Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Aplicação da Teoria de Resposta ao Item no Exame Nacional do Ensino Médio

Camila Theresa Oliveira Rosa e Sousa
Victor Scatolin Ferreira Cruz

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade de Brasília, como parte dos requisitos para a obtenção do título de Bacharel em Estatística.

Brasília
2017

Camila Theresa Oliveira Rosa e Sousa
Victor Scatolin Ferreira Cruz

Aplicação da Teoria de Resposta ao Item no Exame Nacional do Ensino Médio

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade de Brasília, como parte dos requisitos para a obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. **Antônio Eduardo Gomes**

Brasília
2017

Agradecimentos

A Deus por ter nos dado saúde e força para superar as dificuldades.

A esta universidade e todo seu corpo docente, além da direção e a administração, que realizam seu trabalho com tanto amor e dedicação, trabalhando incansavelmente para que nós, alunos, pudéssemos contar com um ensino de extrema qualidade.

Ao Prof. Dr. Antônio Eduardo Gomes pela oportunidade e apoio na elaboração deste trabalho.

A todos os professores por nos proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a nós, não somente por terem nos ensinado, mas por terem nos feito aprender. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os nossos eternos agradecimentos.

Aos amigos, companheiros de trabalhos e irmãos na amizade que fizeram parte da nossa formação e que vão continuar presentes em nossas vidas com certeza. Em especial, aos Melhores Amigos da Estatística, Brenda Ribeiro, João Gustavo e Letícia Porfírio. Não podemos deixar de mencionar grandes amigos que nos suportaram e nos ajudaram ao longo do curso, Alfredo Rossi, Gabriel Ramos, Geiziane Oliveira, Guilherme Maia, Mariana Fehr, Matheus Maroneze e Maria Edleuza.

Aos nossos pais, Antônio Carlos Sousa e Junia Oliveira, Moisés Ferreira e Elizelot Scatolin pela determinação e luta na nossa formação e dos nossos irmãos, não medindo esforços para que nós possamos levar nossos estudos adiante.

A família, que nos momentos de nossa ausência dedicados ao estudo, sempre fizeram entender que o futuro é feito a partir da constante dedicação no presente!

É difícil agradecer todas as pessoas que de algum modo, nos momentos serenos e ou apreensivos, fizeram ou fazem parte das nossas vidas, por isso encerramos agradecendo à todos de coração.

Resumo

Neste trabalho estudamos a técnica da Teoria de Resposta ao Item (TRI) para quantificar o conhecimento de matemática dos candidatos do Distrito Federal que prestaram o Exame Nacional do Ensino Médio (ENEM) de 2013 e avaliar se os itens estão calibrados. A TRI é um conjunto de modelos estatísticos que têm como finalidade avaliar os traços latentes, e isso é possível porque o elemento principal da TRI é o item. Os modelos têm como base a relação entre a probabilidade de acertar determinado item dado que o candidato tem proficiência θ. O modelo utilizado nesse trabalho foi o modelo logístico unidimensional de três parâmetros. Estes 3 parâmetros são: discriminação, dificuldade e acerto ao acaso, sendo todos relacionados aos itens. Ao estimar a proficiência dos candidatos foi aplicada a estimação bayesiana marginal utilizando o estimador de máxima verossimilhança marginal. O processo de estimação foi dividido em duas partes. Primeiro foi estimado os parâmetros dos itens e em seguida das proficiências. Feito isso, sucedeu a estimação não paramétrica da curva característica do item aplicando dois testes de adequabilidade de ajuste do modelo: estimador não paramétrico da regressão de Nadaraya-Watson e da regressão Isotônica. O software SAS foi a ferramenta utilizada inicialmente para a análise descritiva, e o software R foi utilizado para a construção da curva característica do item e para cálculos das estimativas de proficiência deste trabalho.

Palavras-chave: Teoria de resposta ao item, ENEM, Proficiência, Regressão de Nadaraya-Watson, Regressão Isotônica.

Sumário

Lista de Figuras

Lista de Tabelas

Introdução e Justificativa

Este trabalho tem como objetivo utilizar a metodologia da Teoria de Resposta ao Item para quantificar o conhecimento dos estudantes que prestaram o ENEM de 2013. Para clarificar, a Teoria de Resposta ao Item é um conjunto de modelos estatísticos que têm como objetivo principal a avaliação educacional. A propositura da Teoria de Resposta ao Item é avaliar as características do estudante que não são possíveis de serem observadas diretamente.

Dentro da Teoria de Resposta ao Item, essas características podem ser chamadas de traço latente. Em exames como o ENEM, a proficiência do estudante que está solucionando o exame pode ser inferida com uma maior precisão. Ou seja, se o estudante vier a realizar duas provas distintas, elaboradas com o mesmo padrão, ele alcançará notas com valores próximos.

Uma das vantagens de utilizar Teoria de Resposta ao Item é que a mesma permite a comparação entre estudantes, isto é, é possível avaliar o desenvolvimento da proficiência de duas escolas ou universidades que são distintas.

A primeira vez que a Teoria de Resposta ao Item foi utilizada no Brasil foi em 1995, na análise dos dados do Sistema de Ensino Básico (SAEB), que tem como objetivo principal realizar uma análise da educação básica brasileira. Desde então, o uso desta metodologia pelos elaboradores de exames tem crescido.

Capítulo 1

Revisão Bibliográfica e Metodologia

1.1 Exame Nacional do Ensino Médio

Nesta sessão, será abordada a história do Exame Nacional do Ensino Médio (ENEM). Vamos entender melhor porque o ENEM foi criado pelo Ministério da Educação e quais eram seus objetivos iniciais. Também entenderemos melhor todas as mudanças que foram realizadas ao longo dos anos, inclusive o porque de adotar Teoria de Resposta ao Item (TRI) como critério de avaliação nas provas realizadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

O ENEM é uma prova realizada pelo INEP, e o INEP é uma autarquia vinculada ao Ministério da Educação do Brasil.

