



Európska únia

Európsky sociálny fond

NÚCEM

NÁRODNÝ ÚSTAV CERTIFIKOVANÝCH
MERANÍ VZDELÁVANIA



Moderné vzdelávanie pre vedomostnú spoločnosť/Projekt je spolufinancovaný zo zdrojov EÚ

IRT - Item response theory

Úvod

- Učitelía (až na vzácne výnimky) poznajú výsledky svojich žiakov vo forme:

Data Matrix for 14 Selected Persons (A–N) on 12 Selected Items (a–l)

<i>Persons</i>	<i>Items</i>												<i>Raw Score</i>
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	
A	✓	✓	✓	X	X	X	X	✓	✓	X	X	✓	6
B	✓	X	✓	✓	X	X	X	X	✓	X	X	X	4
C	✓	✓	✓	X	✓	X	X	✓	✓	✓	✓	✓	9
D	✓	X	✓	✓	X	X	X	X	✓	X	X	✓	5
E	X	✓	✓	X	X	✓	X	✓	✓	✓	✓	✓	8
F	✓	✓	✓	✓	X	X	X	✓	✓	X	X	✓	7
G	✓	X	✓	X	X	✓	X	X	✓	X	✓	✓	6
H	✓	X	✓	X	X	X	X	X	X	X	X	✓	3
I	✓	✓	✓	✓	X	X	X	✓	✓	X	X	✓	7
J	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	X	10
K	✓	X	✓	X	X	✓	X	X	✓	X	✓	✓	6
L	X	✓	✓	X	X	✓	X	✓	✓	✓	✓	✓	8
M	X	X	X	X	X	X	X	X	X	X	X	X	0
N	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12

Zdroj: Bond, Fox – Applying the Rasch model: Fundamental measurement in human sciences

- Preusporiadaná tabuľka I

Selected Data Matrix With Persons Arranged According to Ability
(From Top to Bottom) and Items Arranged by Facility (From Left to Right)

<i>Persons</i>	<i>Items</i>												<i>Ability</i>
	<i>c</i>	<i>i</i>	<i>a</i>	<i>l</i>	<i>b</i>	<i>h</i>	<i>k</i>	<i>d</i>	<i>f</i>	<i>j</i>	<i>e</i>	<i>g</i>	
N	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12
J	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	✓	10
C	✓	✓	✓	✓	✓	✓	✓	X	X	✓	✓	X	9
E	✓	✓	X	✓	✓	✓	✓	X	✓	✓	X	X	8
L	✓	✓	X	✓	✓	✓	✓	X	✓	✓	X	X	8
I	✓	✓	✓	✓	✓	✓	X	✓	X	X	X	X	7
F	✓	✓	✓	✓	✓	✓	X	✓	X	X	X	X	7
K	✓	✓	✓	✓	X	X	✓	X	✓	X	X	X	6
A	✓	✓	✓	✓	✓	✓	X	X	X	X	X	X	6
G	✓	✓	✓	✓	X	X	✓	X	✓	X	X	X	6
D	✓	✓	✓	✓	X	X	X	✓	X	X	X	X	5
B	✓	✓	✓	X	X	X	X	✓	X	X	X	X	4
H	✓	X	✓	✓	X	X	X	X	X	X	X	X	3
M	X	X	X	X	X	X	X	X	X	X	X	X	0
Facility	13	12	11	11	8	8	7	6	6	3	3	2	

- Preusporiadaná tabuľka II

Selected Ordered Data Matrix for Items
and Persons With Sufficient Information

<i>Persons</i>	<i>Items</i>											<i>Ability</i>	<i>n/N%</i>
	<i>i</i>	<i>a</i>	<i>l</i>	<i>b</i>	<i>h</i>	<i>k</i>	<i>d</i>	<i>f</i>	<i>j</i>	<i>e</i>	<i>g</i>		
J	1	1	0	1	1	1	1	1	0	1	1	9	82
C	1	1	1	1	1	1	0	0	1	1	0	8	73
E	1	0	1	1	1	1	0	1	1	0	0	7	64
L	1	0	1	1	1	1	0	1	1	0	0	7	64
I	1	1	1	1	1	0	1	0	0	0	0	6	55
F	1	1	1	1	1	0	1	0	0	0	0	6	55
K	1	1	1	0	0	1	0	1	0	0	0	5	45
A	1	1	1	1	1	0	0	0	0	0	0	5	45
G	1	1	1	0	0	1	0	1	0	0	0	5	45
D	1	1	1	0	0	0	1	0	0	0	0	4	36
B	1	1	0	0	0	0	1	0	0	0	0	3	27
H	0	1	1	0	0	0	0	0	0	0	0	2	18
Facility	11	10	10	7	7	6	5	5	3	2	1		
n/N%	93	83	83	58	58	50	42	42	25	17	08		

- Sú všetci so skóre 5 rovnako dobrí?
- Je položka *d* dobrá?
- Ktoré položky jasne oddeľujú veľmi dobrých, a ktoré veľmi slabých?

- Používanie iba percentuálneho hodnotenia má za následok, že sa výsledky študentov zhromaždia niekde v strede a nemožno adekvátne rozlíšiť výsledky veľmi dobrých a veľmi slabých študentov.
- Je ten istý rozdiel medzi študentmi s úspešnosťami 50% vs. 55% a medzi študentmi s úspešnosťami 90% vs. 95%?
- Thurstone (1920) navrhol previesť celkové skóre (úspešnosť) do prirodzených logaritmov.

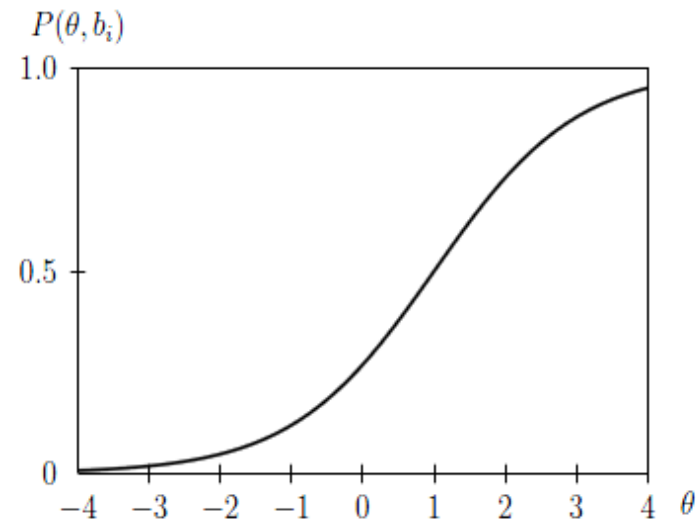
Raw %	Log odds
50%	0
55%	0,087
90%	0,954
95%	1,278

$$L = \log(0,55/0,45) = 0,087$$

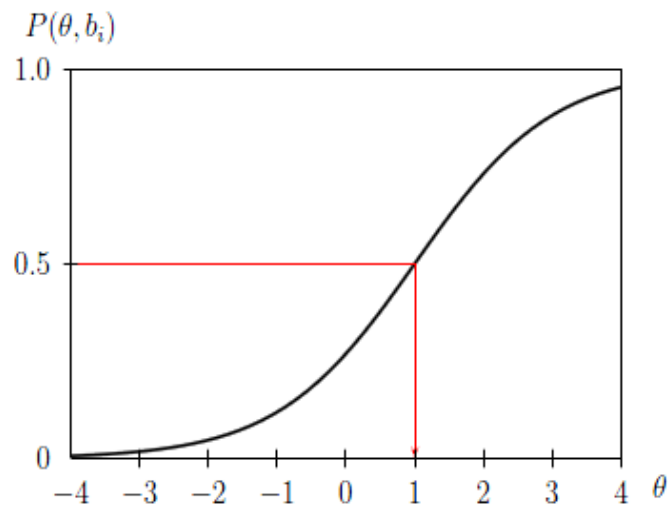
Jednoparametrický logistický model (1PL) – Raschov model

- IRT model predikuje pravdepodobnosť, že určitá osoba zvolí určitú odpoveď na danú položku.
- 1PL model predpovedá pravdepodobnosť správnej odpovede z interakcie medzi abilitou θ_j osoby a parametrom položky b_i

$$P_{ij}(\theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}$$



- Parameter b_i sa nazýva parameter náročnosti (*difficulty parameter*) položky



- Na osi ability/difficulty je to tá hodnota, pre ktorú

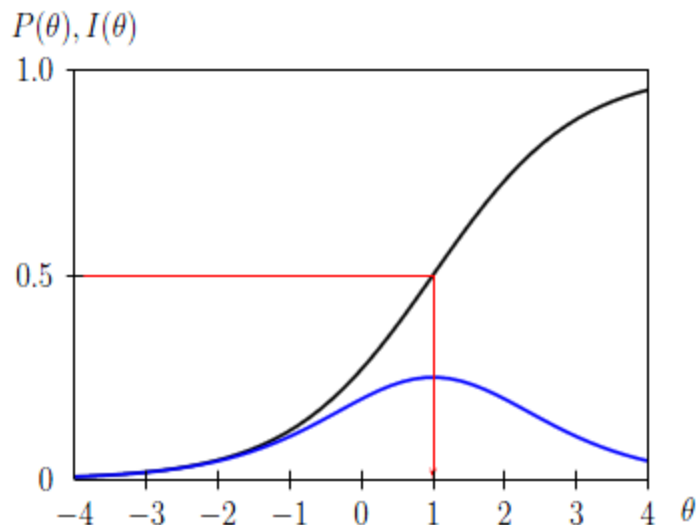
$$P_{ij}(\theta_j - b_i) = \frac{1}{2}$$

