



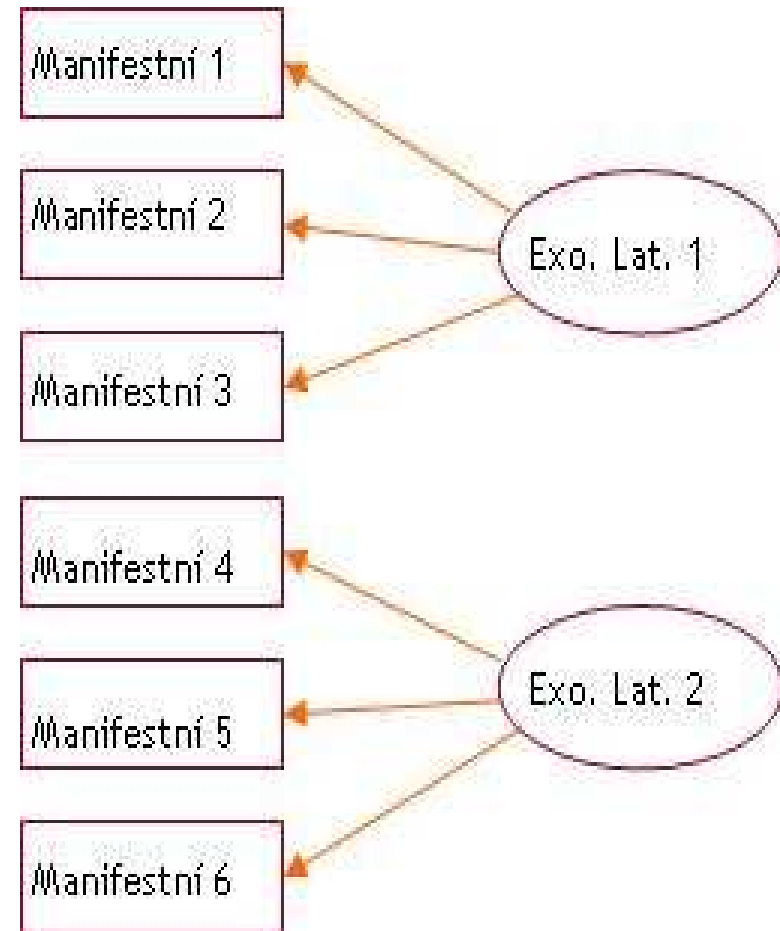
FAKTOROVÁ ANALÝZA

PSY252



LATENTNÍ A MANIFESTNÍ PROMĚNNÉ

- Proměnné, jak je známe, jsou MANIFESTNÍ (...*položky*)
- Pokud spolu MP všechny korelují, můžeme hypotetizovat, že jsou to jen různé projevy jedné PŘÍČINY
- LATENTNÍ PROMĚNNÁ je hypotetickou příčinou odvozenou právě z manifestací (jevů), které spolu nějakým způsobem korelují
- **FAKTOR** je LATENTNÍ PROMĚNNÁ



ÚČEL (EXPLORATORNÍ) FAKTOROVÉ ANALÝZY

Redukovat

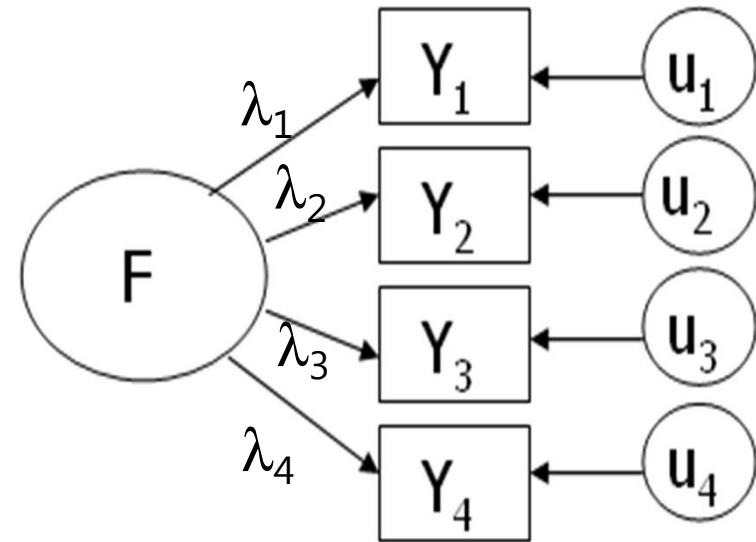
- vyjádřit informaci obsaženou ve více proměnných menším množstvím proměnných - **komponent** (např. k řešení multikolinearity, síly testu apod.)
- faktory pak nemusí nutně reprezentovat nějaký konstrukt – **komponenty**, dimenze
- *Situace*: mám hromadu proměnných, potřebuji to nějak zjednodušit, abych mohl pokračovat v analýze