O ENEM foi criado em 1998 pela gestão do ministro da educação Paulo Renato Souza, no governo Fernando Henrique Cardoso, e tinha como objetivo inicial avaliar a qualidade do ensino médio no país para que assim o Ministério da Educação pudesse realizar mudanças estruturais e pontuais de melhorias no ensino brasileiro de educação. Os primeiros modelos de provas do ENEM realizadas pelo INEP foram utilizados entre 1998 e 2008. A prova possuía 63 questões aplicadas em um único dia de prova.

Em 2004, o objetivo da prova mudou. Deixou de ter somente caráter avaliativo para também se tornar um meio de ingresso em cursos superiores através de bolsas de estudo em faculdades particulares pelo Programa Universidade para Todos (ProUni). O Prouni é um programa do Governo Federal do Brasil criado com o objetivo de conceder bolsas de estudo em instituições particulares de ensino superior.

Em 2009, foi implementado um novo modelo de prova para o Enem, com a proposta de unificar o meio de ingresso nas universidades federais brasileiras. Outra mudança significativa foi que o Enem passou a ser realizado em dois dias de prova, contendo 180 questões objetivas e também uma questão de redação. Também fazendo parte da reformulacão realizada em 2009, foi adotada a Teoria de Resposta ao Item (TRI) para a formulação das provas, tornando possível assim, a comparação de notas obtidas em edições diferentes do exame. Esta implementação ficou a cargo de um dos diretores do INEP, Prof. Dr. Heliton Tavares.

1.2 Teoria de Resposta ao Item

A Teoria Clássica dos Testes (TCT) é uma das ferramentas utilizadas para fazer a avaliação educacional. Ela é baseada nos escores brutos ou padronizados, ou seja, quanto mais acertos maior o domínio do conteúdo do estudante. Por ser uma avaliação pontual a TCT tem várias limitações, que refletem na qualidade dos testes.

A Teoria de Resposta ao Item (TRI) tem como finalidade avaliar os traços latentes, ou seja, características dos estudantes que não são possíveis de serem observadas diretamente. Isso é possível por que o elemento principal da TRI é o item (questão). Logo, cada item tem que ser criado com qualidade, levando em conta três parâmetros: dificuldade, discriminação e probabilidade de acerto ao acaso.

Uma avaliação que utiliza a metodologia da Teoria de Resposta ao Item permite a comparação do conhecimento de duas populações que foram submetidas a provas semelhantes, ou a comparação de estudantes da mesma população que foram submetidos a provas diferentes, desde que as provas tenham itens em comum, permite também a estimação da proficiência (conhecimento) dos candidatos.

Existem vários modelos da TRI dependendo do tipo de item que se queira analisar. O primeiro modelo é o logístico de um parâmetro que é conhecido como modelo de Rasch, adequado para testes onde todos os itens têm o mesmo pode de discriminação e não existe acerto ao acaso. O segundo é o modelo logístico de dois parâmetros, que é semelhante ao primeiro acrescido de mais um parâmetro (discriminação do item). E o terceiro é modelo de três parâmetros (discriminação, dificuldade e acerto ao acaso).

No Exame Nacional do Ensino Médio (ENEM) os itens são considerados dicotômicos (certo ou errado), com poder de discriminação distinto para cada item e possibilidade de acerto ao acaso. Por isso o modelo matemático a ser utilizado é o modelo logístico unidimensional de três parâmetros. Este modelo é dado por:

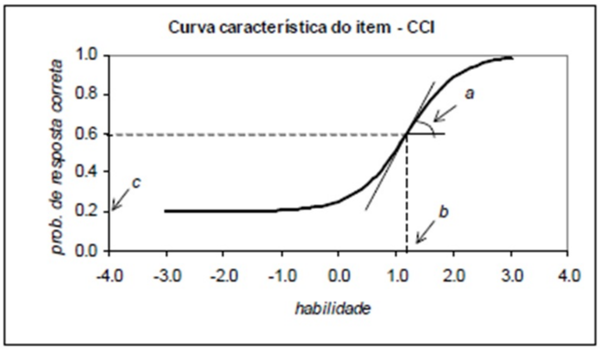
$$\pi_i(\theta_j) = P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i)\left(\frac{1}{1 + \exp^{(-Da_i(\theta_j - b_i))}}\right), \tag{1.1}$$

na qual,

U_{ij} é a variável aleatória indicadora de acerto (ou resposta positiva) do j-ésimo indivíduo ao i-ésimo item;
θ_j representa a habilidade do j-ésimo estudante;
$P(U_{ij} = 1|\theta_j)$ é a probabilidade do estudante j com habilidade θ_j responder corretamente o item i;
a_i é o parâmetro de discriminação do i-ésimo item;
b_i é o parâmetro de dificuldade do i-ésimo item;
c_i é o parâmetro de acerto ao acaso do i-ésimo item;
D é uma constante que permite aproximar $\pi_i(\theta_j)$ pela função distribuição acumulada da distribuição N(0,1) quando D=1,702.

Por meio da Curva Característica do Item (CCI) (Figura 1.1), observa-se melhor a relação entre a probabilidade de acertar um determinado item, dado que o candidato tem a proficiência θ, com os parâmetros do modelo.

Figura 1.1: Curva Característica do Item *(Fonte:Andrade, Tavares e Valle)*



O modelo pressupõe que, quanto maior a proficiência, maior será a probabilidade de acertar um item, ou seja, somente respondentes com proficiência acima do valor do parâmetro b de dificuldade é que terão alta probabilidade de responder corretamente o item. Assim, é possível perceber se o candidato chutou determinado item, pois se ele acertou itens com alto grau de dificuldade, mas errou questões com baixo grau de dificuldade, infere-se que o respondente chutou estas questões consideradas mais difíceis.

