MUNI FSS Katedra psychologie

Faktorová analýza

PSYb2590: Základy psychometriky | **Přednáška 4**

12. 3. 2024 | Petr Palíšek & Hynek Cígler (& Adam Ťápal)

FA v kostce

Pokud

$$\mathbf{R}_k^* = \boldsymbol{\Delta}_k(\boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k\boldsymbol{\Lambda}_k' + \mathbf{I})\boldsymbol{\Delta}_k = \boldsymbol{\Delta}_k\boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k\boldsymbol{\Lambda}_k'\boldsymbol{\Delta}_k + \boldsymbol{\Delta}_k^2$$

a zároveň

$$\boldsymbol{\Sigma}_k^* = \begin{vmatrix} \boldsymbol{\Phi}_k + \mathbf{I} & \boldsymbol{\Phi}_k\boldsymbol{\Lambda}_{sk}' \\ \boldsymbol{\Lambda}_{sk}\boldsymbol{\Phi}_k & \boldsymbol{\Lambda}_{sk}\boldsymbol{\Phi}_k\boldsymbol{\Lambda}_{sk}' + \mathbf{I} \end{vmatrix}$$

pak platí:

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

Metaforické pochopení FA nestačí. Je nezbytně nutné skutečně rozumět analytickému principu.

Uvedené vzorce je bezpodmínečně nutné chápat, znát a umět použít.

Jsou jednoduché; zbytek hodiny se je pokusíme interpretovat.

Konceptuální úvod do FA

- Faktorová analýza a realismus
- Latentní proměnné
- Faktorová analýza
  - Specifikace FA
  - Datový model FA
  - Common Factor Model
  - Druhy FA

Validita v realismu

Atribut existuje a kauzálně způsobuje pozorování

Jak něco takového empiricky testovat?

Velmi obtížné, ale:

1. **Daný předpoklad lze formulovat jako statistický model**

   "Jak by měla vypadat pozorování (data), kdyby byl předpoklad pravdivý?"

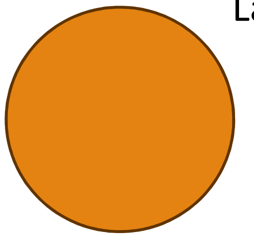
2. **Následně lze porovnat očekávání s pozorováním** – nejsou-li od sebe moc daleko, získáváme podporu pro předpoklad (a v důsledku pro validitu nástroje)

Specifikace očekávání

Typicky se komunikuje symbolicky, symboly ale implikují matematickou strukturu modelu

Stavební kameny:

Latentní proměnná (nepozorovatelná)

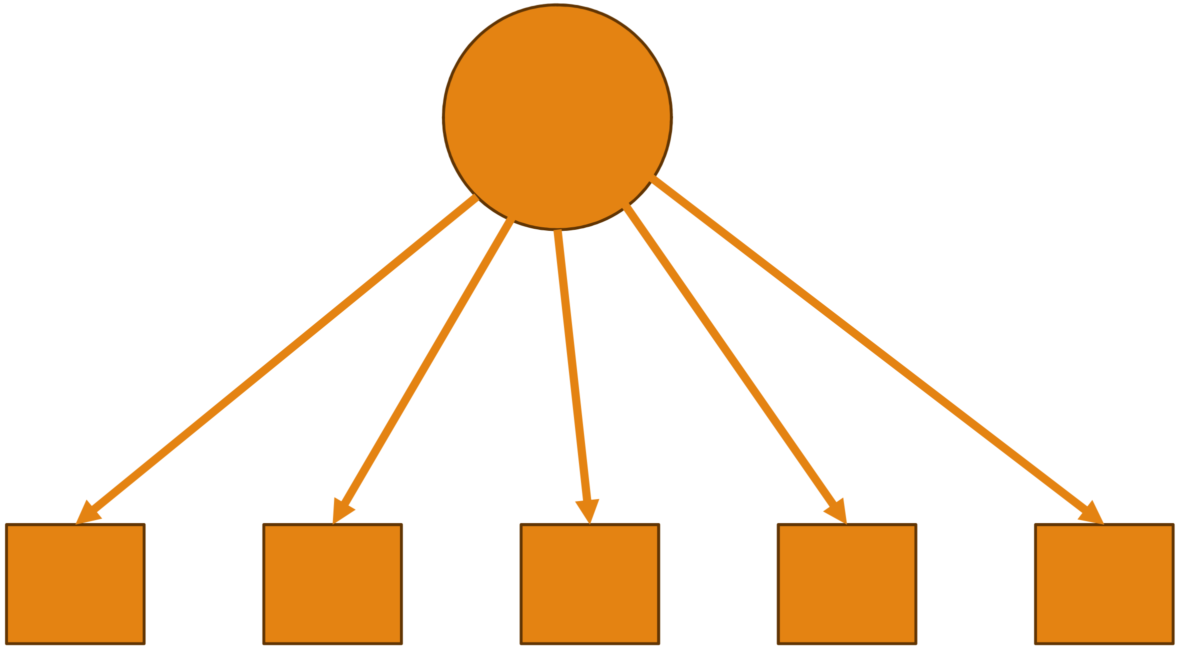


Manifestní proměnná (přímo pozorovatelná)

Kauzální vztah

Korelace (kovariance)

