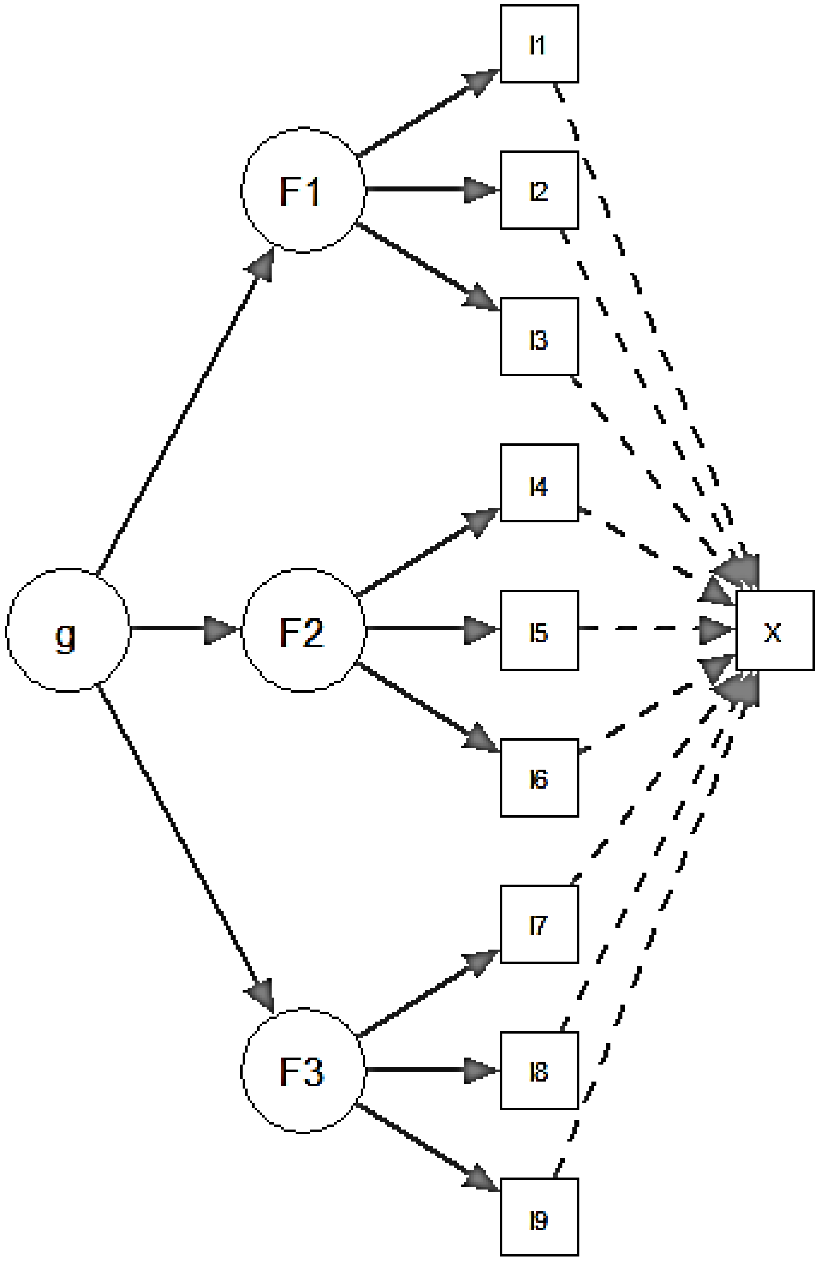
- Rozptyl součtu položek vysvětlený daným faktorem.
- V případě faktoru druhého řádu (g) jsou specifické rozptyly faktorů prvního řádu považovány za chybu.
- **Model based reliabilita:** velmi záleží na definici modelu.

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.

„Dimension-free“ omega:

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



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

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
- **$r_{x\text{-}y}$ – reliabilita rozdílu**
- $SD_{x\text{-}y}$ – SD rozdílu
- $SE_{x\text{-}y}$ – standardní chyba rozdílu
- $CI_{95\%}$ – šířka 95% intervalu spolehlivosti

$r_{xx'}$	$r_{yy'}$	r_{xy}	$r_{x\text{-}y}$	$SD_{x\text{-}y}$	$SE_{x\text{-}y}$	$CI_{95\%}$
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

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

Je evidentní, že korelace testů má opačný vliv na výslednou reliabilitu.