A função de informação do item juntamente com a CCI diz o quanto de informação um item pode mensurar de um traço latente, que no caso deste trabalho é a proficiência do candidato. A função de informação do item é maior quanto maior for o valor do parâmetro de discriminação (a_i) do item e quanto menor for o valor do parâmetro do acerto ao acaso (c_i). Matematicamente, a fórmula utilizada para mensurar essa informação no caso do modelo logístico de 3 parâmetros é

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \tag{1.2}$$

sendo,
$I_i(\theta)$ a informação fornecida pelo i-ésimo item para a habilidade θ;
$P_i(\theta) = P(U_i = 1|\theta)$;
$Q_i(\theta) = 1 - P_i(\theta)$.

A função de informação do teste é a soma das informações fornecidas por cada item, ou seja, ela indica se o teste que será aplicado com todos esses itens geram uma boa informação sobre a proficiência dos candidatos. A função de informação do teste pode ser escrita como

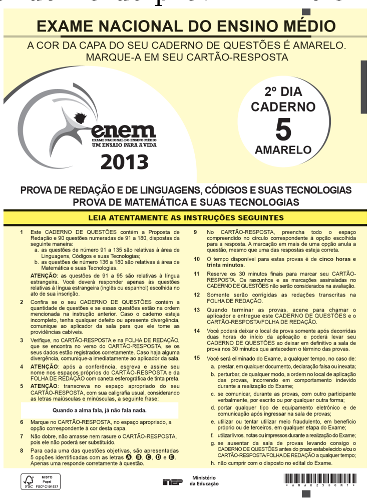
$$l(\theta) = \sum_{i=1}^{I} I_i(\theta) \tag{1.3}$$

1.3 Base de dados

O conjunto de dados que será utilizado neste trabalho é do Exame Nacional do Ensino Médio (ENEM) do ano de 2013 dos candidatos do Distrito Federal. Eles são provenientes do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

A prova do ENEM avalia a proficiência do participante em quatro áreas de conhecimento: Matemática e suas Tecnologias; Linguagens, Códigos e suas Tecnologias; Ciências da Natureza e suas Tecnologias; e Ciências Humanas e suas Tecnologias. Cada uma delas é composta por 45 questões de múltipla escolha. Por serem muitas questões, iremos trabalhar somente com os dados do conhecimento em Matemática e suas tecnologias do caderno de prova amarelo.

Figura 1.2: Caderno de prova amarelo - ENEM 2013



1.4 Estimação da proficiência dos candidatos

A estimação da proficiência será feita pela Estimação Bayesiana Marginal utilizando o estimador de máxima verossimilhança marginal (MVM). Para isso, precisa-se estimar os parâmetros dos itens e das habilidades dos estudantes. A mensuração dos parâmetros é dividida em duas etapas. Primeiro, estima-se o parâmetro dos itens (calibração) e, em seguida, das habilidades.

Antes de começar a estimar o conhecimento dos candidatos, alguns conceitos serão apresentados para que fique mais claro o entendimento da estimação da proficiência.

1.4.1 Estimação Bayesiana Marginal

A Estimação Bayesiana Marginal introduzida por Mislevy (1986) é uma extensão da propositura de Bock & Aitkin (1981). Ela parte do princípio de um conhecimento prévio (distribuição a priori) sobre um parâmetro, gerando em seguida uma nova função denominada distribuição a posteriori, e estima os parâmetros de interesse com base em alguma característica dessa distribuição.

1.4.2 Função de Verossimilhança

A verossimilhança (L) de um conjunto de parâmetros (θ), dada alguma observação (x), é igual à probabilidade daquela observação ter ocorrido dados os valores daqueles parâmetros. Seja $f(\boldsymbol{x}|\theta)$ a função de densidade conjunta de $\boldsymbol{X}$, sendo $\boldsymbol{X}$ uma amostra aleatória. Dado que $\boldsymbol{X}$=(X_1,...,X_n), é observado e igual a $\boldsymbol{x}$=(x_1,..., x_n), a função de (θ) definida por

$$L(\theta) = L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{1.4}$$

é chamada de Função de Verossimilhança. Ela tem a função de transportar a informação dada pela amostra.

1.4.3 Verossimilhança Marginal

A verossimilhança marginal assume uma distribuição de probabilidade para (θ) na população, e integra-se a verossimilhança com relação à distribuição da habilidade. Ou seja,

$$P(\boldsymbol{U_{j.}}|\boldsymbol{\gamma}, \boldsymbol{\eta}) = \int_{\Re} P(\boldsymbol{U_{j.}}|\theta, \boldsymbol{\gamma}, \boldsymbol{\eta}) g(\theta|\boldsymbol{\eta}) d\theta \tag{1.5}$$

sendo que $\boldsymbol{U_{j.}}$=($U_{j1}, ..., U_{jl}$) é o vetor de resposta do j-ésimo indivíduo, $\boldsymbol{\gamma}$ é o conjunto dos parâmetros dos itens, e para o caso de θ seguir uma distribuição Normal, temos $\boldsymbol{\eta} = (\mu, \sigma^2)$, na qual μ é a média e σ^2 é a variância das habilidades dos indivíduos.

1.4.4 Estimador de Máxima Verossimilhança (EMV)

O Estimador de Máxima Verossimilhança (EMV) $\widehat{\theta}$ de θ é o ponto de máximo de L(θ). Para encontrar esse ponto máximo, basta seguir quatro passos. Primeiro calcula-se a função de verossimilhança

$$L(\theta \mid x_1, ..., x_n) = f(x_1 \mid \theta)...f(x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) \tag{1.6}$$

Segundo, aplica-se a função logarítmica em L(θ), pois a função logarítmica é estritamente crescente e o ponto de máximo de l(θ)=log L(θ) coincide com o ponto de máximo de L(θ). Em seguida, deriva-se l(θ) com relação a θ, e por fim iguala-se $\frac{\partial}{\partial\theta} lnL(\theta \mid x_1, ..., x_n)$ a zero, achando assim o EMV, dado pela solução da equação $\frac{\partial}{\partial\theta} lnL(\theta \mid x_1, ..., x_n) = 0$.

