



KONFIRMAČNÍ FAKTOROVÁ ANALÝZA

STANDA JEŽEK FSS:PSY532 2016

stan&apps.fss.muni.cz



CÍLE

- Představit základní principy CFA
- Získat schopnost interpretovat jednoduchý CFA model
- Získat schopnost vytvořit jednoduchý CFA model v R

ODBOČKY A KONSEKVENCE

- Základy statistiky nutných pro porozumění CFA
 - rozptyl a kovariance
 - lineární regrese
 - explorační faktorová analýza
 - položková analýza - vnitřní konzistence škál
- CFA je nejjednodušším strukturním modelem - krok k SEM modelům

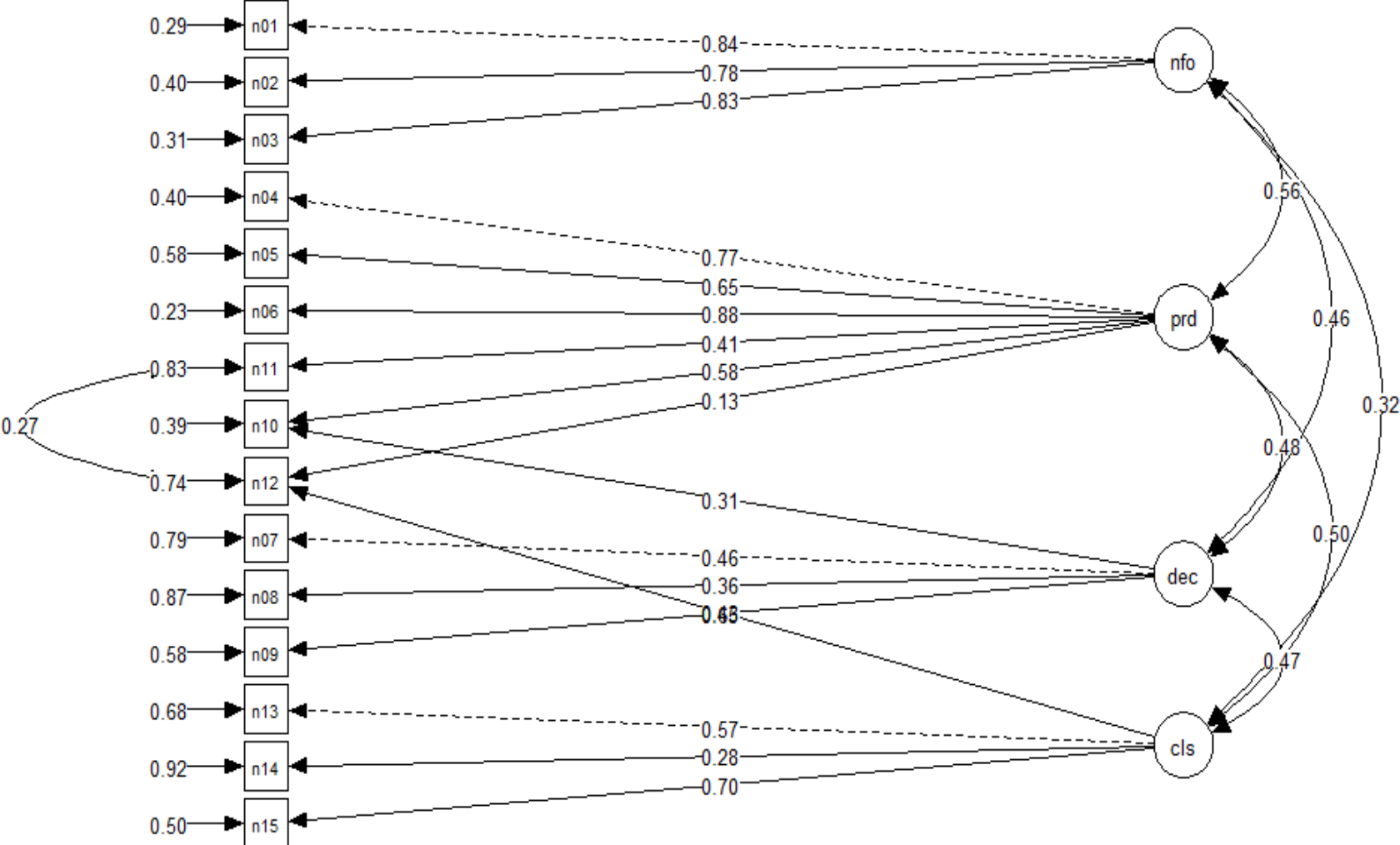
KONFIRMAČNÍ FAKTOROVÁ ANALÝZA CFA

- Cílem je prověřit **plauzibilitu** naší představy o tom, jak **položky/indikátory** společně měří nějaký **konstrukt/y** - **MEASUREMENT MODEL**
- Předpokládá se, že míra konstruktu **lineárně** predikuje/způsobuje odpověď na všechny položky
- Odpověď na položku ale může být způsobena více konstrukty a mnohé nástroje usilují měřit více konstruktů - **faktorů**
- Plauzibilita modelu podporuje víru ve **validitu** a množství konstruktů determinujících odpovědi na položky zase ovlivňuje **reliabilitu**

