

# Teoria de Resposta ao Item

Ana Clara Segal Vidal Pessanha,<sup>1</sup> Gabriela Alcaide,<sup>2</sup> Giovanna Almeida Albuquerque,<sup>3</sup>  
Júlia Du Bois Araújo Silva,<sup>4</sup> Rodrigo Gonçalves Cardoso,<sup>5</sup>

Universidade de São Paulo<sup>1,2,3,4,5</sup>

anaclarasvp@usp.br,<sup>1</sup> gabriela.alcaide@usp.br,<sup>2</sup> gi.albuqq@usp.br,<sup>3</sup> juliaduboisas@usp.br,<sup>4</sup> digo.gcardoso@usp.br<sup>5</sup>

## Resumo

O presente artigo busca descrever a Teoria de Resposta ao Item, especialmente sob perspectiva estatística. Descreve-se o cenário considerado para o experimento, os fundamentos envolvidos e aprofunda-se no processo de inferência e teste de hipótese.

## História da Teoria de Resposta ao Item e que problema ela tenta resolver

Segundo Andrade, Tavares e Valle (2000):

“A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente.”

Ao contrário da TCT (Teoria Clássica dos Testes), onde o resultado de um participante de um teste é dado pela razão entre a quantidade de questões acertadas e a quantidade total de questões, a TRI (Teoria de Resposta ao Item) tem como elementos centrais os itens e não a prova como um todo. (ANDRADE; TAVARES; VALLE, 2000).

A Teoria foi utilizada pela primeira vez no Brasil em 1995 na análise dos dados do Sistema Nacional de Ensino Básico (SAEB), um teste que mede o desempenho dos alunos do ensino fundamental e médio. Em 2009, foi introduzida no Exame Nacional do Ensino Médio (ENEM), a fim de comparar as notas do exame daquele ano com as notas dos exames dos anos seguintes (Ministério da Educação, 2018).

## Descrição do Experimento

O experimento tem como objetivo, baseando-se nas respostas de um indivíduo  $j$ , descobrir sua habilidade  $\theta_j$ . Para isso, será usada a Teoria de Resposta ao Item. Dessa forma, é considerada uma amostra  $J$  de pessoas, que respondem um exame de  $N$  questões. Os indivíduos serão representados pelo índice  $j$ , que pode assumir os valores  $\{1, 2, \dots, J\}$ , e as questões pelo índice  $i$ , que pode assumir os valores

$\{1, 2, \dots, N\}$ .  $\theta_j$  é o que se chama de traço latente do indivíduo no experimento e representa a habilidade individual do indivíduo  $j$ . O traço latente impacta na probabilidade de um certo indivíduo acertar ou errar determinada questão, de acordo com sua probabilidade - se o indivíduo  $j$  possui  $\theta_j$  acima da média, a chance de acertar uma questão difícil é alta, enquanto se ele possui  $\theta_j$  abaixo da média a chance fica baixa.

Os parâmetros de cada questão  $i$  são representados por  $a_i$ ,  $b_i$  e  $c_i$  e indicam respectivamente a discriminação da questão em relação à CCI, a dificuldade do item e o acerto casual.

Outros fatores relevantes para o experimento são a independência local dos itens e a unidimensionalidade, expandidos a seguir:

## Unidimensionalidade

Segundo a teoria de espaço latente, explicada por Andrade e Valle (1998), por trás de cada indivíduo existe um conjunto de traços que interferem com seu desempenho comportamental, criando-se um espaço  $\theta$ , representado por  $q$ , de  $n$  dimensões  $q_1, q_2, \dots, q_n$ , que influenciam o desempenho  $d$  de um indivíduo em conjunto com a sorte. Dessa forma, este comportamento pode ser representado como um vetor de pesos em cada  $\theta$  e é possível inferir que  $d = f(q_1, q_2, \dots, q_n)$ .

Apesar disso, em geral, os modelos da TRI apenas reconhecem uma aptidão como responsável pela realização de um conjunto de tarefas e a considera dominante sobre todas as demais, a qual chamamos de  $\theta_j$ .

## Independência local

Considerando que as aptidões que impactam o resultado do teste estejam constantes, com a exceção da dominante - provinda do conceito de unidimensionalidade -, as respostas de um indivíduo à quaisquer dois itens são independentes, determinando a independência local.