1.4.5 Estimação dos parâmetros dos itens

Apresentados os conceitos, abordaremos neste tópico a demonstração da primeira parte do processo da estimação dos parâmetros das proficiências. Como dito anteriormente, ele é dividido em dois passos. Primeiro, calcula-se os valores dos parâmetros dos itens para depois estimar a proficiência dos candidatos. Levando em conta que não conhecemos as habilidades, é necessário tirá-las da função de verossimilhança. Bock e Lieberman (1970) propuseram a estimação dos itens por máxima verossimilhança marginal. Assim a proficiência passa a não fazer parte da função de verossimilhança. Pela expressão (1.5) e assumindo que existe independência entre as respostas de indivíduos distintos, podemos escrever a probabilidade do vetor de respostas como

$$P(\boldsymbol{U}_{..}|\boldsymbol{\gamma}, \boldsymbol{\eta}) = \prod_{j=1}^{n} P(\boldsymbol{U_{j\cdot}}|\boldsymbol{\gamma}, \boldsymbol{\eta}). \tag{1.7}$$

Pelo fato de existir essa independência, os dados seguem uma distribuição Multinomial. Logo, a função de verossimilhança marginal é dada por

$$L(\boldsymbol{\gamma}, \boldsymbol{\eta}) = \frac{n!}{\prod_{l=1}^{s} R_l!} \prod_{l=1}^{s} P(\boldsymbol{U_{j\cdot}}|\boldsymbol{\gamma}, \boldsymbol{\eta})^{R_l}, \tag{1.8}$$

sendo
R_l : número de ocorrências do l-ésimo padrão;
s : número de padrões observados;
n : $\sum_{l=1}^{s} R_l$.

Por definição, o padrão de resposta observado é uma sequência específica de acertos e erros dos itens.

A estimação dos parâmetros dos itens é feita através da equação

$$\frac{\partial}{\partial\gamma_i} logL(\boldsymbol{\gamma}, \boldsymbol{\eta}) = 0, i = 1, ..., n. \tag{1.9}$$

Aplicando (1.9) para cada parâmetro do item obtemos,

$$a_i : D(1 - c_i) \sum_{j=1}^{s} r_j \int_{\Re} [(U_{ij} - P_i)(\theta - b_i)W_i] g_j^*(\theta) d\theta = 0, \tag{1.10}$$

$$b_i : Da_i(1 - C_i) \sum_{j=1}^{s} r_j \int_{\Re} [(U_{ij} - P_i)W_i] g_j^*(\theta) d\theta = 0, \tag{1.11}$$

$$c_i : \sum_{j=1}^{s} r_j \int_{\Re} [(U_{ij} - P_i)\frac{W_i}{P_i^*}] g_j^*(\theta) d\theta = 0, \tag{1.12}$$

na qual,

$$W_i = \frac{P_i^* Q_i^*}{P_i Q_i}, \tag{1.13}$$

$$P_i^* = [1 + e^{-Da_i(\theta - b_i)}]^{-1}, \tag{1.14}$$

$$Q_i^* = 1 - P_i^*, \tag{1.15}$$

$$g_j^*(\theta) = g(\theta | U_{j}., \gamma, \eta) = \frac{P(U_j.|\theta, \gamma) g(\theta, \eta)}{P(U_j., \gamma, \eta)} \text{ , distribuição a posteriori de } \theta. \tag{1.16}$$

O próximo passo é definir as distribuições a priori para os parâmetros dos itens. Os parâmetros de discriminação (a_i), dificuldade (b_i) e acerto ao acaso (c_i), possuem algumas limitações. O parâmetro a_i tem que ser positivo, b_i pode assumir qualquer valor real e c_i pode ser qualquer valor de 0 a 1 ($0 \leq c_i \leq 1$), pois este se trata de uma probabilidade. Portanto, ao escolher qual distribuição a priori utilizar para cada parâmetro do vetor $\boldsymbol{\gamma_i} = (a_i, b_i, c_i)$, leva-se em conta essas restrições. As distribuições a priori mais adequadas para cada um são: utiliza-se a distribuição Log-normal ou Qui-quadrado para a_i; distribuição Normal para b_i; e distribuição Beta para c_i.

1.4.6 Estimação do parâmetro das proficiências

Após a estimação dos parâmetros dos itens, estima-se as habilidades θ_j, considerando os parâmetros dos itens fixos. Ao adotar que θ_j segue uma distribuição a priori Normal com parâmetros μ e σ^2 conhecidos e considerando a hipótese de independência entre as habilidades de diferentes indivíduos, pode-se dizer que a posteriori para a proficiência do indivíduo j é obtida por (1.16).

A média e a moda da distribuição a posteriori são os estimadores mais utilizados para estimar θ_j, mas abordaremos apenas obtenção da estimação pela média a posteriori (EAP), que consiste em obter

$$\widehat{\theta}_j = E[\theta | \boldsymbol{U_{j}.}, \boldsymbol{\gamma}, \boldsymbol{\eta}] = \frac{\int_{\Re} \theta g(\theta|\boldsymbol{\eta}) P(\boldsymbol{U_j.}|\theta, \boldsymbol{\gamma}) d\theta}{\int_{\Re} g(\theta|\boldsymbol{\eta}) P(\boldsymbol{U_j.}|\theta, \boldsymbol{\gamma}) d\theta}. \tag{1.17}$$

1.5 Estimação não-paramétrica da Curva Característica do Item

Feita a estimação das proficiências, abordaremos neste tópico os testes de adequabilidade de ajuste para verificar se a curva característica do item segue um modelo de 3 parâmetros ajustado.