Specifikace modelu FA



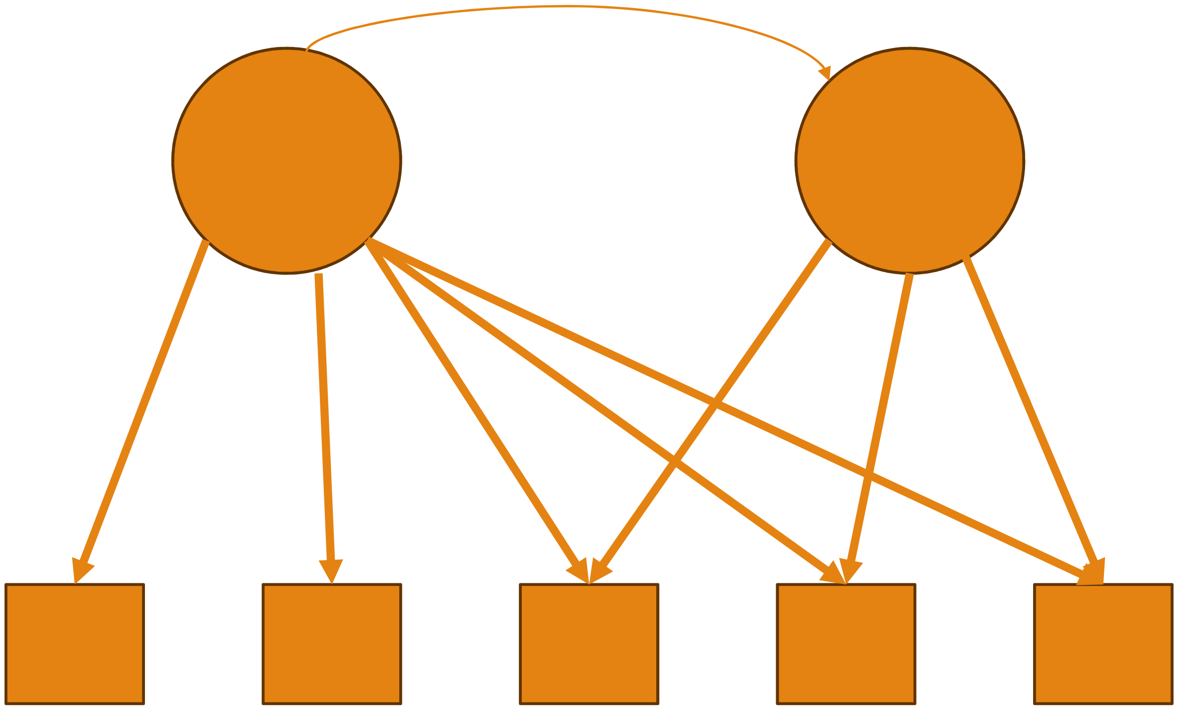
Specifikace modelu FA

Předpoklady:

1. Vztahy (kovariance, korelace) mezi manifestními proměnnými jsou vysvětleny jednou latentní proměnnou
2. Po kontrole efektu latentní proměnné jsou manifestní proměnné vzájemně nezávislé (parciální korelace příslušných MVs jsou po kontrole efektu LV nulové**)**

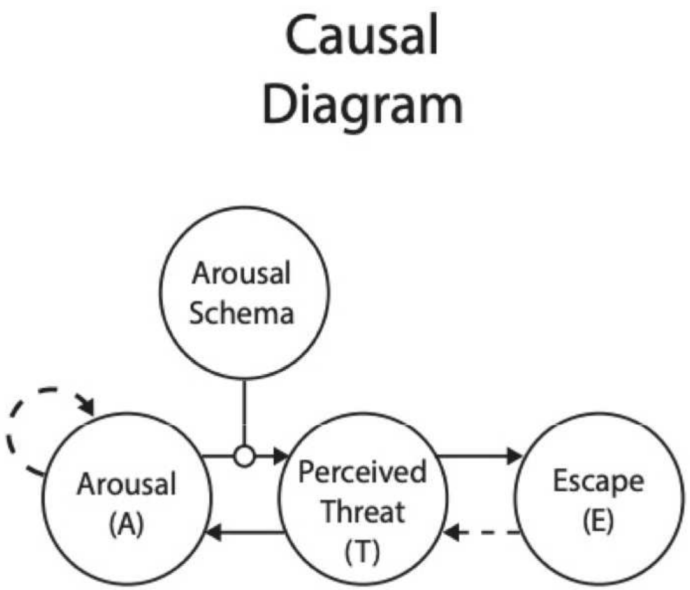
Specifikace modelů (obecně)

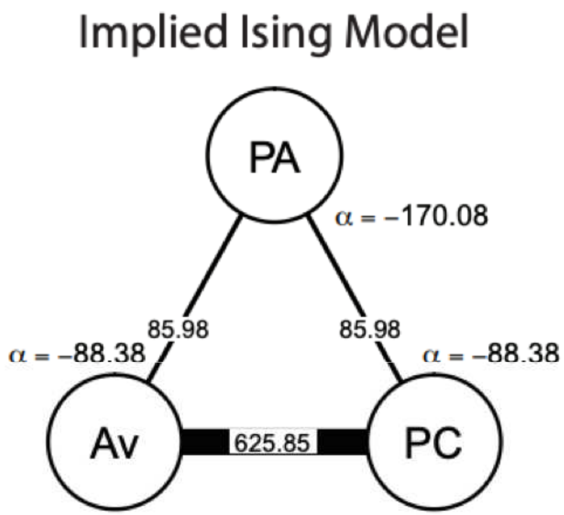
Na základě teorie jde nicméně specifikovat mnoho různých variant např.:



Specifikace modelů (obecně)

Na základě teorie jde nicméně specifikovat mnoho různých variant např.:

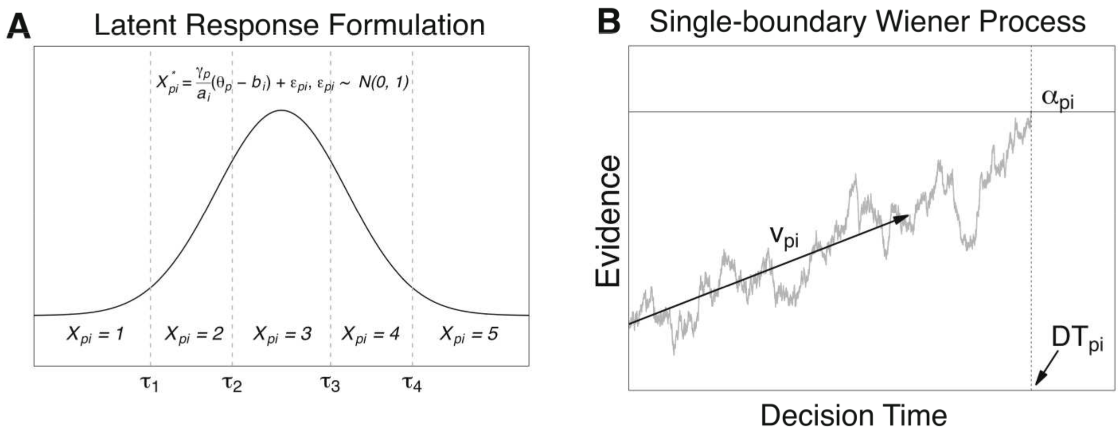




https://psyarxiv.com/km37w/

Specifikace modelů (obecně)