Essa independência resulta no fato que cada item encontra-se exclusivamente em função do  $\theta$  dominante  $\theta_j$ . Portanto, a sequência de respostas de cada item refere-se ao produto das probabilidades de cada item e, caso haja alguma relação entre o resultado dos itens, ela é causada por alguma aptidão não dominante.

## Fundamentos da TRI

De acordo com Andrade, Tavares e Valle (2000), os modelos de resposta ao item, para itens dicotômicos, são classificados em três grupos, considerando a quantidade de parâmetros envolvidos, que podem ser 1, 2 ou 3. Para o modelo de um parâmetro, considera-se somente a dificuldade do item. Para o de dois, leva-se também em conta a discriminação da questão. Por fim, para o de três, adiciona-se a probabilidade de acerto por chute.

Neste artigo, será dada maior ênfase ao modelo de três parâmetros, uma vez que é o mais utilizado. Eles são representados, na literatura, pelas letras  $a_i$ ,  $b_i$  e  $c_i$ , onde ‘ $i$ ’ refere-se ao  $i$ -ésimo item.

O parâmetro ‘ $a_i$ ’ é a capacidade de discriminação (ou a inclinação) do item  $i$ . O parâmetro ‘ $b_i$ ’ trata-se da dificuldade do item  $i$ , através de lógica que será descrita adiante. E o parâmetro ‘ $c_i$ ’ indica a probabilidade de acerto casual, ou seja, a chance de respondentes com baixa habilidade relativa ao item acertarem o mesmo.

Para a descrição que será apresentada adiante, é necessário definir  $\theta_j$ , que se trata da habilidade do  $j$ -ésimo indivíduo. Com isso, pode-se definir  $P(X_{ij} = 1|\theta_j)$  como a chance do indivíduo com habilidade  $\theta_j$  acertar o item  $i$ .

Agora, é possível estudar mais profundamente os parâmetros, através de sua associação com  $P(X_{ij} = 1|\theta_j)$ , a qual é dada pela Curva Característica do Item (CCI).

Quanto às unidades de medida dos parâmetros,  $a_i$  é proporcional à inclinação da CCI no ponto  $b_i$ . O parâmetro  $b_i$  utiliza a mesma unidade da habilidade  $\theta_j$ . O parâmetro  $c_i$ , por sua vez, não possui unidade, uma vez que trata-se de uma probabilidade.

O parâmetro  $a_i$  advém da derivada da tangente da CCI no ponto de inflexão (momento em que a curvatura tem seu sinal invertido). Uma vez que a curva é crescente (conforme a habilidade do respondente aumenta, a probabilidade de responder corretamente ao item também aumenta), a derivada da tangente no ponto citado será sempre positiva, ou seja, o parâmetro  $a_i$  nunca será negativo. Quando a curva é pouco íngreme, haverá valores baixos de  $a_i$ , indicando que o item tem pequena capacidade de distinguir respondentes com habilidades consideravelmente diferentes. Por outro lado, para curvas íngremes,  $a_i$  será alto, o que demonstra que o item difere claramente os respondentes conforme suas habilidades.

O parâmetro  $c_i$  é a chance de um respondente com baixa habilidade  $\theta_j$  acertar o item  $i$ . Ou seja, a probabilidade de um chute resultar em acerto.

O parâmetro  $b_i$  retrata a dificuldade do item  $i$  ao representar o nível de habilidade necessário para  $P(X_{ij} = 1|\theta_j) = \frac{1+c_i}{2}$ .

## Distribuição de Probabilidade do Experimento

A distribuição de probabilidade na TRI descreve a relação entre o traço latente de um dado respondente e a probabilidade deste acertar um item específico no teste. Para tal mensuração, utiliza-se um modelo logístico unidimensional de três parâmetros para itens dicotômicos.

O modelo logístico é uma função matemática comumente usada para modelar relações entre variáveis. A unidimensionalidade diz respeito a como as variáveis trabalhadas são tratadas/interpretadas – o modelo considera apenas uma dimensão de cada variável, ou seja, cada variável será representada por um único número ao longo de uma escala contínua. Por fim, o modelo é dicotômico, pois na TRI, os itens são considerados como variáveis que podem assumir apenas dois valores possíveis, correto (1) ou incorreto (0).