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

Práce s chybou měření

Využití reliability
při praktické diagnostice

Statisticky významný rozdíl

Klinicky významný rozdíl

Chyba predikce

A další...

Otázky spojené s chybou měření

Respondentovi naměřím výšku 178 cm.
Jaké otázky si mohu položit?

- Kolik měří právě teď?
- Kolik bude měřit příště?
- Kolik mu můžu naměřit příště, pokud se jeho výška nezmění?
- Kolik mu musím naměřit příště, abych mohl konstatovat, že se jeho výška změnila?

Kromě toho naměřím i jeho hmotnost 65 kg.
Jaké další otázky si mohu položit?

- Je „vyšší než těžší“?
- Je „vyšší než těžší“ oproti jiným respondentům?

Chyba měření (v CTT)

Chyba měření (SE) popisuje směrodatnou odchylku rozložení pozorovaných proměnných okolo měřené hodnoty.

- Z hlediska CTT: Rozložení pozorovaných skórů okolo pravého skóru.

Na základě toho, čeho chybu zjišťujeme, musíme:

- Zvolit správnou střední hodnotu.
- Zvolit správný postup pro výpočet SE.

Základní vzorec (chyba měření): $\sigma_e = \sigma_x\sqrt{1 - r_{xx'}}$

- Pouze úprava základního vzorce $r_{xx'} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$.

Chyba měření (v CTT)

Chyba měření (SE) popisuje směrodatnou odchylku rozložení pozorovaných proměnných okolo měřené hodnoty.

- Z hlediska CTT: Rozložení pozorovaných skórů okolo pravého skóru.

Na základě toho, čeho chybu zjišťujeme, musíme:

- Zvolit správnou střední hodnotu.
- Zvolit správný postup pro výpočet SE.

vysvětlený rozptyl

Základní vzorec (chyba měření): $\sigma_e = \sigma_x\sqrt{1 - r_{xx'}}$

- Pouze úprava základního vzorce $r_{xx'} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$.

Chyba měření (v CTT)

Chyba měření (SE) popisuje směrodatnou odchylku rozložení pozorovaných proměnných okolo měřené hodnoty.

- Z hlediska CTT: Rozložení pozorovaných skórů okolo pravého skóru.

Na základě toho, čeho chybu zjišťujeme, musíme:

- Zvolit správnou střední hodnotu.
- Zvolit správný postup pro výpočet SE.

vysvětlený rozptyl

„nevysvětlený" rozptyl

Základní vzorec (chyba měření): $\sigma_e = \sigma_x\sqrt{1 - r_{xx'}}$

- Pouze úprava základního vzorce $r_{xx'} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$.

Chyba měření (v CTT)

Chyba měření (SE) popisuje směrodatnou odchylku rozložení pozorovaných proměnných okolo měřené hodnoty.

- Z hlediska CTT: Rozložení pozorovaných skórů okolo pravého skóru.

Na základě toho, čeho chybu zjišťujeme, musíme:

- Zvolit správnou střední hodnotu.
- Zvolit správný postup pro výpočet SE.

vysvětlený rozptyl

„nevysvětlený“ rozptyl

převod rozptylu na SD

Základní vzorec (chyba měření): $\sigma_e = \sigma_x\sqrt{1 - r_{xx'}}$

- Pouze úprava základního vzorce $r_{xx'} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$.

Chyba měření (v CTT)

Chyba měření (SE) popisuje směrodatnou odchylku rozložení pozorovaných proměnných okolo měřené hodnoty.

- Z hlediska CTT: Rozložení pozorovaných skórů okolo pravého skóru.

Na základě toho, čeho chybu zjišťujeme, musíme:

- Zvolit správnou střední hodnotu.
- Zvolit správný postup pro výpočet SE.