Na základě teorie jde nicméně specifikovat mnoho různých variant např.:



$$\nu_{pi} = |\theta_p - b_i|$$

$$\alpha_{pi} = s \cdot \frac{\gamma_p}{a_i}$$

$$X^*_{pi} = \frac{\gamma_p}{a_i}(\theta_p - b_i) + \epsilon_{pi}, \quad \epsilon_{pi} \sim N(0, 1)$$

$$X_{pi} = \begin{cases} 1 & \text{if } X^*_{pi} \leq \tau_1 \\ 2 & \text{if } \tau_1 < X^*_{pi} \leq \tau_2 \\ 3 & \text{if } \tau_2 < X^*_{pi} \leq \tau_3 \\ 4 & \text{if } \tau_3 < X^*_{pi} \leq \tau_4 \\ 5 & \text{if } \tau_4 < X^*_{pi} \end{cases}$$

$$T_{pi} = DT_{pi} + t_{0p}$$

FIGURE 1.

Illustration of Model 1: Single-accumulator model. The model is a combination of the latent response formulation and the single-boundary Wiener process.

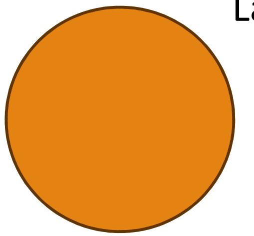
https://bpb-us-w2.wpmucdn.com/u.osu.edu/dist/6/60429/files/2023/07/6b99f784-3793-4461-9ddc-11483cec75da.pdf

Specifikace FA modelů

V případě faktorové analýzy si ale vystačíme s:

Stavební kameny:

Latentní proměnná (nepozorovatelná)



Manifestní proměnná (přímo pozorovatelná)

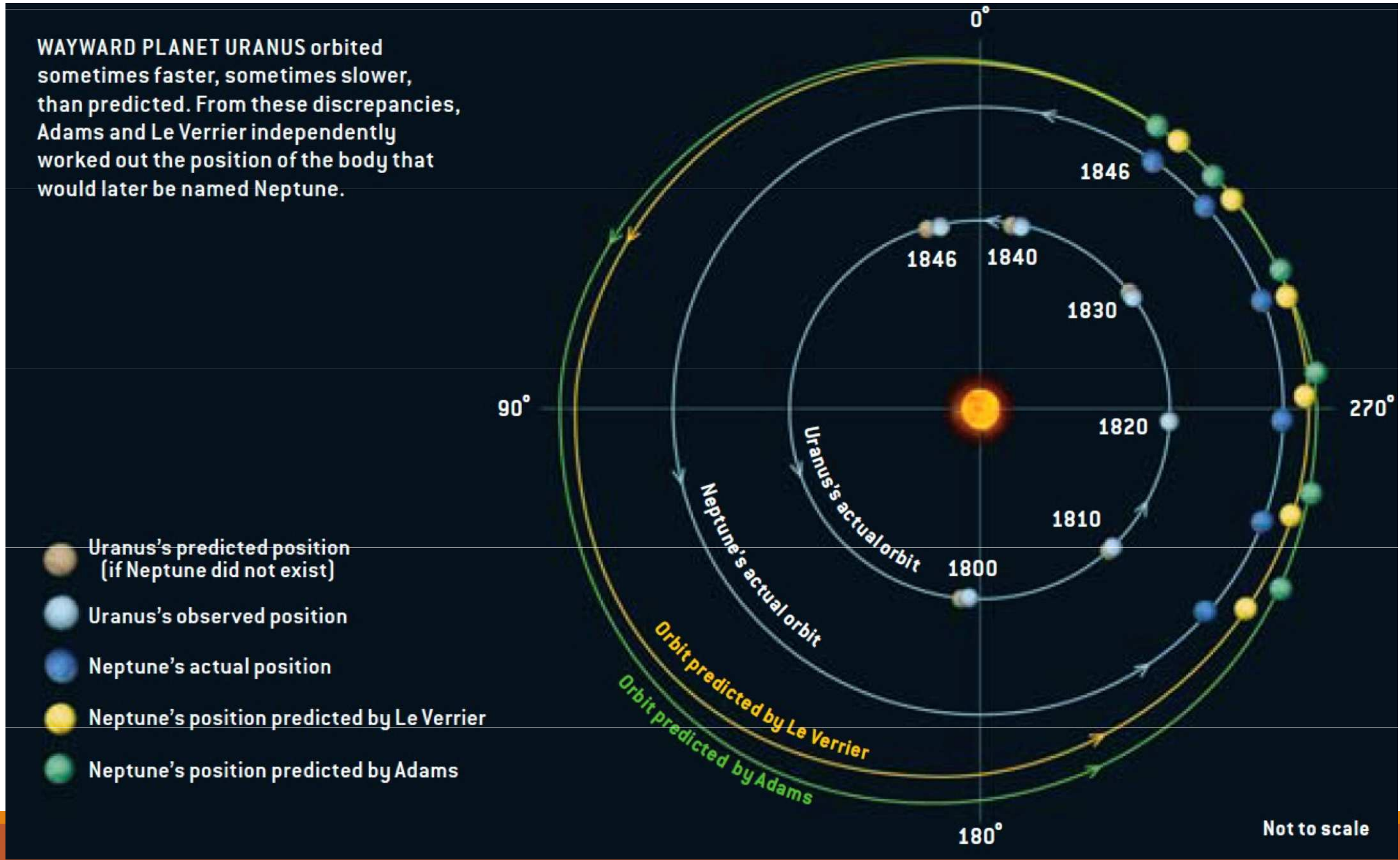
Kauzální vztah

Korelace (kovariance)

Take home message

1. Faktorová analýza je jeden ze způsobů testování teoretických očekávání spojených se vztahem mezi atributem a nástrojem

