Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

# Comparações entre a Teoria de Resposta ao Item e a Teoria Clássica de Medidas para Banco de Itens Simulados

**Anna Cristina Rezende Braga** [1]
**Augusto Sousa da Silva Filho** [2]
**Livingstone Augusto Eller** [3]
**Lorrayne Marciel Rocha** [4]
**Petrina Teixeira Santos** [5]

## 1 Introdução

A Teoria de Resposta ao Item surgiu da necessidade de superar as limitações da apresentação de resultados somente através de percentuais de acertos, e ainda da dificuldade de comparar resultados de diferentes testes. Na teoria clássica dos testes, os resultados dependem do particular conjunto de questões que compõem a prova e dos indivíduos que a fizeram, ou seja, as análises e interpretações estão sempre associadas à prova e ao grupo de indivíduos, logo a comparação entre grupos de indivíduos somente é possível quando eles são submetidos às mesmas provas, logo fica difícil fazer comparações quando diferentes indivíduos fazem provas diferentes. Para solucionar este problema aplica-se a TRI, que é um conjunto de modelos matemáticos que considera o item como unidade básica e procura representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e do(s) traço(s) latente(s) do indivíduo. A ideia básica da TRI consiste no emprego de modelos, geralmente paramétricos, nos quais os parâmetros representem características importantes dos itens, sendo, assim, interpretáveis. Um dos grandes obstáculos da TRI são os cálculos matemáticos utilizados para a determinação dos parâmetros do modelo, além da grande dificuldade de obtenção de banco de itens que sejam considerados grandes para que a TRI possa ser aplicada, desta forma, apresentaremos neste pôster bancos de itens simulados, as técnicas computacionais para a obtenção dos parâmetros dos modelos, além de comparação com os métodos clássicos de correção dos itens.

## 2 Material e Métodos

São apresentadas as técnicas comumente usadas para analisar itens provenientes de testes ou banco de itens.

[1] IBS Business School de Minas Gerais - FGV. e-mail: *annacristinarb@gmail.com*
[2] IBS Business School de Minas Gerais - FGV. e-mail: *augustofilho@yahoo.com.br*
[3] IBS Business School de Minas Gerais - FGV. e-mail: *livins.eller@gmail.com*
[4] IBS Business School de Minas Gerais - FGV. e-mail: *lorrayne.maciel@gmail.com*
[5] IBS Business School de Minas Gerais - FGV. e-mail: *petrina.marzzagao@gmail.com*

## 2.1 Teoria Clássica de Medidas

Os primeiros trabalhos de Sperarman [3], inseriam uma modelagem estatística para estimação das habilidades através do total de pontos em testes (escores individuais), com forma axiomática final devido aos trabalhos de Novick [2], ficando conhecido como Análise Clássica de Itens. Desta forma, nos modelos clássicos são introduzidos dois construtos: um erro de medida e um escore verdadeiro. O erro de medida pode ser definido como a diferença entre o escore verdadeiro e o observado. O escore verdadeiro pode ser definido como um valor esperado dos seus escores em vários testes. Logo, o modelo clássico assume que os erros de medida são aleatórios com média zero e não correlacionados entre si e com os escores verdadeiros. Logo, os escores observados e os erros de medida são linearmente relacionados.

$$X = T + E \tag{1}$$

Onde:

**i)** X é o escore observado;

**ii)** T é o escore verdadeiro (a habilidade);

**iii)** E é um erro de medida.

Essa metodologia faz uso da estatística descritiva, coeficiente de proporções e correlação com o objetivo de medir a qualidade dos itens, e praticamente nenhuma estatística inferencial, no entanto, fornecem resultados úteis como a fórmula de Kuder-Richardson e a fórmula de Sperman-Brow, ambas utilizadas para calcular a fidedignidade de um teste, ou seja, busca-se verificar a estabilidade dos resultados de um teste.