Základní vzorec (chyba měření): $\sigma_e = \sigma_x\sqrt{1 - r_{xx'}}$

- Pouze úprava základního vzorce $r_{xx'} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$.

vysvětlený rozptyl

„nevysvětlený“ rozptyl

převod rozptylu na SD

převod ze z-skóre na hrubého skóre

Chyba měření (v CTT)

Takto spočítanou chybu měření mohu použít pro konstrukci intervalu spolehlivosti.

$$CI_i = \mathrm{E}(X) \pm z_i \sigma_e$$

- $\mathrm{E}(X)$ = očekávaná hodnota, okolo které interval konstruuji.
- σ_e = chyba měření
- z_i = kvantil normálního rozdělení

Kvantily normálního rozdělení:

- 95% CI: $z_{95\%} \cong 1{,}96$
- 90% CI: $z_{90\%} \cong 1{,}64$
- 80% CI: $z_{80\%} \cong 1{,}28$
- 68% CI: $z_{68\%} \cong 1{,}00$

Drobný zádrhel:

V testu (M = 100, SD = 15) naměříme náhodně vybranému respondentovi hrubý skór 70 IQ.

- Jaká je nejpravděpodobnější úroveň jeho pravého skóre?
- Jaký skór naměříme s nejvyšší pravděpodobností příště?

Řešení: Regresní model CTT

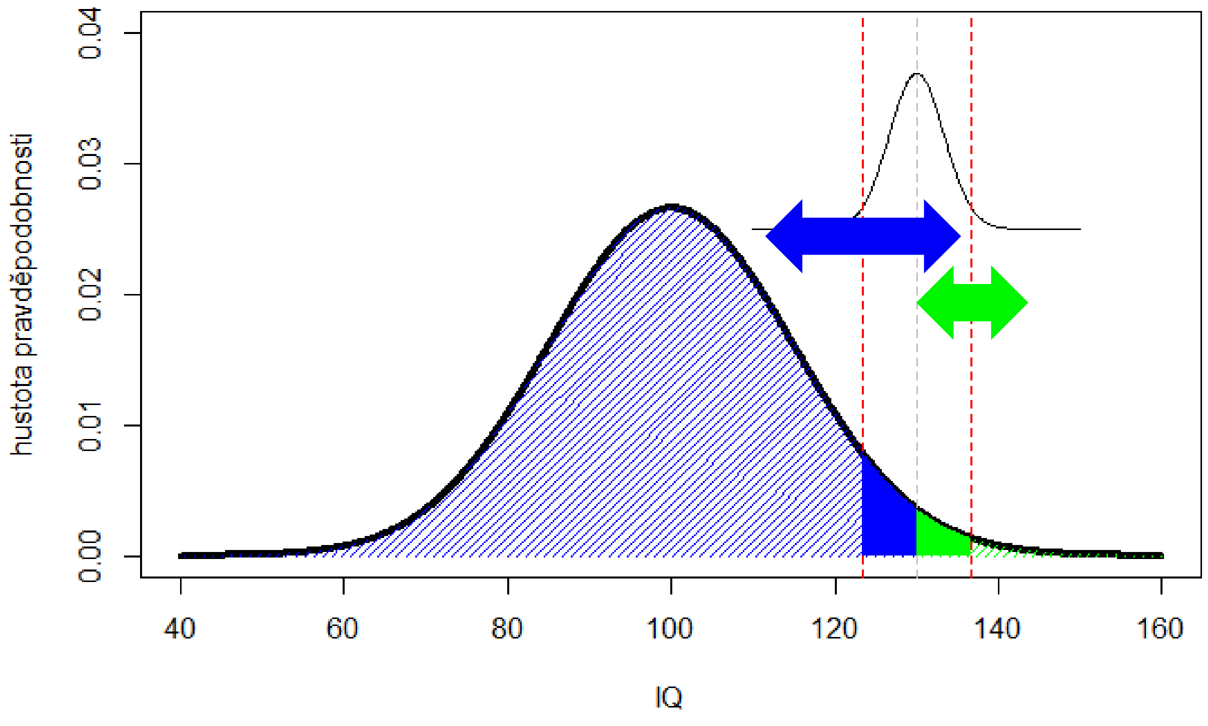
Naměřil jsem klientovi IQ 70 v inteligenčním testu.