Hledat (kauzální) faktory ... konstrukty

- **faktory**, které nejlépe vystihnou korelace mezi proměnnými
- smysluplné faktory reprezentující konstrukty
- budování teorie (hledání struktury), měření (usuzování z vnějších projevů na rys)
- *Situace 1*: zjišťuji, jakými faktory je syceno odpovídání na položky dotazníku
- *Situace 2*: hledám obecnější konstrukty, které stojí za širší paletou měřítek výkonu (IQ)

FAKTOROVÝ MODEL COMMON FACTOR MODEL

FA je „jako“ regrese s obrácenými šipkami.



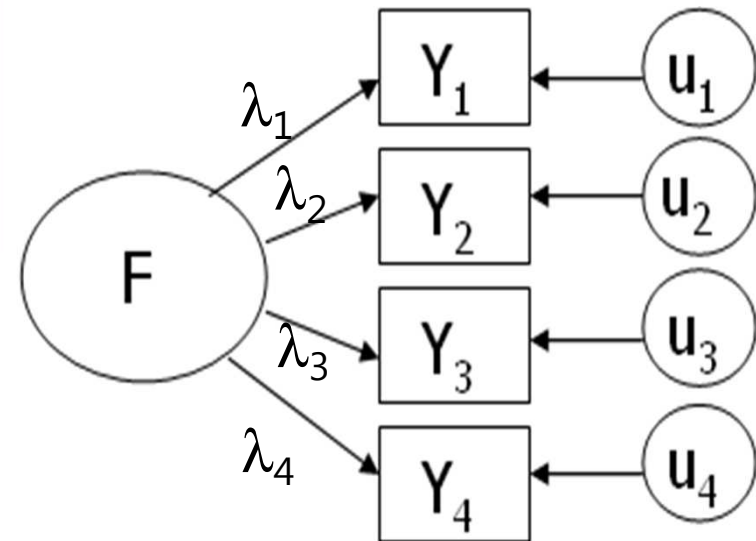
- Faktor(y) vysvětlují/predikují MP
- Proměnné/faktory jsou **standardizované**

■ Faktorový náboj - λ

- Standardizovaný regresní koeficient, jímž faktor predikuje Y_i (položka **syti** faktor)
- Je-li jen 1 faktor, nebo pokud jsou faktory nekorelované, $\lambda = r_{FY_i}$
- λ^2 vyjadřuje, jak velkou část rozptylu Y_i faktor F vysvětluje

