



Universidade de Brasília
Departamento de Estatística

Teoria da Resposta ao Item Aplicada à Avaliação das Questões da Prova de
Geografia do Vestibular da UnB de 2014

Guilherme Silva de Souza

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2023

Guilherme Silva de Souza

**Teoria da Resposta ao Item Aplicada à Avaliação das Questões da Prova de
Geografia do Vestibular da UnB de 2014**

Orientador: Prof. Antonio Eduardo Gomes

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Agradecimentos

É importante para mim mostrar a minha gratidão a todos que me apoiaram ao longo dessa jornada, então, faço questão de nomeá-los aqui.

Em primeiro lugar, agradeço a Deus por seu amparo, sua proteção e suas bênçãos sobre mim, para que eu continuasse forte e não desistisse dos meus sonhos.

Agradeço também aos meus pais, Uelton Souza e Cíntia Souza, que sempre torceram por mim, são minha base emocional e protetora, cuidaram e cuidam de mim e dos meus irmãos e sempre fizeram questão de investir nos nossos estudos.

Aos meus irmãos, Gustavo Souza e Gabriel Souza, por serem as pessoas que mais me conhecem e entendem e por tornarem a minha vida mais leve e fácil de seguir.

Aos meus melhores amigos, Lucas Lisboa, Nowras Otmen, Ícaro Farias, Pedro Nóbrega, Larissa Brecht e Ian Alves, por me acompanharem desde a adolescência e serem meus refúgios em momentos de tensão ao longo da vida.

A minha namorada, Rayssa Zangerolami, por me incentivar a me esforçar ao máximo em tudo que faço e por todo seu carinho nos meus momentos de ansiedade e fraqueza.

Aos meus amigos do curso, João Pedro Pupe, Caio Cavalcante, João Victor Melo, Lucas Menezes e Isabela Harumi, por me alegrarem em meio ao caos da universidade e me ajudarem nos estudos e nas atividades da graduação.

Por fim, ao meu orientador, Professor Antonio Eduardo Gomes, por me fazer apreciar o mundo da Teoria da Resposta ao Item, por aceitar entrar neste projeto comigo e por todo o amparo necessário.

Resumo

Este trabalho tem como objetivo analisar as respostas aos itens da prova de geografia do vestibular da UnB de 2014, realizado pela banca CEBRASPE, a partir de métodos e ajustes de Teoria da Resposta ao Item. O foco será nos Modelos Logísticos de 2 e de 3 parâmetros e no Modelo de Resposta gradual e suas duas possíveis ordenações, sendo “não resposta”, “errado” e “certo” a ordenação que se espera melhor ajuste, devido à penalização dada pela banca ao errar uma questão.

Para o estudo e a análise dos itens, foi utilizado o software R, contando com a assistência de tabelas e gráficos para um melhor entendimento do assunto, para uma visualização mais clara dos parâmetros e para uma interpretação mais concisa das estimativas.

Palavras-chave: Teoria da Resposta ao Item; Modelos Logísticos; Modelo de Resposta Gradual; Vestibular UnB; CEBRASPE.

Abstract

This work aims to analyze the responses to the geography items of the UnB (University of Brasília) 2014 entrance exam, conducted by the CEBRASPE board, using methods and adjustments of Item Response Theory. The focus will be on the 2-parameter and 3-parameter Logistic Models, as well as the Graded Response Model, and their two possible orderings, with “no response”, “wrong” and “correct” being the expected ordering that provides the best fit due to the penalty applied by the board for answering a question incorrectly.

For the study and analysis of the items, the R software was used, with the assistance of tables and graphs to enhance understanding of the subject, provide a clearer visualization of the parameters, and offer a more concise interpretation of the estimates.

Keywords: Item Response Theory; Logistic Models; Graded Response Model; UnB Entrance Exam; CEBRASPE.

Lista de Tabelas

1	Parâmetros do ML3	21
2	Parâmetros do ML2	24

Lista de Figuras

1	Divisão das Questões do Vestibular	10
2	Curva Característica do Item	11
3	Distribuição das Respostas por Item	14
4	Boxplots Escore do Vestibular	15
5	Comparação de Escores	16
6	Histogramas de Distribuição dos Escores	16
7	Proporção de Respostas Corretas por Item	17
8	Curvas Características dos Itens (CCI)	20
9	Coeficiente de Correlação Ponto-Bisserial	20
10	Dificuldade x Proporção de Acerto no ML3	22
11	Curvas Características dos Itens no ML3	22
12	Proficiência estimada x Escore padronizado no ML3	23
13	Proficiência estimada x Escore vestibular padronizado no ML3	23
14	Comparações entre o ML2 e o ML3	25
15	Comparação das Medianas	26
16	Proficiência Estimada MRG x Dicotômicos (errado, NR, certo)	27
17	Discriminação e Dificuldade MRG (errado, NR, certo)	28
18	Proficiência Estimada MRG x Dicotômicos (NR, errado, certo)	29
19	Curvas Características do MRG (NR, errado, certo)	31
20	Discriminação e Dificuldade (NR, errado, certo)	31

Sumário

1 Introdução	8
2 Metodologia	9
2.1 Conjunto de dados	9
2.2 Coeficiente de Correlação Ponto-Bisserial	10
2.3 Teoria da Resposta ao Item	11
2.3.1 Curva Característica do Item	11
2.3.2 Modelos Logísticos para Itens Dicotômicos	12
2.3.3 Estimativa dos Parâmetros	12
2.3.4 Modelo de Resposta Gradual	13
3 Resultados	14
3.1 Análise Descritiva	14
3.2 Ajuste dos Modelos Logísticos	21
3.2.1 Modelo Logístico de 3 Parâmetros	21
3.2.2 Modelo Logístico de 2 Parâmetros	24
3.3 Ajuste do Modelo de Resposta Gradual	27
4 Conclusão	32
Referências	33

1 Introdução

No Brasil e em boa parte do mundo, é comum a utilização de provas e exames como um modo de avaliação e de classificação de estudantes e, mesmo que haja controvérsias sobre sua eficiência, é um sistema bastante complexo e com métodos avaliativos e comparativos bastante rebuscados.

Um método criado nesses moldes foi a Teoria da Resposta ao Item (TRI), que se apresenta como uma evolução da Teoria Clássica dos Testes (TCT) e tem como objetivo analisar cada item da prova, para assim, ter resultados sobre as habilidades e os traços latentes do respondente, isto é, analisar características não observáveis, como o conhecimento.

A TRI apresenta duas suposições a serem seguidas, a primeira é a unidimensionalidade, que é ter um traço latente dominante e todos os itens estarem medindo essa mesma habilidade. A segunda suposição é a independência local, que, mantendo o traço latente constante, a resposta dos indivíduos a um item deve ser independente à resposta a outros itens.