- $r_{xx'} = 0{,}8.$
- Jaká je nejpravděpodobnější hodnota jeho „pravého“ IQ?

$X = T + e$

$\sigma_X^2 = \sigma_T^2 + \sigma_e^2$

Správná odpověď je 76.



Regresní model CTT

Naměřené hodnoty se pohybují kolem pravé hodnoty, nikoliv naopak.

- Výsledkem je tzv. **regrese k průměru**.
- Intervaly spolehlivosti jsou „asymetrické“ kolem naměřené hodnoty.

Regresní model CTT:

- $X = \tau + e$

Odhad pravého skóre:

$\mathrm{E}(T|x) = r_{xx'}x + (1 - r_{xx'})M_x$, kde

- $\mathrm{E}(T|x)$: očekávané pravé skóre na základě pozorovaného
- $r_{xx'}$: reliabilita
- M_x : průměrné skóre; ($(1 - r_{xx'})M_x$ je „intercept“)
- Čím větší reliabilita, tím větší vliv pozorovaného skóre a menší vliv průměru (a naopak).

Shrnutí: Důležité prvky práce s SE

Co je očekávanou hodnotou, okolo které interval konstruuji?

- Pozorované skóre?
- Odhad pravého skóre?
- Nula (pro rozdíl dvou skórů)?

Jak spočítám chybu pro daný účel/diagnostickou otázku?

Jaký odhad reliability nejlépe použiju pro daný účel?

Důležité vzorce

$$SE = \sigma_e = \sigma_x \sqrt{1 - r_{xx'}}$$

$$\sigma_\tau = \sqrt{r_{xx'}} \sigma_x$$

- SD pravého skóre je $\sqrt{r_{xx'}}$ krát menší než SD pozorovaného skóre
- pozor, regrese k průměru je ale $r_{xx'}$ krát!

$$\mathrm{E}(T|x) = r_{xx'} x + (1 - r_{xx'}) M_x$$

$$\sigma_{A \pm B}^2 = \sigma_A^2 \pm 2\sigma_{AB}^2 + \sigma_B^2 = \sigma_A^2 \pm 2 r_{AB} \sigma_A \sigma_B + \sigma_B^2$$

Funkce norm.dist a norm.inv v Excelu pro výpočt přesné p-hodnoty.

Scénář 1: Standardní chyba měření

Pokud jsme naměřili pozorované skóre X, jaké jiné alternativní X jsme mohli rovněž naměřit?

Slouží pro popis chyby měření a intervalu spolehlivosti jednoho jediného měření.

Velikost chyby:

$$\sigma_e = \sigma_x \sqrt{1 - r_{xx'}}$$

Středová hodnota: odhad pravého skóre

$$\mathrm{E}(T|x) = r_{xx'} x + (1 - r_{xx'}) M_x$$

Scénář 2: Chyba odhadu pravého skóre

Pokud jsme naměřili pozorované skóre X, jaká je chyba odhadu pravého skóre τ?

Vzorec je stejný, jen namísto SD pozorovaného skóre použijeme odhad SD pravého skóre:

Velikost chyby:

$$\sigma_{e(\tau)} = \sigma_\tau \sqrt{1 - r_{xx'}} = \sigma_x \sqrt{r_{xx'}} \sqrt{1 - r_{xx'}}$$

Středová hodnota:

$$\mathrm{E}(T|x) = r_{xx'} x + (1 - r_{xx'}) M_x$$

Někteří autoři tento postup doporučují, ale potíž s interpretací.

- Zajímá nás chyba na škále použité při konstrukci norem. Zpravidla tedy nepoužitelné.
- Nicméně např. WISC-5[UK] – pro standardizaci na IQ použil právě σ_τ
    - Standardizace $IQ = 15 \frac{(X - M_x)}{\sigma_x \sqrt{r_{xx'}}} + 100$ namísto běžného $IQ = 15 \frac{(X - M_x)}{\sigma_x} + 100$

Scénář 3: Standardní chyba predikce

Naměřil jsem X. V jakém rozsahu bude ležet příští měření, pokud se úroveň atributu nezmění?