FAKTOROVÝ MODEL

COMMON FACTOR MODEL



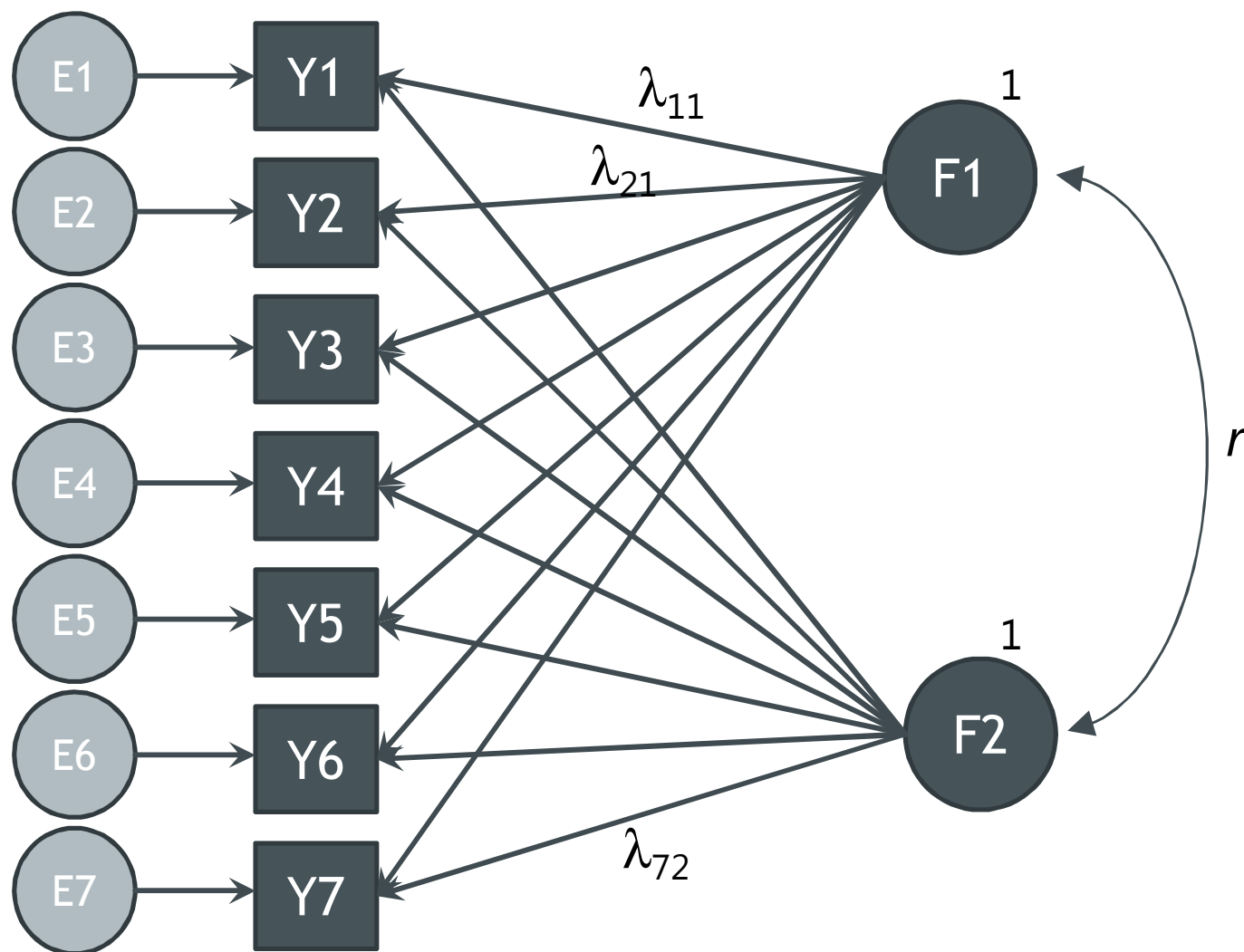
Složky rozptylu Y_i ve faktorovém modelu:

■ Komunalita (h^2):

- Sdílený (faktorový) rozptyl
- Část rozptylu vysvětlená společnými faktory, tj. prostřednictvím faktoru/ů sdílená s ostatními proměnnými
- h^2 = suma čtverců *nábojů* na všech faktorech -zde λ_1^2
- Sdílený rozptyl je patrný z korelační matice Y; jsou-li korelace Y nenulové, potom Y sdílí nějakou část rozptylu

■ Unicita (u^2): jedinečnost, nesdílená část rozptylu Y_i

- Unicita = 1 – komunalita
- Unicita je rozptyl vlastního faktoru U -zahrnuje „vlastní rozptyl“ proměnné & chybu měření
- Unicity jsou nekorelované – parciální korelace mezi položkami, kontrolujeme-li všechny faktory = 0 ... lokální nezávislost položek ... korelují-li položky, pak jedině díky tomu, že jsou syceny společnými faktory



V EFA je faktor sycen vždy všemi položkami!
 Nekonečně mnoho řešení – rotace.