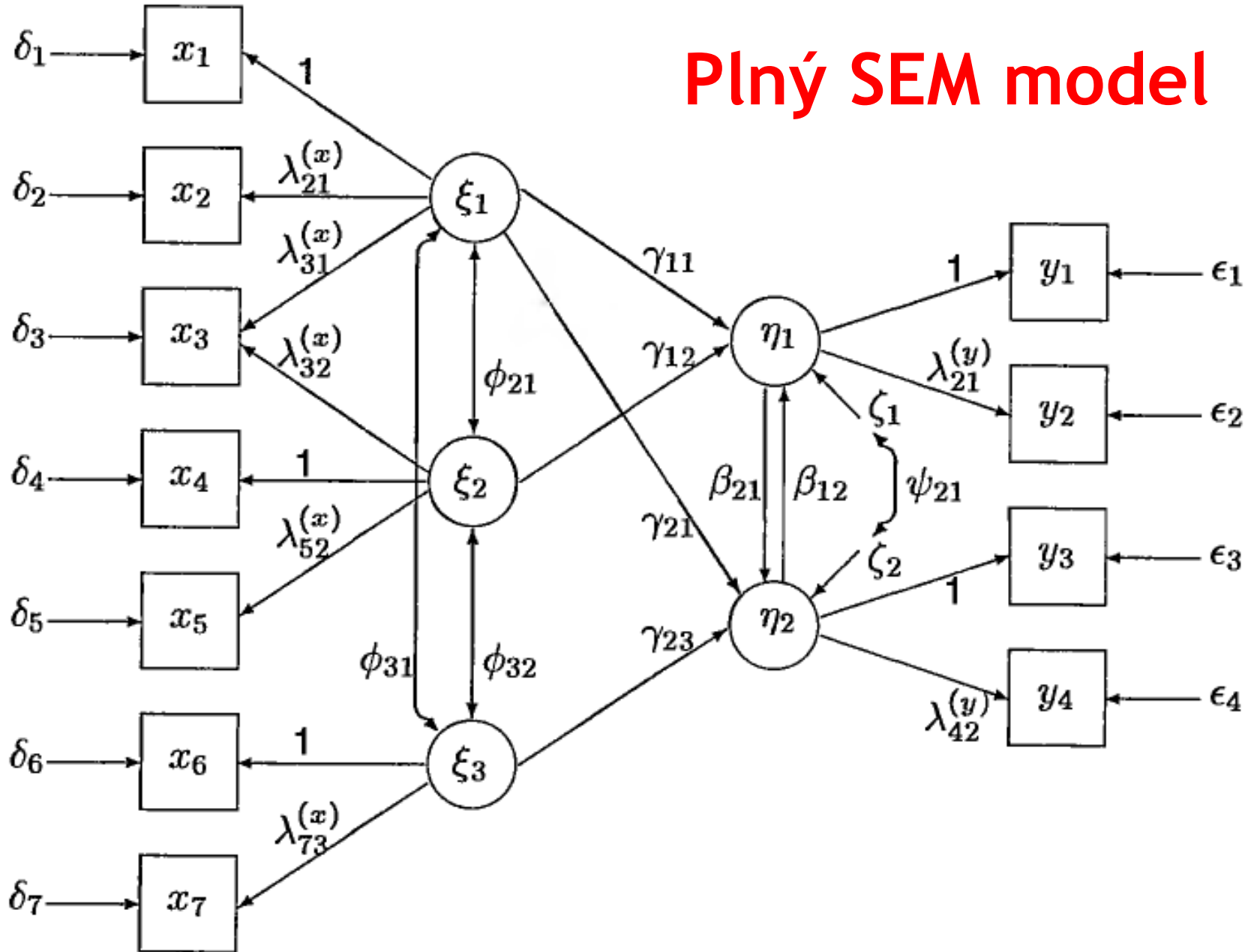
KONCEPTUÁLNÍ PRVKY MODELU

- Model je budován ze vztahů mezi měřenými (manifestními) proměnnými (položkami).
- Hypotetizujeme faktory (latentní proměnné, konstrukty), které dokáží vysvětlit vztahy mezi položkami.
- Na vstupu je tedy kovarianční (korelační) matice manifestních proměnných a na výstupu je tatáž matice, vypočítaná na základě modelu. Shoda/rozdíl mezi nimi je jedním z hlavních ukazatelů toho, jak je model dobrý (shoda modelu s daty).
- Kovariance mezi dvěma položkami může být v jednoduchém modelu vysvětlena 3 způsoby:
 - Odpovědi na obě položky jsou způsobovány stejným faktorem
 - Položky sytí jiné faktory, ovšem tyto faktory jsou zkorelované
 - 2 položky sdílí rozptyl, který nesdílí s žádnými dalšími položkami - korelace reziduí (lokální závislost)

CFA model



Plný SEM model



INTERPRETACE PARAMETRŮ CFA MODELU

- Faktorové náboje vyjadřují vliv faktoru na jednotlivé položky. Interpretují se jako regresní koeficienty.
 - Pokud položka sytí jen jeden faktor, standardizovaný koeficient vyjadřuje korelaci mezi položkou a faktorem a umocněna na druhou vyjadřuje podíl rozptylu položky vysvětlený faktorem
 - Pokud položka sytí více faktorů, standardizované koeficienty zohledňují korelaci mezi faktory a pak je nelze umocňovat na druhou. Rozptyl položky vysvětlený faktory pak zjišťujeme spíše ze standardizovaného rezidua (1-s.r.)
- Vztahy mezi faktory a vztahy mezi rezidui jsou kovariance (ve standardizovaném modelu korelace)
- Každý parametr má svou S.E., t (popř. z) a p

FAKTOROVÁ RELIABILITA (\approx VNITŘNÍ KONZISTENCE)

■ Raykovovo ρ =
$$\frac{(\text{suma nábojů})^2 (\text{rozptyl faktoru})}{(\text{suma nábojů})^2 (\text{rozptyl faktoru}) + (\text{suma rezid. rozp.})}$$