Pelo fato de a TRI ser baseada nos itens e não no questionário como um todo, isso permite análises estatísticas que a TCT não permitia. Assim como é descrito por Araujo, Andrade e Bortolotti (2009), a TRI apresenta vantagens como a comparação entre respondentes que não necessariamente tenham feito a mesma prova e também, os itens e os indivíduos são analisados em uma mesma escala, o que possibilita a comparação de características dos dois.

Além disso, há a possibilidade de comparação temporal entre os respondentes, isto é, a partir dos modelos de TRI, pode-se comparar duas provas aplicadas em anos diferentes e medir as probabilidades do respondente acertar, errar ou não responder itens de acordo com seu traço latente.

Então, neste projeto, foi utilizada a TRI para analisar a prova de geografia o vestibular de 2014 da Universidade de Brasília (somente as questões dos tipos A e C), que é um dos processos seletivos mais concorridos do Distrito Federal, isso porque a UnB é reconhecida como uma das melhores universidades do Brasil.

Este estudo foi escolhido porque a análise de avaliações e das características dos vestibulandos é algo extremamente útil para o entendimento do processo escolar dos estudantes, assim como para possíveis melhoras nas questões do vestibular, para que assim meçam melhor o conhecimento dos participantes.

2 Metodologia

Este estudo será iniciado a partir de uma análise descritiva e exploratória dos dados, seguido da utilização da Teoria Clássica dos Testes (TCT), que é um método que não permite a comparação entre indivíduos que responderam provas diferentes. Então, para acabar com essa limitação, será iniciada uma análise a partir da Teoria de Resposta ao Item (TRI).

Na TRI, serão analisados os itens em específico e, corroborando com este estudo, autores referem que, em TRI, há uma análise da habilidade do respondente em comparação com o acerto ou não da questão e o esperado é que quanto maior for a habilidade ou o conhecimento do indivíduo, maior deverá ser a probabilidade de ele acertar o item (ANDRADE; TAVARES; VALLE, 2000).

Na primeira parte dessa análise de TRI, será utilizado o Modelo Logístico de 3 Parâmetros (discriminação, dificuldade e probabilidade de acerto ao acaso) e depois, o Modelo Logístico de 2 Parâmetros (sem acerto ao acaso).

Por fim, será aplicado o Modelo de Resposta Gradual (MRG) de Samejima (1969), que é um modelo capaz de avaliar o grau de importância dos itens e os classificar. Primeiramente, será utilizada a ordem de respostas “errado”, “não resposta” e “certo”, porém, como nas provas do Cebraspe há o fator de correção (penalização quando se erra a questão), será testado também o modelo em que a categoria “não resposta” vem antes da categoria “errado”.

2.1 Conjunto de dados

O conjunto de dados deste estudo foi analisado a partir do vestibular da UnB de 2014, que teve as questões distribuídas segundo o quadro da Figura 1.

No total são 300 questões, mas neste trabalho será analisada somente a prova de geografia, realizada no primeiro dia. Foram realizadas, ao total, 24 questões de geografia. Todavia, 3 delas foram anuladas, então, a análise se baseou somente em 21 questões, sendo elas 19 do tipo A (certo ou errado) e 2 do tipo C (múltipla escolha), sendo eles os itens 11 e 21 e, ao total, foram analisadas respostas de 7.232 candidatos.

É importante ressaltar que os itens do Vestibular não foram construídos utilizando técnicas de TRI, então talvez as suposições de unidimensionalidade e independência local não sejam contempladas.

Data	Prova	Disciplinas-Foco	Nº de itens	Duração
1º Dia 7/6/2014	Conhecimentos - Parte I	Língua Espanhola, Língua Francesa ou Língua Inglesa	30	300min
	Conhecimentos - Parte II	Língua Portuguesa e Literaturas de Língua Portuguesa, Geografia e História, Artes (Artes Cênicas, Artes Visuais e Música), Filosofia e Sociologia	120	
	Redação em Língua Portuguesa	-	-	
2º Dia 8/6/2014	Conhecimentos - Parte III	Biologia, Física, Química e Matemática	150	300min

Figura 1: Divisão das Questões do Vestibular

2.2 Coeficiente de Correlação Ponto-Bisserial

O coeficiente de correlação ponto-bisserial é calculado a partir da correlação de Pearson e tem como objetivo evidenciar o nível de correlação linear entre as respostas binárias para o item e o escore dos indivíduos respondentes. Espera-se que o valor desse coeficiente seja positivo, pois assim, demonstra coerência na correlação.

A fórmula deste coeficiente é dada por

$$\rho_i^{PB} = \frac{\sum_{j=1}^n U_{ij}T_j - (\sum_{j=1}^n U_{ij})(\sum_{j=1}^n T_j)/n}{\sqrt{[\sum_{j=1}^n (U_{ij} - \bar{U}_i)^2][\sum_{j=1}^n (T_j - \bar{T})^2]}} = \frac{n_i\bar{T}_A - n_i\bar{T}}{n\sqrt{p_i(1-p_i)}S_T} = \left(\frac{\bar{T}_A - \bar{T}}{S_T}\right) \sqrt{\frac{p_i}{1-p_i}}, \quad (2.2.1)$$

onde

- U_{ij} é a resposta do j -ésimo indivíduo para o i -ésimo item (1 = acerto, 0 = erro);
- T_j é o escore do j -ésimo indivíduo;
- n é o número de respondentes;
- n_i é o número de respondentes que acertaram o i -ésimo item;
- \bar{T}_A é o escore médio dos respondentes que acertaram o i -ésimo item;
- \bar{T} é o escore médio dos respondentes;
- p_i é a proporção de respondentes que acertaram o i -ésimo item;
- S_T é o desvio padrão dos escores.

2.3 Teoria da Resposta ao Item

As técnicas utilizadas na TRI visam descobrir a relação entre uma característica não observável do respondente (traço latente, conhecimento) e o item respondido. Esse modelo apresenta processos, conceitos e resultados que ajudam a comparar os indivíduos e seus níveis de proficiência, por exemplo. Então, este estudo será iniciado com uma abordagem dicotômica dos itens e com melhorias nos ajustes ao longo do processo.

2.3.1 Curva Característica do Item

A Curva Característica do Item é uma ótima forma de se relacionar o nível de proficiência θ (habilidade) do respondente com a probabilidade de ele acertar o item. Um exemplo dessa curva é apresentado na Figura 2.