## 2.2 Teoria de Resposta ao Item

Nos modelos considerados neste texto, um item será caracterizado por parâmetro referente às seguintes características: discriminação, dificuldade e probabilidade de acerto casual. Outro parâmetro de suma importância nos modelos da TRI é a proficiência do indivíduo, também conhecida como habilidade cognitiva ou somente habilidade. Os modelos da Teoria de Resposta ao Item, ao levar em consideração as características dos itens na produção das habilidades, permitem que a proficiência seja medida com maior precisão e que a escala produzida seja interpretada. Outra vantagem da TRI é possibilitar o cálculo da proficiência de um aluno mesmo que este não tenha respondido a todos os itens, o que é feito sem a necessidade de tratamento adicional dos dados. A distinção principal entre modelos de teoria de resposta ao item refere-se à suposição sobre o relacionamento entre a seleção de opções de uma resposta e o nível do traço latente. Desta forma, é possível perceber a existência de dois tipos de processos de resposta: o acumulativo e o de desdobramento. Com isso, foram desenvolvidos modelos de natureza acumulativa e de natureza de desdobramento para dados dicotômicos ou binários e politômicos, nominal ou graduados, modelos não paramétricos e modelos paramétricos, modelos unidimensionais e multidimensionais.

# 3 Resultados e Discussão

## 3.1 Modelo Logístico de 3 Parâmetros - ML3

A teoria de resposta ao item é uma família de modelos matemáticos e a principal diferença entre entre estes modelos refere-se à suposição quanto ao relacionamento entre a seleção de opções de uma resposta e o nível do traço latente. Dos modelos para itens com resposta dicotômicas unidimensionais, destaca-se o modelo logístico de três parâmetros (ML3), dado por:

$$P\left(U_{ij}=1|\theta_j\right)=c_i+(1-c_i)\frac{1}{1+e^{-Da_i\left(\theta_j-b_i\right)}} \tag{2}$$

Onde, $i=1,2,\ldots,I$ representa os I itens propostos para avaliar o traço latente considerado e $j=1,2,\ldots,n$ representa os n elementos que compõem a amostra, que podem ser indivíduos, empresas, etc.

A descrição da equação 2 é apresentada a seguir:

**i)** $P\left(U_{ij}=1|\theta_j\right)$ representa a probabilidade do j-ésimo indivíduo escolhido ao acaso com grau de proficiência $\theta_j$ responder corretamente ao i-ésimo item;

**ii)** $U_{ij}$ representa a variável dicotômica que assume o valor 1 quando o j-ésimo indivíduo responde corretamente ao item $i$, e assume zero quando o j-ésimo respondente não responde acertadamente ao item $i$;

**iii)** $\theta_j$ representa o grau de proficiência (traço latente) do j-ésimo respondente;

**iv)** $a_i$ é o parâmetro correspondente ao índice de discriminação;

**v)** $b$ é o parâmetro correspondente ao grau de dificuldade do item;

**vi)** $c_i$ parâmetro que representa a probabilidade de acerto ao acaso;

**vii)** $D$ representa o fator de escala usado para aproximar a função logística da ogiva Gaussiana com valor 1,7.

## 3.2 Resultados Preliminares

Utilizando os packages do R, (ltm, irtoys), pode-se construir banco de itens com tamanhos variáveis e realizar comparações entre as metodologias clássicas de correção do item e a TRI. Desta forma, temos:

```
install.packages("ltm")  #para instalar
library(ltm)             #para chamar o que foi instalado
```

Este procedimento é para fazer a instalação do pacote e para ativar o ltm. Agora, com o pacote instalado, iremos trabalhar com uma base de dados que faz parte do pacote ltm. Abaixo é apresentado apenas os 6 primeiros dados simulados.

```
  Item 1 Item 2 Item 3 Item 4 Item 5
1      0      0      0      0      0
2      0      0      0      0      0
3      0      0      0      0      0
4      0      0      0      0      1
5      0      0      0      0      1
6      0      0      0      0      1
```