Latentní proměnná



Latentní proměnná

Consider the following sentence: "Einstein would not have been able to come up with his $e = mc^2$ had he not possessed such an extraordinary intelligence." What does this sentence express? It relates observable behavior (Einstein's writing $e = mc^2$) to an unobservable attribute (his extraordinary intelligence), and it does so by assigning to the unobservable attribute a causal role in bringing about Einstein's behavior. In psychology, there are many constructs that play this type of role in theories of human behavior; examples are constructs like extraversion, spatial ability, self-efficacy, and attitudes. Such variables are usually referred to as *latent variables*. It is common to investigate the structure and effect of unobservables like intelligence through the analysis of interindividual differences data by statistically relating covariation between observed variables to latent variables. This is done, for example, in the widely used factor model. The idea is that although

https://www.researchgate.net/publication/10758565_The_Theoretical_Status_of_Latent_Variables

Latentní proměnná

Víc názorů, dle Borsbooma (2008) je rozdíl mezi MV (manifestní p.) a LV (latentní p.) **epistemologický, ne ontologický**

Jinými slovy – proměnné nejsou nutně latentní, nebo manifestní, jde vždy jen o to, jak přesně je dokážeme pozorovat

MVs tak jsou vlastně speciálním případem LV, protože jdou přesně pozorovat

https://www.researchgate.net/publication/232051996_Latent_Variable_Theory

Latentní proměnná

Hlavním rozdílem je vztah mezi **skutečnou pozicí člověka** na **měřené** proměnné (*variable structure*) a **pozorovanou pozicí člověka** na **pozorované** proměnné (*data structure*)

Manifestní proměnná: Pozice člověka na měřené proměnné se do pozorované proměné překládá přímo (**deterministicky**)

Latentní proměnná: Vztah mezi skutečnou pozicí člověka v atributu se do jeho*její pozice překládá **pravdpodobnostně**

Teoreticky je možné, že se z dnešní LV časem stane MV (?)

https://www.researchgate.net/publication/232051996_Latent_Variable_Theory

Latentní proměnná

Dle Borsbooma (2008) jsou LVs poměrně obyčejné, tajemné jsou naopak MVs

O mnoha proměnných nemá smysl neuvažovat jako o LV, ačkoliv by to šlo – jejich pojetí jako MV je akceptovatelné zjednodušení (v psychologii třeba pohlaví, věk)

Dle Borsbooma (2003) **je důležité LV neztotožňovat s výstupem analýzy** (tzv. *operacionální LV*), jde vždy o teorií popsanou proměnnou, která koresponduje s něčím reálným

LVs můžou mít více forem – nominální, spojitá,… V FA jde vždy o spojitou LV (lidé se liší v míře atributu)

Take home message

1. Faktorová analýza (FA) je jeden ze způsobů testování teoretických očekávání spojených se vztahem mezi atributem a nástrojem

2. FA očekává latentní proměnnou, o které jde uvažovat jako o obtížně pozorovatelné, ale existující proměnné, která způsobuje pozorování

Historie FA

Počátky u Spearmana, který se snažil vysvětlit korelace inteligenčních subtestů

Později ji zdokonaloval Thurstone (kniha *Vectors of Mind*)

Následně se stala důležitým nástrojem pro rozvoj psychologických teorií

V 70. letech Jöreskog přišel s “konfirmační” FA, paralelně s tím se rozvíjely modely z Teorie odpovědi na položku (IRT), které umožnily další aplikace ve vzdělávání (Rasch, Birnbaum)

Dnes je FA členem široké rodiny mnoha různých modelů s latentní proměnnou

Common Factor Model

Řekněme, že nás zajímá **schopnost hrát šachy**.

Nemůžeme ji pozorovat přímo, jde tedy o latentní proměnnou.

Projevuje se skrze jednotlivé šachové partie.

Common Factor Model

Řekněme, že nás zajímá **schopnost hrát šachy**.

Nemůžeme ji pozorovat přímo, jde tedy o latentní proměnnou.

Projevuje se skrze jednotlivé šachové partie.

MVs (“položky”):

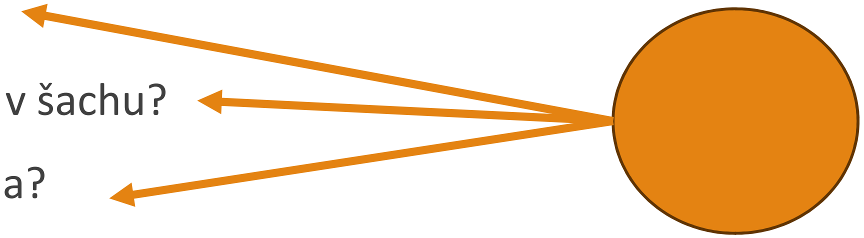
- kolikrát z 10 her porazil*a cvičenou opici?

- kolikrát z 10 her porazil*a okresního mistra v šachu?

- kolikrát z 10 her porazil*a Magnuse Carlsena?

- ...

- ...

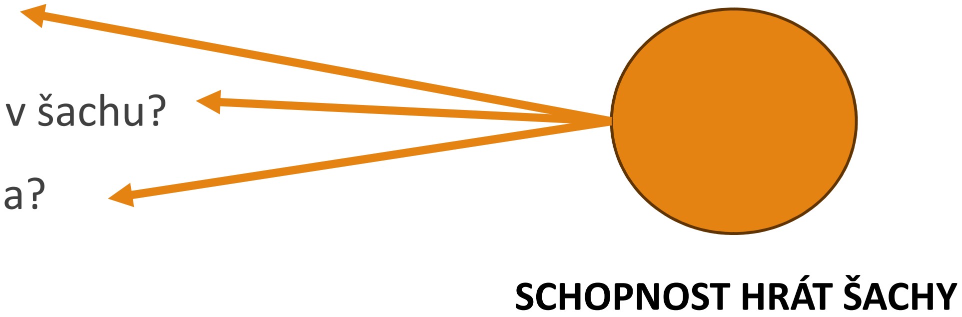


SCHOPNOST HRÁT ŠACHY

Common Factor Model

<u>MVs ("položky"):</u>

- kolikrát z 10 her porazil*a cvičenou opici?
- kolikrát z 10 her porazil*a okresního mistra v šachu?
- kolikrát z 10 her porazil*a Magnuse Carlsena?



