

# Teorie odpovědi na položku

---

PSY028 | JARO 2018 | BLOK 4

MULTIDIMENZIONÁLNÍ IRT MODEL Y

# Ekvivalence IRT a SEM

---

<http://m-clark.github.io/docs/sem/item-response-theory.html>

<https://github.com/tomatitito/grm-plots>

# Další modely

---

Continous Response Model (poprvé Mellenbergh, 1994)

- Určený např. pro VAŠ (vizuální analogová škála) nebo pro jiná data

Item Response Time Model (např. van der Linden, různé parametrizace)

- Větší množství modelů, které zohledňují i čas respondenta

Ipsative Item Response Models

- Různé parametrizace pro ipsativní položky, vč. multidimenzionálních
- (např. <http://www.shlsolutionpartner.com/au/resources/NEWOPQ32TechManualsupplement.pdf>)
- Parametrizace i pro VAŠ (Beta Unfolding Model, Noel 2013)

Multidimenzionální a FA IRT modely

Přesah do kognitivního modelování (IRT jsou jednoduché kognitivní modely).

Celkově velké množství upravených modelů pro specifická použití, člověk si ale běžně vystačí s těmi základními zmíněnými výše.

# Multidimenzionální IRT (MIRT)

---

Předpoklad lokální nezávislosti zachován, ale rozptyl je vysvětlován více faktory.

- Má tedy méně předpokladů než klasické IRT.
- Na rozdíl od CFA/SEM neumožňuje reziduální korelace, ty jsou proto řešeny specifickými faktory.
- Neumožňuje rovněž regresní vztahy mezi latentními proměnnými, pouze jejich korelace (proto zejm. bifaktorové modely).

# Multidimenzionální IRT (MIRT)

---

Dva hlavní typy:

- Kompenzatorní – vysoká úroveň jednoho rysu může kompenzovat nízký druhý rys.
  - They jsou aditivní (na stejné škále):  $\theta_g = \theta_A + \theta_B$ .
  - běžnější, jednodušší
- Non-kompenzatorní – vysoká úroveň rysu vyžaduje vysokou úroveň všech dílčích rysů.
  - They nejsou aditivní (vznikají např. součinem):  $\theta_g = \theta_A \theta_B$
  - Málo používané, řada komplikací.
  - Lze použít např. pro parametrizaci teorie vědomostních prostorů.

# Multidimenzionální IRT (MIRT)

---

McDonaldův MIRT založený na normální ogivě

- Technicky vzato faktorová analýza s nelineární parametrizací.
- Tzv. probit parametrizace.

vs. Reckaseho logistický model.

- Protože normální ogiva je blízká logistické funkci, výsledky jsou v praxi velmi podobné.
- Logistický model dnes jednoznačně vede (McDonaldův model se zpravidla odhaduje prostřednictvím ordinální CFA).
- Tzv. logit parametrizace.

# MIRT: Latentní rysy

---

Model může být exploratorní (EFA MIRT) nebo konfirmační (CFA MIRT).

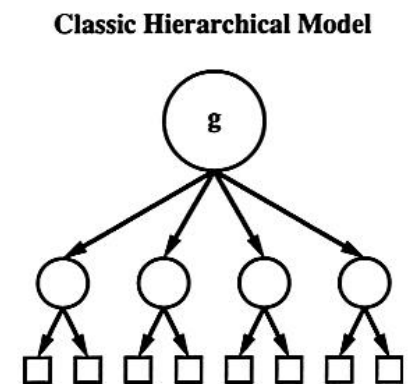
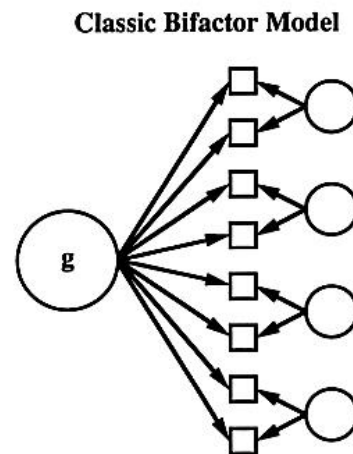
- Rotace u exploračních modelů stejně jako v EFA.

Každé osobě je přiřazen vektor latentních rysů, pro každou dimenzi jeden.

- Mohou být korelované nebo nekorelované.

Namísto hierarchických modelů jako v SEM se používá bifaktorový model.

Explorační MIRT model lze rotovat stejně, jako EFA.