Esta base de dados contém as respostas de 1000 pessoas que responderam a 5 perguntas de um teste simulado por computador. Este é um exemplo típico de uma aplicação da TRI em um exame educacional. Primeiramente realizaremos uma análise clássica e em seguida realizaremos uma análise seguindo a Teoria de Resposta ao Item para o conjunto de dados simulados. A primeira parte de uma análise clássica dos dados é a realização de uma análise descritiva do conjunto de dados. A análise descritiva faz com que os resultados obtidos sejam facilmente aceitos pelos pesquisadores, pois a análise descritiva irá validar de uma forma mais simples os resultados obtidos pela inferência estatística. Desta forma, para realizar uma estatística descritiva dentro da TRI, deveremos escrever no R, o seguinte comando:

```
descript(simulado)

Descriptive statistics for the 'SIMULADO' data-set

Sample:
 5 items and 1000 sample units; 0 missing values

Proportions for each level of response:
           0     1  logit
Item 1 0.076 0.924 2.4980
Item 2 0.291 0.709 0.8905
Item 3 0.447 0.553 0.2128
Item 4 0.237 0.763 1.1692
Item 5 0.130 0.870 1.9010
```

Acima, temos a proporção de erros (0) e acertos (1) para os cinco itens. Logo, para o **item 1**, é possível afirmar que 92,4% acertaram este item e 7,6% erraram este item. O **item 3**, apresenta 55,3% de acertos e 44,7% de erros. Desta análise inicial é possível afirmar que o **item 1** apresentou o maior percentual de acertos e o **item 3** apresentou o maior percentual de erros, ou seja, é possível concluir que o **item 1** é o mais fácil e o **item 3** é o mais difícil.

A próxima saída de dados apresenta a **frequência para os scores totais.**

```
Frequencies of total scores:
     0  1  2   3   4   5
Freq 3 20 85 237 357 298
```

Desta tabela de frequência é possível destacar que 3 alunos não conseguiram acertar nenhum item, 20 alunos acertaram apenas 1 item, 85 alunos acertaram 2 itens, 237 alunos acertaram 3 itens, 357 alunos acertaram 4 itens e 298 alunos acertaram aos 5 itens da prova.

O coeficiente de correlação bisserial $(\rho_b)$ é uma medida de associação entre uma variável dicotomizada e uma variável contínua, e é definido por:

$$\rho_b = \rho_{pb}\frac{\sqrt{p\,(1-p)}}{h(p)} \tag{3}$$

A interpretação da correlação bisserial afirma que quanto maior for o coeficiente da correlação, maior é a discriminação do item, ou seja, o item será considerado mais difícil.

```
Point Biserial correlation with Total Score:
       Included Excluded
Item 1   0.3618   0.1128
Item 2   0.5665   0.1531
Item 3   0.6181   0.1727
Item 4   0.5342   0.1444
Item 5   0.4351   0.1215
```

Portanto, o **item 3** é considerado o item mais difícil, pois apresenta o maior coeficiente de correlação ponto bisserial (0.6181) Included, ou seja, incluindo este item, da mesma forma, que o **item 1**, apresenta o menor valor para o coeficiente de correlação ponto bisserial (0.3618), ou seja, pode ser considerado o item mais fácil. Aqui é importante destacar que o item que apresenta o maior coeficiente de correlação bisserial será o item com maior valor de **discriminação**. O coeficiente alfa de Cronbach é utilizado para medir a consistência interna do instrumento de medida, e é definido por:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum s_i^2}{s_T^2}\right) \tag{4}$$

Onde:

**a)** $n$ é o número de itens;

**b)** $\sum s_i^2$ é a soma das variáveis dos $n$ itens;

**c)** $s_T^2$ é a variância global dos escores dos testes;

**Este coeficiente varia de 0 a 1. Quanto mais próximo de 0 menor será a consistência e quanto mais próximo de 1 maior a consistência do teste.**

```
Cronbach's alpha:
                  value
All Items        0.2950
Excluding Item 1 0.2754
Excluding Item 2 0.2376
Excluding Item 3 0.2168
Excluding Item 4 0.2459
Excluding Item 5 0.2663
```