Adam: (9, 1, 0)

Bára: (10, 10, 1)

Cecil: (3, 0, 0)

Common Factor Model

Rozdílná schopnost hrát šachy se projeví **rozdílnou úspěšností** na **různě obtížných** položkách

Z definice LV víme, že tento vztah nebude dokonalý, ale **zatížený chybou**

Na úrovni jedné položky tedy, v FA:

$$Y_{Carlsen} = obtížnost + náboj * schopnost + chyba$$

Nebo:

$$Y_{Carlsen} = \mu + \lambda\theta + u$$

Common Factor Model

$$Y_{Carlsen} = \mu + \lambda\theta + u$$

Y … **odpověď na položku** (MV) označenou dolním indexem

μ … **obtížnost položky** (nebo taky její průsečík / průměr)
Proč? Porazit Magnuse Carlsena je těžší než porazit cvičenou opici

λ … **faktorový náboj** (těsnost vztahu mezi LV a MV)
Proč? Každá položka měří daný atribut různě dobře

u … **chyba**
Proč? Vztah mezi LV a MV není dokonalý

Common Factor Model

$$Y_{Carlsen} = \mu + \lambda\theta + u$$

Jde o regresi!

$$Y_{Carlsen} = M + bX + e$$

Chyba v FA

Co je myšlenou chybou?

Nedokonalost vztahu mezi LV a MV je daná:

1. Náhodnými (nesystematickými) vlivy (**chybový rozptyl**)
2. Systematickými vlivy specifickými pro danou MV (**specifický rozptyl**)

Oba zdroje chyby nemůžeme rozplést, pokud dvě MV nesdílí specifický rozptyl (tj. nemají něco společného nad rámec efektu LV)

Společně tak mluvíme o **unikátním rozptylu**, který je složený z chybového a specifického.

Chyba v FA

Dále pokud:

$$Y = \mu + \lambda\theta + u$$

Tak lze dovodit, že chyba *u* je to, co zůstává nevysvětlené LV:

komunalita

Multidimenzionalita

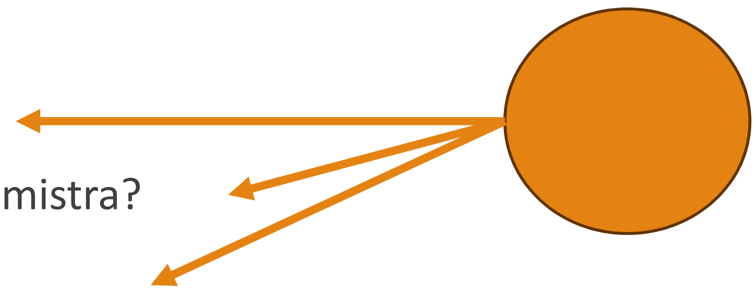
Faktorový model může obsahovat více LVs.

Nový atribut: **boxerská zdatnost**

MVs:

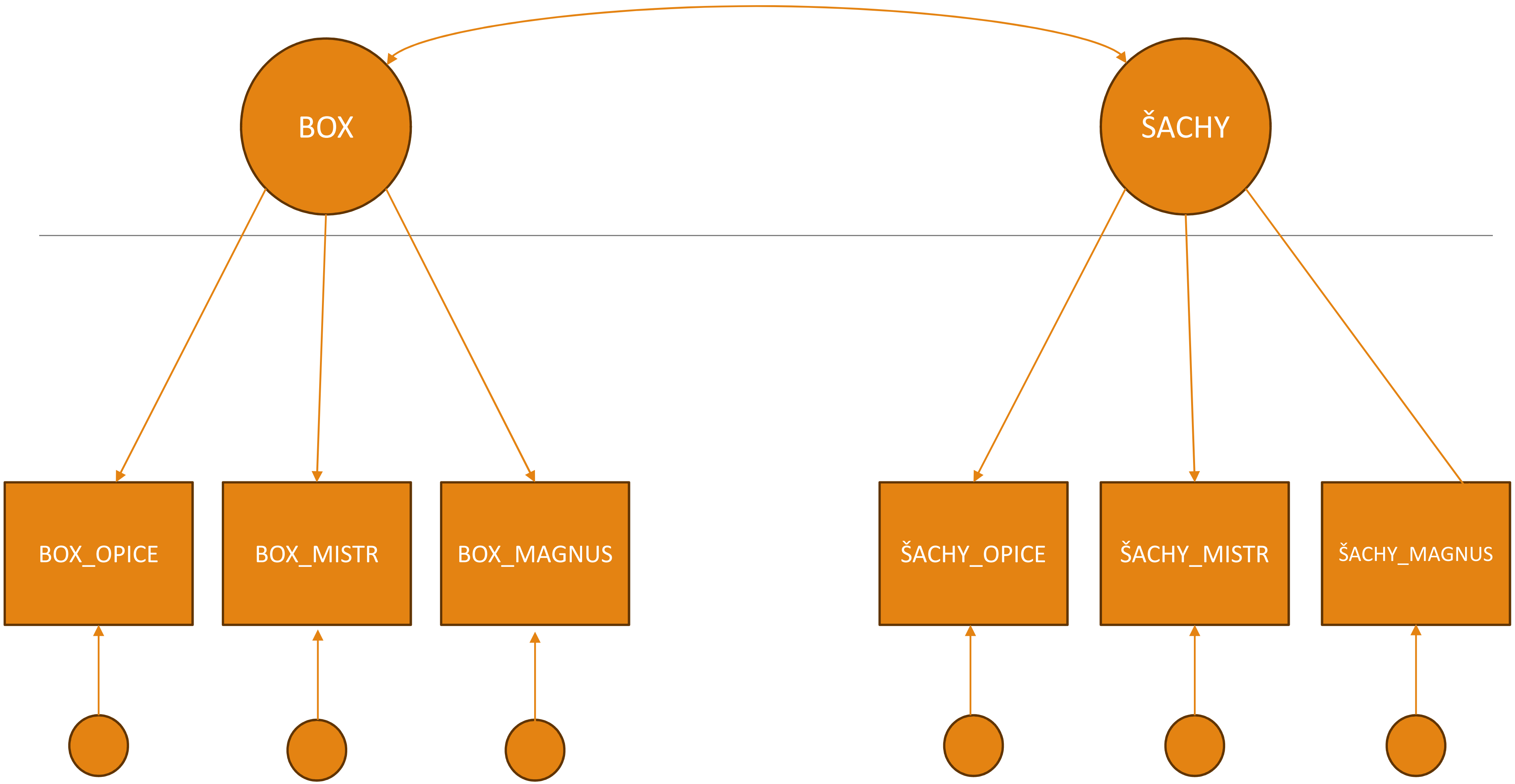
- Kolikrát z 10 boxerských zápasů porazil*a cvičenou opici?
- Kolikrát z 10 boxerských zápasů porazil*a okresního šachového mistra?
- Kolikrát z 10 boxerských zápasů porazil*a Magnuse Carlsena?

