

# Faktorová analýza

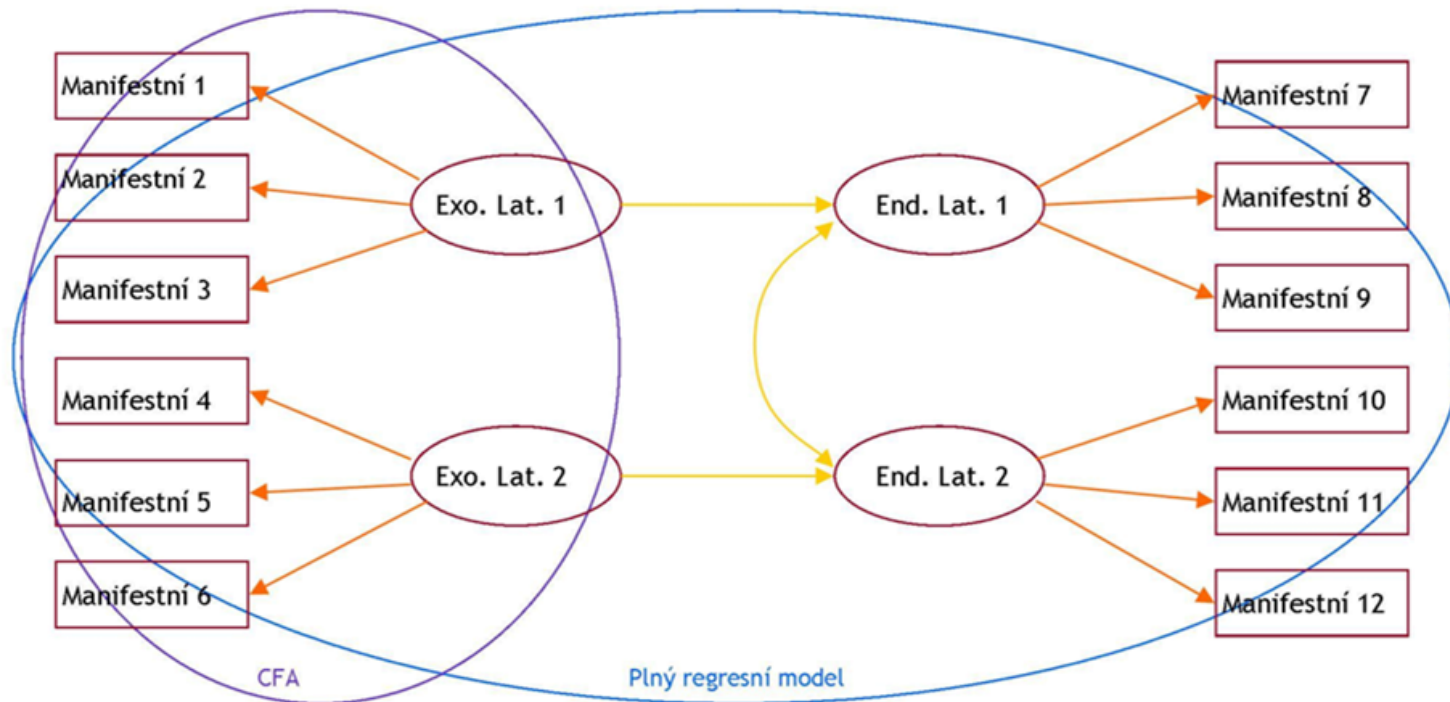
PSY252 Statistická analýza dat v psychologii II