O valor de Alfa de Cronbach para todos os itens foi igual a 0,2950. Este valor indica ao mesmo tempo que o questionário pode indicar falta de confiabilidade como o fato de os itens (perguntas) do questionário não estarem medindo o mesmo construto ou mesma dimensão (unidimensional).

```
Pairwise Associations:
   Item i Item j p.value
1       1      5   0.565
2       1      4   0.208
3       3      5   0.113
4       2      4   0.059
5       1      2   0.028
6       2      5   0.009
7       1      3   0.003
8       4      5   0.002
9       3      4   7e-04
10      2      3   4e-04
```

De acordo com o gráfico da Figura 1, classifica a proporção de acertos em relação ao escore total de acertos. Logo é possível afirmar que o item 1 é o mais fácil e o item 3 é o mais difícil. Para o mesmo conjunto de dados será feita uma análise nos moldes da Teoria de Resposta ao Item e os resultados serão apresentados em seguida.

Agora, iremos estimar um modelo logístico de dois parâmetros para a base de dados SIMULADO, usando o pacote "irtoys".

Para isso deveremos proceder da seguinte forma:

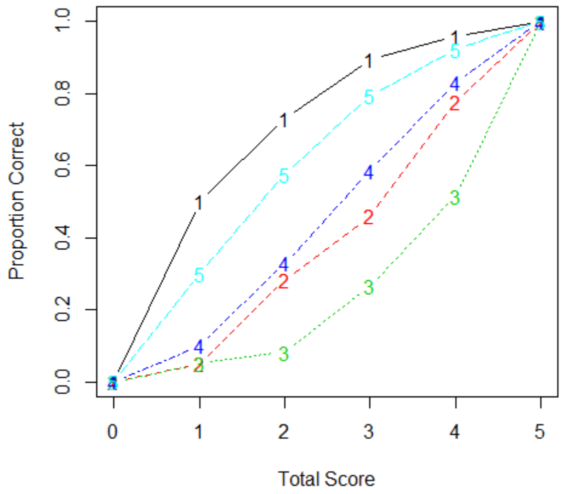


Figura 1: Gráficos para todos os itens

1. Realizar a calibração dos itens. Utilizaremos para isso a função "**est**". Essa função apresenta mais opções do que a apresentada no pacote **ltm**. Para estimar os parâmetros dos itens de um modelo pela TRI o pacote **est** usa fragmentos de programas importantes da TRI, como por exemplo o BILOG-MG, ICL ou LTM. A sua sintaxe é apresentada a seguir:

```
livins.par=est(livins.itens,model="2PL",engine="ltm")
livins.par

$est
             [,1]       [,2] [,3]
Item 1 0.8253717 -3.3597333    0
Item 2 0.7229498 -1.3696501    0
Item 3 0.8904752 -0.2798981    0
Item 4 0.6885500 -1.8659193    0
Item 5 0.6574511 -3.1235746    0
```

O objeto **livins.par** contém os resultados da função **est** aplicada nos itens sobre a prova simulada. No objeto foram armazenados os resultados da calibração de um modelo com 2 parâmetros (a e b). A opção (**nqp = 20**) indica o número de pontos de quadratura utilizados na estimação. Para (**model="2PL"**), obtivemos como resultado final os parâmetros estimados na escala logística. O objeto **livins.par** apresenta os valores dos parâmetros $a$ (discriminação) e $b$ (dificuldade) de cada item em $[,1]$ e em $[,2]$.

```
             [,1]       [,2] [,3]
```

```
Item 1 0.8253717 -3.3597333    0
Item 2 0.7229498 -1.3696501    0
Item 3 0.8904752 -0.2798981    0
Item 4 0.6885500 -1.8659193    0
Item 5 0.6574511 -3.1235746    0
```