FAKTOROVÝ MODEL

COMMON FACTOR MODEL

Faktory

- Jsou latentní proměnné, jsou lineární kombinací (=váženým součtem) Y – jejich hodnoty (faktorové skóry) nás však zajímají jen někdy
- Jsou syceny všemi položkami, Y (jen v EFA), jde jen o to, jak moc
- „Síla“ faktoru je vyjádřena tím, jak velkou část sdíleného rozptylu položek faktor vysvětluje
- Mohou korelovat (šikmé, kosé, oblique) nebo být nekorelované (ortogonální) - volíme na základě teoretických očekávání
 - Ortogonální faktory jedinečně přispívají k vysvětlení Y , představují „nezávislé dimenze“
 - Korelované faktory jsou obvykle realističtější očekávání, mohou lépe predikovat jednotlivé položky (Y_j)



POSTUP FAKTOROVÉ ANALÝZY

1. Zvážení předpokladů FA
2. Volba typu analýzy - FA vs PCA
3. Volba počtu faktorů
4. Extrakce
5. Rotace
6. Interpretace

1. PŘEDPOKLADY POUŽITÍ FA

Smysluplná korelační matice

- Alespoň ordinální úroveň měření
- Smysluplně vysoké korelace (min. Bartlettův test sféricity)

Dostatečný počet Y vzhledem k předpokládanému počtu faktorů

- 3 při předpokladu jediného faktoru
 - $k*2$ při předpokladu k faktorů
 - KMO test of sampling adequacy
-
- Adekvátní počet měření
 - FA je nesmírně citlivá – „výběrová chyba korelací na entou“
 - Absolutní minimum velikosti vzorku je $N > 5*MP$ a současně $N > 20*k$, ideálně od $N > 20*MP$

2. VOLBA TYPU ANALÝZY - FA VS. PCA

- PCA pro účely redukce a čiré explorace
 - Faktoruje veškerý rozptyl, včetně unicit (komunality na počátku = 1)
 - Výsledkem jsou „komponenty“ ne faktory
 - Lze počítat faktorové skóry
- FA (common factor analysis)
 - Faktoruje jen sdílený rozptyl (komunality < 1)
 - Výsledkem jsou faktory, které mohou být konstrukty
 - V psychologii metoda primární volby

Výsledky PCA a FA vypadají velmi podobně, proto dochází snadno k záměně.

3. VOLBA POČTU FAKTORŮ

Máme-li teoretický předpoklad, použijme ho!

Tradiční techniky

- Cattelův scree plot - bod zlomu - 1
- Kaiserovo kritérium - Eigenvalues > 1
- Někdy též podle vysvětleného rozptylu (čistá redukce)

Pokročilé techniky

- Př. parallel analysis

Dobré je vyzkoušet více možností

4. EXTRAKCE

ODHAD PARAMETRŮ FAKTOROVÉHO MODELU

- Nejčastěji ML - maximum-likelihood
 - Též GLS, ULS
- Iterační, „slepý“ postup volící takové parametry, které maximalizují při určitém způsobem nastaveném modelu shodu mezi vstupní a výstupní korelační maticí
- Existuje řada okolností, které mohou způsobit, že algoritmus vyprodukuje neplatné řešení - nejčastěji omezení dat, popř. extrahování příliš mnoha faktorů

4. EXTRAKCE

ODHAD PARAMETRŮ FAKTOROVÉHO MODELU

Posouzení kvality extrahovaného modelu

- Shoda mezi vstupní korelační maticí a maticí vyplývající z modelu
 - Počet reziduí větších než 0,05 (méně než 10%)
 - Test dobré shody - chíkvadrát
 - Nedostatek shody naznačuje, že vztahy jsou složitější.
- Komunality - položky s nízkými komunalitami (pod 0,2-0,3) jsou kandidáty na vyřazení

Nároky se v různých kontextech liší.