Parametry z nestandardizovaného řešení

alias McDonaldovo omega

POSTUP CFA ANALÝZY

- Specifikace modelu podle teorie (na papíře a pak v počítači) a posouzení identifikace parametrů
- Odhad parametrů modelu
- Prověření shody modelu s daty
- Případná respecifikace či specifikace alternativního modelu
- Interpretace finálního modelu, konfrontace s teorií
- Postup: často od 1F modelu. Ale jiní doporučují od nejkompexnějšího.

SPECIFIKACE MODELU

V lavaan se model definuje v textovém objektu.

1. Které položky mají být predikovány kterými faktory

- `prda =~ nfc04+nfc05+nfc06+nfc10+nfc11+nfc12`

2. Kovariance reziduí položek

- `nfc11~~nfc12`

- Jde o lokální závislosti, metodové efekty ... netoužíme po nich

3. Korelace mezi faktory

- ON/OFF nastavuje se až při odhadu parametrů modelu

- `cfa(..., orthogonal=TRUE)`

- Jednotlivě v definici modelu: `nfo~~prda`

4. Metrika faktorů - faktor musí mít škálu

- 1. Defaultně ji přebírá od první položky `prda =~`

- `1*nfc04+nfc05+nfc06+...`

- 2. Alt. lze fixovat rozptyl faktoru `prda ~~ 1*prda` (pak `prda =~`

- `NA*nfc04+...`)

POČET ODHADOVANÝCH PARAMETRŮ

Typ parametru	Počet
Volné náboje položek na faktorech	Počet položek - počet faktorů + crossloadings
Reziduální rozptyly položek	Počet položek (P)
Rozptyly faktorů	Počet faktorů (F)
Korelace faktorů	Každá 1 parametr max $F(F-1)/2$
Reziduální kovariance	Každá 1 parametr
	Celkem=

Počet odhadovaných parametrů by neměl přesáhnout počet kovariancí vstupujících do analýzy $df_{NULL} = P(P+1)/2$.

Stupně volnosti modelu $df_M = df_{NULL}$ - počet odhadovaných parametrů
Pokud model obsahuje i průměry, pak vše stoupne ještě o počet vstupních a odhadovaných průměrů (tj. obvykle počet položek).