Também são apresentado os erros padrões das estimativas:

```
$se
          [,1]       [,2] [,3]
[1,] 0.2580641 0.86694584    0
[2,] 0.1867055 0.30733661    0
[3,] 0.2326171 0.09966721    0
[4,] 0.1851659 0.43412010    0
[5,] 0.2100050 0.86998187    0
```

2. Agora deveremos utilizar os objetos **livins.itens** e **livins.par** para determinar as estimativas das habilidades $\hat{\theta}$ de cada um dos indivíduos. Para isso deveremos trabalhar com a função **eap** disponibilizada pelo pacote **irtoys**.

Desta forma, para determinarmos as habilidades de todos os indivíduos $\hat{\theta}$, deveremos usar:

```
livins.sco=eap(livins.itens,ip=livins.par$est,qu=normal.qu())
```

Como este banco de dados foi aplicado para 1000 alunos, será apresentado somente os seis primeiros valores:

```
head(livins.sco}
```

Desta forma, temos:

```
           est       sem n
[1,] -1.893406 0.7957686 5
[2,] -1.893406 0.7957686 5
[3,] -1.893406 0.7957686 5
[4,] -1.474010 0.8011759 5
[5,] -1.474010 0.8011759 5
[6,] -1.474010 0.8011759 5
```

No objeto **livins.sco** a primeira coluna representa os valores das estimativas das habilidades de todos os alunos, a segunda coluna representa o erro padrão das estimativas e a terceira coluna representa o número de respostas de cada aluno.

Com as informações do objeto **livins.sco** é possivel construir o gráfico da função da resposta ao item. O comando usa a função **irf**.

```
plot(irf(livins.par$est),label=T)
```

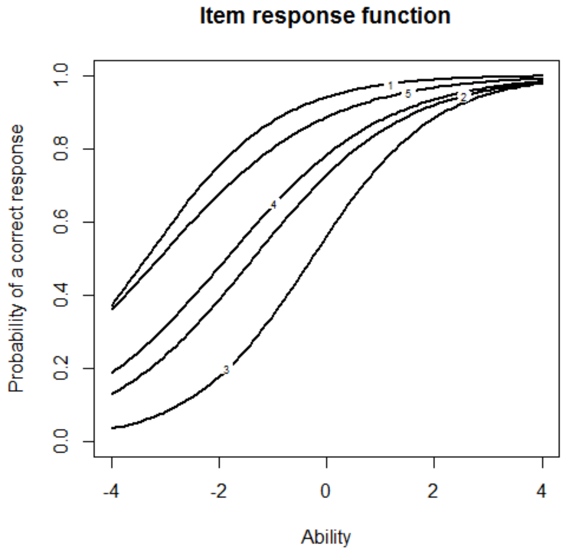


Figura 2: IRF - Função de Resposta ao Item

# 4 Conclusão

O objetivo deste pôster é apresentar a utilização da Teoria Clássica de Correções de Itens aplicadas em itens simulados comparados com a Teoria de Resposta ao Item, com modelos logísticos de 2 e 3 parâmetros. Com isso, iremos descrever que a TRI é uma ferramenta que não vem substituir a Teoria Clássica de Correção dos Itens, mais vem dar apoio para que o pesquisador tenha mais uma ferramenta adequada para uma melhor avaliação de seus estudos.

# Referências

[1] ASSUNÇÃO, C.N.B. **Estimação dos parâmetros de modelos de teoria de resposta ao item e aplicações**. 1999. 124 p. Dissertação (Mestrado em Estatística), Universidade Federal de Minas Gerais, 1999.

[2] NOVICK, M.R. **The axioms and principal results of classical test theory**, v. 3, n. 1, p. 1-18, 1966.

[3] SPEARMAN, C. **The proof and measurement of association between two things. By C. Spearman, 1904.**, Am J Psychol, v. 100, n. 3-4, p. 441-471, 1987.