Informačná funkcia položky - IIF

- IIF sa vzťahuje na presnosť, s akou môžeme odhadnúť abilitu.
- Každá položka v teste poskytuje nejakú informáciu o abilitе testovaného, jej množstvo závisí na tom, ako tesne sa náročnosť položky spája s abilitou osoby

$$I_i(\theta, b_i) = P_i(\theta, b_i)Q_i(\theta, b_i).$$

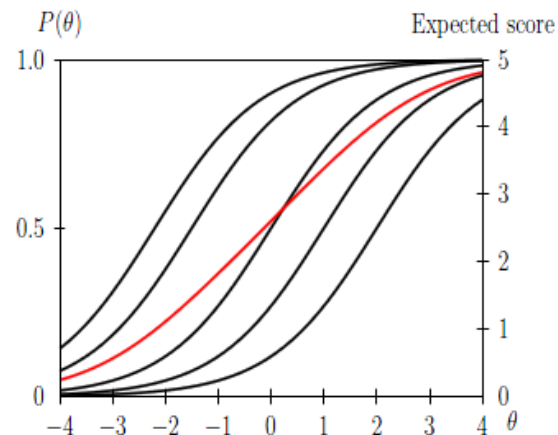
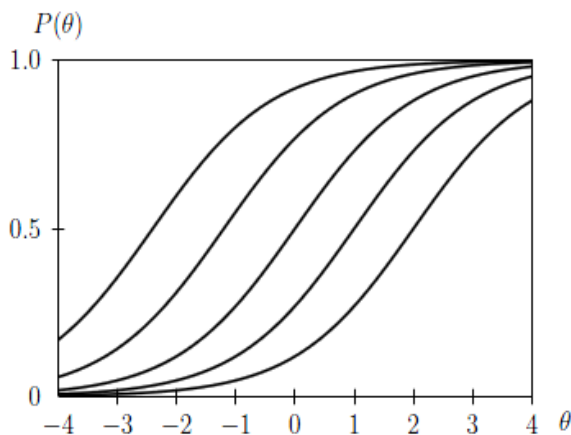
- Maximum je 0,25 a to v bode, kde sa abilita rovná náročnosti položky.



- Položka poskytuje najviac informácie pre osoby, ktorých abilita je rovnaká, ako náročnosť položky.
- Ak chceme dosiahnuť dobré merania, tak potrebujeme položky rôznej náročnosti.

The test response function - TRF

- TRF predpovedá očakávané skóre v teste ako funkciu latentnej ability a parametrov položiek.
- Je súčtom položkových odpoved'ových funkcií (IRF).
- Potrebujeme veľa položiek, ktorých náročnosť je rozptýlená pozdĺž celého intervalu ability.

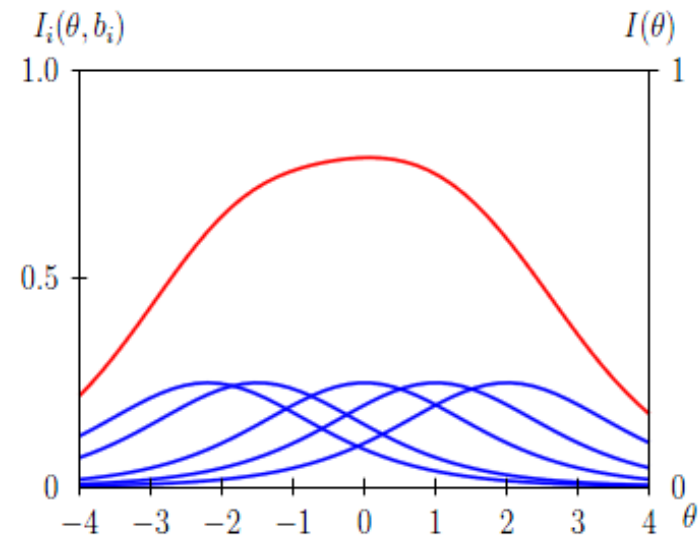


- IRF funkcia pre položky s náročnosťami (-2.2, -1.5, 0.0, 1.0, 2.0).

The test information function

- Je súčtom položkových informačných funkcií

$$I_j(\theta_j) = \sum_i I_{ij}(\theta_j, b_i).$$



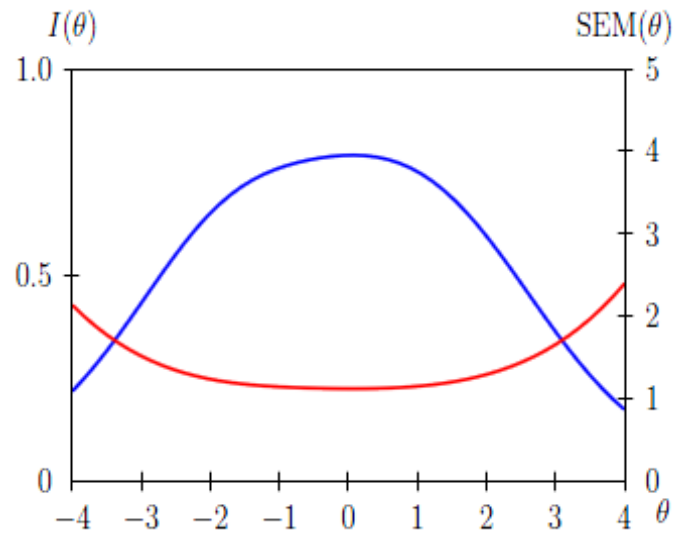
- Test ako celok je viac informatívny, ako jeho jednotlivé položky.
- TIF predikuje presnosť, s akou môžeme merať určitú hodnotu latentnej ability.
- Nevieme merať abilitu, ale vieme, akú presnosť dosiahneme na ktorejkoľvek úrovni.

Presnosť a chyba merania

- Presnosť je opakom chyby (error).
- Porovnáme ju s odhadom ability alebo na stanovenie intervalu spoľahlivosti pre odhad ability

$$\text{Var}(\hat{\theta}) = \frac{1}{I(\hat{\theta})}.$$

- SEM (*standard error of measurement*)



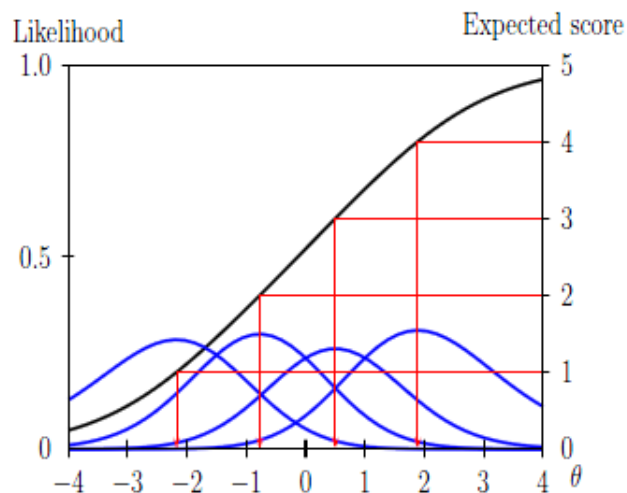
Odhad ability v 1PL

- Metóda maximálnej vierohodnosti
- Predpoklad *lokálnej nezávislosti* – odpovede na jednotlivé položky v teste sú navzájom nezávislé pre danú úroveň ability

$$L(\theta) = \prod_i P_i(\theta, b_i)^{u_i} Q_i(\theta, b_i)^{1-u_i} \quad u_i \in \{0; 1\}$$

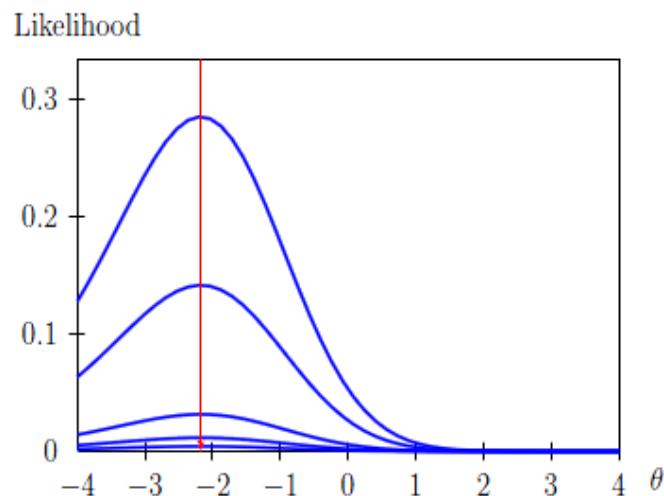
- Pravdepodobnosť odpoved'ového vzoru pre ability θ a položkové parametre
- Odhad ability v 1PL závisí iba od toho, koľko položiek bolo zodpovedaných správne.