9.12.2009

# Latentní a manifestní proměnné

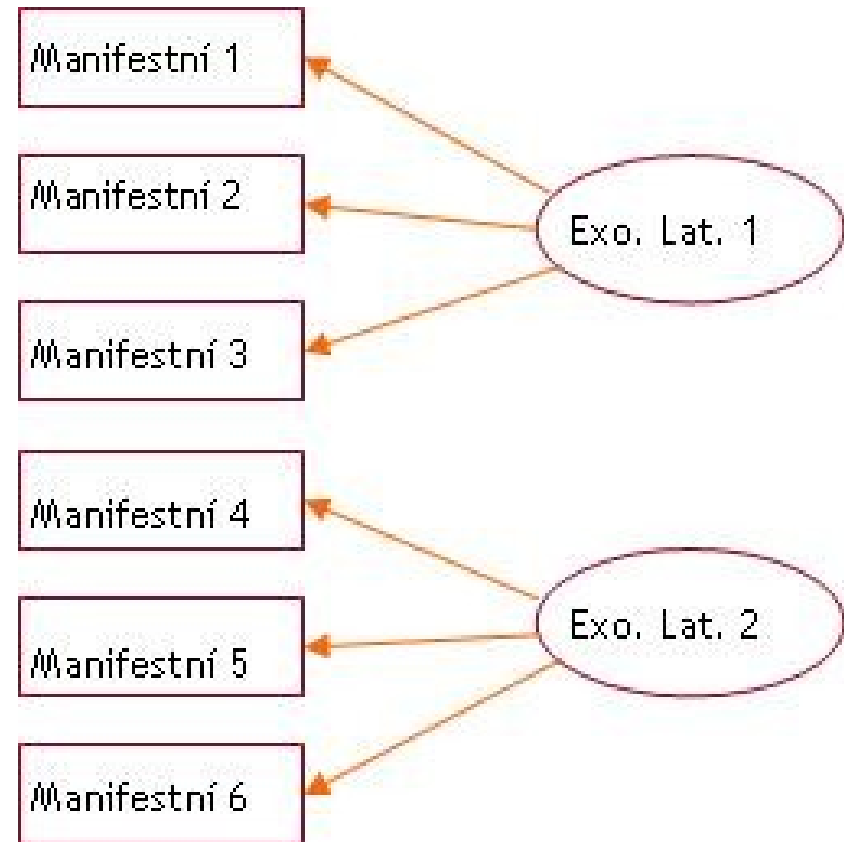
## ► Perspektiva CTT:

- (pro)Jevy, které spolu nějakým způsobem souvisejí, mají stejnou podstatu, jsou určeny stejnou *latentní proměnnou*
- LP je hypotetický konstrukt, odvozený právě z manifestací (jevů), které spolu nějakým způsobem kovariují
- *Manifestní proměnné* tedy sdílí nějakou část svého celkového rozptylu



# Faktorový model

- ▶ Dvě složky rozptylu manifestní proměnné ve faktorovém modelu:
  - ▶ Komunalita: sdílený (faktorový) rozptyl – určený latentní proměnnou, „společný“ s ostatními proměnnými
  - ▶ Unicity: jedinečnost, vlastní část rozptylu proměnné
  - ▶ Unicity = 1 – komunalita
  - ▶ CTT: Unicity zahrnuje jednak část „vlastního rozptylu“ proměnné, jednak chybu měření
- ▶ Sdílený rozptyl je patrný z korelační matice MP; jsou-li korelace MP nenulové, potom MP sdílí nějakou část



# Faktorový model

- ▶ Korelující proměnné mohou být nahrazeny jednou proměnnou, která je jejich lineární kombinací – faktorem.
- ▶ Faktorový náboj ( $F_x$ ,  $F_y$ ) je interpretován jako korelace původní proměnné s daným faktorem.
- ▶ Komunalita  $h^2 = F_x P1^2 + F_y P1^2$  je faktorový rozptyl položky, podíl rozptylu položky vyčerpaný daným faktorovým řešením.
- ▶ „Dobrá struktura“ je požadavek na jasnost faktorové matice. Každá položka by měla vysoko skórovat v právě jednom faktoru, každý faktor by měl obsahovat dva nebo více vysokých faktorových nábojů.