Para a estimação não paramétrica da CCI, obtemos uma estimativa inicial de θ_j, onde $j = 1, ..., n$. Em seguida, agrupam-se as proficiências em intervalos e estima-se a probabilidade de resposta positiva para indivíduos com proficiência estimada em cada intervalo como sendo a proporção de respostas positivas para o item dos indivíduos com proficiência estimada em cada intervalo. A proporção é calculada da seguinte maneira:

$$\hat{P_m} = \frac{1}{Nm} \sum_{j:\widehat{\theta}_j \in Lm} U_{ij}, \tag{1.18}$$

na qual,
U_{ij} é a variável aleatória indicadora de acerto (ou resposta positiva) do j-ésimo indivíduo ao i-ésimo item;
L_m é o m-ésimo intervalo de proficiências estimadas, onde m=(1,...,M), onde M é o número de intervalos;
N_m é o número de indivíduos com proficiência estimada $\widehat{\theta}_j \in L_m$, j=1,...,n, para o i-ésimo item.

Calcula-se $\widehat{\theta}_j = \Phi^{-1}(T_j/I)$, desde que $T_j \neq 0$ e $T_j \neq I$, ou seja, isso não vale para indivíduos que obtiveram escore igual a zero ou escore total.

A estatística do teste é a distância entre o modelo paramétrico ajustado aos dados e o estimador não paramétrico, como mostra a fórmula abaixo:

$$Q = \int_{\Re} \left| \widehat{P}_i(\theta) - \widehat{F}_{ac}(\theta) \right| \phi(\theta) d\theta, \tag{1.19}$$

na qual $\widehat{P}_i(\theta)$ é a curva característica do item ajustada para o modelo de 3 parâmetros, $\widehat{F}_{ac}(\theta) = \widehat{P_m}$ e ϕ é a função densidade da distribuição $N(0, 1)$.

Quanto maior Q, maior é a probabilidade da curva de ajuste se afastar do modelo não paramétrico.

Após a definição da estatística do teste podemos testar as seguintes hipóteses:
H_0: a CCI para o i-ésimo item segue o modelo de 3 parâmetros ajustado
H_a: a CCI para o i-ésimo item não segue o modelo de 3 parâmetros ajustado

Para a avaliação do p-valor do teste, é utilizado um procedimento bootstrap paramétrico. Assim, gera-se w amostras de U_{ij} a partir do modelo de 3 parâmetros ajustados para o item; e calcula-se o p-valor do teste como sendo a proporção dos w-valores de Q calculados em cada amostra que estão acima dos valores de Q calculados para amostra original. Se o p-valor for muito pequeno há evidências para rejeitar H_0.

Os testes utilizados para testar essas hipóteses foram: regressão de Nadaraya-Watson e regressão Isotônica.

1.5.1 Regressão de Nadaraya-Watson

O estimador não paramétrico utilizado foi a regressão de Nadaraya-Watson, que é calculada da seguinte forma:

$$\widehat{\pi}_i(\theta) = \frac{\sum_{j=1}^{n} u_{ij} k(\frac{\widehat{\theta}_j - \theta}{h})}{\sum_{j=1}^{n} k(\frac{\widehat{\theta}_j - \theta}{h})}, \tag{1.20}$$

na qual k é uma função densidade de probabilidade simétrica em torno do zero e unimodal, e h é o parâmetro de suavização. Com relação ao parâmetro de suavização da regressão de Nadaraya-Watson, há algumas propostas para sua determinação. A mais utilizada é a validação cruzada.

1.5.2 Regressão Isotônica

A técnica da regressão isotônica, com referência no livro de *Barlow et al. (1972)*, baseia-se em encontrar uma função não decrescente que minimize a soma de quadrados dos erros, ou seja, que minimize a distância entre a curva suavizada e os pontos. Um ponto importante e que difere a regressão isotônica do teste anterior é o fato de utilizar a informação a priori sobre uma possível relação de ordem na variável resposta. Assim, na regressão isotônica os parâmetros devem seguir essa restrição de ordem.

Escrevendo de forma matemática queremos encontrar $g^*(x_j)$, em que $j = 1, ..., n$, que minimiza

$$\sum_{j=1}^{n} [y_j - f(x_j)]^2 w_j \tag{1.21}$$

entre todas as funções f não decrescentes em x_j, $j = 1, ..., n$, isto é,

$$f(x_1) \leq f(x_2) \leq ... \leq f(x_n), \tag{1.22}$$

na qual w_j é o peso da j-ésima observação. Sendo que (x_j, y_j) é uma amostra aleatória, $j = 1, ..., n$, e $y_j = U_{ij}$, $x_j = \widehat{\theta}_j$, $x = \theta$. Ou seja,

$$g^* = \arg\min_{f \in F} \sum_{j=1}^{n} [y_j - f(x_j)]^2 w_j, \tag{1.23}$$

em que F é o conjunto das funções não decrescentes em x.

Para obter g^* temos que seguir três passos:

1. Constrói-se o diagrama de somas acumuladas (DSA) formado pelos pontos

$$P_j = (w_j, G_j) = (\sum_{l=1}^{j} w_l, \sum_{l=1}^{j} y_l w_l) \tag{1.24}$$

   com, $j = 1, ..., n$ e $P_0 = (0, 0)$.

2. Obtém-se a função minorante convexa máxima do DSA.

3. $g^*(x_j)$ é dado pelo vetor da derivada à esquerda da função minorante convexa máxima no ponto w_j.

A suavização da regressão isotônica se dá pela seguinte fórmula:

$$\tilde{\pi}(\theta) = \sum_{j=1}^{n} k(\frac{\hat{\theta}_j - \theta}{h})[\hat{\pi}(\hat{\theta}_j) - \hat{\pi}(\hat{\theta}_{j-1})] \tag{1.25}$$

em que $k(y)$ é uma função de distribuição acumulada da distribuição $N(0, 1)$, e $\hat{\pi}(\theta_j)$=$g^*(x_j)$.