# Reckaseho (1983) model

---

V případě dvou faktorů:

$$p(x_i = 1 | \theta_1, \theta_2) = \frac{e^{a_{i1}(\theta_1 - b_{i1}) + a_{i2}(\theta_2 - b_{i2})}}{1 + e^{a_{i1}(\theta_1 - b_{i1}) + a_{i2}(\theta_2 - b_{i2})}}$$

- Tzv. tradiční IRT parametrizace

Což lze přepsat na:

$$p(x_i = 1 | \theta_1, \theta_2) = \frac{e^{(a_{i1}\theta_1 + a_{i2}\theta_2) + \gamma_i}}{1 + e^{(a_{i1}\theta_1 + a_{i2}\theta_2) + \gamma_i}}$$

- Kde  $\gamma_i = -(a_{i1}b_{i1} + a_{i2}b_{i2})$  z původního zápis.
- Tzv. **slope-intercept parametrizace** (proto dvoje summary v MIRTu!).

Pro n-faktorový model pak:

$$p(x_i = 1 | \theta_1, \dots, \theta_n) = \frac{e^{\sum_{j=1}^n (a_{ij}\theta_j) + \gamma_i}}{1 + e^{\sum_{j=1}^n (a_{ij}\theta_j) + \gamma_i}}$$



# MIRT: diskriminace

---

Pro kombinaci každého rysu (1-a položky ( $i$ )) je vlastní diskriminační parametr  $a_{ki}$ .

Diskriminační „síla“ položky je zpravidla součtem diskriminačních parametrů:

$$\mu_{ki} = \sqrt{\sum_{k=1}^{N_i} a_{ki}^2}$$

Typicky se používá 2PL MIRT model.

- V případě Raschova MIRT modelu  $a_{ki}=1$ , a tedy každou položku sytí právě jeden faktor (a všechny stejně). Jde tedy vlastně jen o souběžný odhad více Raschových modelů najednou.
- Stejný výsledek, jako separátní odhad a následný součet chí-testů, jen korelace faktorů je odhadnuta na úrovni latentní úrovni.

# MIRT: Ostatní

---

Namísto charakteristické křivky testu je definovaná „charakteristická plocha testu“.

- Ale její výpočet je analogický.

Obdobně pak „informační plocha“ testu...

- ... vzniká součtem informačních ploch položek.
- Zajímá nás rovněž, ve směru které dimenze chceme diskriminovat, podle toho se může odhad informační funkce lišit.

Software:

- R
  - MIRT package v R pro full-information FA, skvělý balíček
  - Funkce irt.fa v balíčku psych
- Komerční SW
  - Mplus, TESTFACT, flexMIRT,