Dessa forma, de acordo com SILVA (2014), a  $P(X_{ij} = 1|\theta_j)$  é dada por:

$$P(X_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta_j - b_i)}}, i = 1, 2, \dots, I; j = 1, 2, \dots, n$$

em que:

$I$  é o número de itens;

$n$  é o número de respondentes;

$\theta_j$  é o traço latente;

$a_i$  é o parâmetro de discriminação do item  $ik$ , representando o quanto o dado item consegue distinguir/diferenciar indivíduos com diferentes traços latentes;

$b_i$  é o parâmetro de dificuldade do item, no qual um item é considerado mais difícil à medida que um traço latente mais elevado daí-se por necessário a fim de acertá-lo;

$c_i$  é a probabilidade de indivíduos com baixos traços de latência acertarem o item  $i$ , resposta casual. O acerto casual corresponde à resposta correta oriunda de outros traços do indivíduo, que não aquele considerado no traço latente, ou ainda, àquela oriunda de uma resposta aleatória (“chute”).

Por se tratar de uma modelagem logística,  $D$  é uma constante que é definida como 1.

O gráfico de tal equação produz CCI, representada na Figura 1:

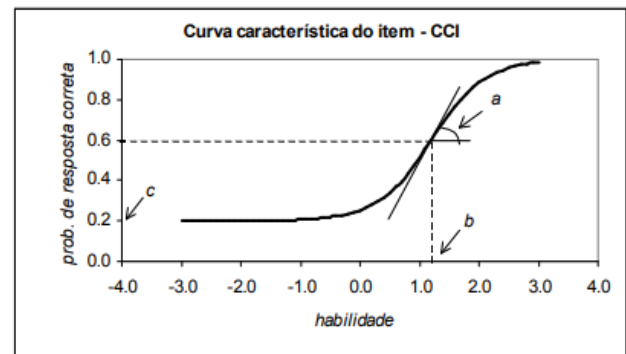


Figure 1: CGI - FONTE: Andrade (2000)

Desse modo, percebe-se que indivíduos com maior traço latente possuem maior probabilidade de acertar um dado item, e tal relação não é estabelecida de forma linear.

“De fato, pode-se perceber a partir do gráfico acima que a CCI tem forma de ‘S’ com inclinação e deslocamento na escala de habilidade definidos pelos parâmetros do item” (ANDRADE, 2000).

## Teoria de como fazer a inferência

No processo de estimação da TRI, existem dois grupos de parâmetros que possivelmente devem ser descobertos: os parâmetros  $a_i$ ,  $b_i$  e  $c_i$  dos itens e as habilidades  $\theta_j$  dos respondentes. As respostas de cada indivíduo aos itens são informações já conhecidas.

Dessa maneira, seria possível agir em três casos distintos: conhecemos os parâmetros do item, e desejamos inferir a habilidade do respondente; conhecemos a habilidade do respondente, e desejamos inferir os parâmetros do item; ou desejamos inferir ambas as informações.

No presente artigo, será abordado o primeiro caso: inferir a habilidade do  $j$ -ésimo respondente considerando que já conhecemos os parâmetros do item (parâmetros calibrados anteriormente), seguindo a abordagem proposta por Andrade, Tavares e Valle (2000). Assim, será utilizado o Método da Máxima Verossimilhança, com Algoritmo Newton-Raphson como processo iterativo.

Para isso, será necessário definir algumas notações e variáveis aleatórias. Sendo  $X_{ij}$  a variável aleatória que representa a resposta do  $j$ -ésimo indivíduo ao item  $i$ , uma observação genérica dela será  $x_{ij}$ . Além disso, teremos  $P_{ij} = Pr(X_{ij}|\theta_j)$  e  $Q_{ij} = 1 - P_{ij}$  (seu complemento). Por fim,  $X_j$  representa o vetor das respostas do  $j$ -ésimo indivíduo a todos os itens.

O valor desejado é  $Pr(\theta_j|X_j)$ . Dessa forma, é necessário calcular o Likelihood (ou Verossimilhança)  $L(X_j|\theta_j)$ , ou seja, a chance de o  $j$ -ésimo indivíduo responder aos itens da forma que respondeu, dada sua habilidade.