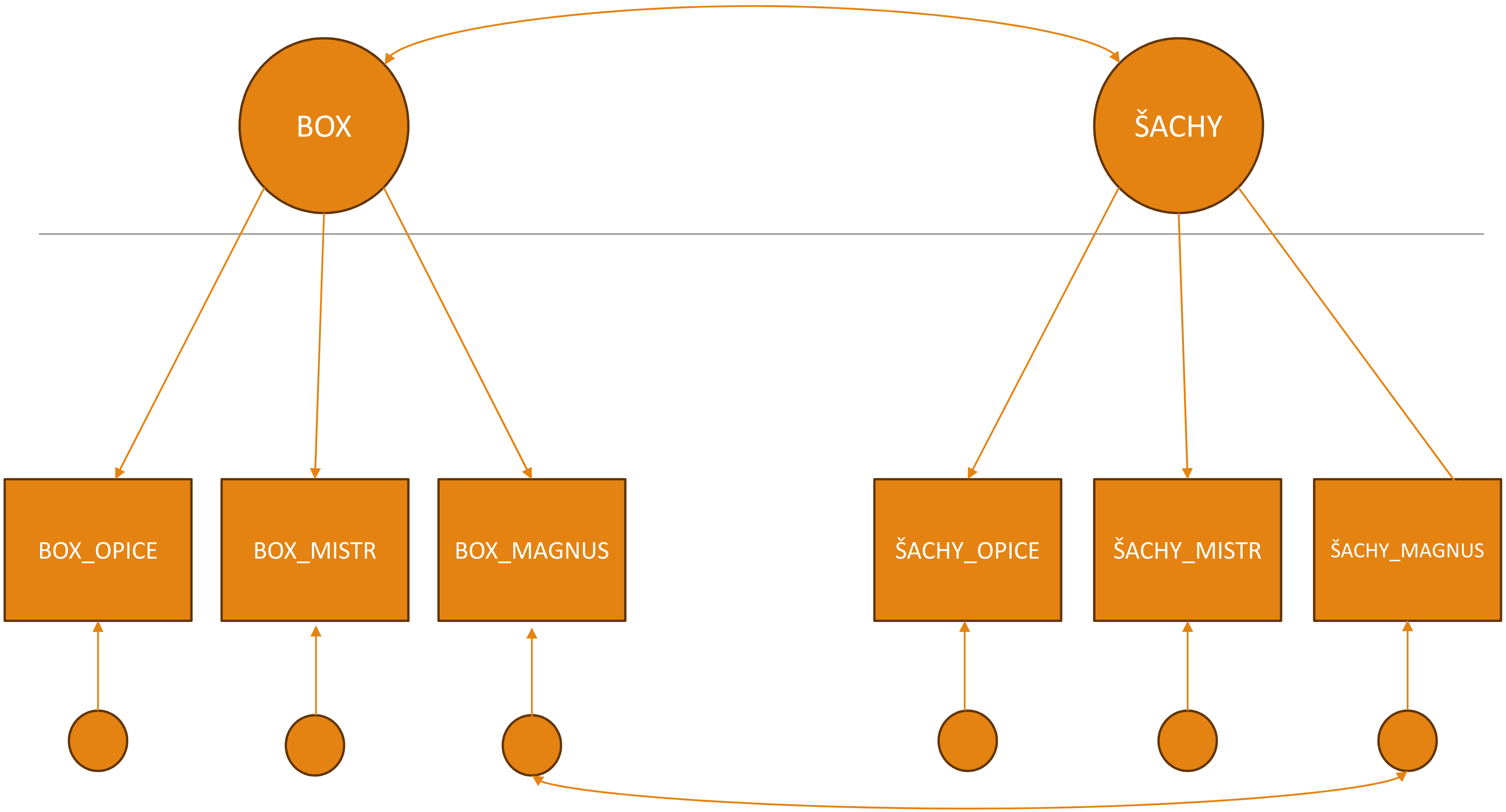
BOXERSKÁ ZDATNOST



Multidimenzionalita

- Kolikrát z 10 boxerských zápasů porazil*a cvičenou opici?
- Kolikrát z 10 boxerských zápasů porazil*a okresního šachového mistra?
- Kolikrát z 10 boxerských zápasů porazil*a Magnuse Carlsena?
- Kolikrát z 10 her porazil*a cvičenou opici?
- Kolikrát z 10 her porazil*a okresního mistra v šachu?
- Kolikrát z 10 her porazil*a Magnuse Carlsena?


BOX
ŠACHY
BOX_OPICE
BOX_MISTR
BOX_MAGNUS
ŠACHY_OPICE
ŠACHY_MISTR
ŠACHY_MAGNUS


BOX
ŠACHY
BOX_OPICE
BOX_MISTR
BOX_MAGNUS
ŠACHY_OPICE
ŠACHY_MISTR
ŠACHY_MAGNUS

Multidimenzionalita

Měli jsme:

$$Y_{Carlsen\_ŠACHY} = \mu + \lambda\theta + u$$

Jak se to změní v multidimenzionálním modelu?

Multidimenzionalita

Měli jsme:

$$Y_{Carlsen\_ŠACHY} = \mu + \lambda\theta + u$$

Jak se to změní v multidimenzionálním modelu?

$$Y_{Carlsen\_ŠACHY} = \mu + \boldsymbol{\lambda\theta_{box}} + \lambda\theta_{šachy} + u$$

Kolika je ale rovno $\boldsymbol{\lambda\theta_{box}}$?

Multidimenzionalita

Měli jsme:

$$Y_{Carlsen\_ŠACHY} = \mu + \lambda\theta + u$$

Jak se to změní v multidimenzionálním modelu?

$$Y_{Carlsen\_ŠACHY} = \mu + \boldsymbol{\lambda\theta_{box}} + \lambda\theta_{šachy} + u$$

Kolika je ale rovno $\boldsymbol{\lambda\theta_{box}}$?

$$\boldsymbol{\lambda\theta_{box} = 0 * \theta_{box} = 0}$$

Boxerská zdatnost totiž u šachové položky nemá hrát žádnou roli!