IDENTIFIKACE PARAMETRŮ MODELU

- Máme dost informace pro jedinečné stanovení hodnoty každého parametru?
- Lze stanovit matematicky, ale pro běžné smrtníky jsou pravidla+pokus-omyl
- Pravidla
 - $df_M \geq 0$
 - Pro každý faktor máme 2 a více jedinečných položek (3, když máme jen 1F)
 - S korelovanými rezidui a crossloadings nároky na počet položek stoupají
 - Více <http://davidakenny.net/cm/identify.htm>
- Empirická underidentifikace
 - Špatný model, malé N, moc slabé/moc silné vztahy
- Problémy s identifikací program neidentifikuje - projeví se neschopností odhadu, nesmyslnými hodnotami parametru, obrovskými s.e.

ODHAD PARAMETRŮ MODELU

- V lavaan funkce cfa
- Vedle specifikace modelu a dat se zde nastavuje řada dalších věcí
 - Metoda odhadu: estimator = ML/MLR pro spojité, WLSMV pro kategorické
 - <http://lavaan.ugent.be/tutorial/est.html>
 - Zda jsou položky ordinální: ordered =
 - Zda mají faktory korelovat: orthogonal=
 - Co s chybějícími hodnotami: missing=„FIML“ vs. „listwise“
 - ...

UKAZATELE FITU - ABSOLUTNÍ

Pouze parametry správného modelu jsou správné!

- Absolutní fit - odpovídá model datům dostatečně?
 - chí-kvadrát test -test shody mezi pozorovanou a modelem implikovanou kovarianční maticí. Nechceme signifikantní rozdíl. Na větších vzorcích velmi konzervativní ☹️.
 - RMSEA - Čím menší, tím lepší. Chceme $<0,08$. Horní mez 90%intervalu spolehlivosti by neměla přesahovat $0,10$. Trestá za komplexitu.
 - CFI - Liberální, čím vyšší tím lepší. Chceme $>0,95$
 - TLI (NNFI) - Jako CFI, jen přísnější.
 - SRMR - vychází ze standardizovaných reziduí, čím menší, tím lepší, chceme $<0,08$.

UKAZATELE FITU - RELATIVNÍ

Odpovídá datům lépe model A než model B?

- BIC - Bayesian information criterion.
Zohledňuje vzorek i složitost modelu. Čím nižší, tím lépe.
- AIC - Akaike IC - jako BIC, jen nepenalizuje za vzorek

SROVNÁVÁNÍ MODELŮ LIKELIHOOD-RATIO TESTEM

- chíkvadrát LR test - rozdíl χ^2 mezi dvěma vnořenými(nested) modely (má χ^2 rozložení s $df_A - df_B$ stupni volnosti) - jen ML odhad, pro robustní nutná korekce
 - Model B je vnořený do A, když může vzniknout zafixováním jedno či více parametrů modelu A
 - `anova(fit.A, fit.B)`
 - `semTools::compareFit`
- Srovnávání modelů
- Testování signifikance parametrů

UKAZATELE FITU - CODA

- Ukazatelů je tolik, že někdy badatele vede k tomu, že si vybírají ty, které jim vychází, a zamlčují ty, které nevychází, jak by měly.
- Vhodným průvodcem je web Davida Kennyho, zde konkrétně stránka <http://davidakenny.net/cm/fit>

PŘÍPRAVA DAT - PŘEDPOKLADY

- Velikost vzorku - velká
 - Různá pravidla:
 - $N:q > 10:1$; $q_{\max} \approx v(v+1)/2$
 - $N > 200$ (pod 100 si koledujeme o problém)
- Normalita - předpoklad metody maximální věrohodnosti
 - Univariační: Problém: Šikmost $> |3|$, Strmost $> |8|$
 - i vícerozměrné normální rozložení: linearita/eliptičnost scatterplotů - psych::mardia, nebo MVN
 - Odchytky od normality řeší různé metody odhadu
- Čistá data s vyřešenými outliery

OBVYKLÉ HYPOTÉZY ŘEŠENÉ SROVNÁVÁNÍM MODELŮ

- kongenerické vs. tau-ekvivalentní vs. paralelní položky
- unidimenzionalita vs multidimenzionalita
- korelované vs. nekorelované faktory