Uma vez que as respostas de indivíduos diferentes são independentes e que existe independência local, pode-se calcular a verossimilhança por:

$$L(\theta_j; X_j) = \prod_{j=1}^n \prod_{i=1}^I P(X_{ij} = x_{ij}|\theta_j)$$

Como  $x_{ij}$  pode assumir os valores 0 ou 1, tem-se:

$$\begin{aligned} P(X_{ij} = x_{ij}|\theta_j) &= P(X_{ij} = 1|\theta_j)^{x_{ij}} P(X_{ij} = 0|\theta_j)^{1-x_{ij}} \\ &= P_{ij}^{x_{ij}} \times Q_{ij}^{1-x_{ij}} \end{aligned}$$

E, com isso, reescrever a expressão para verossimilhança como:

$$L(\theta_j; X_j) = \prod_{j=1}^n \prod_{i=1}^I P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}}$$

Por ser uma expressão de cálculo complexo, será adotada a log-verossimilhança:

$$\log L(\theta_j; X_j) = \sum_{j=1}^n \sum_{i=1}^I x_{ij} \log P_{ij} + (1 - x_{ij}) \log Q_{ij}$$

Resta, agora, descobrir  $\theta_j$  que maximiza a expressão. Para isso, é necessário derivá-la e igualar a expressão a 0:

$$\frac{\partial \log L(\theta_j)}{\partial \theta_j} = D \sum_{i=1}^I a_i(1 - c_i)(x_{ij} - P_{ij})W_{ij}$$

$$D \sum_{i=1}^I a_i(1 - c_i)(x_{ij} - P_{ij})W_{ij} = 0$$

Onde:

$$W_{ij} = \frac{P_{ij}^* Q_{ij}^*}{P_{ij} Q_{ij}}$$

$$P_{ij}^* = 1 + e^{-Da_i(\theta_j - b_i)}^{-1}$$

$$Q_{ij}^* = 1 - P_{ij}^*$$

Essa equação não possui uma solução explícita para  $\theta_j$ . Por isso, será aplicado o algoritmo iterativo Newton-Raphson.

Seja  $\hat{\theta}_j^{(t)}$  a estimativa para  $\theta_j$  na  $t$ -ésima estimação e  $\hat{\theta}_j^{(t+1)}$  a estimativa para  $\theta$  na  $t+1$ -ésima estimação.

Então, pelo Algoritmo de Newton-Raphson, teremos:

$$\hat{\theta}_j^{(t+1)} = \hat{\theta}_j^{(t)} - [H(\hat{\theta}_j^{(t)})]^{-1} h(\hat{\theta}_j^{(t)})$$

Onde:

$$h(\theta_j) = \frac{\partial \log L(\theta)}{\partial \theta_j}$$

e

$$H(\theta_j) = \frac{\partial^2 \log L(\theta)}{\partial \theta_j^2}$$

## Aplicação prática da Inferência

Para efeitos ilustrativos esse artigo se debruçará na inferência das habilidades de um determinado respondente ao ENEM. Para tal feito, serão utilizados os dados fornecidos pelo Governo Federal, ENEM 2012, a respeito das respostas de um respondente aleatório  $j$ , além dos parâmetros ( $a_i$ ,  $b_i$ ,  $c_i$ ) dos 180 itens que compõem o exame.

A prova é dividida em 4 áreas de conhecimento: Ciências Humanas e suas Tecnologias (CH); Ciências da Natureza e suas Tecnologias (CN); Linguagens, Códigos e suas Tecnologias (LC); Matemática e suas Tecnologias (MT). Para este presente estudo analisaremos o desempenho do respondente  $j$  na área de Ciências da Natureza e suas Tecnologias, a fim de demonstrar como o processo de inferência funciona, e traremos um panorama geral da sua performance nas demais áreas do conhecimento.

Para realizar a análise dos dados e obter os resultados desejados, utilizamos um código na linguagem de programação Python, que utiliza as bibliotecas Numpy e Sympy.