- Odpoveďové vzory: (TFFFF), (TTFFF), (TTTFF), (TTTTF)



Odhad ability maximálnou vierohodnosťou

- Funkcie vierohodnosti pre odpoveďové vzory s celkovým skóre 1



Rasch model – polytomous case model PCM

- Likertova škála, stupnica od 0 – 5, problem – solving, ...
- Odvodenie PCM (Masters 1982) z Raschovho dichotomického modelu pre susedný pár skóre kategórií
pre prípad 3 kategórií (0, 1, 2)

$$p_{0/0,1} = \Pr(X = 0 / X = 0 \text{ or } X = 1) = \frac{\Pr(X = 0)}{\Pr(X = 0) + \Pr(X = 1)} = \frac{1}{1 + \exp(\theta - \delta_1)}$$

$$p_{1/0,1} = \Pr(X = 1 / X = 0 \text{ or } X = 1) = \frac{\Pr(X = 1)}{\Pr(X = 0) + \Pr(X = 1)} = \frac{\exp(\theta - \delta_1)}{1 + \exp(\theta - \delta_1)}$$

$$p_{1/1,2} = \Pr(X = 1 / X = 1 \text{ or } X = 2) = \frac{\Pr(X = 1)}{\Pr(X = 1) + \Pr(X = 2)} = \frac{1}{1 + \exp(\theta - \delta_2)}$$

$$p_{2/1,2} = \Pr(X = 2 / X = 1 \text{ or } X = 2) = \frac{\Pr(X = 2)}{\Pr(X = 1) + \Pr(X = 2)} = \frac{\exp(\theta - \delta_2)}{1 + \exp(\theta - \delta_2)}$$

PCM pravdepodobnosti pre všetky kategórie

- Uvažujeme všetky kategórie súčasne

$$p_0 = \Pr(X = 0) = \frac{1}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))}$$

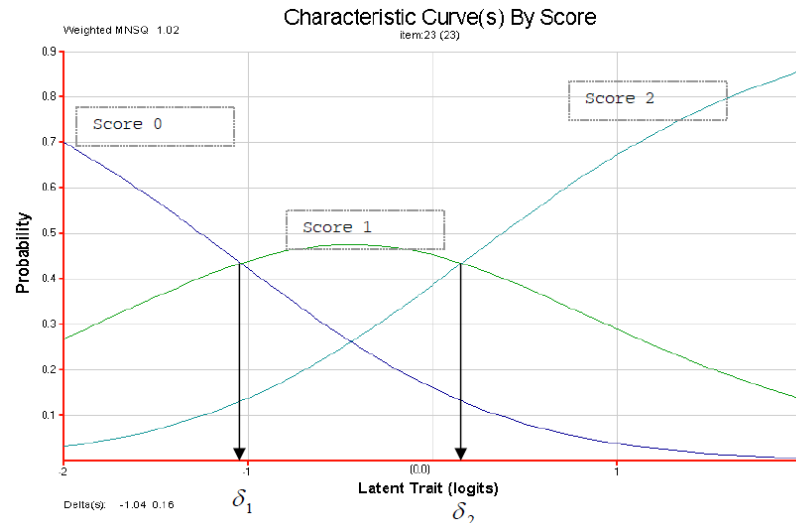
$$p_1 = \Pr(X = 1) = \frac{\exp(\theta - \delta_1)}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))}$$

$$p_2 = \Pr(X = 2) = \frac{\exp(2\theta - (\delta_1 + \delta_2))}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))}$$

$$\Pr(X_{ni} = x) = \frac{\exp \sum_{k=0}^x (\theta_n - \delta_{ik})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^h (\theta_n - \delta_{ik})}$$

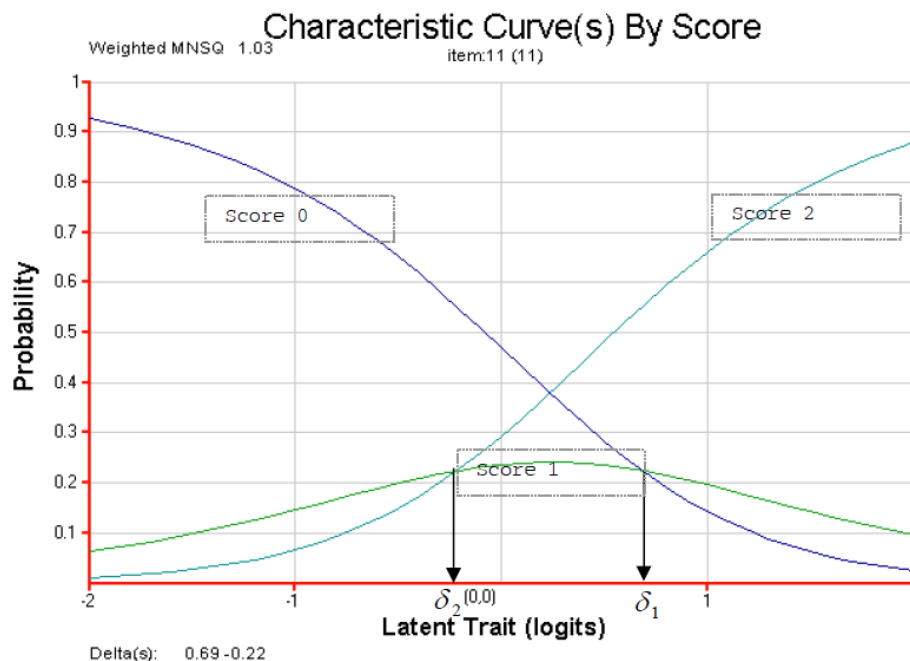
Interpretácia parametrov δ_k

- δ_k nie je parameter náročnosti pre zisk k bodov!



- δ_k je bod, v ktorom pravdepodobnosť byť v kategórii $k-1$, je rovnaká, ako pravdepodobnosť byť v kategórii k .
- Od $-\infty$ do δ_1 je najpravdepodobnejšie skóre 0.
- Od δ_1 do δ_2 je najpravdepodobnejšie skóre 1.
- Od δ_2 do ∞ je najpravdepodobnejšie skóre 2.

Problémy s interpretáciou parametrov δ_k



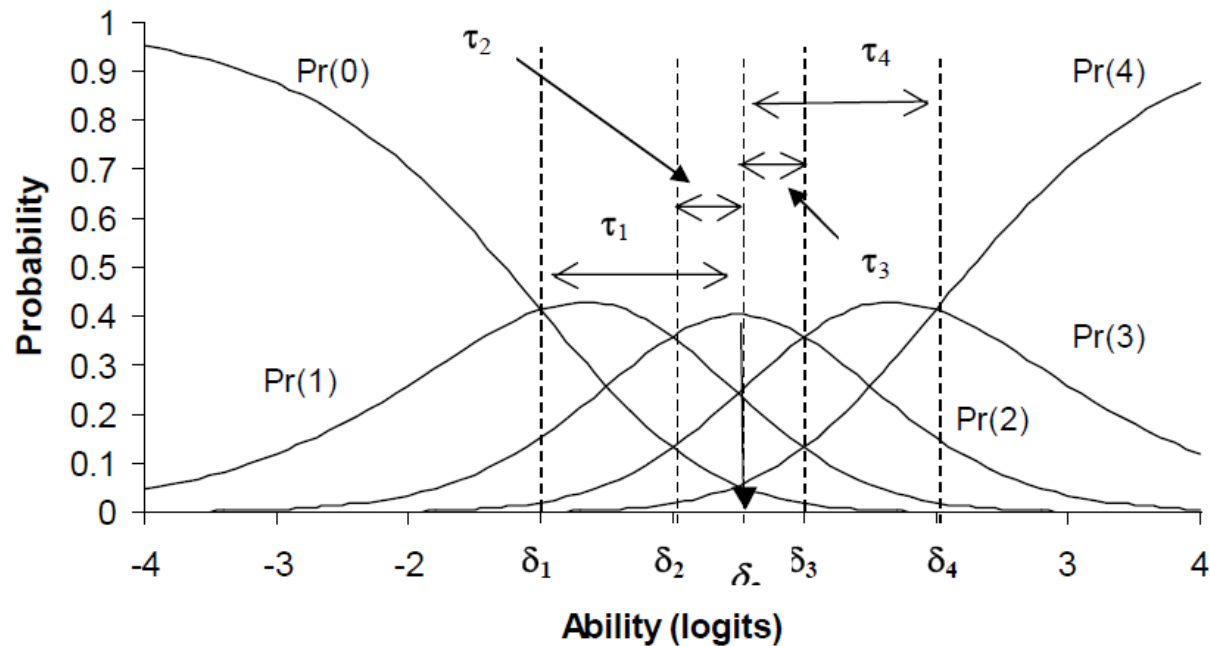
- Parametre δ_1 , δ_2 nie sú usporiadané.
- Tento jav je nevýhodou použitia parametrov δ pri interpretácii odpovedí vo vzťahu k abilitě.
- Hodnoty δ závisia od náročnosti jednotlivých krokov.