Korelační matice	P1	P2	P3	P4
P1	1	-,14	,74	,08
P2	-,14	1	,19	,59
P3	,74	,19	1	,17
P4	,08	,59	,17	1

Faktorová matice	F1	F2	$h^2$
P1	,73	-,60	,89
P2	,47	,77	,81
P3	,87	-,35	,87
P4	,58	,66	,77

# Explorativní a konfirmatorní FA

---

- ▶ Explorativní faktorová analýza je analytický postup, jehož smyslem je nalézt optimální matici faktorových nábojů, které maximálně zjednoduší korelační matici
  - ▶ = při co nejmenším počtu faktorů vysvětlí co největší podíl celkového rozptylu
- ▶ „Redukce korelační matice“
- ▶ Metoda maximální věrohodnosti (maximum likelihood)
  - ▶ Vlastní FA; zdůrazňuje specifické faktory
- ▶ Analýza hlavních komponent (principal components)
  - ▶ Postupný rozbor sdílených rozptylů; zdůrazňuje g-faktor
- ▶ Konfirmatorní faktorová analýza je test hypotézy o korelační a faktorové matici
  - ▶ Matice je nulová
  - ▶ Matice má konkrétní strukturu
- ▶ Kdykoliv je to možné, měli bychom se snažit o použití CFA – tedy formulovat hypotézy, spíše než dojit data
  - ▶ Software nám bohužel nevychází vstříc
    - ▶ STATISTICA – SEPATH
    - ▶ SPSS – AMOS
    - ▶ LISREL, M+ a další

# Předpoklady použití FA

---

- ▶ Možnost vytvoření smysluplné korelační matice:
  - ▶ Ordinální úroveň měření
  - ▶ Rozložení proměnných nesmí být extrémně šikmé
- ▶ Proměnné musí pocházet zhruba ze stejné domény
  - ▶ Až na speciální případy nemá smysl analyzovat jednotlivé položky osobnostního dotazníku společně se součtovými skóry jiného či proměnnými úplně jiného charakteru (výsledky výkonového testu)
- ▶ Smysluplný počet faktorů:
  - ▶ 3 při předpokladu jediného faktoru
  - ▶  $k \cdot 2$  při předpokladu  $k$  faktorů (jinak nemůže vzniknout Thurstonova struktura)
- ▶ Adekvátní počet měření
  - ▶ Málo je málo a moc je moc ☺
    - ▶ Absolutní minimum velikosti vzorku je  $N > 5 \cdot m_p$  a současně  $N > 20 \cdot k$ , ideálně od  $N > 20 \cdot m_p$
    - ▶ Extrémně velké soubory poskytují nepříjemně přesné odhady parametrů:
      - V CFA paradoxně dochází k zamítnutí jakéhokoliv modelu
      - V ML nikdy nevyjde uspokojivě test dobré shody
      - Proto se zavádí tzv.  $\chi^2/df$  ratio:  $\chi^2/df$  by měl poskytovat hodnotu okolo 2, nikdy více než 5

# Komunality

**Communalities**

	Initial	Extraction
LI07 Rodiče jsou nespokojeni s mým školním prospěchem.	1,000	,633
LI09 Mám pocit, že nemůžu zvládnout učivo tak, jak bych měl/a.	1,000	,668
LI10 Často nosím poznámky.	1,000	,668
LI11 Mám neshody a problémy s některými z učitelů.	1,000	,636
LI12 Škola mě nebaví, nejraději bych už do školy nechodil/a.	1,000	,537
LI13 V naší třídě jsou špatné vztahy.	1,000	,519
LI14 Někteří z mých vrstevníků se ke mně chovají nepřátelsky.	1,000	,564
LI16 S většinou svých vrstevníků si nerozumím.	1,000	,541
LI17 Schází mi dobrý kamarád/kamarádka.	1,000	,349

Extraction Method: Principal Component Analysis.