### Inferência em Ciências da Natureza e suas Tecnologias

O primeiro passo é obter os parâmetros de cada questão e também as respostas do indivíduo  $j$  a cada uma delas. Para o  $j$ -ésimo respondente considerado em nosso artigo, as respostas obtidas na área de Ciências da Natureza e suas Tecnologias foram:

CN : 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0,  
1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0

Agora será necessário calcular, para cada questão, a probabilidade:

$$P(X_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta_j - b_i)}}$$

$$P_{46j} = 0.097112509 + \frac{1 - 0.097112509}{1 + e^{-1.1136143(\theta_j - 1.7022573)}}$$

...

$$P_{90j} = 0.115248448 + \frac{1 - 0.115248448}{1 + e^{-2.1216341(\theta_j - 2.2942219)}}$$

Para obtermos uma boa estimativa do valor de  $\theta$ , utilizaremos a equação conhecida como Log-Likelihood, e para isso substituiremos os parâmetros pelos  $P_{ij}$  pelos calculados, os  $x_{ij}$  pelos dados observados na área do conhecimento selecionada e utilizaremos a seguinte equação para encontrar o valor de  $Q$  em cada questão:  $Q = 1 - P_{ij}$ .

$$\log L(\theta) = \sum_{i=46}^{90} (x_i \log P_i + (1 - x_i) \log Q_i)$$

A partir dessa equação foi possível obter a figura 2. Com o gráfico gerado pelo código em Python, podemos concluir que o  $\theta$  que maximiza a equação está em torno de 2,5.

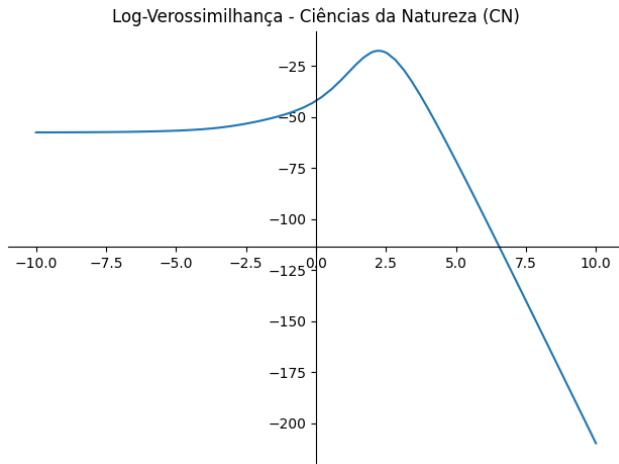


Figure 2: Log-Verossimilhança - Ciências da Natureza (CN)

Para encontrarmos de fato o  $\theta$  que maximiza a equação, seria necessário derivá-la e igualá-la a zero.

$$\theta : D \sum_{i=46}^{90} a_i(1 - c_i)(x_i - P_i)(W_i) = 0$$

Como antecipado na teoria sobre como a inferência poderia ser realizada, essa equação não possui uma solução explícita para  $\theta_j$ , utilizaremos então o método iterativo de Newton-Raphson com margem de erro de  $10^{-5}$ , e, como analisado no gráfico de Log-Likelihood, começaremos com uma estimativa de  $\hat{\theta} = 2, 5$ .

$$\hat{\theta}_j^{(t+1)} = \hat{\theta}_j^{(t)} - \frac{\frac{\partial \log L(\hat{\theta})}{\partial \hat{\theta}}}{\frac{\partial^2 \log L(\hat{\theta})}{\partial^2 \hat{\theta}}}$$

sendo

$$\frac{\partial \log L(\hat{\theta})}{\partial \hat{\theta}}$$

a margem de erro.

Depois de realizar apenas 4 iterações utilizando a fórmula acima, obtivemos o seguinte resultado:  $\hat{\theta} = 2.34835489759920$ . Vale ressaltar que a baixa quantidade de iterações se deve ao fato de que tivemos uma boa estimativa inicial, extraída da figura 02.

### Inferência em todas as áreas do conhecimento

Para realizar a inferência em todas as áreas do conhecimento, utilizaremos a mesma sequência de passos que foi realizada na área de Ciências da Natureza e suas tecnologias. Os dados desse mesmo indivíduo  $j$  para as outras áreas do conhecimento são:

CH : 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1,

1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1

LC : 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0,

1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1

MT : 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1,

1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1

Calculamos o  $P_{ij}$  para cada uma das questões dentro de cada área do conhecimento com a mesma fórmula e código utilizada nas questões de Ciências da Natureza e suas Tecnologias. Para obtermos uma boa estimativa de theta, utilizamos o Log-Likelihood para as questões específicas de cada área, obtendo o seguinte gráfico:

Sendo que a cor azul representa MT, a cor laranja representa CN, a cor verde representa CH, a cor vermelha representa LC. Com isso, conseguimos estimar valor aproximados para os  $\theta$  de cada área

$$\hat{\theta}_{MT} \simeq 2, 4$$

$$\hat{\theta}_{CH} \simeq 2, 4$$

$$\hat{\theta}_{LC} \simeq 2, 5$$

Utilizando as mesmas duas equações utilizadas para calcular  $\hat{\theta}_{CN}$  (Derivada do Log-Likelihood e Newton-Raphson), calculamos o valor dos outros theta, e obtivemos os seguintes resultados:

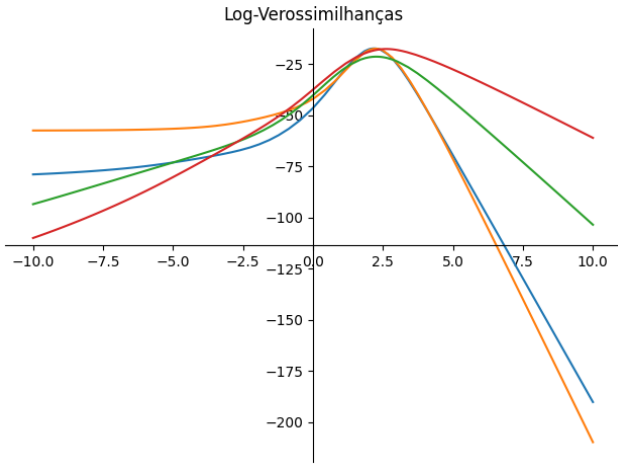


Figure 3: Log-Verossimilhanças - Todas as Áreas do Conhecimento

$$\hat{\theta}_{MT} \cong 2,19378942480157$$

$$\hat{\theta}_{CH} \cong 2.19566703363362$$

$$\hat{\theta}_{LC} \cong 2.56676828718506$$

O algoritmo iterativo de Newton-Raphson possuía uma margem de erro de  $10^{-5}$  e precisou ser utilizado 4 vezes para Matemática e suas Tecnologias, 4 vezes para Ciências Humanas e suas Tecnologias e 3 vezes para Linguagens, Códigos e suas Tecnologias.

Além disso, existe ainda a possibilidade de gerar um gráfico que relacione a quantidade de questões corretas com o valor do estimador  $\hat{\theta}$ .

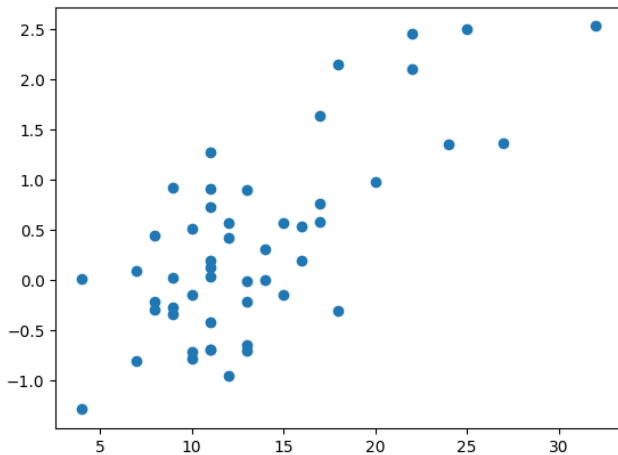


Figure 4: Relação entre acertos e estimadores theta em CN

Analisando o seguinte gráfico, que utiliza os acertos e os  $\hat{\theta}_j$  de CN, é possível concluir que as informações estão de fato relacionadas. Deste modo, quanto maior o valor do estimador  $\hat{\theta}_j$  maior a quantidade de acertos.

## Teste de Hipótese

Um teste de hipótese se propõe a averiguar a validade de uma afirmação feita sobre uma população com base em uma amostra desta. Na TRI os testes de hipótese são utilizados a fim de averiguar se as respostas reais (frequências observadas,  $f_o$ ) dos participantes são bem representadas pelo modelo ajustado (frequências esperadas,  $f_e$ ) (SOUSA,2018).