5. VOLBA ROTACE

- FA má ve skutečnosti nekonečně mnoho ekvivalentně dobrých řešení ve smyslu míry determinace manifestních proměnných faktory
- Extrakce vyprodukuje ortogonální faktory (unrotated solution) podle následujícího schématu:
 - 1. faktor, který vysvětlí nejvíc rozptylu (nejvíce sycen proměnnými)
 - 2. faktor, který vysvětlí nejvíc rozptylu nevysvětleného 1. faktorem
 - 3. a další faktory, které vysvětlí nejvíc rozptylu nevysvětleného předchozími faktory
- Nerotované řešení je polotovar, který obvykle nelze interpretovat
- **ROTACE** je volba takového řešení (faktorové matice), které usnadní interpretaci

5. VOLBA ROTACE

ORTOGONÁLNÍ ROTACE

- faktory/komponenty zůstávají kolmé, nezkorelované
- Efektivita pro další analýzy
- VARIMAX - maximalizace nábojů proměnné na jednom faktoru (nejblíže jednoduché struktury při zachování ortogonality)

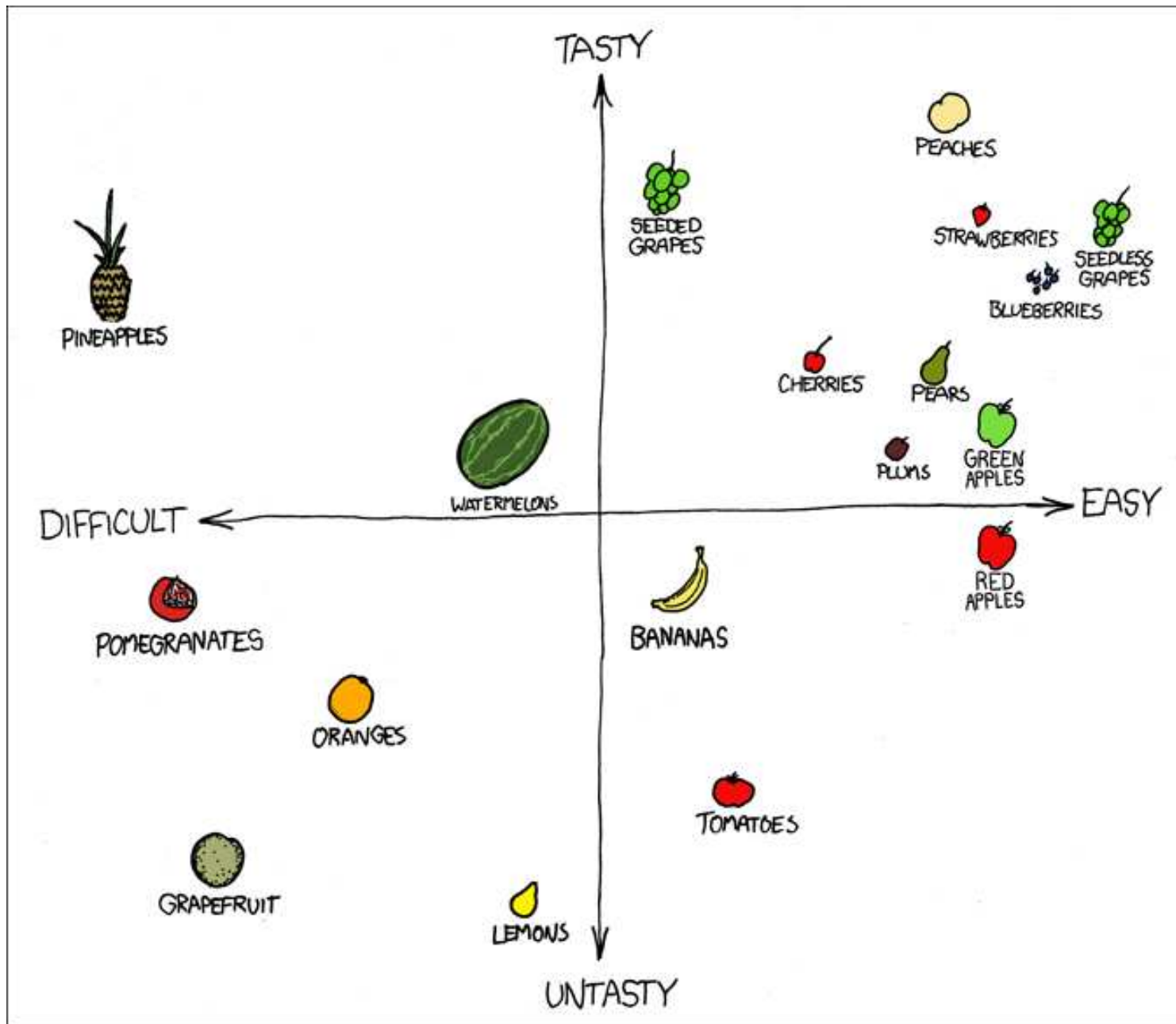
ŠIKMÉ ROTACE

- Faktory mohou korelovat, ale hůře se interpretují
- Výhoda, pokud to implikuje teorie
- Blíže jednoduché struktury
- DIRECT OBLIMIN
- Náboje jsou pak v „Pattern matrix“. Korelace položky s faktorem již není totéž, co náboj, a tak jsou ve „Structure matrix“