- *„Zlepšil se klient v terapii?" „Je účinný výukový program?"*

Velikost chyby:

$$\sigma_{pred} = \sigma_x \sqrt{1 - r_{xx'}^2}$$

- $r_{xx'}^2$ - druhá mocnina (test-retest) reliability.
- jde o úpravu $\sigma_{pred} = \sqrt{\sigma_e^2 + \sigma_{e(\tau)}^2}$, tedy rozdíl chyby odhadu pravého skóru a chyby měření

Středová hodnota = očekávaný skór při retestu: odhad pravého skóre:

$$\mathrm{E}(T|x) = r_{xx'}x + (1 - r_{xx'})M_x$$

Scénář 4: Statisticky významný rozdíl

Tradiční označení pro rozdíl dvou nezávislých testů jedné osoby; případně rozdíl dvou osob.

Jaká je očekávaná odlišnost v měření dvěma testy?

- „Dosáhla vyššího skóru Anežka nebo Bedřich?“ „Je Cyril vyšší nebo těžší?“
- Musí být ve stejných jednotkách.

Velikost chyby:

$$\sigma_{e(A-B)} = \sqrt{\sigma_{e(A)}^2 + \sigma_{e(B)}^2} = \sigma_{ab}\sqrt{2 - r_{aa'} - r_{bb'}}$$

- Pokud jde o měření jediným testem (dvěma testy se stejnou reliabilitou), lze zjednodušit:

$$\sigma_{e(A-B)} = \sqrt{2}SE = \sigma_x\sqrt{2}\sqrt{1 - r_{xx'}}$$

Středová hodnota:

- Jde o rozdíl a očekávaný rozdíl je zpravidla žádný rozdíl, **proto zpravidla 0**.
- To není úplně pravda; pokud $r_{aa'} \neq r_{bb'}$, pak je střední hodnotou $E(\tau_A' - \tau_B') = \sqrt{r_{AA'}}(A - M) - \sqrt{r_{BB'}}(B - M)$, ale výsledek bude velmi podobný. Zanedbejte.

Scénář 5: Klinicky významný rozdíl

Liší se dva skóry téhož respondenta více či méně než u „běžných“ respondentů?

- To, že se skóry liší, neznamená, že se liší více, než bychom čekali u náhodně vybraného člověka.
- Klinické hypotézy: *„Rozkolísaný profil schopností...“*, *„Je rozdíl ‚klinicky‘ významný?“* atd.

Příklad:

- **Statisticky významný rozdíl:** *„Člověk má vyšší váhu než výšku (ve standardních jednotkách, např. IQ skórech)“*.
- **Klinicky významný rozdíl:** *„Člověk má vyšší váhu, než by odpovídalo jeho výšce, je tedy obézní.“*

Scénář 5: Klinicky významný rozdíl

Více postupů. Nejjednodušší používá pouze korelaci a je zcela shodný s postupem pro chybu predikce.

Odhad chyby:

$$\sigma_{A-B} = \sigma_{AB}\sqrt{1 - r_{AB}^2}$$

- r_{AB} je korelace testů A a B, σ_{AB} je směrodatná odchylka obou testů (musí být shodná)

Středová hodnota:

$$\mathrm{E}(B|A) = r_{AB}A + (1 - r_{AB})M_{AB}$$

Scénář 6: Více měření

Lze testovat, zda má klient celkově „rozkolísaný profil“.

- Např.: *„Liší se subtesty ve WAIS-III od celkového IQ více, než bychom čekali?“*
- Analogie F-testu u lineární regrese s více prediktory.

Poskytují jen některé diagnostické metody, není pravidlem.

Technicky vzato není ideální interpretovat „profil“, pokud test celkového rozdílu není signifikantní na zvolené *p*-hladině.

Ruční výpočet je příliš náročný.