### Formulação das hipóteses:

- Hipótese nula ( $H_0$ ): Hipótese a ser testada. Representa a ausência de diferença entre o esperado e o real. No que diz respeito ao TRI, o  $H_0$  representa a capacidade do modelo representar fielmente/acuradamente os dados reais. A hipótese nula prevê a igualdade entre as frequências esperadas e observadas.
- Hipótese alternativa ( $H_1$ ): Hipótese que contradiz a hipótese nula, representando a presença de uma diferença. A hipótese alternativa prevê a desigualdade entre as frequências esperadas e observadas.

### Escolha o nível de significância:

O nível de significância ( $\alpha$ ) é uma medida crucial ao realizar um teste de hipótese. De acordo com SOARES (2018), o nível de significância representa a probabilidade de rejeitar incorretamente a hipótese nula quando ela é realmente verdadeira.

### Medida estatística de teste:

Existem diferentes estatísticas de teste para diferentes tipos de dados e para diferentes hipóteses a serem testadas. No que tange a Teoria de Resposta ao Item, é comumente empregado nos cálculos de teste de hipótese a medida estatística Qui quadrado ( $X^2$ ), por ser considerada a medida que melhor se adequa à pergunta de pesquisa e aos dados envolvidos na dada teoria. Tal medida é, principalmente, utilizada para comparar os dados observados com os valores esperados sob uma determinada distribuição teórica (SOARES, 2018). Desse modo, na TRI, o  $X^2$  testa a adequação do modelo, podendo ser calculado por:

$$\chi^2 = \sum \frac{(O_{ig} - E_i)^2}{E_i}$$

, ou ainda, especialmente para modelos dicotômicos, como:

$$\chi_i^2 = \sum_{g=1}^G \frac{N_g(O_{ig} - E_i)^2}{E_{ig}(1 - E_{ig})}$$

onde:

$O_{ig}$  é a proporção de respostas corretas para o  $i$ -ésimo item no intervalo  $g$  de proficiências;

$E_{ig}$  é a proporção esperada de respostas corretas para o  $i$ -ésimo item, baseada no valor da CCI na mediana das proficiências estimadas dado os scores obtidos pelos respondentes dentro do intervalo  $g$ ;

$N_g$  é o número de indivíduos com proficiência dentro do intervalo  $g$ ;

$G$  é o número de intervalos/subgrupos em que as proficiências foram divididas;

Desse modo, se dá por necessário a divisão dos respondentes em intervalos de proficiência. Esses intervalos podem ser definidos de várias maneiras, mas comumente são usados decís (10 intervalos), quartis (4 intervalos)

No que permeia a definição das proporções, tanto as obtidas quanto as esperadas, tais devem ser calculadas individualmente para cada item  $i$  e cada intervalo de proficiência  $g$ .

A proporção das frequências observadas é obtida por:

$$\frac{\text{Número de respostas corretas no intervalo } g \text{ para o item } i}{N_g}$$

Enquanto que a proporção das frequências esperadas é definida pelo valor de  $P(X_{ig} = 1 | \theta_{(med)g})$ .

### Grau de liberdade e valor crítico

O grau de liberdade (degree of freedom,  $df$ ) é uma medida que se refere à flexibilidade de escolha de diferentes valores após levar em consideração certas restrições ou condições, como o número total de observações.

“Esta estatística  $[X^2]$  tem distribuição [...] com  $G - m$  graus de liberdade, em que  $m$  é o número de parâmetros estimados para o item” (SOARES, 2018)

No que tange o cálculo de  $X^2$  na TRI especificamente,  $m$  costuma assumir o valor de  $m = 1$ .

A partir do grau de liberdade juntamente com o nível de significância, podemos, a partir da tabela Chi-Square Distribution (Turney, 2023) definir o valor crítico.

O valor crítico,  $X_{\alpha}^2$ , é um ponto específico em uma distribuição de probabilidade que divide a área de rejeição da área de aceitação em um teste de hipótese. Em outras palavras, é um limite que determina quando rejeitar a hipótese nula com base nos resultados observados.

Desse modo, geralmente, quanto menor o nível de significância escolhido, mais rigoroso será o teste de hipótese e mais distante o valor crítico estará do centro da distribuição.

Através do valor crítico se tem a região crítica do teste referente à qualidade do modelo.