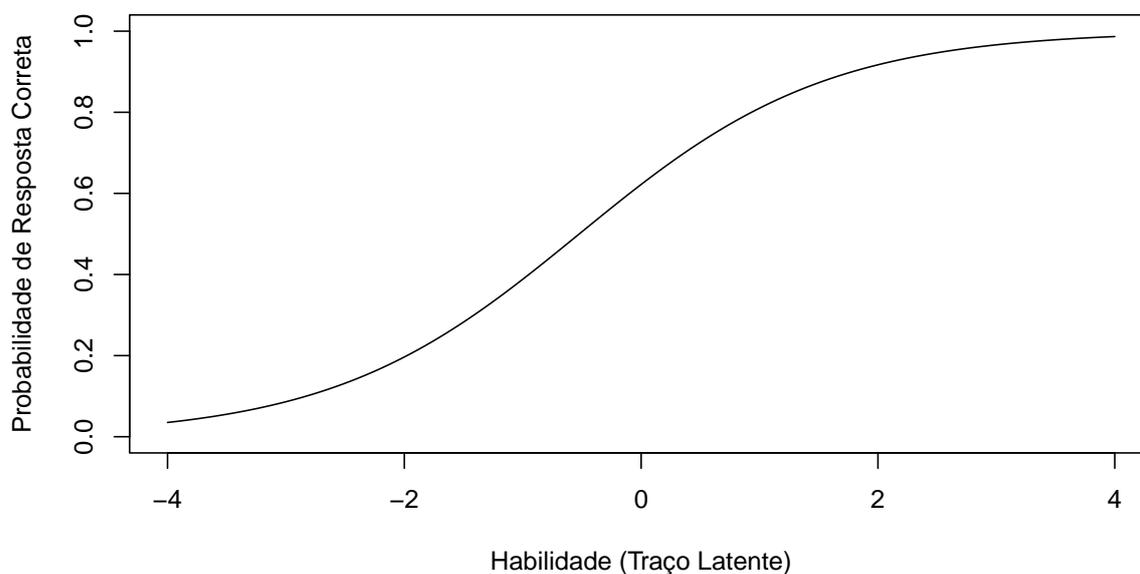


Figura 2: Curva Característica do Item

Percebe-se que, quanto maior é a habilidade do respondente, maior é a probabilidade de ele acertar o item, o que é o esperado nos modelos de TRI.

2.3.2 Modelos Logísticos para Itens Dicotômicos

Os Modelos Logísticos para Itens Dicotômicos devem ser os primeiros ajustes em um modelo de TRI, pois são essenciais para a análise a partir dos parâmetros que apresentam.

- Modelo Logístico de 2 Parâmetros

O ML2 é um modelo que apresenta os parâmetros de discriminação (a_i) e de dificuldade (b_i), sendo que o primeiro representa a inclinação da Curva Característica do Item no ponto b_i e diferencia um respondente com alto nível de proficiência de um com baixo nível, enquanto o segundo demonstra se o item é considerado fácil ou difícil. No espaço paramétrico, espera-se que o a_i seja maior que 0.5 e que nunca seja negativo, enquanto o b_i não tem um intervalo específico, mas valores altos indicam um item difícil e valores baixos um item mais fácil.

A equação 3.3.1 representa a probabilidade de um respondente j com nível de proficiência θ acertar um item i e é conhecida como Função de Resposta do Item (FRI).

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, i = 1, 2, \dots, I; j = 1, 2, \dots, n \quad (2.3.1)$$

- Modelo Logístico de 3 Parâmetros

Já para o ML3, há a adição de um novo parâmetro, a probabilidade de acerto ao acaso (c_i), que representa a probabilidade de um respondente com baixa proficiência acertar o item, em que, no espaço paramétrico, são apresentados valores entre 0 e 1 e espera-se valores baixos, dada a penalização realizada pelo CEBRASPE. A FRI do ML3 é dada por

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, i = 1, 2, \dots, I; j = 1, 2, \dots, n. \quad (2.3.2)$$

2.3.3 Estimativa dos Parâmetros

Assim como é descrito por Anjos e Andrade (2012), a estimação dos parâmetros e da habilidade deve ser feita pelo Método de Máxima Verossimilhança e otimizada por métodos iterativos. Essa estimação é de extrema importância e pode ser realizada no ML2 e no ML3.

2.3.4 Modelo de Resposta Gradual

O Modelo de Resposta Gradual (MRG), adequado para itens politômicos, ordena as respostas dos indivíduos, em primeiro lugar, como “errado”, “não-resposta” e “certo”. Todavia, como já fora mencionado neste trabalho, a banca Cebraspe penaliza o erro, então deve-se esperar que só responda a questão quem realmente acha que sabe, o que leva a uma nova ordenação, com “não-resposta” antes de “errado”.

Essa ordenação segue a lógica de indivíduo com menor nível de habilidade para maior e isso permite ao pesquisador saber mais sobre o respondente do que simplesmente se ele acertou ou não a questão. Então, tem-se que a probabilidade de um candidato j com proficiência θ escolher uma categoria k ou alguma outra superior para o item i é dada por

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k})}}, i = 1, 2, \dots, I; j = 1, 2, \dots, n; k = 1, 2, \dots, m. \quad (2.3.3)$$

3 Resultados

Neste capítulo, serão apresentados os resultados estudados e analisados a partir do Software R.

3.1 Análise Descritiva

Para uma análise preliminar, é necessária a utilização da Figura 3, que apresenta a distribuição das respostas por item.

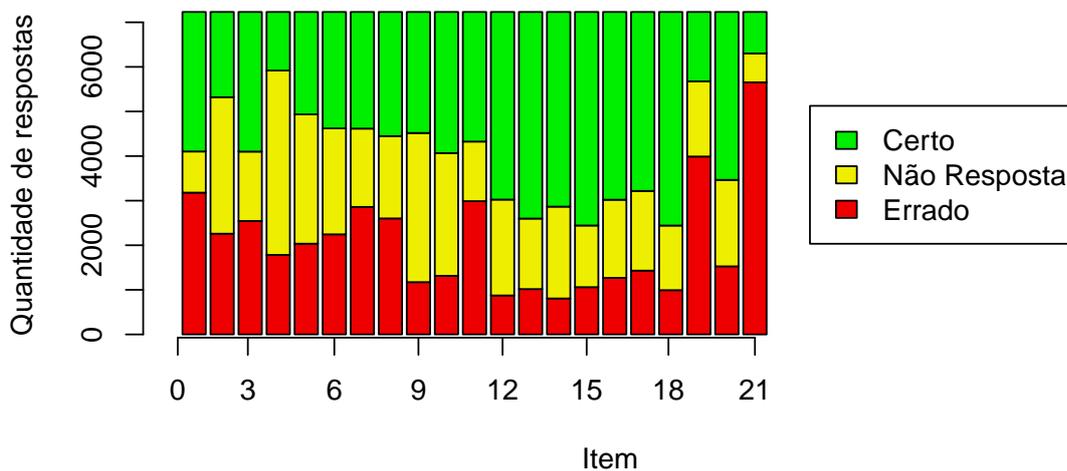


Figura 3: Distribuição das Respostas por Item

Nota-se que os itens com mais respostas erradas foram o 21, o 19 e o 1. Enquanto os itens com maior número de acertos foram o 18, o 15 e o 13. Já em relação à maior quantidade de não resposta, os itens que mais se destacaram foram o 4, o 9 e o 2.