Multidimenzionalita

Je ale možné modelovat i tzv. **crossloading**, tedy situaci, kdy jednu MV způsobuje více LVs

MVs

- Kolikrát z 10 zápasů v šachboxu porazil*a cvičenou opici?
- Kolikrát z 10 zápasů v šachboxu porazil*a okresního šachového mistra?
- Kolikrát z 10 zápasů v šachboxu porazil*a Magnuse Carlsena?

Take home message

1. Faktorová analýza (FA) je jeden ze způsobů testování teoretických očekávání spojených se vztahem mezi atributem a nástrojem (**modelů měření**)

2. FA očekává **latentní proměnnou**, o které jde uvažovat jako o **obtížně pozorovatelné, ale existující proměnné, která způsobuje pozorování**

3. V FA můžeme specifikovat řadu modelů měření, včetně multidimenzionality

Common Factor Model

Dosud uváděný model predikoval hodnotu jednoho člověka na jedné MV

$$Y_{Carlsen\_ŠACHY} = \mu + \boldsymbol{\lambda\theta_{box}} + \lambda\theta_{šachy} + u$$

Obsahuje ale příliš mnoho neznámých – je potřeba najít nějaké řešení, které obejde potřebu znát latentní skór konkrétních osob.

Faktorová analýza proto stojí na **modelování kovariancí mezi MVs (covariance structure).**

I náš model proto musí počítat se všemi MVs a LVs najednou.

Základní pojmy

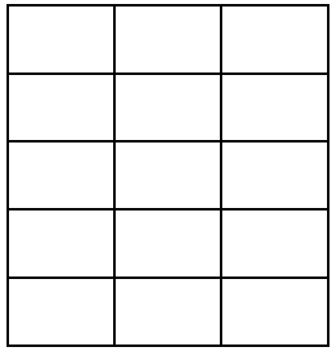
- Jaká je typická podoba dat v případě faktorové analýzy?

- Multivariační data – data pro soubor osob, větší množství manifestních (měřených, pozorovaných) proměnných (např. skóry z testů, škál, položek...)

Datová matice:

Co sloupec, to proměnná

Co řádek, to osoba



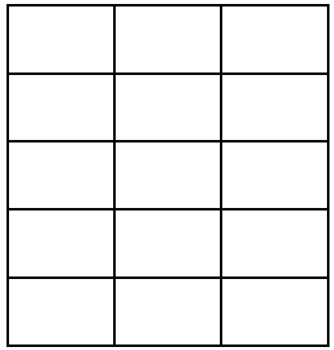
Základní pojmy

- Jednotlivé buňky v datové matici představují skór dané osoby na dané manifestní proměnné
- Fundamentální premisa faktorové analýzy: Tyto skóry nejsou nějakými náhodnými hodnotami, ale vykazují určité systematické aspekty, kterými se můžeme zabývat

Co sloupec, to proměnná

Datová matice:

Co řádek, to osoba



Základní pojmy

Datová matice:

$$\begin{array}{cc} & \textit{p sloupců (proměnných)} \\ \boldsymbol{X} = & \begin{bmatrix} x_{11} & x_{12} & & x_{1p} \\ & & x_{ij} & \\ x_{N1} & x_{N2} & & x_{Np} \end{bmatrix} \end{array} \quad \textit{N řádků (osob)}$$

Skór osoby *i* na proměnné *j*

Základní pojmy

Čeho si můžeme na těchto datech všimnout?

- Variabilita každé proměnné napříč osobami (**rozptyl / SD**)
- Kovariance dvou proměnných napříč osobami (**kovariance / korelace**)

x_{11}	x_{12}		x_{1p}
		x_{ij}	
x_{N1}	x_{N2}		x_{Np}

Základní pojmy

Korelační matice:

$$
\boldsymbol{R} = \begin{array}{c}
p \text{ manifestních proměnných} \\
\begin{bmatrix}
1 & r_{12} & r_{13} & \cdots & \cdots & \cdots & r_{1p} \\
r_{21} & 1 & r_{23} & & & & r_{2p} \\
r_{32} & r_{32} & 1 & & & & r_{3p} \\
\vdots & & & \ddots & r_{kj} & & \\
\vdots & & & r_{jk} & \ddots & & \\
\vdots & & & & & \ddots & \\
r_{p1} & r_{p2} & r_{p3} & & & & 1
\end{bmatrix}
\end{array}
\begin{array}{l} p \text{ manifestních} \\ \text{proměnných} \end{array}
$$

Pozn.: Na obrázku je korelační matice (na diagonále jsou jedničky $r_{jj} = 1$, mimo diagonálu korelace r_{jk}), faktorová analýza ale velmi často pracuje s kovarianční maticí, kde na diagonále je rozptyl (σ_{jj}^2) a mimo diagonálu kovariance (σ_{jk}) příslušných proměnných. Typicky EFA pracuje s korelační, zatímco CFA s kovarianční maticí. Z kovarianční matice lze získat korelační matici snadno, $r_{jk} = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}^2 \sigma_{kk}^2}}$. Naopak to nefunguje, protože korelační matice nenese informaci o rozptylech.

Model kovarianční struktury

Jednoduchý model:

$$Y_{Carlsen\_ŠACHY} = \mu + \lambda\theta_{box} + \lambda\theta_{šachy} + u$$

Je proto (pomocí krásné matematiky založené na vlastnostech jednotlivých částí modelu) nutné formulovat model kovarianční struktury všech MVs a LVs:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\phi}\boldsymbol{\Lambda}' + \boldsymbol{D}_{\boldsymbol{\psi}}$$

Model kovarianční struktury

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{D}_{\boldsymbol{\psi}}$$

- **Σ** (sigma) je matice korelací / kovariancí mezi manifestními proměnnými
- **Λ** (lambda) je matice faktorových nábojů (apostrof značí transpozici)
- **Φ** (phi / fí) je matice korelací / kovariancí mezi (obecnými) faktory. Faktory být korelované nemusí – v takovém případě lze říci, že faktory jsou tzv. *ortogonální*
- **D**$_\psi$ (D-psi / D-psí) je matice rozptylů unikátních faktorů (a případně reziduálních kovariancí).
- ...jak možná správně tušíte, k faktorové analýze nepotřebujete syrová data, ale korelace / kovariance mezi MVs.