τ a $\delta_{\bullet}$

- Alternatívou parametrizácie PCM sú τ a $\delta_{\bullet}$.

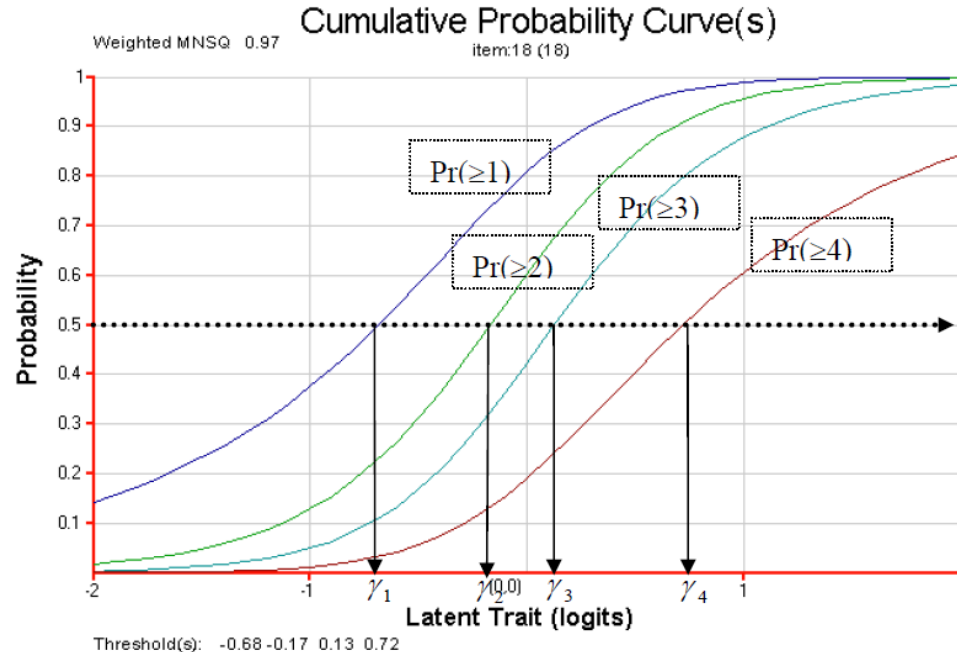
$$\delta_{\bullet} = \frac{1}{m} \sum_{k=1}^{m_i} \delta_k$$

$$\tau_k = \delta_{\bullet} - \delta_k$$



Thurstonove hranice - γ

- Pre kategóriu skóre sa definuje ako abilita, pri ktorej pravdepodobnosť dosiahnutia tohto alebo vyššieho skóre je práve 0,5.

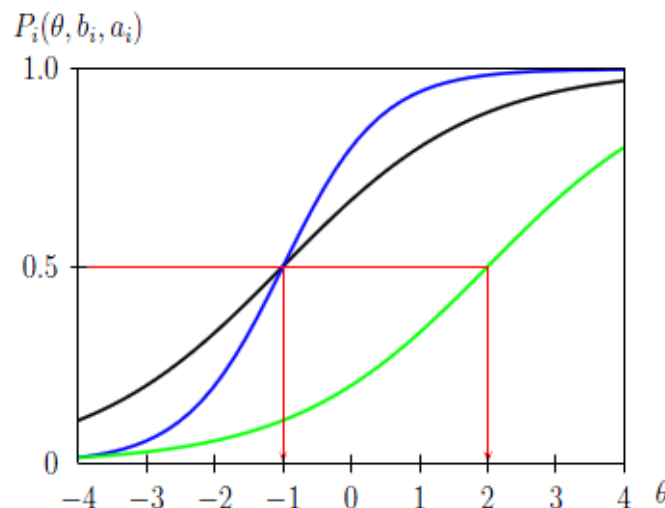


- Deliaci bod pre rozdelenie kontinua ability do „oblastí skóre“
- γ_i je miera náročnosti pre skóre i

Dvojparametrický logistický model 2PL

- Parameter a_i - diskriminačný parameter i – tej položky (parameter sklonu)

$$P_{ij}(\theta_j, b_i, a_i) = \frac{\exp[a_i(\theta_j - b_i)]}{1 + \exp[a_i(\theta_j - b_i)]}$$



- Dve položky majú rovnakú náročnosť -1.0, ale pretínajú sa. Modrá je strmšia ako čierna, pretože má väčšiu hodnotu diskriminačného parametra.
- Zelená krivka má rovnakú hodnotu diskriminačného parametra ako modrá, ale väčšiu náročnosť.

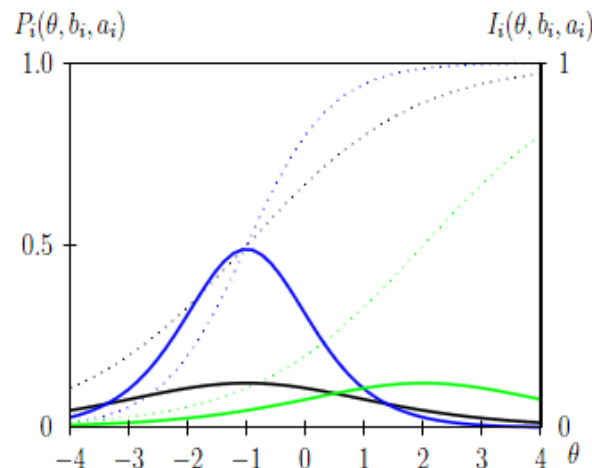
Test response function

- Súčet IRF cez všetky položky v teste
- TRF je v normálnom prípade rastúcou funkciou ability (položky so zápornou diskrimináciou sú zlé položky).
- Ak je krivka príliš široká, tak očakávané skóre nie je veľmi citlivé na rozdiely v skutočnej abilitě.
- Test lepšie „diskriminuje“ – rozlišuje medzi osobami s rôznou abilitou.

The item information function - IIF

$$I_j(\theta_j) = \sum_i I_{ij}(\theta_j, b_i, a_i) = \sum_i a_i^2 P(\theta, b_i, a_i) Q(\theta, b_i, a_i)$$

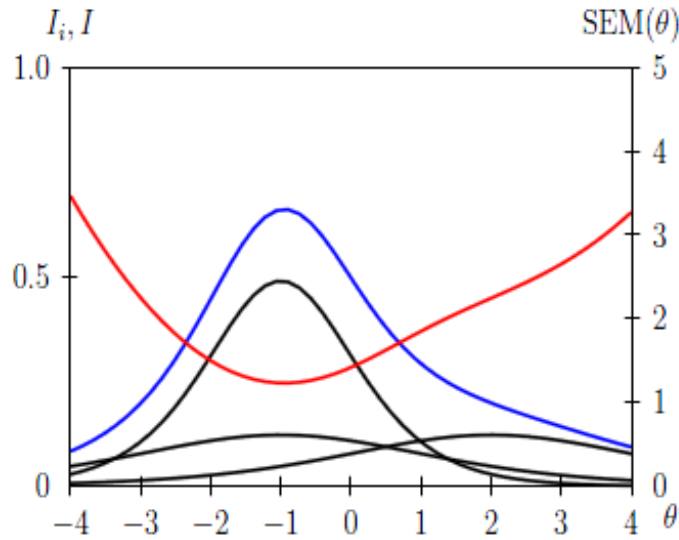
- Vplyv diskriminačného parametra je silnejší
Diskriminačný parameter menší ako 1 môže znižovať hodnotu IF dramaticky, kým diskriminačný parameter nad 1 jej hodnotu podstatne zvýši.



- Pri veľkej diskriminácii položka poskytuje viac informácie o abilitate.
- Položky s nízkou hodnotou diskriminačného parametra sú menej informatívne

SEM in 2PL

$$\text{SEM}(\theta) = \sqrt{1 / \sum_i a_i^2 P_i(\theta, b_i, a_i) Q_i(\theta, b_i, a_i)}.$$



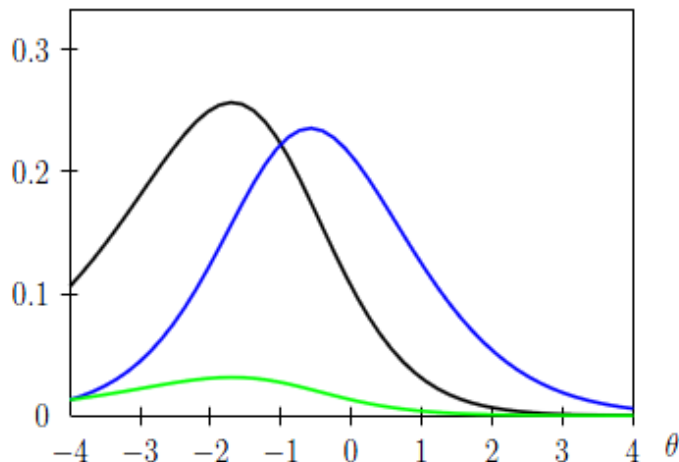
Modrá – TIF
Červená – SEM
Čierne – IRF položiek v teste

- Ak test obsahuje málo položiek, náročnosti sú rozložené nerovnomerne a diskriminácie sú veľmi rozdielne, tak testová TIF a SEM dominujú na jedinej vysoko diskriminujúcej položke.