A partir desses dados, pode-se ter uma noção da dificuldade dos itens, pois o esperado é que os itens com uma maior quantidade de respostas erradas ou não resposta sejam mais difíceis que os itens que apresentaram um maior número de acertos.

No Vestibular da UnB, o escore é calculado da seguinte forma: Para questões do tipo A, errado = -1, não resposta = 0, certo = 1 e para questões do tipo C, errado = -0.667, não resposta = 0, certo = 2. Então, a partir disso, é possível fazer os bloxpots da Figura 4.

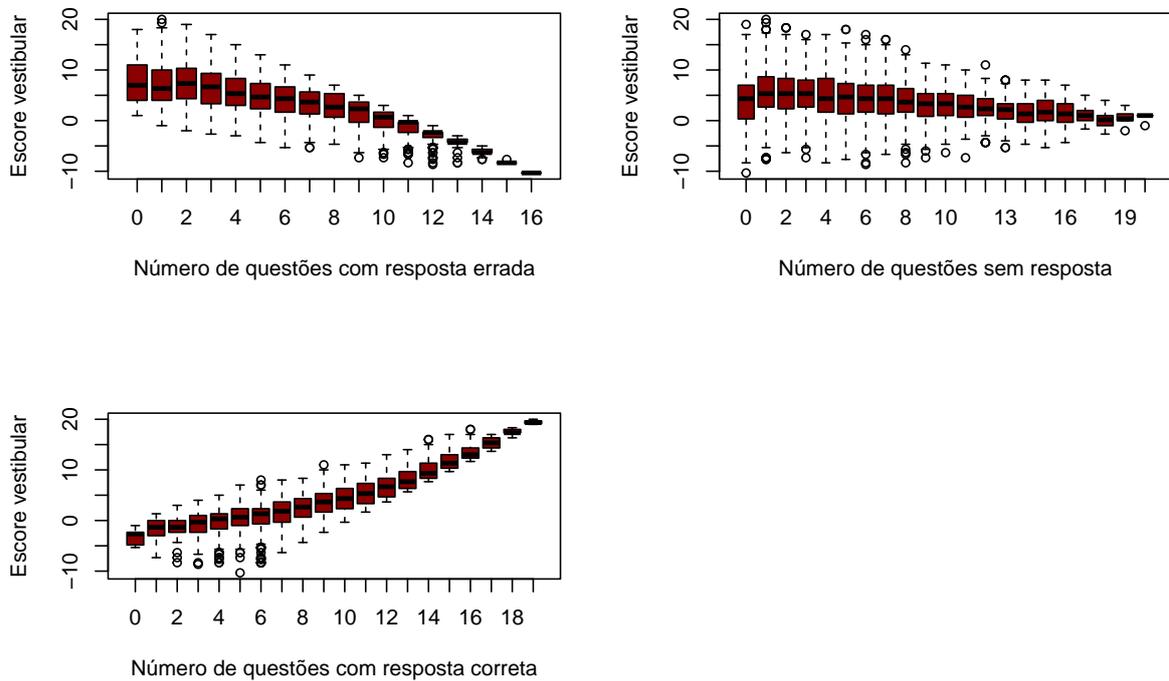


Figura 4: Boxplots Escore do Vestibular

A partir da análise dos bloxplots, é notória a relação dos escores escolhidos pela banca Cebraspe e o fato do respondente acertar, errar ou não responder os itens. E, além disso, percebe-se que, nos três casos, a distância entre o escore mínimo e o máximo é maior no começo, isto é, com menos itens analisados. Também é importante frisar que há um maior número de outliers nos maiores números de itens errados e nos menores números de itens acertados.

Agora, será realizada uma codificação dicotômica, na qual respostas erradas e não respostas receberão o valor 0 e respostas certas receberão o valor 1.

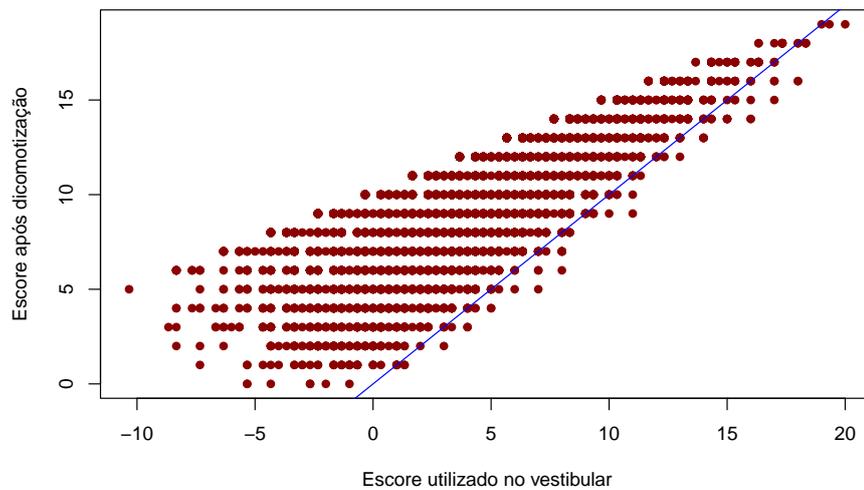


Figura 5: Comparação de Escores

Na Figura 5, há uma representação da comparação entre o escore do vestibular e o escore após a dicotomização e, a partir desse gráfico, nota-se que há uma variabilidade grande no escore utilizado no vestibular para candidatos com o mesmo escore dicotomizado. Isto é decorrência da penalização adotada no escore do vestibular.