Capítulo 2

Resultados e Discussão

2.1 Resultados

Depois de realizadas todas as etapas no tratamento dos dados e também finalizadas as análises descritivas, iniciamos os cálculos dos parâmetros. Após os parâmetros terem sido calculados, foi realizado o teste de adequabilidade. Através dele foi descoberto que para alguns itens o modelo não-paramétrico não se aproximava do modelo paramétrico.

Abaixo, com a tabela 2.2, temos a estimativas dos parâmetros para as 45 questões referentes a prova de matemática do ENEM de 2013. Vale ressaltar que estas estimativas foram calculadas através do software R, utilizando a função tpm, do pacote LTM ("Latent Trait Models").

Tabela 2.1: Estimativas dos Parâmetros

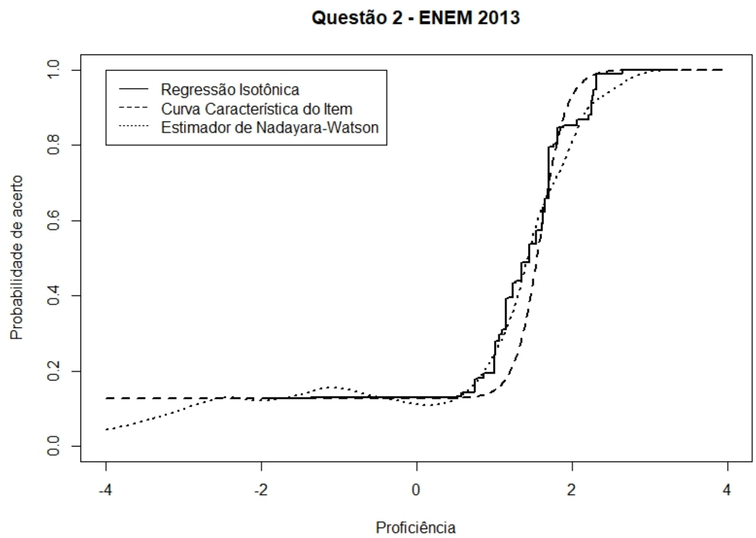
	Acerto Casual	Dificuldade	Discriminação
Questão 1	0,1280242	2,203261	3,282423
Questão 2	0,1263489	1,600773	3,58905
Questão 3	0,1691969	1,382541	2,732857
Questão 4	0,00112627	-0,6407266	1,549941
Questão 5	0,1060969	-0,361886	1,835063
Questão 6	0,2416132	2,008754	6,179447
Questão 7	0,1076189	1,75297	2,83522
Questão 8	0,1589516	1,768507	3,034073
Questão 9	0,1306532	1,703345	4,006224
Questão 10	0,1581965	1,922758	3,204293
Questão 11	0,2138953	0,7014648	2,675136
Questão 12	0,178563	0,9280518	2,65187
Questão 13	0,1612532	0,526628	2,534022
Questão 14	0,005356573	-0,02223527	1,228127
Questão 15	0,1138477	18,20373	0,1149797
Questão 16	0,06291263	1,568739	2,545352
Questão 17	0,2217809	0,8954247	1,491224
Questão 18	0,09937846	1,483124	2,882806
Questão 19	0,08485819	0,8509253	0,8773932
Questão 20	0,2138715	-9,441045	-0,3432438
Questão 21	0,09202271	2,087003	3,779586
Questão 22	0,187695	2,011951	2,289638
Questão 23	0,181987	2,12547	3,028183
Questão 24	0,185485	1,158371	2,2098
Questão 25	0,119913	2,431284	1,235122
Questão 26	0,2593503	17,30287	0,1364004
Questão 27	0,1126714	2,026333	3,390804
Questão 28	0,2547265	1,962555	2,586053
Questão 29	0,2088186	2,240731	5,31299
Questão 30	0,1506275	1,932997	2,990251
Questão 31	0,1531787	2,675752	2,00442
Questão 32	0,002572649	-5,742139	-0,3435109
Questão 33	0,1126714	1,257872	0,6728973
Questão 34	0,1468346	0,8990545	2,546176
Questão 35	0,1113687	-0,03228337	1,521718
Questão 36	0,03780544	52,98853	0,03497029
Questão 37	0,2517573	2,621595	2,395915
Questão 38	0,1491496	-4,123213	-1,041135
Questão 39	0,2111825	2,22714	2,131945
Questão 40	0,2360828	2,719245	1,113031
Questão 41	0,09998384	-3,871787	-1,072227
Questão 42	0,1450543	2,312045	2,277411
Questão 43	0,2107267	1,741143	1,903197
Questão 44	0,1267286	2,051237	1,60893
Questão 45	0,1132302	2,678573	1,414727

Quando se iniciou a averiguação dos resultados, foi observado que na maioria das questões o resultado foi positivo, ou seja, os modelos não paramétricos construídos se aproximaram dos paramétricos.

Entretanto, existiram casos distintos, onde os itens não seguiram o comportamento que era esperado.

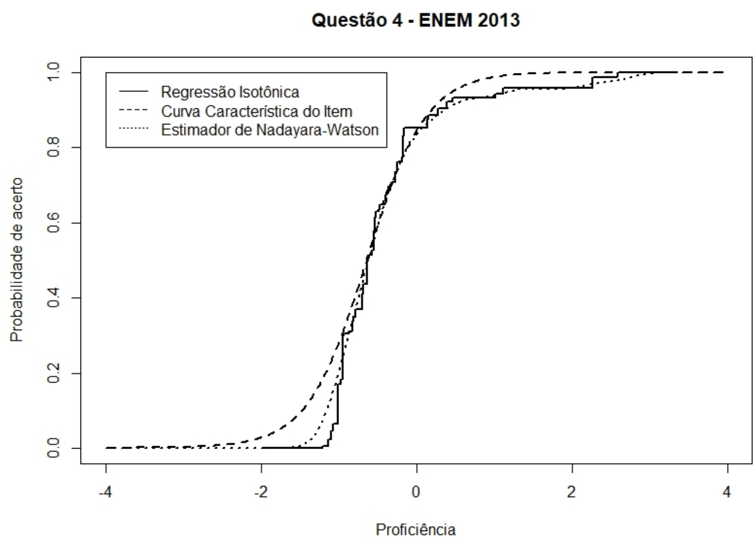
Abaixo serão colocados os gráficos de algumas questões, para elucidar melhor o que foi dito acima. As curvas não paramétricas nos gráficos seguintes, se referem a regressão de Nadaraya-Watson e a regressão isotônica.