TRABLŠŮTINK

- Co se může pokazit?
 - Model se nedopočítá - algoritmu se nepodaří spolehlivě odhadnout všechny parametry
 - Model se dopočítá, ale s chybovými hlášeními o nedůvěryhodnosti odhadů parametrů (matrix not positive definite)
 - Některé parametry mají nepřipustné hodnoty (záporné rozptyly, standardizované parametry > 1)
- Co za tím může být?
 - Příliš složitý model (pro daná data) >> ubrat parametry
 - Data nesplňující požadavky - normalita, outlieři >> transformace, balíčky, jiné metody odhadu
 - Příliš špatný model
 - Příliš malý vzorek

DALŠÍ PRVKY SPECIFIKOVÁNÍ MODELU POMOCÍ OPERÁTORU *

Pomocí * můžeme dále konkretizovat specifikaci modelu

■ Labels - pojmenování parametru

■ `nfo =~ nfc01 + naboj1*nfc02 + naboj2*nfc03`

■ Fixing - nastavení pevné hodnoty parametru

■ `nfo =~ 1*nfc01 + nfc02 + nfc03` (to, co se dělá automaticky)

■ `nfo =~ 1*nfc01 + 2*nfc02 + 3*nfc03`

■ Nastavení startovacích hodnot (když se odhad nedaří)

■ `nfo =~ nfc01 + start(0.8)*nfc02 + nfc03`

■ Equality constraints - nastavení rovnosti dvou nebo více parametrů

■ řešení problémů s identifikací (odhaduje se jeden parametr místo více)

■ př. `nfo =~ nfc01+p23*nfc02+p23*nfc03` (p23 je label). Alternativně
`nfo =~ nfc01+nfc02+equal („nfo~="nfc02"")*nfc03`

SAMOSTATNÉ CVIČENÍ

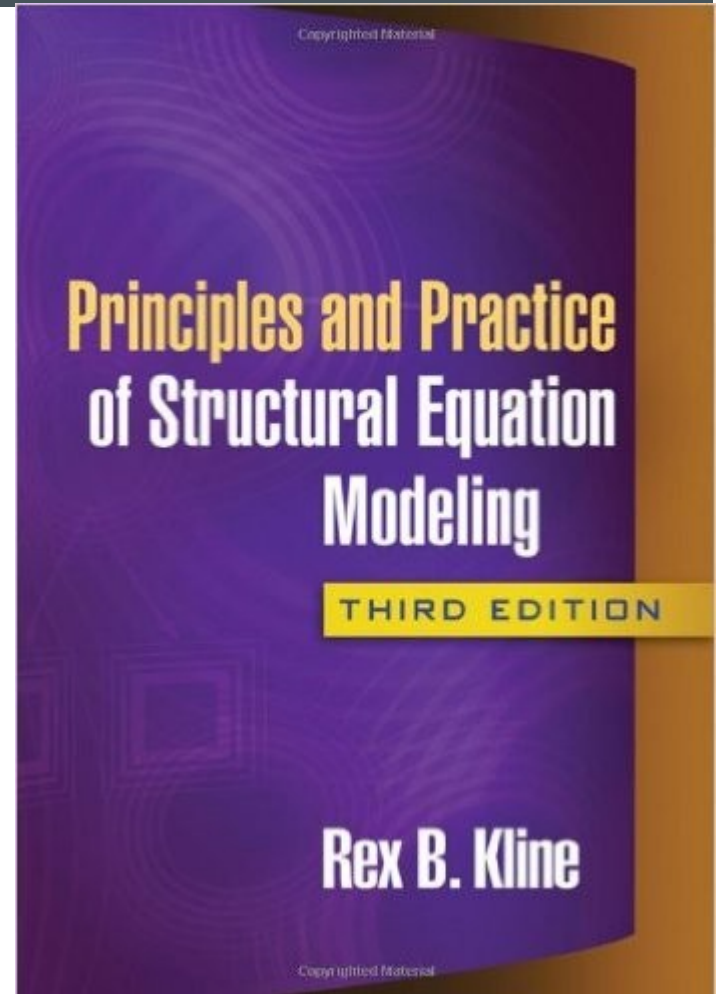
Na datech z ECR-R, která jste analyzovali pomocí EFA vytvořte

1. teoretický model (1-6 AV, 7-9 AX korelované)
2. model, který byste udělali na základě vašich výsledků EFA
3. nejlépe fitující model, který s pomocí modifikačních indexů dokážete
4. nejlepší model. U něj zkuste argumentovat, proč je nejlepší, a interpretujte jeho parametry.

