

## Тестово оценяване на базата на шаблона на тестовия отговор

Д. Атанасов<sup>1</sup>, В. Стоименова<sup>2</sup>, Д. Димитров<sup>3</sup>

1. Нов български университет
2. СУ "Св. Климент Охридски"
3. NCA, KSA.

1. Въведение
2. Проблематика
3. Реализация
4. Ресурси

# Въведение

# Въведение

- ▶ Оценяване на големи популации

# Въведение

- ▶ Оценяване на големи популации
- ▶ Оценяване чрез относително големи тестове

# Въведение

- ▶ Оценяване на големи популации
- ▶ Оценяване чрез относително големи тестове
- ▶ Формализиране на оценяването

- ▶ оценяване на способностите

- ▶ оценяване на способностите
- ▶ използване на шаблона на отговорите



- ▶ оценяване на способностите
- ▶ използване на шаблона на отговорите
- ▶ оценяване на параметрите на тестовите въпроси

# Най-често използвани подходи

# Най-често използвани подходи

- ▶ Тестов бал

# Най-често използвани подходи

- ▶ Тестов бал
- ▶ Rasch, IRT

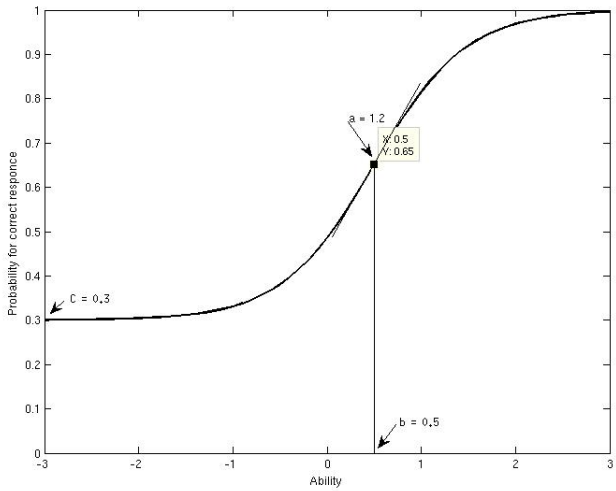
# Най-често използвани подходи

- ▶ Тестов бал
- ▶ Rasch, IRT
- ▶ други

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-aD(\theta-b)}},$$

където

- ▶  $\theta$  - ability levels,
- ▶  $P(\theta)$  - probability of correct response from a person with ability  $\theta$ ,
- ▶  $b$  - difficulty,
- ▶  $a$  - discrimination,
- ▶  $c$  - guessing.
- ▶  $D$  - constant,  $D = 1.72$ .



# Оценяване на параметрите

- ▶ **Неизместеност и състоятелност**



# Оценяване на параметрите

- ▶ Неизместеност и състоятелност
- ▶ ML, JML

# Оценяване на параметрите

- ▶ Неизместеност и състоятелност
- ▶ ML, JML
- ▶ EM

# Оценяване на параметрите

- ▶ Неизместеност и състоятелност
- ▶ ML, JML
- ▶ EM
- ▶ Проблеми

# Реализация

$$X_i = \{X_{ij} = 0, 1\}_{j=1}^J; i = 1, \dots, N$$

# Реализация

$$X_i = \{X_{ij} = 0, 1\}_{j=1}^J; i = 1, \dots, N$$

$\delta_j$  - характеристика на тестовата задача

# Реализация

$$X_i = \{X_{ij} = 0, 1\}_{j=1}^J; i = 1, \dots, N$$

$\delta_j$  - характеристика на тестовата задача

$$D_i = \frac{\sum_{j=1}^J \delta_j X_{ij}}{\sum_{j=1}^J \delta_j}$$

# Реализация

$$X_i = \{X_{ij} = 0, 1\}_{j=1}^J; i = 1, \dots, N$$

$\delta_j$  - характеристика на тестовата задача

$$D_i = \frac{\sum_{j=1}^J \delta_j X_{ij}}{\sum_{j=1}^J \delta_j}$$

$\delta_j$  могат да се интерпретират като "трудност"

# Реализация

$$X_i = \{X_{ij} = 0, 1\}_{j=1}^J; i = 1, \dots, N$$

$\delta_j$  - характеристика на тестовата задача

$$D_i = \frac{\sum_{j=1}^J \delta_j X_{ij}}{\sum_{j=1}^J \delta_j}$$

$\delta_j$  могат да се интерпретират като "трудност"

$$\delta_j = 1 - d_j; d_j = \frac{\sum_{i=1}^N X_{ij}}{N}$$



- ▶  $D_i$  и  $\delta_j$  се намират на една скала

- ▶  $D_i$  и  $\delta_j$  се намират на една скала
- ▶  $D_i, \delta_j \in [0, 1]$

- ▶  $D_i$  и  $\delta_j$  се намират на една скала
- ▶  $D_i, \delta_j \in [0, 1]$
- ▶ имат логична интерпретация