SIMPLE STRUCTURE

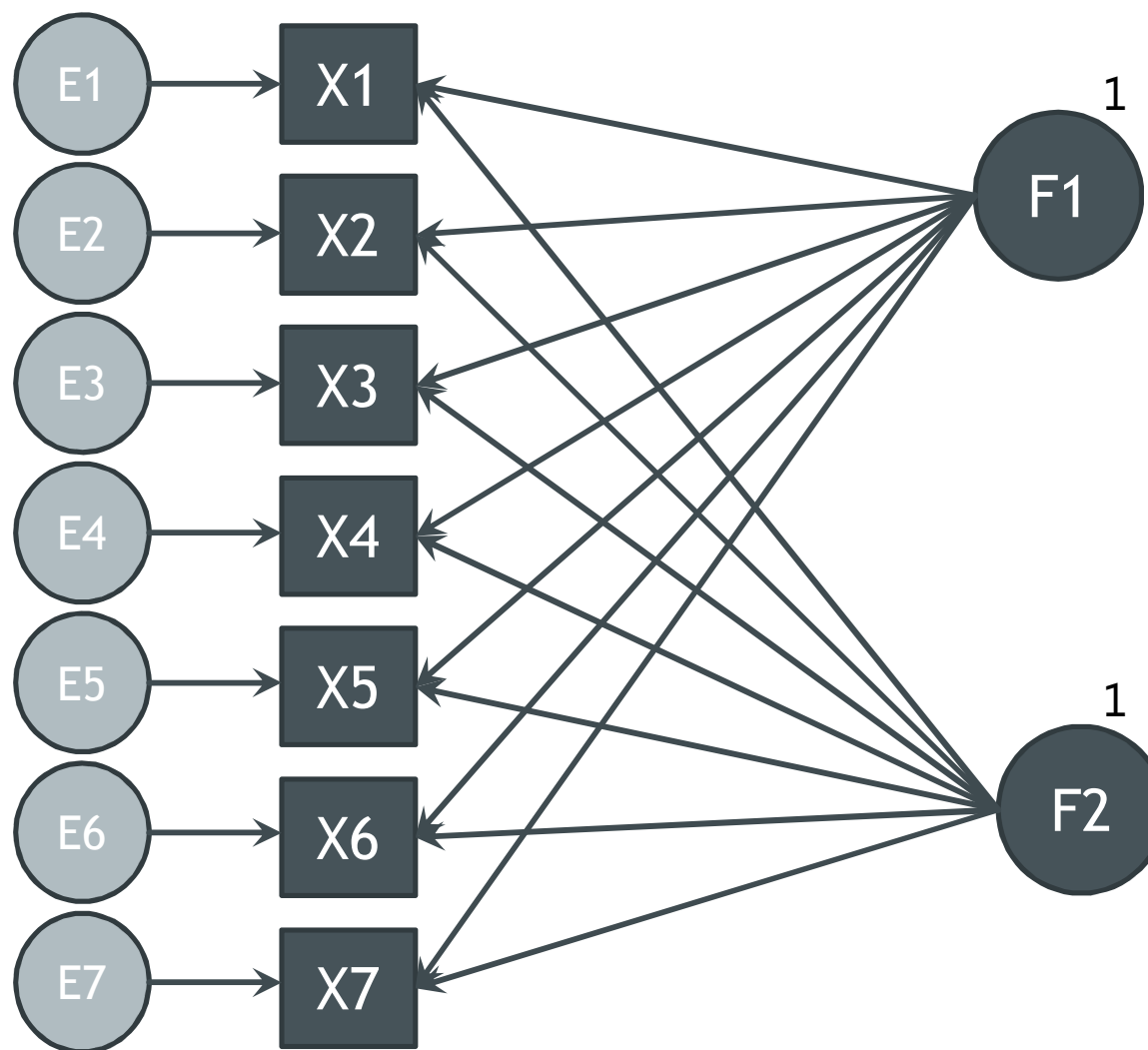
KÝŽENÝ VÝSLEDEK ROTACE

1. Každá položka je sycena jen jedním faktorem ... tj. odpověď je způsobována jen tím, co chceme měřit.
2. Každá položka má vysokou komunalitu - vliv měřeného konstruktů na odpovědi je velký.
3. V modelu nechybí významný faktor ovlivňující odpovědi na položku (specification error)
4. I při malém počtu faktorů jsou vztahy mezi položkami modelem dobře reprodukovány (test dobré shody)



FA SHRNU TÍ

- FA je primárně explorační metoda
 - Nekonečně mnoho řešení, jimž je třeba hermeneuticky vdechnout smysl
 - Volba různého počtu faktorů a rotací vedena snahou o nalezení interpretovatelného řešení, které představuje plauzibilní model toho, co stojí „za“ pozorovanými vztahy mezi
 - Opakování analýzy (třeba 10x), dokud nedává smysl
- Co s výsledky
 - Někdy jsou cenné samy o sobě - píšeme zprávu o faktorové struktuře
 - Jindy je výsledek podkladem pro tvorbu součtových skóre z položek sytících společný faktor



V EFA je faktor sycen vždy všemi položkami!
Nekonečně mnoho řešení – rotace.



VNITŘNÍ KONZISTENCE - RELIABILITA

S náhodnou chybou měření bojujeme opakováním měření

- **opakovanou administrací testu - r_{tt}**

Nelze-li opakovat - test rozpůlíme a považujeme to za opakování :: split-half

- **korelováním polovin testu - r** (pomocí Spearman-Brownova vzorce přepočítáme korelaci na reliabilitu)