Figura 2.1: Curva Característica do Item - Questão 2



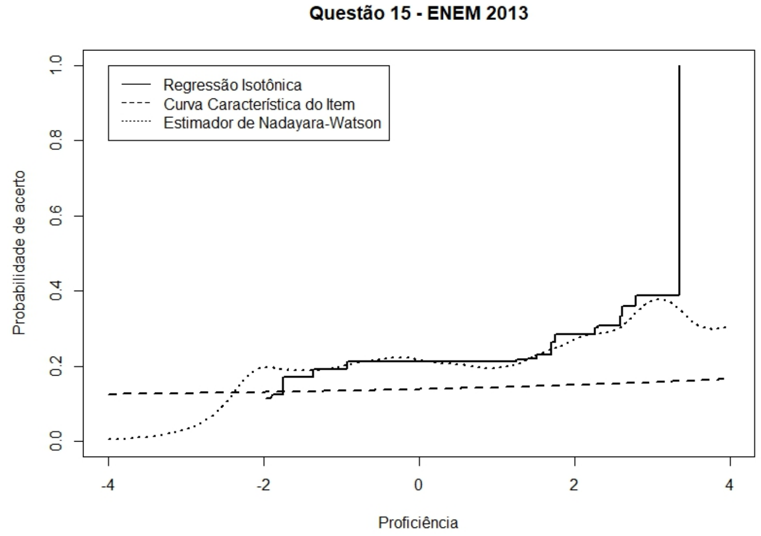
Comentário: A questão 2 foi bem ajustada pelos modelos não paramétricos. Observando a curva característica do item (CCI), que é nossa curva paramétrica, vemos que tanto a regressão de Nadaraya-Watson quanto a regressão isotônica convergiram bem durante toda a curva da questão.

Figura 2.2: Curva Característica do Item - Questão 4



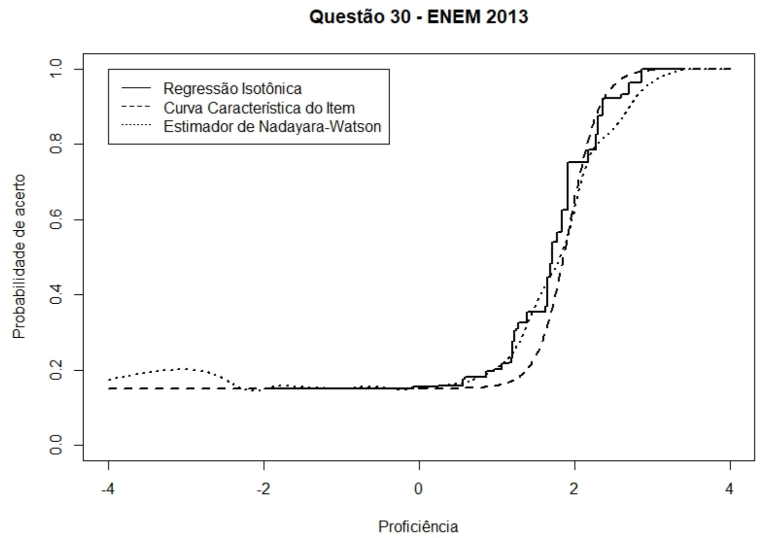
Comentário: A questão 4 foi bem ajustada pelos modelos não paramétricos. Observando a curva característica do item (CCI), enxergamos algo bem parecido com a questão 2, porém o que é interessante ao analisar essas duas questões, é que na questão 4, o gráfico tem a crescente na probabilidade de acerto com uma proficiência muito menor do que na questão 2.

Figura 2.3: Curva Característica do Item - Questão 15



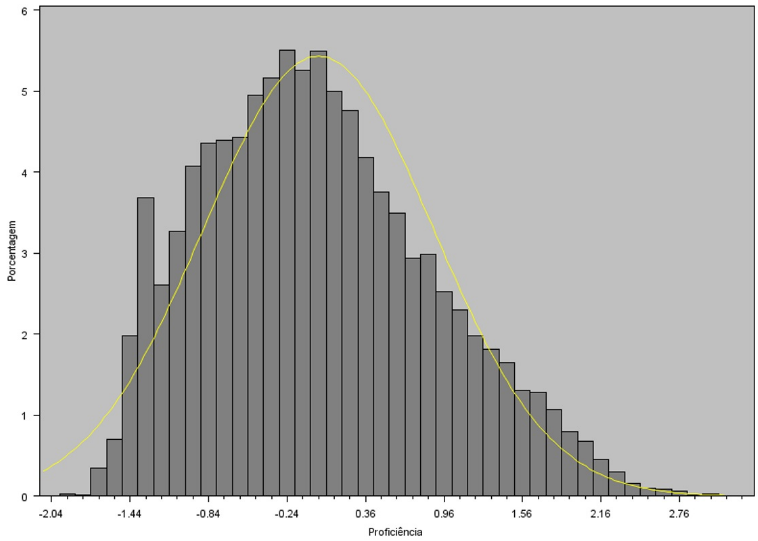
Comentário: Dentre todos as questões, a questão 15 foi a pior ajustada, isso se deve ao fato de que a curva característica do item foi extremamente divergente das regressões de Nadaraya-Watson e sobretudo da regressão isotônica.