# Оценяване на $\delta_j$

- ▶ чрез  $d_j$  - зависи от извадката

# Оценяване на $\delta_j$

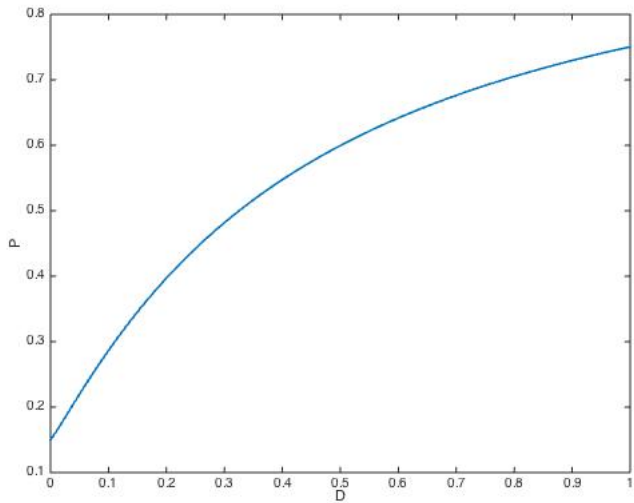
- ▶ чрез  $d_j$  - зависи от извадката
- ▶ чрез бутстрапинг

	$d_j$	btsrp_min	btsrp_mean	btsrp_mode	btsrp_median	btsrp_max	btsrp_SE
0.9	0.8956	0.8433	0.8956	0.8933	0.8966	0.9433	0.0162
0.8	0.7896	0.7033	0.7883	0.7933	0.79	0.8666	0.0230
0.7	0.69	0.62	0.6909	0.6866	0.69	0.7633	0.0255
0.6	0.6126	0.5166	0.6121	0.6133	0.613	0.6933	0.0265
0.5	0.503	0.4133	0.5025	0.5066	0.503	0.59	0.0267
0.4	0.3956	0.3133	0.3965	0.39	0.396	0.5	0.0267
0.3	0.2863	0.1833	0.2856	0.2833	0.286	0.3766	0.0249
0.2	0.2006	0.13	0.2009	0.21	0.2	0.2833	0.0221
0.1	0.1113	0.0533	0.1102	0.1133	0.11	0.1633	0.0173

## Параметризиране на $\delta_j$

$$P(X_{ij} = 1 \mid D_i, \alpha, \beta, \gamma) = 1 - \frac{1 - \gamma}{1 + \left(\frac{D_i}{\beta}\right)^\alpha}.$$

- ▶  $\beta$  - difficulty,
- ▶  $\alpha$  - discrimination,
- ▶  $\gamma$  - guessing.

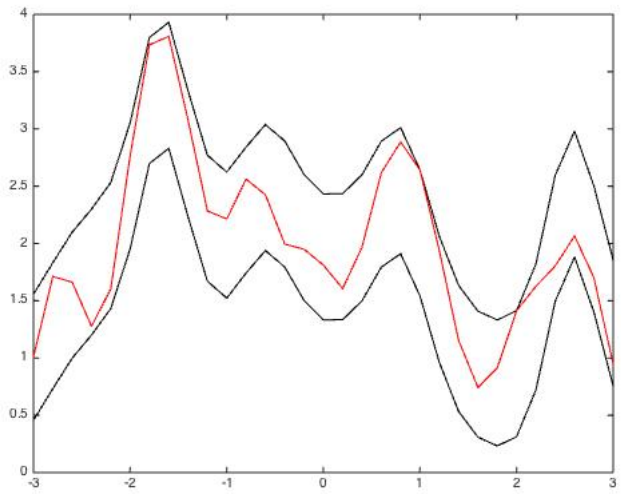




- ▶ IIF  $I(D) = \frac{(P')}{P(1-P)}$
- ▶ TIF

- ▶ IIF  $I(D) = \frac{(P')}{P(1-P)}$
- ▶ TIF
- ▶ Банка от тестови задачи

- ▶ IIF  $I(D) = \frac{(P')}{P(1-P)}$
- ▶ TIF
- ▶ Банка от тестови задачи
- ▶ Съставяне на "оптимални" тестове





Baker F. B. (1987). Methodology review: Item parameter estimation under one-, two-, and three-parameter logistic models. *Applied Psychological Measurement*, 11, 111-143.



Baker, Frank B.; Kim, Seock-Ho (2004). *Item Response Theory: Parameter Estimation Techniques* (2nd ed.).



Bock R. D., Aitkin, M. (1981). Marginal Maximum Likelihood estimation of item parameters. An application of an EM algorithm. *Psychometrika*, 46, 443 – 459.



Dimitrov, D. M. (2015a). *A Method of Scoring Tests With Binary Items and Its Application in Test Equating*. Riyadh, Saudi Arabia: National Center for Assessment in Higher Education (Research Note: RN 1-2015)



Dimitrov, D. M., & Atanasov, D. V. (2013). *TEQNCA: A computer program for IRT true score equating of multiple test forms*. Riyadh, KSA: National Center for Assessment in Higher Education.



Domingue, B. W., & Dimitrov, D. M. (2015). *A comparison of IRT theta estimates and Delta Scores from the perspective of additive conjoint measurement*. National Center for Assessment in Higher Education (Research Report: RN 4-2015).