# MIRT: Ostatní

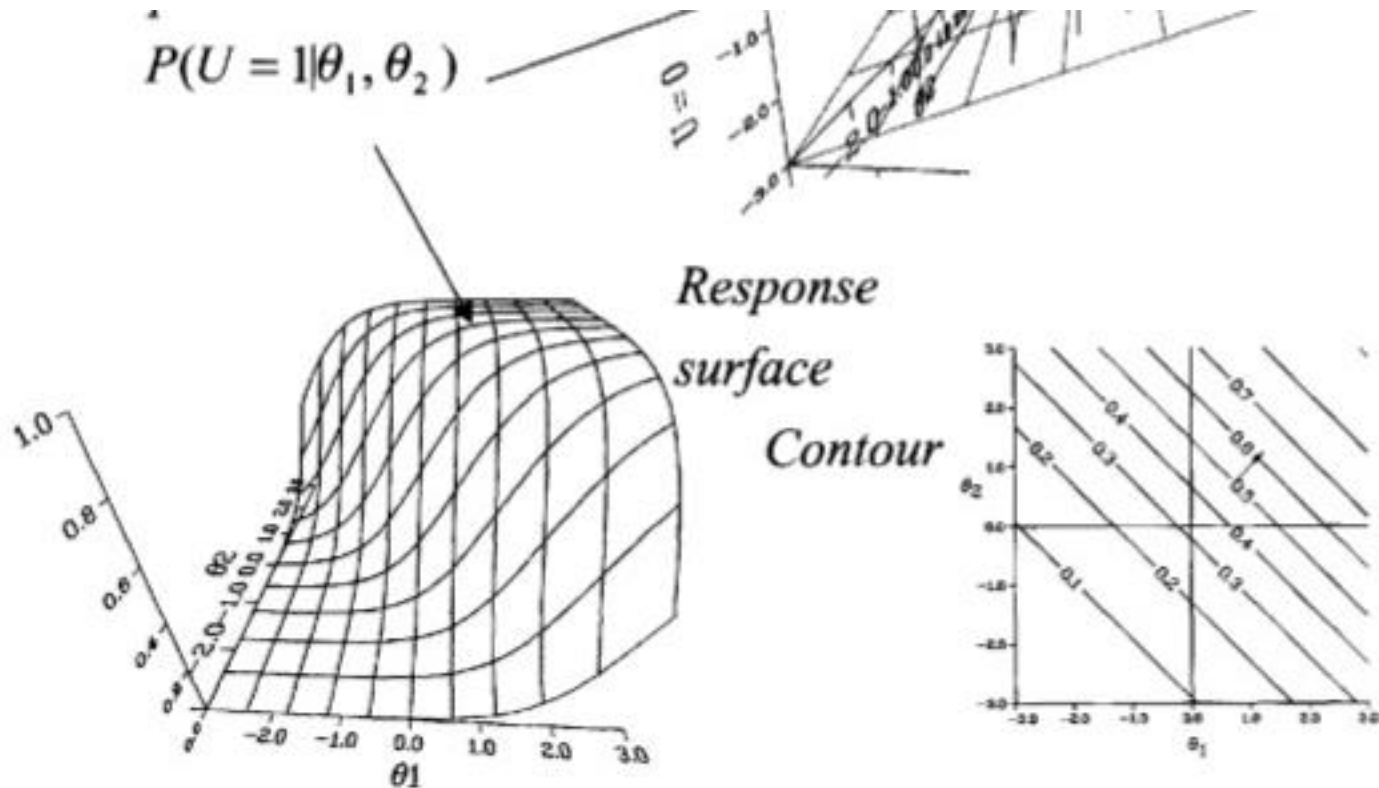


FIG. 1.3. (Cont.)

# Test equating

---

Vyvážení obtížnosti jednotlivých forem testu.

- V high stakes testech jednorázové vyvážení – sjednocení obtížností a srovnání probandů napříč formami testu.
- V psychologických metodách vyvážení skóru paralelních forem a vyvinutí rovnocenných nástrojů.
- **Linking** (prosté srovnání měřítek) vs. **equating** (srovnání odhadů).

Předpoklad: Obě formy měří stejný konstrukt (otázka validity).

GRE, SAT: od konce 80./začátku 90. let je (v USA) IRT equating high stakes testů normou.

Typické kroky: volba designu, sběr dat, samotná transformace.

# Test equating

---

Tři klasické (CTT) způsoby:

- Vyvažování na základě průměru (M) – testy musí mít stejné rozptyly, data musí být normálně rozdělená.  $x_2 = x_1 + \bar{X}_2 - \bar{X}_1$
- Lineární vyvažování (M, SD) – rozptyly se mohou lišit, data musí být normální.  $x_2 = \bar{X}_2 + \frac{\sigma_2}{\sigma_1}(x_1 - \bar{X}_1)$  (transformace přes z-skór)
- Equipercilové vyvažování – varianty jsou upraveny tak, aby tentýž skór měl v obou variantách stejný percentil. Výsledkem je stejné rozdělení dat, je silně závislé na vzorku (použitelné jen u velkých souborů).
  - Používá se i pro standardizaci nenormálních skórů na normální.
  - Percilové vyvažování není vyvažování, percentil z principu ztrácí část informace. Žádné zvláštní požadavky na data.

IRT vyvažování bylo prvními hromadnými aplikacemi IRT do praxe.

# IRT equating: Sběr dat

---

## Designy s jednou výzkumnou skupinou

- Skupinu rozdělíme náhodně na dvě (tři...) podskupiny.
- **Counterbalancing** – Jeden test administrujeme jedné skupině dvakrát (střídáme pořadí). Data musejí být sebrána vždy ve stejném čase!
- **Náhodné skupiny** – každé osobě administrujeme test jen jednou.

## Design s více skupinami:

- Dvě nezávislé/nenáhodné skupiny, ale oba testy mají společné položky (tzv. „kotvu“ – **anchor test**), které slouží ke kalibraci.
- Ta může, ale nemusí být zahrnuta pro zjištění celkového skóru.
- Kotev může být více („planned missing data design“).
- Položky mohou být prekalibrované a sloužit k odhadu nových parametrů do staré škály.

# IRT equating: Transformace skóru

---

## Kalibrace „do stejné škály“

- a) IRT analýza proběhne naráz nad všemi daty. IRT nevdí chybějící data!
- b) IRT parametry odhadneme z analýzy prvního datasetu a „zafixujeme“ pro analýzu druhého.

## Metrická transformace.

- Využívá odhadnuté transformační koeficienty.
- Větší množství postupů, typické pro high stakes testy.
  - Předkalibrované položky, na které vyvažujeme aktuální sběr dat.
- Např. lineární transformace (viz výše), ICC linking a další.
- Obecně např. nedojde ke změně, když přičteme konstantu k parametrům obtížnosti a odhadům latentních rysů.