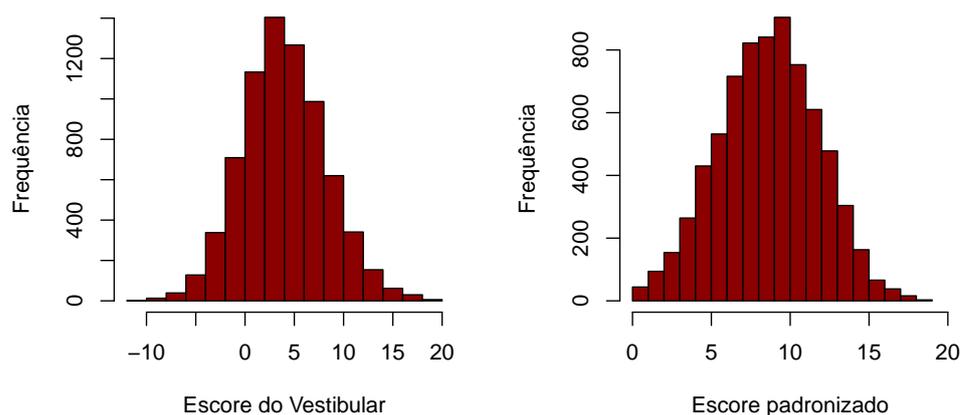


Figura 6: Histogramas de Distribuição dos Escores

Nos histogramas apresentados na Figura 6, nota-se que a distribuição do escore do vestibular e do escore padronizado (após dicotomização) são simétricas, o primeiro apresenta média em torno de 4 e valor máximo de 20, já o segundo tem média em torno de 9 e valor máximo de 19.