KAM DÁL

- Multi-group CFA - testování invariance napříč skupinami
- Korektní metody odhadu pro ordinální proměnné a proměnné ne normálně rozložené
- SEM modely - kauzální vztahy mezi latentními proměnnými

- <http://www.amazon.com/Principles-Practice-Structural-Equation-Methodology/dp/1606238760>



ÚČEL

MULTIGROUP CFA/SEM

- Moderační hypotézy - liší se mezi skupinami vztahy mezi proměnnými (latentními či manifestními) - třeba korelace faktorů
- Hypotézy o odlišné struktuře v různých skupinách (př. počet faktorů)
- Řešení otázky invariance modelů měření (measurement invariance)

ZÁKLADNÍ PRINCIPY

MULTIGROUP CFA/SEM

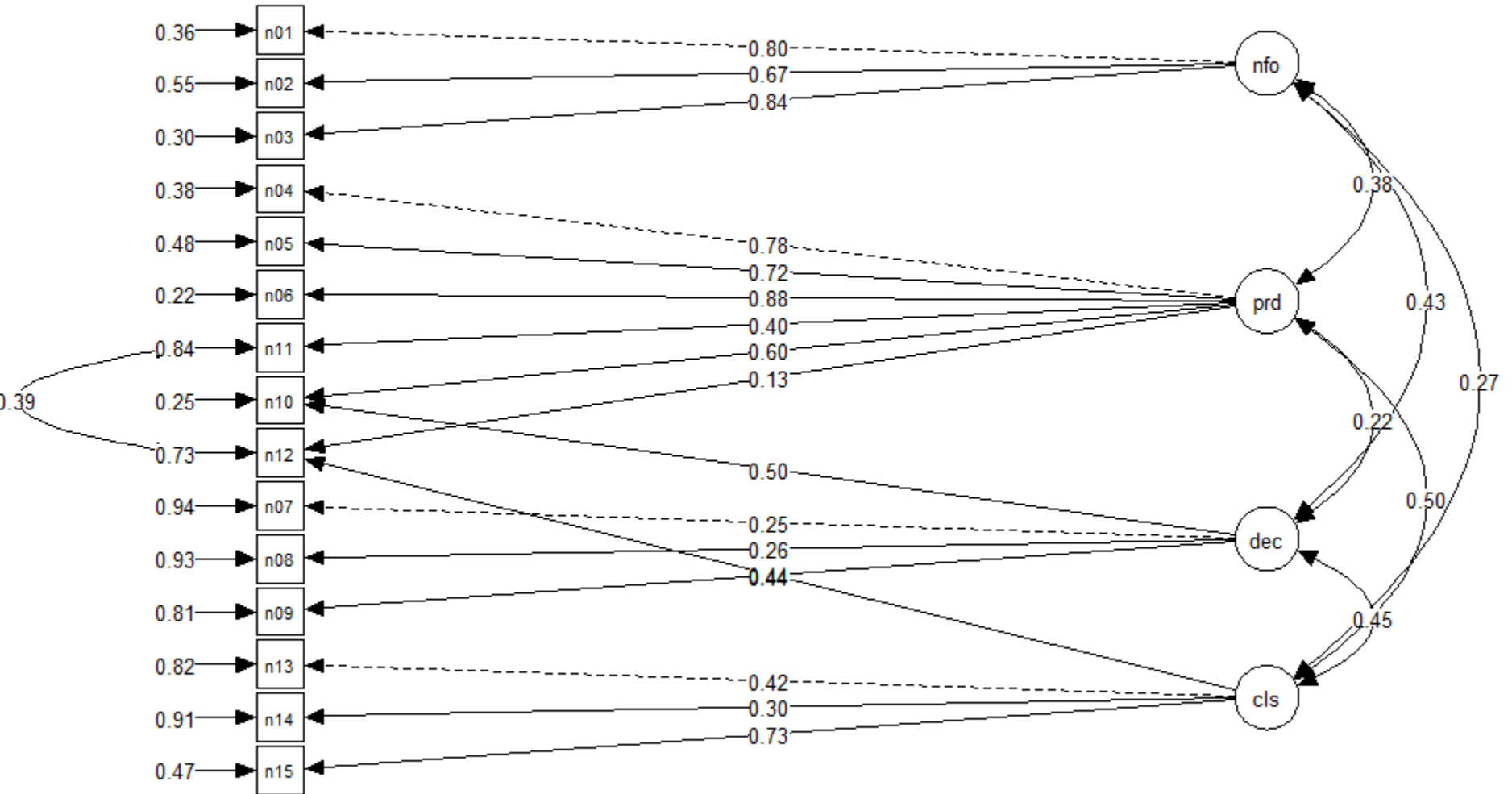
- Model se odhaduje **současně** ve 2+ skupinách
 - ... jedny společné ukazatele fitu
- Lze stanovit, v čem jsou modely stejné a v čem se mezi skupinami mají lišit
- Hypotézy o shodě parametrů mezi skupinami ověřujeme porovnáváním fitu modelů, v nichž jsou parametry jednou specifikované jako odlišné a pak jako shodné napříč skupinami