Odhad ability v 2PL

- v 1PL modeli pre odhad ability platilo: $\sum_i P(\hat{\theta}, b_i) = \sum_i u_i$
- V 2PL modeli pre odhad ability platí: $\sum_i a_i P(\hat{\theta}, b_i, a_i) = \sum_i a_i u_i$
- Rovnaké hodnoty skóre už nemusia viesť k rovnakému odhadu ability!
Nezáleží už iba na tom, koľko bolo správne zodpovedaných, ale aj ktoré.

Likelihood



modrá: $a = 1$ $b = -1$

čierna: $a = 0,5$ $b = -1$

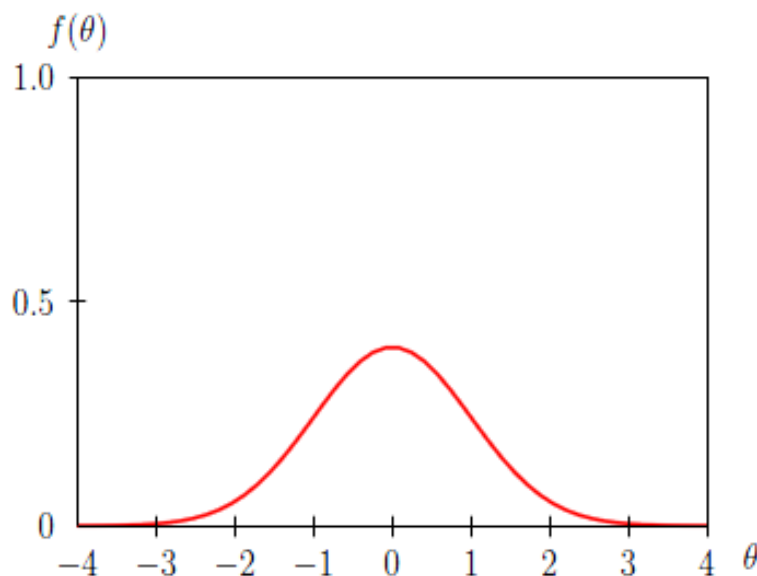
zelená: $a = 0,5$ $b = 2$

odpoveďový vzor: (TFF, FTF, FFT)

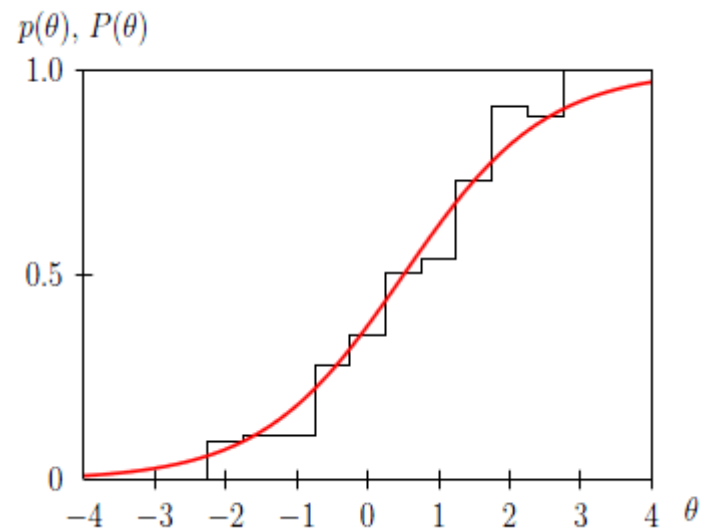
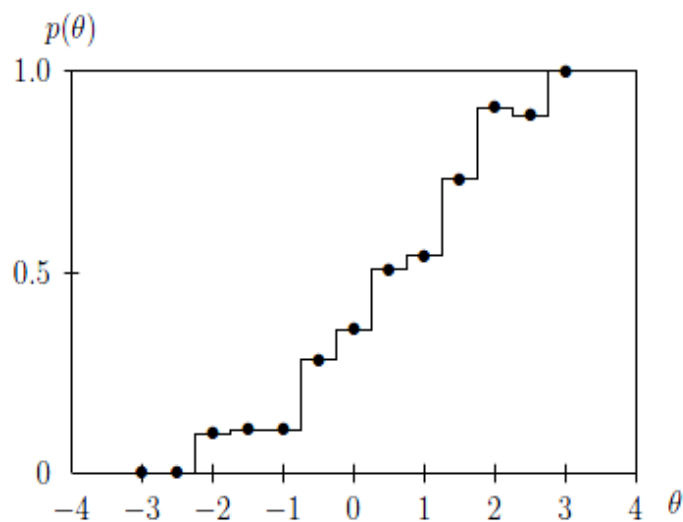
farba určuje, ktorá položka bola zodpovedaná správne

Odhad parametrov položiek

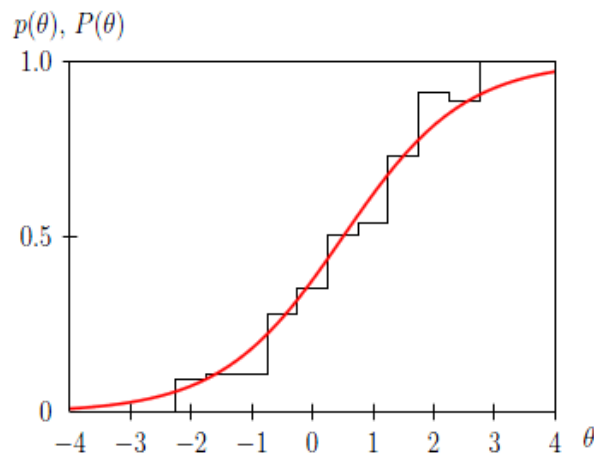
- Predpokladáme, že abilita osôb má normálne rozdelenie.
- Predpokladáme, že poznáme ability osôb.
- Snažíme sa odhadnúť parametre logistickej krivky.



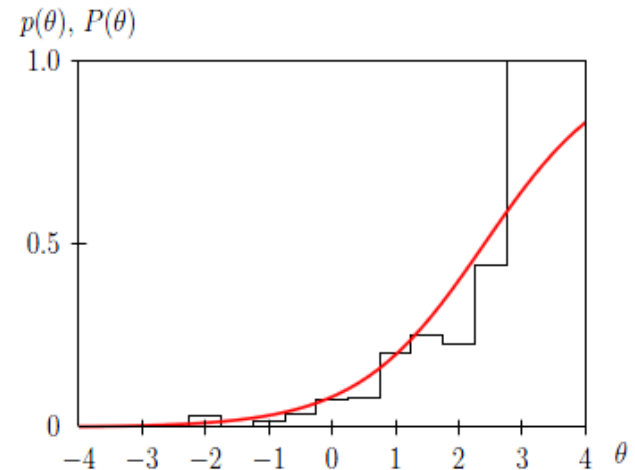
- Vzorka 1000 osôb, ktorým dáme stredne ťažkú položku ($b = 0,5$)



- Čo sa stane, ak tej istej skupine dáme ťažkú položku ($b = 2.4$)?
- Dostaneme ten istý odhad ability?