**Communalities<sup>a</sup>**

	Initial	Extraction
LI07 Rodiče jsou nespokojeni s mým školním prospěchem.	,305	,494
LI09 Mám pocit, že nemůžu zvládnout učivo tak, jak bych měl/a.	,243	,447
LI10 Často nosím poznámky.	,129	,189
LI11 Mám neshody a problémy s některými z učitelů.	,198	,579
LI12 Škola mě nebaví, nejraději bych už do školy nechodil/a.	,230	,322
LI13 V naší třídě jsou špatné vztahy.	,220	,343
LI14 Někteří z mých vrstevníků se ke mně chovají nepřátelsky.	,257	,448
LI16 S většinou svých vrstevníků si nerozumím.	,230	,339
LI17 Schází mi dobrý kamarád/kamarádka.	,147	,163

Extraction Method: Maximum Likelihood.

a. One or more communality estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

- ▶ **Přehled komunalit...**
  - ▶ V PC vždy vyšší. požadavek alespoň 0,7 teoreticky!
  - ▶ V případě ML pozor na tzv. nevlastní řešení (Heywoodův případ) – faktorová matice je problematická
    - ▶ Znamená, že některá z položek „vyčnívá“, je sama o sobě faktorem

# Eigenvalue & Explained variance

## ▶ Přehled vysvětleného rozptylu

### ▶ Eigen value – vlastní hodnota

- ▶ Suma eigenvalues vždy rovna počtu položek 😊
- ▶ Vypovídá o poměru rozptylu vysvětleného daným faktorem/komponentou vzhledem k celku

### ▶ Obdobně procentuální údaj

### ▶ Eigenvalue je obvykle kritériem volby počtu interpretovaných faktorů/komponent

- ▶ Eigen > 1
- ▶ Nemá smysl interpretovat faktory, které vysvětlují méně než „jednu“ proměnnou

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2,108	23,421	23,421
2	1,926	21,402	44,822
3	1,079	11,992	56,814
4	,903	10,031	66,845
5	,716	7,954	74,798
6	,635	7,058	81,857
7	,615	6,834	88,691
8	,520	5,782	94,473
9	,497	5,527	100,000

Extraction Method: Principal Component Analysis.

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2,108	23,421	23,421
2	1,926	21,402	44,822
3	1,079	11,992	56,814
4	,903	10,031	66,845
5	,716	7,954	74,798
6	,635	7,058	81,857
7	,615	6,834	88,691
8	,520	5,782	94,473
9	,497	5,527	100,000

Extraction Method: Maximum Likelihood.



# Je řešení smysluplné?

## Test dobré shody a reprodukovávaná matice

- ▶ Test dobré shody (pouze ML)
  - ▶ Test hypotézy o residuální matici
  - ▶ !!! Testujeme hypotézu o tom, že residuální matice je nulová – tedy naším požadavkem je dojít k *neprůkaznému testu*
  - ▶ V praxi problematické, na velkých souborech je test vždy průkazný a na malých průkaznost nespolehlivá
  - ▶ Proto požadavek na Chi<sup>2</sup>/df ratio okolo hodnoty 2
- ▶ Residuální matice by neměla obsahovat věcně významné korelace (dejme tomu do 10%)

Goodness-of-fit Test		
Chi-Square	df	Sig.
28,821	12	,004

Residual <sup>b</sup>				
LI07 Rodiče jsou nespokojení s mým školním prospěchem.			,000	
LI09 Mám pocit, že nemůžu zvládnout učivo tak, jak bych měl/a.	,000			-
LI10 Často nosím poznámky.	,011	-,011		
LI11 Mám neshody a problémy s některými z učitelů.	,002	,000		-
LI12 Škola mě nebaví, nejraději bych už do školy nechodil/a.	-,007	,007		
LI13 V naší třídě jsou špatné vztahy.	,017	-,029		-
LI14 Někteří z mých vrstevníků se ke mně chovají nepřátelsky.	-,013	,000		
LI16 S většinou svých vrstevníků si nerozumím.	-,010	,023		-
LI17 Schází mi dobrý kamarád/kamarádka.	,021	,009		

Extraction Method: Maximum Likelihood.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 5 (13,0%) nc