“Assim, devemos rejeitar a hipótese  $H_0$  de que o modelo proposto se ajusta bem aos dados, quando a estatística  $X_i^2$  é maior do que um valor crítico  $X_{\alpha}^2$ ” (SOARES, 2018)

Desse modo, caso  $X_i^2 \geq X_{\alpha}^2$ ,  $H_0$  há de ser rejeitado, contudo, caso contrário,  $H_0$  há de ser aceita.

### Cálculo e resultado do teste

Para calcular o  $X^2$ , usamos um código em Python com dados do ENEM fornecidos pelo professor, focando nas questões de Ciências da Natureza, e aplicamos funções para o cálculo da inferência desenvolvidas pelo grupo. Para calcular o Grau de Liberdade ( $df$ ), usamos um subconjunto da amostra de respondentes escolhido aleatoriamente, dividimos os valores dos  $\theta$ s obtidos em quartis e obtivemos  $df = 3$ . A partir disso, e, adotando o nível de significância

( $\alpha$ ) de 0,05, após análise da tabela de Qui-Quadrado, encontramos o valor crítico de 7,815.

Por fim, utilizando a fórmula descrita anteriormente, calculamos o valor de  $X^2$  e, como resultado, apenas um dos valores foi maior do que  $X_{\alpha}^2$ . Apesar disso, como as frequências observadas se aproximam bem das frequências esperadas, a hipótese nula  $H_0$  não pode ser refutada, ou seja, não podemos descartar a hipótese de que a TRI é uma boa modelagem para os dados do ENEM.

### Conclusão

A TRI apresentou-se como uma técnica eficiente para inferência das habilidades de indivíduos com base em uma série de respostas a itens definidos. Apresenta embasamento teórico satisfatório, considera ampla gama de fatores e possui bom critério avaliativo.

### Referências

- ANDRADE, D. F. d.; TAVARES, H. R.; VALLES, R. d. C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. [S. l.]: SINAP, fev. 2000. P. 3–18. Disponível em: <https://docs.ufpr.br/~niveam/micro%20da%20sala/bom/LivroTRI.pdf>.
- EDUCAÇÃO, M. da. **MEC vai esclarecer sobre TRI à Justiça Federal do Ceará**. [S. l.]: Ministério da Educação. Disponível em: <http://portal.mec.gov.br/component/tags/tag/34530>. Acesso em: 9 maio 2024.
- RABELO, M. **Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro**. 1. ed. [S. l.]: Sociedade Brasileira de Matemática, 2013.
- REVELLE, W. The “New Psychometrics” - Item Response Theory. In: [s. l.: s. n.]. (An introduction to psychometric theory with applications in R). Disponível em: <https://www.personality-project.org/r/book/Chapter8.pdf>. Acesso em: 13 maio 2024.
- SILVA, E. P. d. **Teoria de resposta ao item: análise de atitude dos graduandos em relação às disciplinas de estatística**. Nov. 2014. Tese (Doutorado). Disponível em: [https://bdm.unb.br/bitstream/10483/10076/1/2014\\_20EtiennePereiradaSilva.pdf](https://bdm.unb.br/bitstream/10483/10076/1/2014_20EtiennePereiradaSilva.pdf). Acesso em: 13 maio 2024.
- SOARES, D. J. M. **Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de Matemática básica**. 2018. Tese (Doutorado). Disponível em: <https://www.locus.ufv.br/bitstream/123456789/18404/1/texto%20completo.pdf>. Acesso em: 27 maio 2024.
- TÔRRES, F. C. **Uma Aplicação da Teoria de Resposta ao Item em um Simulado de Matemática no Modelo ENEM**. 2015. Tese (Doutorado). Disponível em: [http://www.realp.unb.br/jspui/bitstream/10482/19102/1/2015\\_FredericoCarvalhoT%c3%b4rres.pdf](http://www.realp.unb.br/jspui/bitstream/10482/19102/1/2015_FredericoCarvalhoT%c3%b4rres.pdf). Acesso em: 27 maio 2024.
- TURNERY, S. **Chi-Square ( $X^2$ ) Table — Examples & Downloadable Table**. [S. l.: s. n.], maio 2022. Disponível em: <https://www.scribbr.com/statistics/chi-square-distribution-table/>. Acesso em: 27 maio 2024.