Model kovarianční struktury

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ \lambda_4 & 0 \\ \lambda_5 & 0 \\ 0 & \lambda_6 \\ 0 & \lambda_7 \\ 0 & \lambda_8 \\ 0 & \lambda_9 \\ 0 & \lambda_{10} \end{bmatrix} \qquad \mathbf{\Phi} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{D_\Psi} = \begin{bmatrix} \xi_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \xi_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \xi_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \xi_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \xi_5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \xi_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \xi_7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi_8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi_9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi_{10} \end{bmatrix}$$

$$\mathbf{\Lambda}' = \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_6 & \lambda_7 & \lambda_8 & \lambda_9 & \lambda_{10} \end{bmatrix}$$

Common Factor Model

Faktorová analýza se zabývá vysvětlením kovariancí / korelací mezi MVs.

- V případě jediné dimenze je kovariance položek a, b rovna $\sigma_{ab} = \lambda_a \lambda_b \phi$, kde ϕ je rozptyl faktoru.
- V případě korelační matice a jediného a standardizovaného faktoru platí $r_{ab} = \lambda_a \lambda_b$.

Znamená to, že **nepotřebujeme znát skóry osob na LVs** (které stejně neznáme a znát nemůžeme – jsou nepozorované a *neurčitelné* [*indeterminate*]).

Common Factor Model

Není nutné předchozí rovnice chápat – v praxi se bez nich obejdete, jinak se jim blíže věnujeme v B kurzu *Introduction to Factor Analysis*

Je ale zásadní:

1. vědět, co představují jednotlivé části common factor modelu (jednotlivé matice)
2. odnést si, že FA vysvětluje kovariance / korelace pomocí jiných parametrů (např. faktorových nábojů), které mají teoretickou interpretaci
3. uvědomit si, že rovnice common factor modelu doslova znamená, že obě strany jsou ekvivalentní, tedy že $\boldsymbol{\Lambda\Phi\Lambda'} + \boldsymbol{D_\psi}$ je jen překladem (jiným vyjádřením) kovarianční matice mezi MVs.

Take home message

1. Faktorová analýza (FA) je jeden ze způsobů testování teoretických očekávání spojených se vztahem mezi atributem a nástrojem (**modelů měření**)

2. FA očekává **latentní proměnnou**, o které jde uvažovat jako o **obtížně pozorovatelné, ale existující proměnné, která způsobuje pozorování**

3. V FA můžeme specifikovat řadu modelů měření, včetně multidimenzionality

4. FA **modeluje kovarianční strukturu** mezi MVs tím, že ji **zjednodušuje na latentní strukturu** složenou z **faktorových nábojů**, **korelací mezi faktory** a **residuálních rozptylů** (kovariancí)

O co nám tedy ve FA jde?

- Cílem je **odhalit, pochopit, popsat a/nebo otestovat** latentní strukturu, která způsobuje korelace mezi MVs.

- Chceme tedy identifikovat (nebo ověřit) **počet a charakter** (význam) faktorů, které způsobují pozorované korelace mezi manifestními proměnnými.
- Jinými slovy, chceme přijít na to, kolik LVs ovlivňuje naše MVs a **odhadnout sílu a směr** (+/-) **faktorových nábojů**.
- Velikost a směr faktorových nábojů nám napomáhají v určení podstaty LV (dle vazby mezi konkrétní LV a konkrétními MV, viz definice validity)

Dva cíle FA

Konfirmační cíl

Známe latentní strukturu a testujeme, nakolik odpovídá pozorované kovarianční struktuře

“**Kdyby** byla pravda, že inteligenční subtesty způsobuje právě jeden g-faktor, jaké kovariance mezi inteligenčními subtesty bychom museli pozorovat?”

Výsledkem je vyjádření **shody modelu** (očekávání) **s daty** (pozorováním) **a test teorie**

Explorační cíl

Latentní strukturu neznáme, snažíme se ji najít pomocí pozorované kovarianční struktury

“**Kdyby** byla pravda, že inteligenční subtesty způsobuje několik LVs, jaká latentní struktura nejlépe odpovídá pozorované kovarianční struktuře?”

Výsledkem je vyjádření **shody modelu** (očekávání) **s daty** (pozorováním), ale **nikoliv test teorie**

Dva druhy FA

“Konfirmační” FA

Umožňuje specifikovat latentní strukturu a testovat, nakolik odpovídá pozorované kovarianční struktuře

Ale latentní struktura jde testovat i exploračně (explorační “tunění modelu”)!

Lepší název je proto **restricted FA**, protože vyžaduje specifikovat a umožňuje testovat restrikce, bez ohledu na cíl

“Explorační” FA

Umožňuje najít latentní strukturu pomocí pozorované kovarianční struktury

Ale i EFA vyžaduje specifikovat aspoň počet faktorů, takže jde použít konfirmačně!

Lepší název je proto **unrestricted FA**, protože vyžaduje jen specifikaci počtu faktorů

Take home message

1. Faktorová analýza (FA) je jeden ze způsobů testování teoretických očekávání spojených se vztahem mezi atributem a nástrojem (**modelů měření**).

2. FA očekává **latentní proměnnou**, o které jde uvažovat jako o **obtížně pozorovatelné, ale existující proměnné, která způsobuje pozorování.**

3. V FA můžeme specifikovat řadu modelů měření, včetně multidimenzionality.

4. FA **modeluje kovarianční strukturu** mezi MVs tím, že ji **zjednodušuje na latentní strukturu** složenou z **faktorových nábojů**, **korelací mezi faktory** a **residuálních rozptylů** (kovariancí).

5. **Cílem FA může být konfirmace, nebo explorace**. Použít k tomu jdou dva typy FA – **restricted a unrestricted**, tradičně zvané *konfirmační* (CFA) a *explorační* (EFA).

Některé otázky k FA

1. Je možné FA provést, i když LV neexistuje / nezpůsobuje MVs?
2. Znamená dobrá shoda FA modelu s daty, že latentní struktura odpovídá realitě?
3. Jaký je vztah mezi FA a reifikací?
4. Jaká jsou využití FA mimo psychometriku?

https://www.researchgate.net/publication/321253793_What_is_the_p_-factor_of_psychopathology_Some_risks_of_general_factor_modeling