# Komponentová/faktorová matice

Component Matrix<sup>a</sup>

	Component		
	1	2	3
LI07 Rodiče jsou nespokojení s mým školním prospěchem.	,760		-,230
LI12 Škola mě nebaví, nejraději bych už do školy nechodil/a.	,702		-,204
LI09 Mám pocit, že nemůžu zvládnout učivo tak, jak bych měl/a.	,653		-,487
LI11 Mám neshody a problémy s některými z učitelů.	,609		,510
LI16 S většinou svých vrstevníků si nerozumím.		,732	
LI14 Někteří z mých vrstevníků se ke mně chovají nepřátelsky.		,729	,172
LI13 V naší třídě jsou špatné vztahy.		,708	-,132
LI17 Schází mi dobrý kamarád/kamarádka.		,587	
LI10 Často nosím poznámky.	,482		,660

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Factor Matrix<sup>a</sup>

	Factor		
	1	2	3
LI07 Rodiče jsou nespokojení s mým školním prospěchem.	,666		,221
LI11 Mám neshody a problémy s některými z učitelů.	,603		-,462
LI12 Škola mě nebaví, nejraději bych už do školy nechodil/a.	,548		,134
LI09 Mám pocit, že nemůžu zvládnout učivo tak, jak bych měl/a.	,542		,387
LI10 Často nosím poznámky.	,365		-,235
LI14 Někteří z mých vrstevníků se ke mně chovají nepřátelsky.		,656	-,128
LI13 V naší třídě jsou špatné vztahy.		,581	
LI16 S většinou svých vrstevníků si nerozumím.		,581	
LI17 Schází mi dobrý kamarád/kamarádka.		,395	

Extraction Method: Maximum Likelihood.

a. 3 factors extracted. 10 iterations required.

- ▶ Primární výstup PC/ML, obsahuje matici nerotovaných faktorových nábojů
- ▶ Kontrola požadavků na dobrou strukturu! Za nepodstatné lze považovat pouze náboje pod 0,1
- ▶ Pokud není jasná dobrá struktura, rotujeme.

# Rotovaná matice

Rotated Component Matrix <sup>a</sup>			
	Component		
	1	2	3
LI14 Někteří z mých vrstevníků se ke mně chovají nepřátelsky.	,736		,148
LI16 S většinou svých vrstevníků si nerozumím.	,735		
LI13 V naší třídě jsou špatné vztahy.	,701	,104	-,131
LI17 Schází mi dobrý kamarád/kamarádka.	,587		
LI09 Mám pocit, že nemůžu zvládnout učivo tak, jak bych měl/a.		,814	
LI07 Rodiče jsou nespokojení s mým školním prospěchem.		,764	,215
LI12 Škola mě nebaví, nejraději bych už do školy nechodil/a.		,704	,202
LI10 Často nosím poznámky.			,816
LI11 Mám neshody a problémy s některými z učitelů.		,239	,759

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 4 iterations.

Rotated Factor Matrix <sup>a</sup>			
	Factor		
	1	2	3
LI14 Někteří z mých vrstevníků se ke mně chovají nepřátelsky.	,658		,119
LI16 S většinou svých vrstevníků si nerozumím.	,582		
LI13 V naší třídě jsou špatné vztahy.	,579		
LI17 Schází mi dobrý kamarád/kamarádka.	,395		
LI09 Mám pocit, že nemůžu zvládnout učivo tak, jak bych měl/a.		,667	
LI07 Rodiče jsou nespokojení s mým školním prospěchem.		,657	,241
LI12 Škola mě nebaví, nejraději bych už do školy nechodil/a.		,514	,235
LI11 Mám neshody a problémy s některými z učitelů.		,183	,737
LI10 Často nosím poznámky.		,139	,411

Extraction Method: Maximum Likelihood.  
Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 4 iterations.

- ▶ Rotovaná matice je výsledek pokusu „vyčistit“ řešení při zachování stávajících dimenzí
- ▶ Pokud ani rotovaná matice nedává smysl, opouštíme FA jako řešení daného problému...