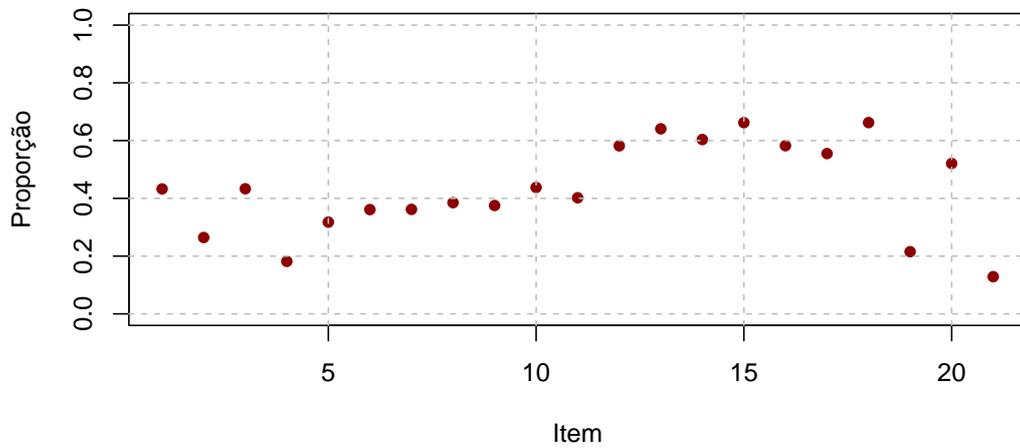
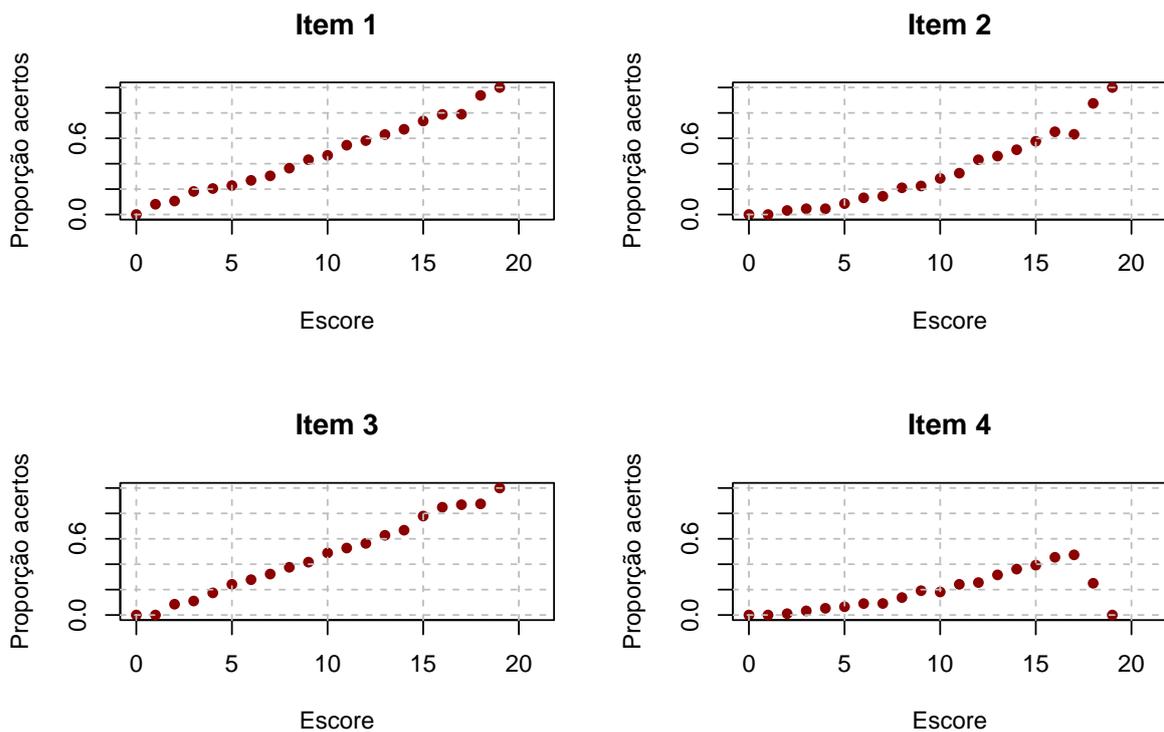
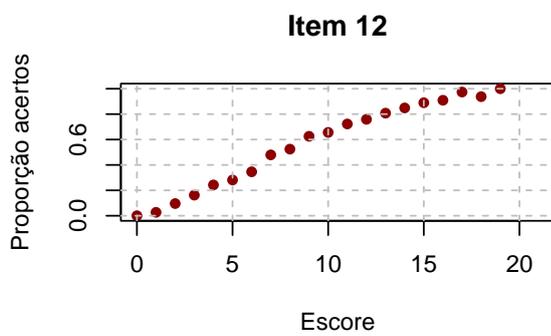
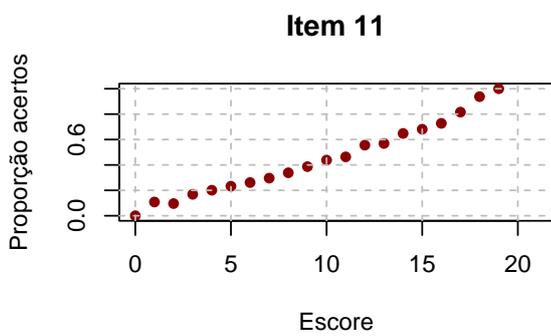
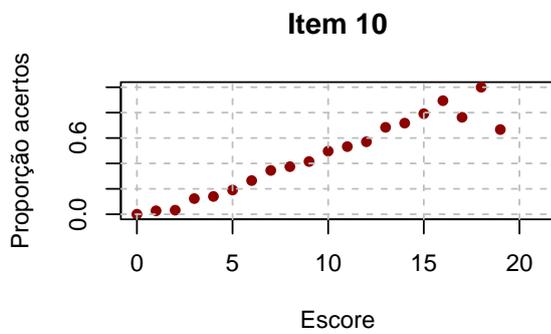
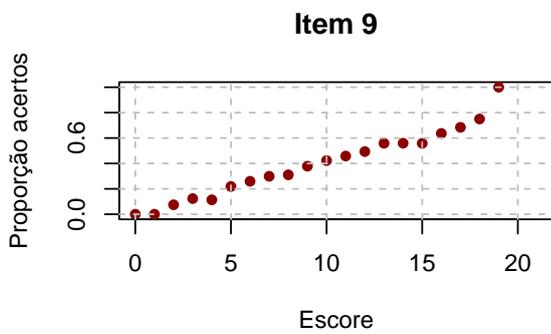
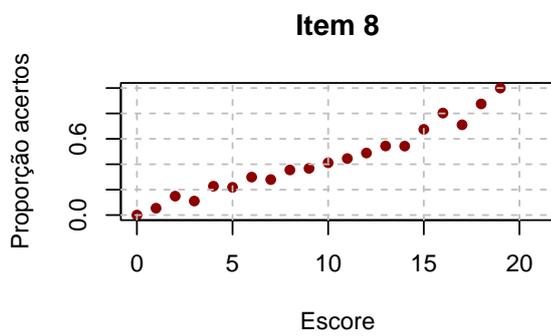
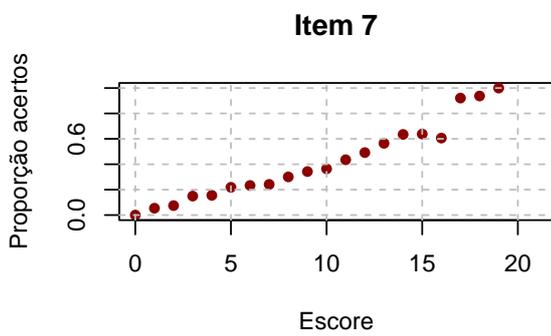
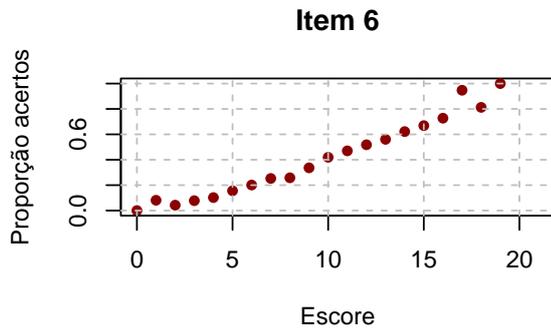
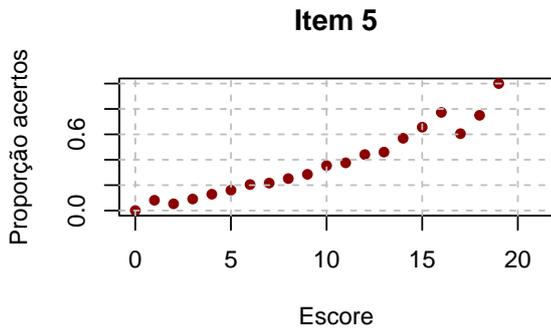
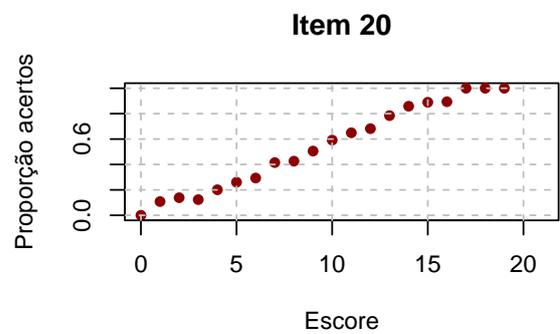
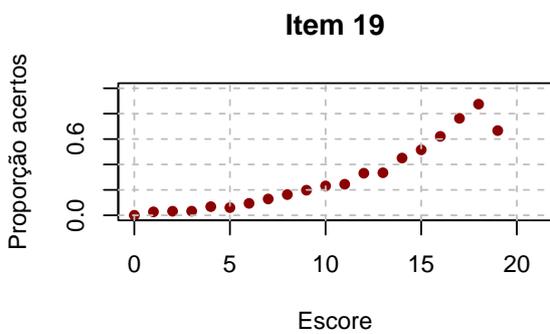
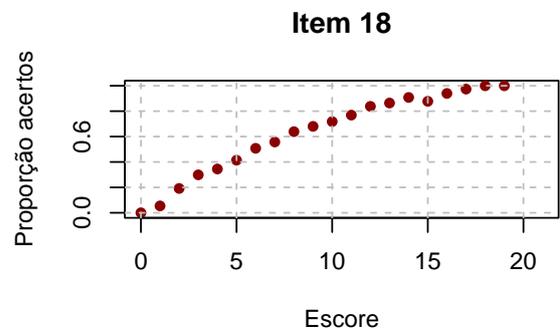
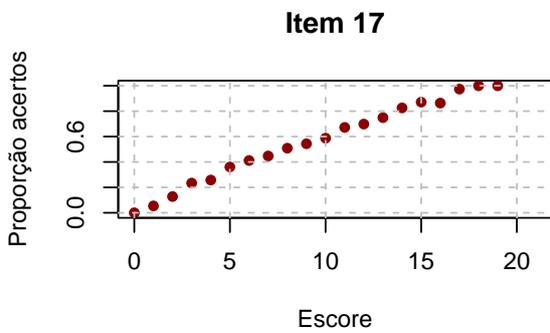
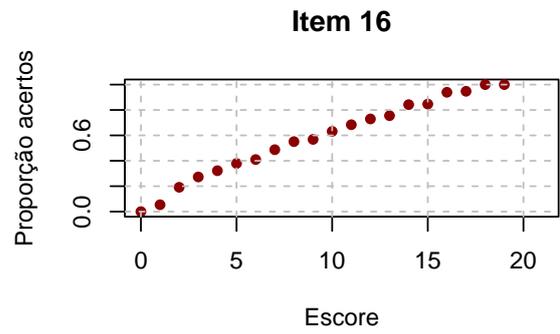
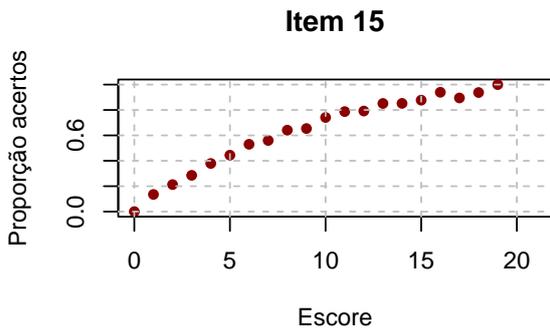
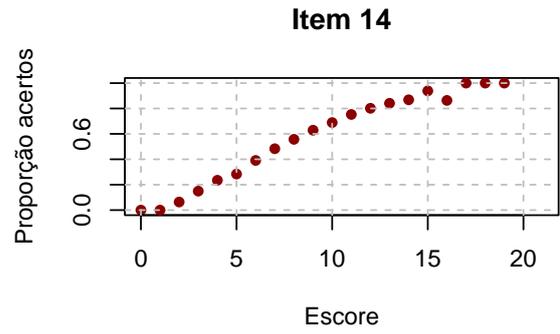
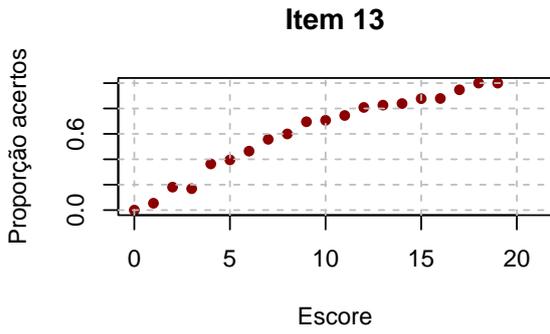