Půlit lze mnoha způsoby a lze i čtvrtit ... až ho rozdělíme na jednotlivé položky.

- **korelováním položek testu - konzistence - α**

Uvažujeme-li test=položka, administrace sady položek je opakovaným administrováním testu

- Administrace v jednom čase > řada náhodných vlivů zůstává konstantní > primárním uvažovaným zdrojem chyby jsou **specifika/nedokonalosti jednotlivých položek** (\approx unicity).
- Do jaké míry měří testy-položky totéž?

CRONBACHOVA ALFA

- Alfa je odhadem reliability (jedním z mnoha)
 - Podíl rozptylu hypotetické true-scorové škály (složené ze všech myslitelných položek), který by naše škála vysvětlila.
 - Korelace škály se všemi možnými stejně dlouhými škálami měřícími týž konstrukt
 - Průměrná korelace všech myslitelných půlení testu
 - Alfa je projevem unidimensionality (nikoli důkazem)
- $$r_{tt} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k s_i^2}{s_t^2} \right)$$

Obvykle požadovaná hodnota pro výzkum $>0,7$, pro diagnostiku $>0,9$
(záleží na tom, jak velkou šíři intervalu spolehlivosti můžeme tolerovat).

Cronbachova alfa je primárně funkcí

- průměrné korelace mezi položkami (čím vyšší, tím vyšší alfa)
- počtu položek (čím vyšší, tím vyšší alfa)



ÚKOL

- Long2 data
- Položky i01 až i13.
- Vytvořte faktorový model. Na jeho základě vytvořte součtové škály z položek a určete vnitřní konzistenci takto vytvořených škál.