ZÁKLADNÍ PRINCIPY

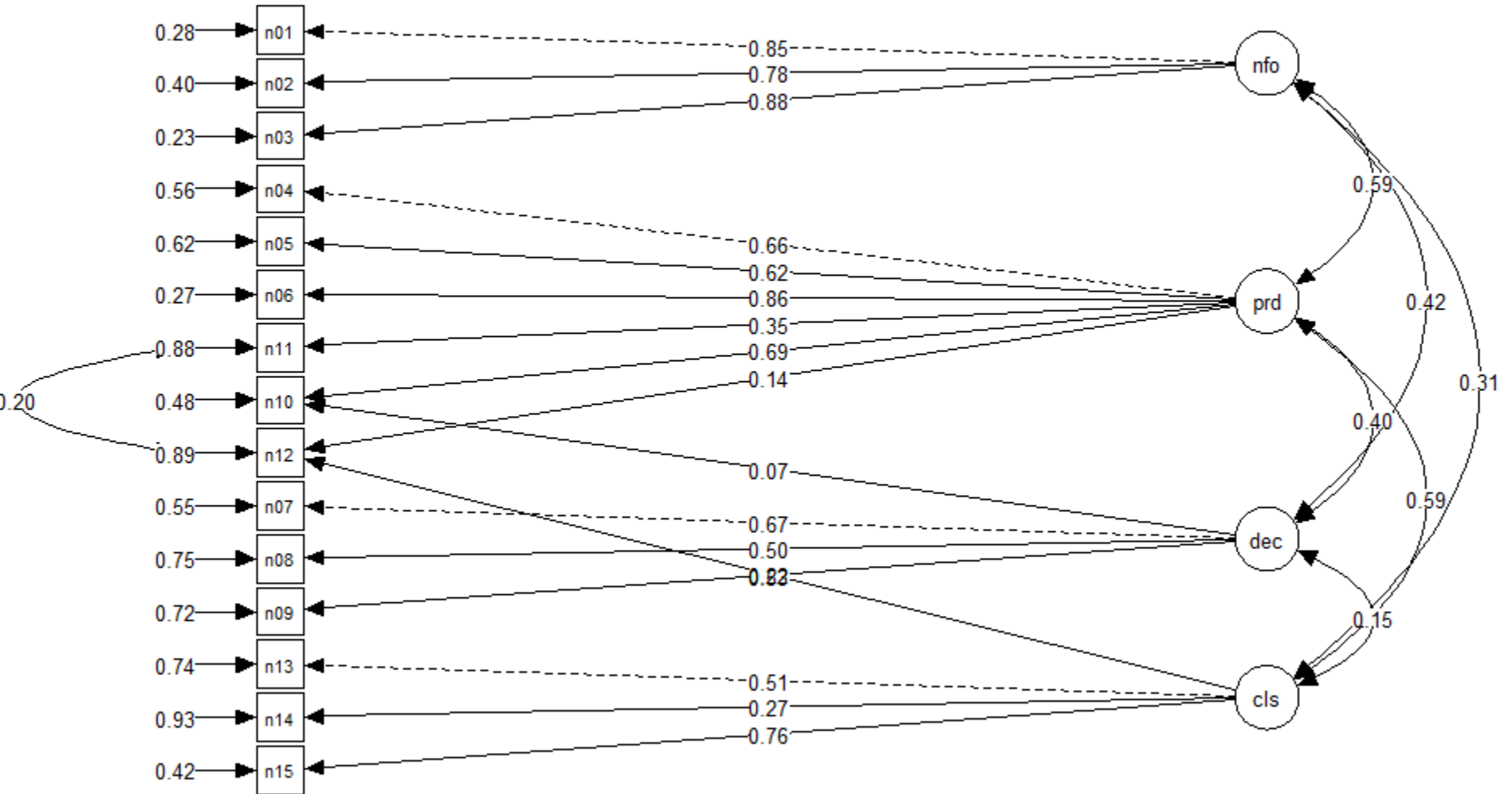
MULTIGROUP CFA/SEM

- Model se odhaduje **současně** ve 2+ skupinách
 - ... jedny společné ukazatele fitu
- Lze specifikovat, v čem jsou modely stejné a v čem se mezi skupinami mají lišit
- Hypotézy o shodě parametrů mezi skupinami ověřujeme porovnáváním fitu modelů, v nichž jsou parametry jednou specifikované jako odlišné a pak jako shodné napříč skupinami