Figura 7: Proporção de Respostas Corretas por Item

A Figura 7, que representa a proporção de respostas corretas por item, mostra que a maior parte dos itens teve menos de 60% de acerto, os únicos que passaram disso foram os itens 13, 14, 15 e 18, enquanto o item 21 foi o que apresentou menor proporção de acerto, com aproximadamente 13%.







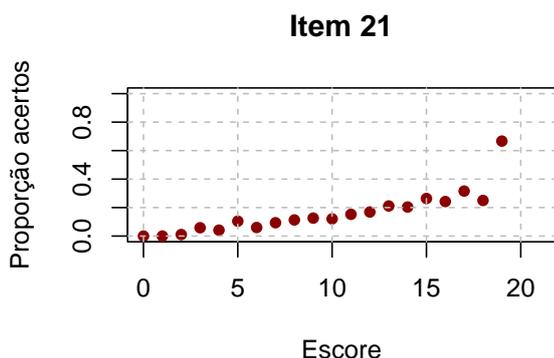


Figura 8: Curvas Características dos Itens (CCI)

A partir da análise das Curvas Características Itens (CCI) de TCT no formato não paramétrico, apresentadas na Figura 8, nota-se que a maioria dos itens segue a premissa de que, quanto maior o escore, maior é a proporção de acertos, porém, os itens 4 e 19 apresentam uma diminuição da proporção de acertos em escores muito altos, o que deve ser estudado (provavelmente, devido à baixa frequência de respondentes com alto escore, o que acarreta em alta oscilação na proporção de acertos para algum item entre respondentes com alto escore). Além disso, os itens 4 e 21 têm proporções de acerto bem baixas para todos os scores, o que dá indícios de serem itens com uma dificuldade maior.

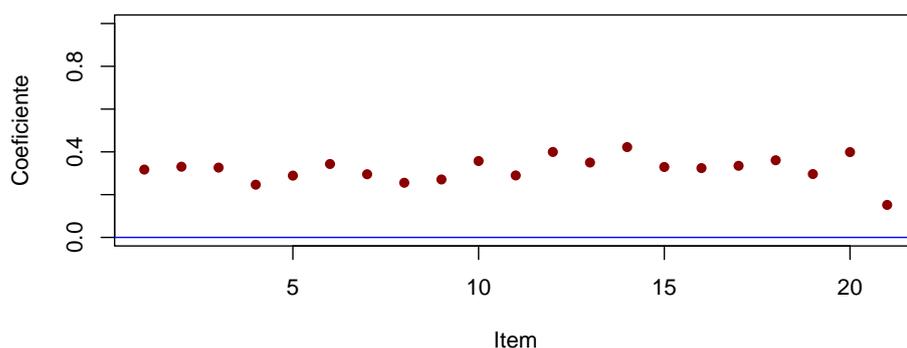


Figura 9: Coeficiente de Correlação Ponto-Bisserial

Ao analisar o Coeficiente de Correlação Ponto-Bisserial, apresentado na Figura 9, é perceptível que nenhum dos itens apresenta um coeficiente com valor menor que zero, o que é bom para a análise e mostra coerência. Todavia, é importante frisar que o item 21 é o que tem o menor valor para o coeficiente, que é, aproximadamente, 0.15.

3.2 Ajuste dos Modelos Logísticos

3.2.1 Modelo Logístico de 3 Parâmetros

Para se iniciar os ajustes deste Modelo, é necessária a análise dos seus parâmetros apresentados na Tabela 1.

Item	Acerto ao Acaso	Dificuldade	Discriminação
1	0,00	0,54	0,54
2	0,00	1,86	0,59
3	0,00	0,51	0,57
4	0,00	3,86	0,40
5	0,00	1,81	0,45
6	0,00	1,06	0,58
7	0,07	1,91	0,44
8	0,03	1,86	0,29
9	0,00	1,74	0,30
10	0,00	0,45	0,61
11	0,05	1,30	0,41
12	0,00	-0,45	0,84
13	0,00	-0,98	0,64
14	0,00	-0,52	0,95
15	0,00	-1,30	0,55
16	0,23	0,30	0,66
17	0,00	-0,42	0,53
18	0,00	-1,05	0,71
19	0,00	2,31	0,60
20	0,15	0,32	1,06
21	0,00	10,27	0,19

Tabela 1: Parâmetros do ML3

É possível notar que a maioria dos itens apresenta uma baixa probabilidade de acerto ao acaso, o que pode ser explicado pela penalização do CEBRASPE, porém o item 16 apresenta um valor relativamente alto (0,23). É notório também que, em geral, os itens apresentam baixo índice de discriminação, o que já é um argumento para a não adequação do ML3. Além disso, nota-se que os itens mais difíceis são os itens 4, 19 e 21, sendo este último o mais difícil, já os mais fáceis são os itens 13, 15 e 18, sendo o item 15 o mais fácil.