$b = 0.5$



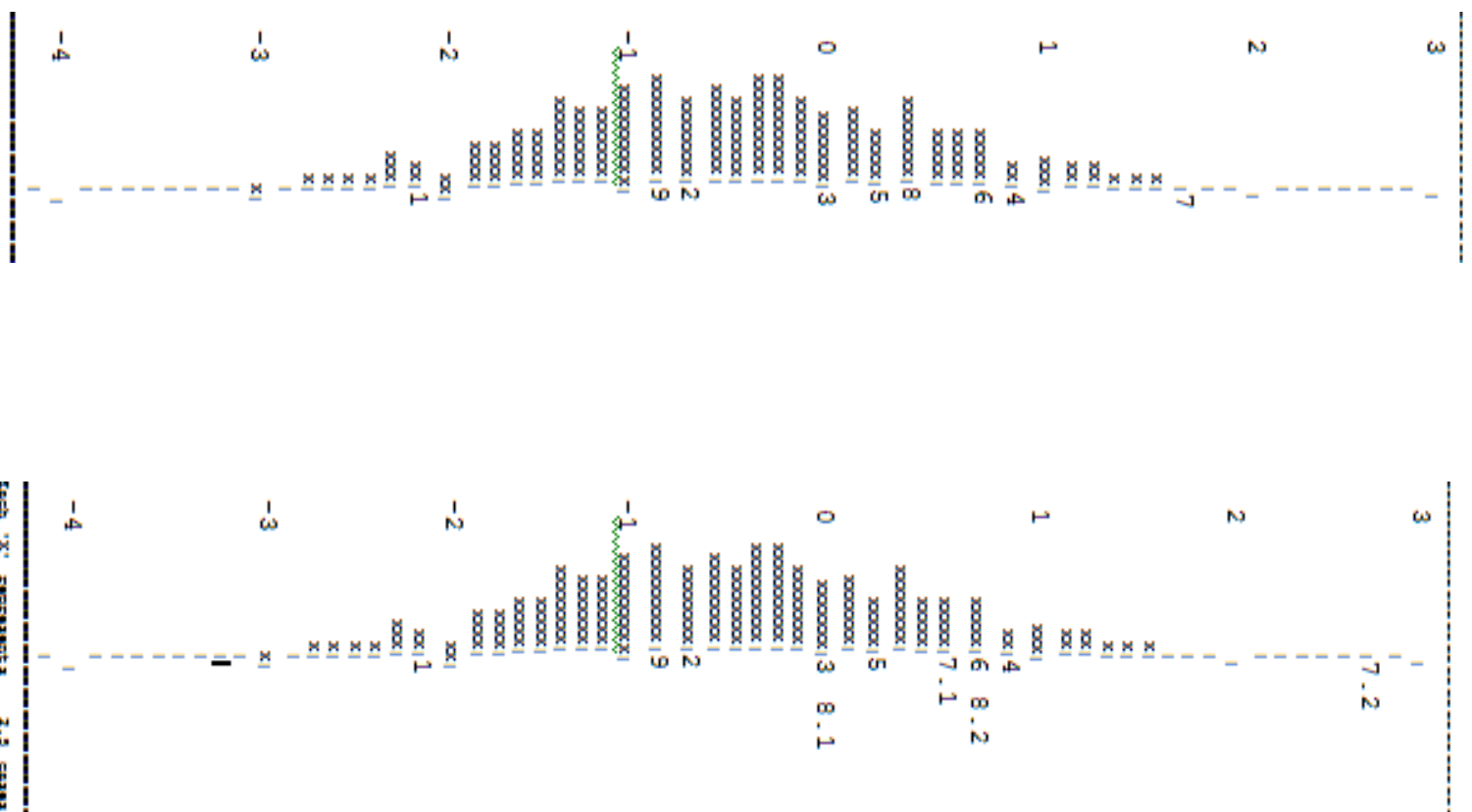
$b = 2.4$

- Parametre položiek dokážeme odhadovať z akejkolvek časti IRC (item response curve), čo znamená, že parametre položiek dokážeme odhadovať z akejkolvek skupiny osôb – ***skupinová invariancia***.

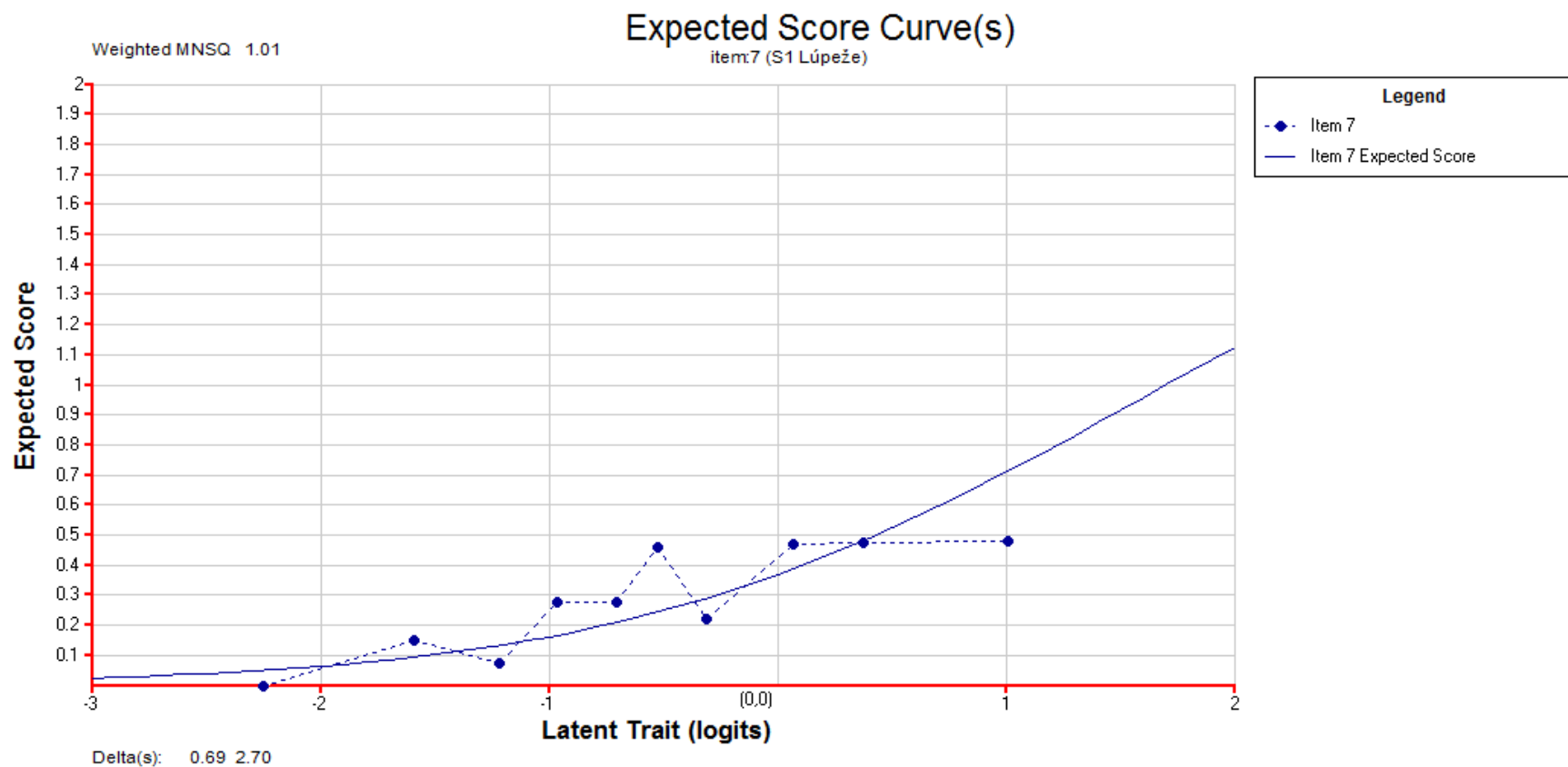
Niektoré úlohy z projektu

- [do prezentácie\K1 hodnotenie_PISA kódy.docx](#)
- Výsledky písomnej práce
[do prezentácie\item 4 \(S10 V8sledky písomnej práce\) - ICC.png](#)
[do prezentácie\item 4 \(S10 V8sledky písomnej práce\) - Item Information.png](#)
- Lúpeže
[do prezentácie\item 7 \(S1 Lúpeže\) - CCC.png](#)
[do prezentácie\item 7 \(S1 Lúpeže\) - Expected.png](#)
[do prezentácie\item 7 \(S1 Lúpeže\) - Item Information.png](#)
- Podpora prezidentovi
[do prezentácie\item 8 \(S17 Podpora prezidentovi\) - ICC.png](#)
[do prezentácie\item 8 \(S17 Podpora prezidentovi\) - Expected.png](#)
[do prezentácie\item 8 \(S17 Podpora prezidentovi\) - Item Information.png](#)