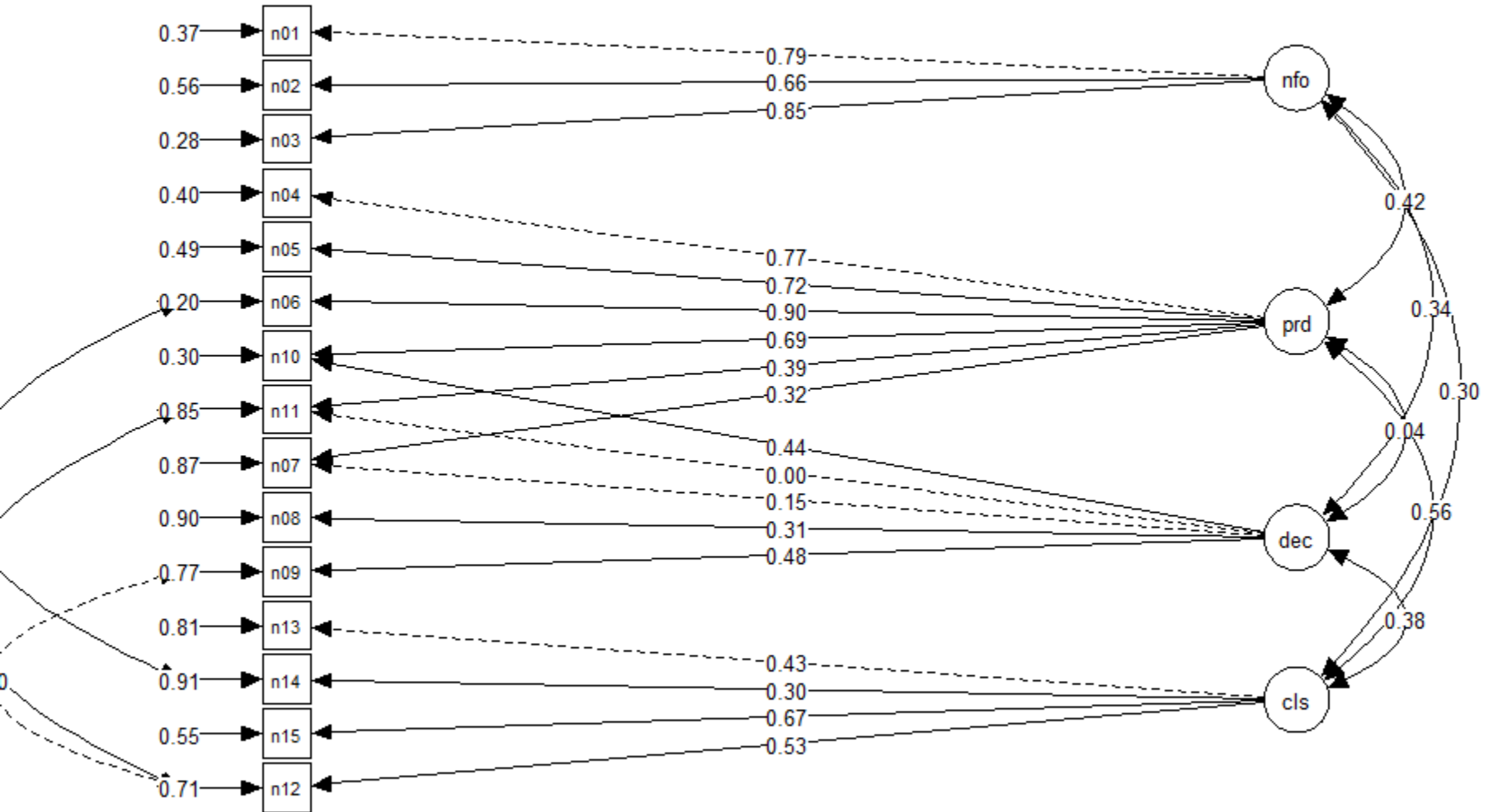
Studenti psicologie



Ne-Studenti psicologie



Studenti psicologie - tuzing



Ne-Studenti psicologie - tuzing