Figura 2.4: Curva Característica do Item - Questão 30



Comentário: Análogo às questões 2 e 4, a questão 30 também foi bem ajustada, porém observa-se que há divergências na calda superior em relação a curva característica do item e a regressão de Nadaraya-Watson, e também no ponto em que ambas começam a aumentar a probabilidade de acerto da questão.

Figura 2.5: Histograma - Proficiência dos candidatos



Comentário: O histograma mostrado acima, representa a proficiência estimada dos candidatos analisados. Apenas visualizando o histograma, é possível sugerir que a proficiência estimada dos candidatos segue uma distribuição Normal, porém ao realizar o teste de Kolmogorov-Smirnov, na qual testamos as seguintes hipóteses

H_0: Os dados seguem uma distribuição Normal
H_a: Os dados não seguem uma distribuição Normal

verificamos que o D observado ($D_{obs} = 0,03765$) era maior que o D crítico ($D_{crit} = 0.0095$), com isso rejeitamos a hipótese de normalidade dos dados com 95% de confiança.

Capítulo 3

Conclusão

Nesse trabalho foram utilizadas 3 técnicas distintas, sendo elas, uma paramétrica (curva característica do item) e duas não paramétricas (regressão de Nadaraya-Watson e regressão isotônica). Todas com o objetivo de melhor quantificar a proficiência do estudante no ENEM 2013.

É possível afirmar que os testes realizados para os modelos de teoria de resposta ao item atenderam as expectativas. Ao analisar os resultados, vimos que existem questões que a curva característica do item construída é bem ajustada em relação aos gráficos dos modelos não paramétricos. Entretanto existem questões, como a questão 15, que os testes não atendem as expectativas esperadas.

Tivemos 4 questões com o parâmetro de discriminação negativo. Conforme tabela 3.1, podemos observar que as questões 20, 32, 38, 41 possuem valores negativos de acordo os cálculos dos parâmetros realizados pelo R, dessa maneira elas necessitariam ser reavaliadas. Vale ressaltar, que a estimativa da curva característica do item para estes testes poderia ter um aspecto mais próximo do que se espera dos itens do ENEM utilizando uma amostra de respondentes maior do que a que foi utilizada neste trabalho.

Tabela 3.1: Questões com parâmetro de discriminação negativo

ENEM 2013	Discriminação
Questão 20	-0,3432438
Questão 32	-0,3435109
Questão 38	-1,041135
Questão 41	-1,072227

Referências Bibliográficas

[1] Andrade, Tavares e Valle **Teoria da Resposta ao Item: Conceitos e Aplicações.** SINAPE. 2000.

[2] Barlow, R. E., Bartholomew **Statistical Inference Under Order Restrictions.** 1972.

[3] Bock, R. D. and Lieberman, M. **Fitting a response model for n dichotomously scored items**. Psychometrika, 35, 179-197.1970.

[4] Mislevy, R. J. **Bayes modal estimation in item response models.** Psychometrika, 51, 177-195.1986.

[5] Bock, R. D. and Aitkin, M. **Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm.** Psychometrika, 46, 433-459. 1981.

Apêndice A

Obtenção da equação de estimação dos parâmetros

Para obtermos a equação de estimação para o parâmetro de discriminação, a_i, temos que:

$$\begin{aligned}
\frac{\partial logL(\gamma,\eta)}{\partial a_i} &= \sum_{j=1}^{s}\int_{\Re}\left[\left[U_{ji}-P_i\left(\frac{\partial P_i}{\partial a_i}\right)\frac{W_i}{P_i^*Q_i^*}\right]g_j^*(\theta)d(\theta)\right] \\
&= \sum_{j=1}^{s} r_j\int_{\Re}\left[(U_{ji}-P_i)D(1-c_i)(\theta-b_i)P_i^*Q_i^*\frac{W_i}{P_i^*Q_i^*}\right]g_j^*(\theta)d(\theta) \\
&= D(1-c_i)\sum_{j=1}^{s} r_j\int_{\Re}[(U_{ij}-P_i)(\theta-b_i)W_i]g_j^*(\theta)d\theta
\end{aligned} \tag{A.1}$$

Para o parâmetro de dificuldade, b_i, temos que:

$$\begin{aligned}
\frac{\partial logL(\gamma,\eta)}{\partial b_i} &= \sum_{j=1}^{s}\int_{\Re}\left[(U_{ij}-P_i)\left(\frac{\partial P_i}{\partial b_i}\right)\left(\frac{W_i}{P_i^*Q_i^*}\right)\right]g_j^*(\theta)d(\theta) \\
&= \sum_{j=1}^{s}\int_{\Re}\left[(U_{ij}-P_i)(-1)Da_i(1-c_i)P_i^*Q_i^*\left(\frac{W_i}{P_i^*Q_i^*}\right)\right]g_j^*(\theta)d(\theta) \\
&= Da_i(1-C_i)\sum_{j=1}^{s} r_j\int_{\Re}[(U_{ij}-P_i)W_i]g_j^*(\theta)d\theta
\end{aligned} \tag{A.2}$$

Para o parâmetro de acerto ao acaso, c_i, temos que:

$$\begin{aligned}
\frac{\partial logL(\gamma,\eta)}{\partial c_i} &= \sum_{j=1}^{s}\int_{\Re}\left[(U_{ij}-P_i)\left(\frac{\partial P_i}{\partial c_i}\right)\left(\frac{W_i}{P_i^*Q_i^*}\right)\right]g_j^*(\theta)d(\theta) \\
&= \sum_{j=1}^{s}\int_{\Re}\left[(U_{ij}-P_i)Q_i^*\left(\frac{W_i}{P_i^*Q_i^*}\right)\right]g_j^*(\theta)d(\theta) \qquad \text{(A.3)} \\
&= \sum_{j=1}^{s}r_j\int_{\Re}\left[(U_{ij}-P_i)\left(\frac{W_i}{P_i^*}\right)\right]g_j^*(\theta)d\theta
\end{aligned}$$