- Item – person map

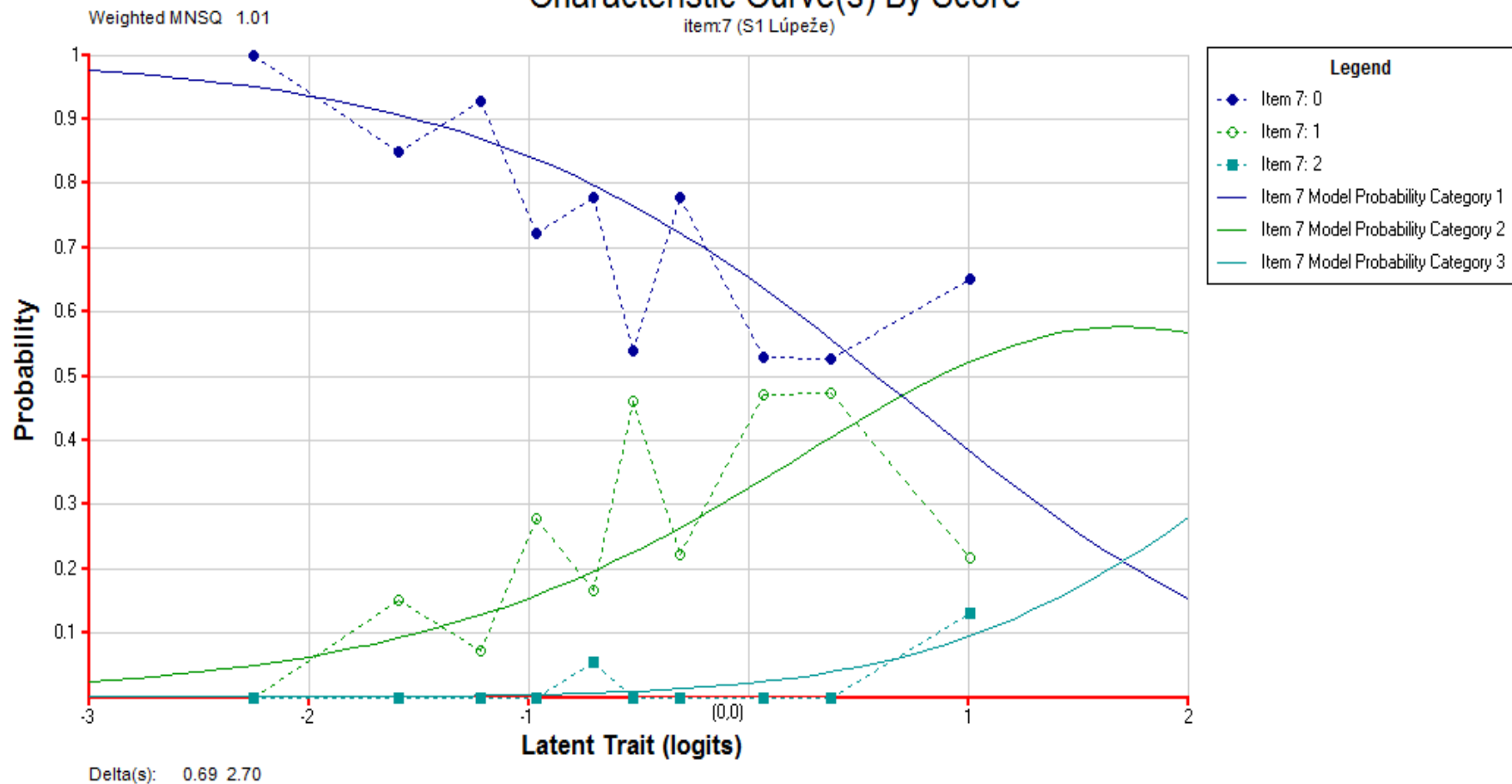


Lúpeže

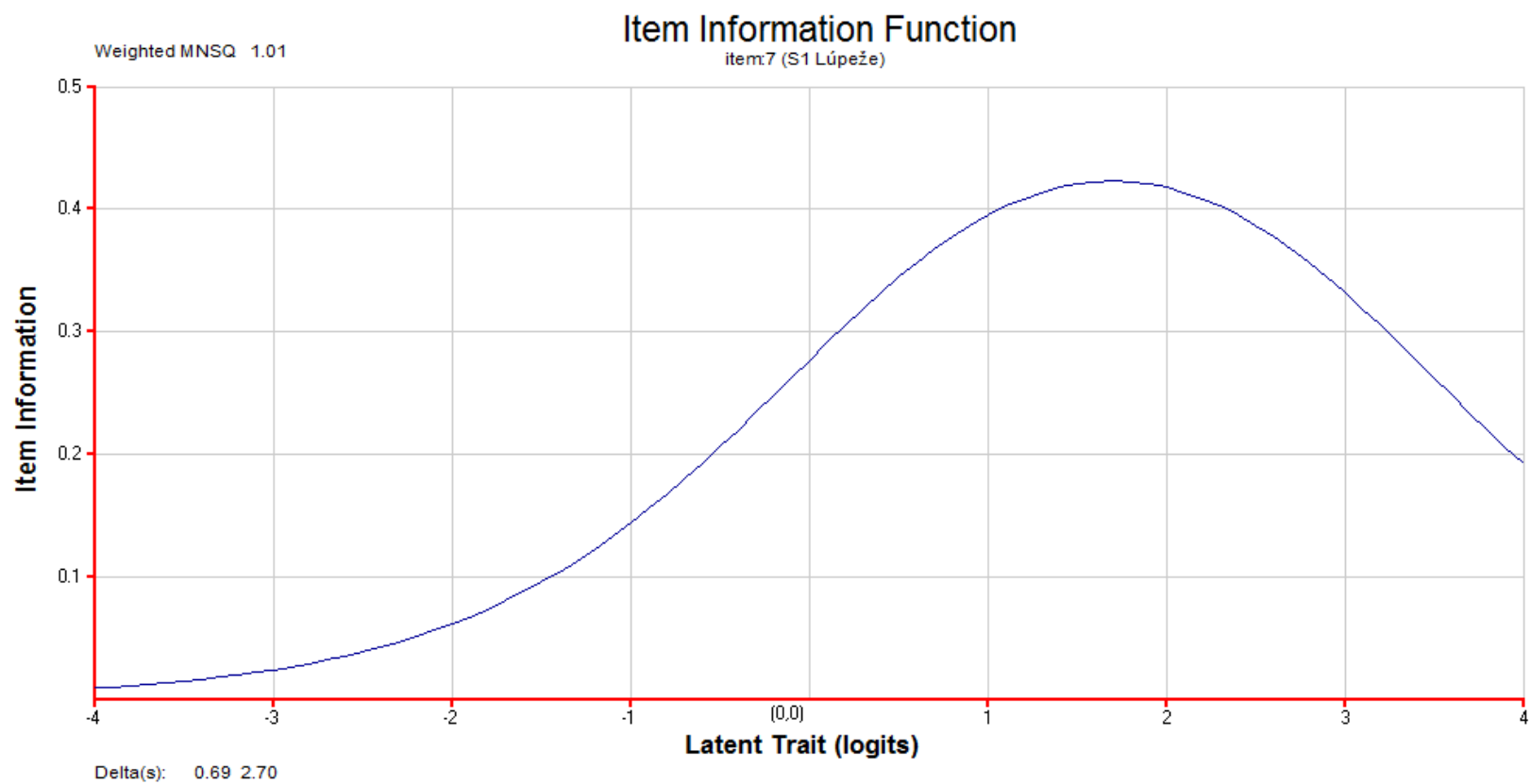


Lúpeže

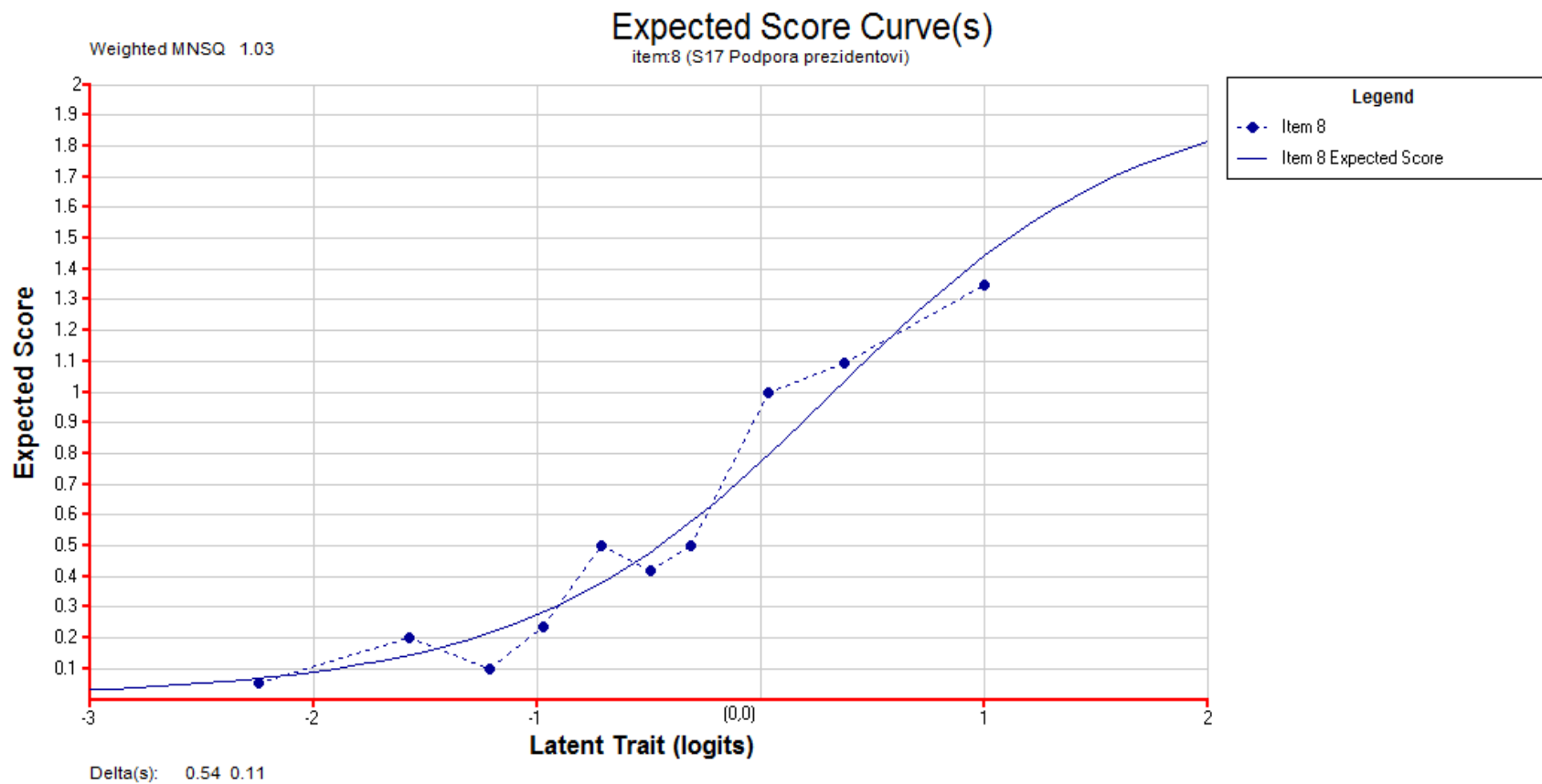
Characteristic Curve(s) By Score



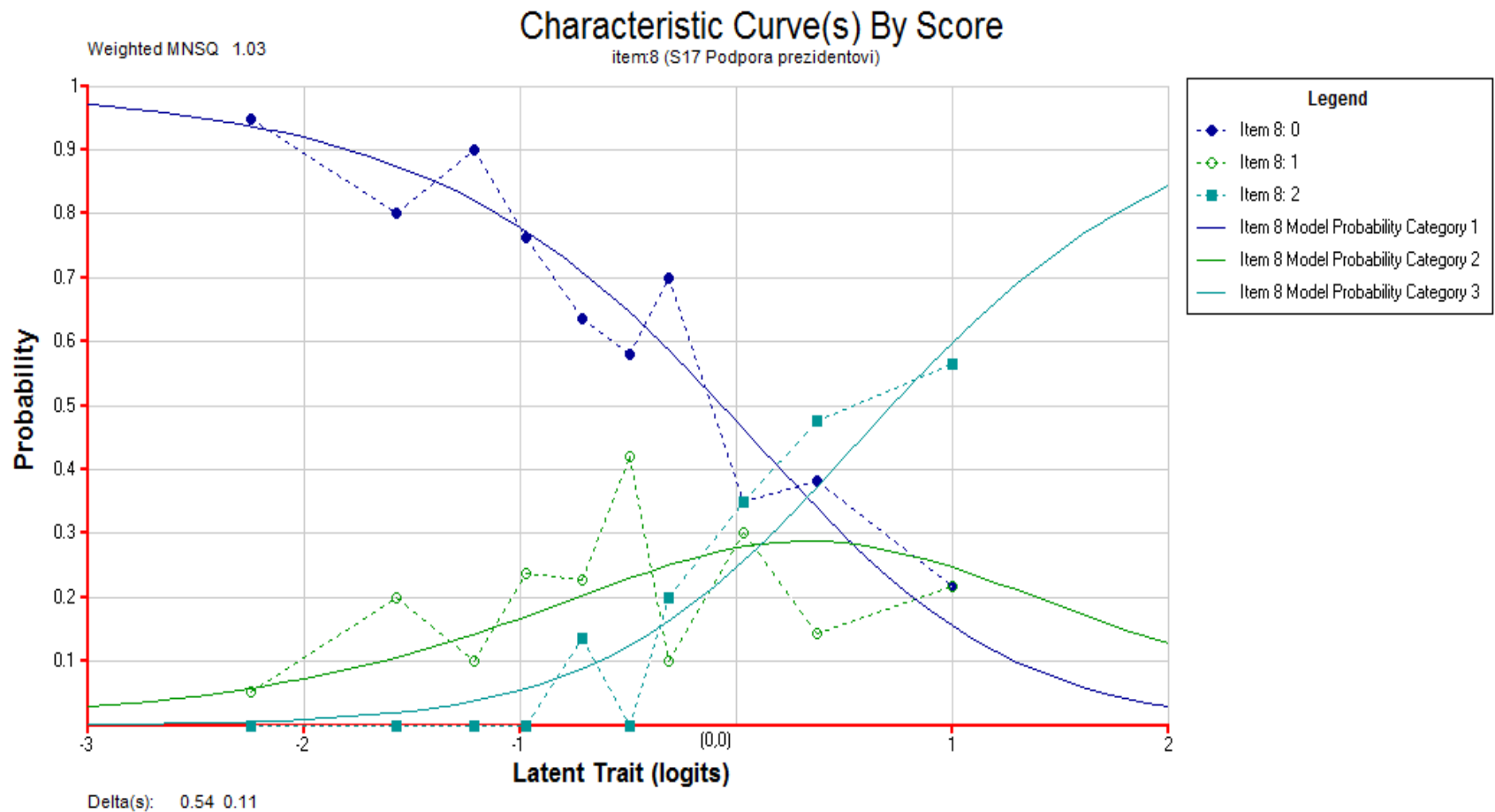
Lúpeže



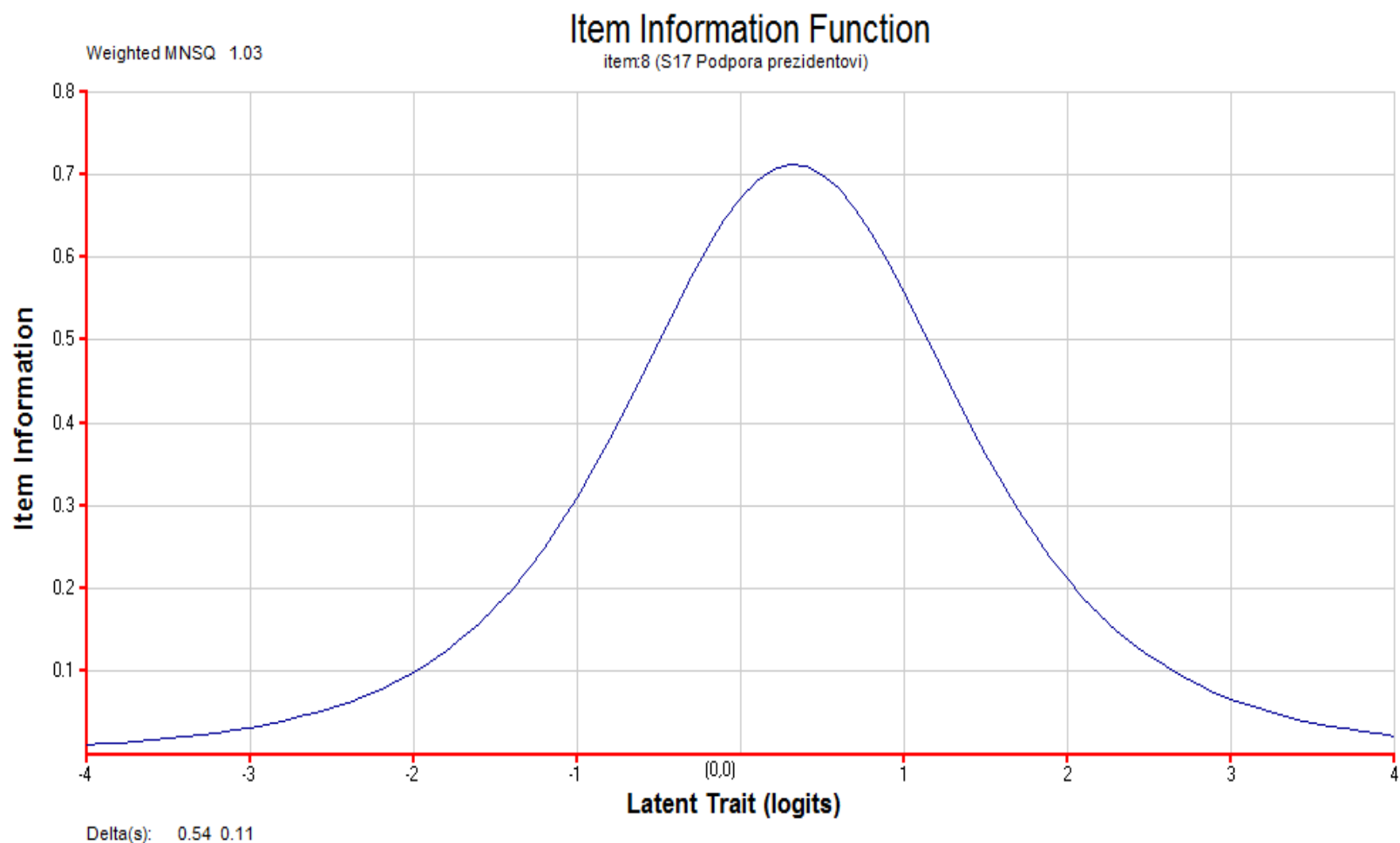
Podpora prezidentovi



Podpora prezidentovi



Podpora prezidentovi



Záver

- Základnou myšlienkou IRT je objaviť modely, ktoré vyjadrujú pomocou parametrov vlastnosti, rys (trait) osoby a vlastnosti položiek súčasne a oddelene.
- Parametre položiek môžu byť odhadnuté zo vzorky osôb, ktorá má náhodilé rozdelenie – skupinová invariancia.
- Pre dané kalibrované položky môže byť odhadnutá latentná abilita pomocou ľubovoľného testu: ľahkého, stredne ťažkého, ťažkého.
- IRT porovnáva osoby s položkami a nie s inými osobami (CTT).

Literatúra

- Baker F. B. – The basics of item response theory, Wisconsin, 2001
- Bond T. G., Fox Ch. M. – Applying the Rasch model: Fundamental measurement in the human sciences, London, 2001

Ďakujem za pozornosť