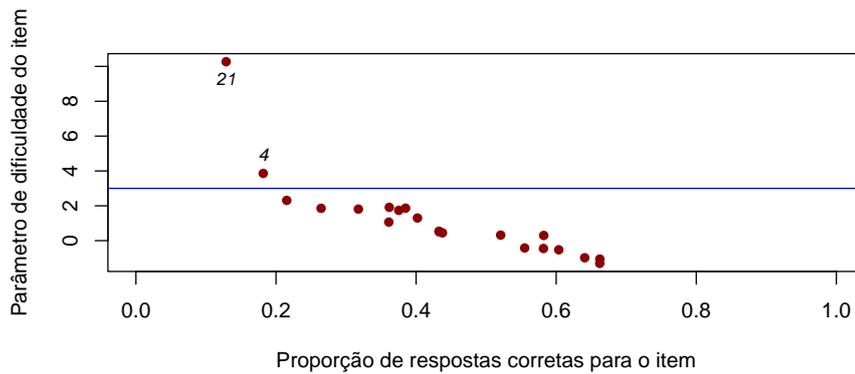


Figura 10: Dificuldade x Proporção de Acerto no ML3

Na Figura 10, é apresentado o parâmetro de dificuldade do item em comparação à proporção de respostas corretas para o item e, ao analisar este gráfico, é perceptível que os itens 4 e 21 destoaram dos outros, sendo os mais difíceis e com menor proporção de acerto, principalmente o item 21, que apresentou uma grande discrepância do resto. Por outro lado, como esperado, à medida que a dificuldade do item diminui, a proporção de acertos aumenta.

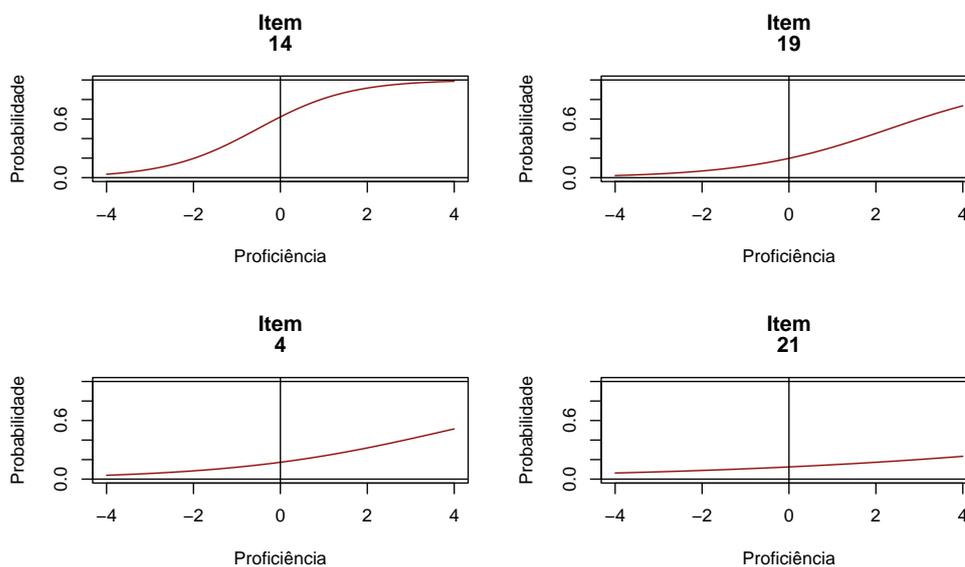


Figura 11: Curvas Características dos Itens no ML3

A Figura 11 mostra as Curvas Características de 4 itens, o item 14 seria o resultado ideal, enquanto os itens 19, 4 e 21 são os mais problemáticos, com alta dificuldade e menor índice de discriminação.

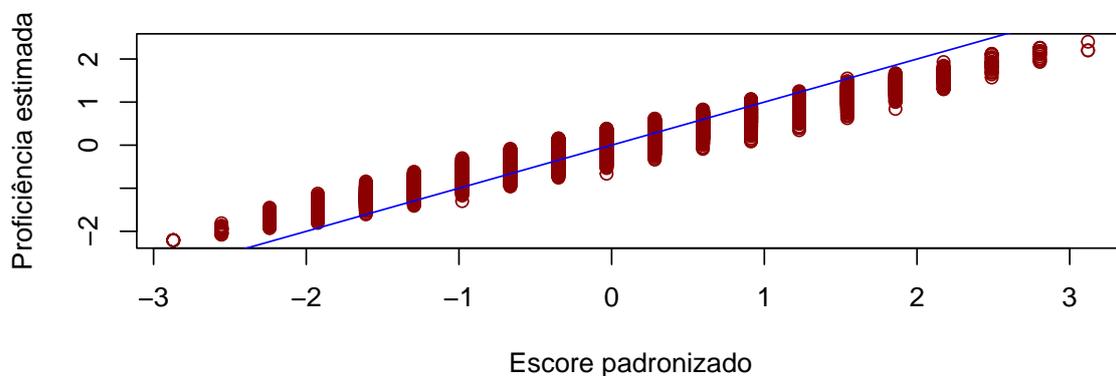


Figura 12: Proficiência estimada x Escore padronizado no ML3

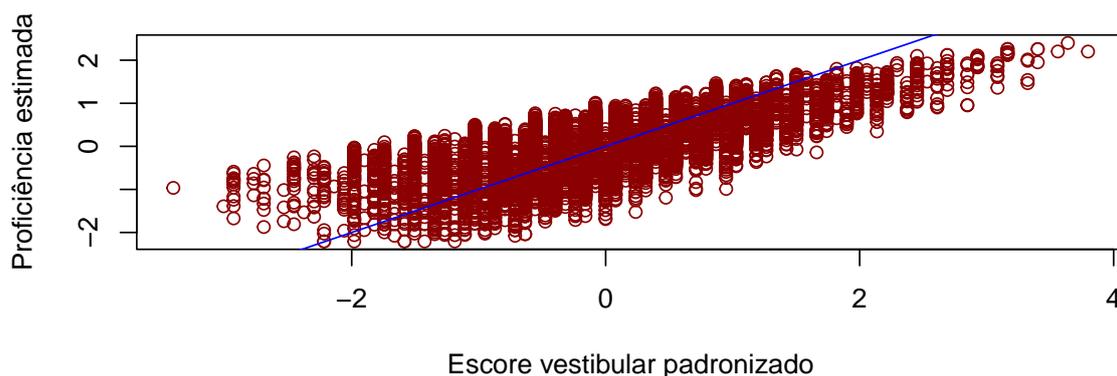


Figura 13: Proficiência estimada x Escore vestibular padronizado no ML3

Na Figura 12, é demonstrada a relação entre a proficiência estimada e o escore padronizado (ajuste usando TRI), enquanto na Figura 13 é a proficiência estimada com o escore do vestibular padronizado. É perceptível como o primeiro gráfico mostra uma correlação maior entre escore e proficiência do que no segundo, o coeficiente de correlação de Pearson calculado na primeira foi de 0.98, bem alto, enquanto no segundo foi de 0.75. Isso mostra uma coerência no ajuste do Modelo, porém, em razão dos outros fatores anteriormente apresentados, como baixo índice de discriminação e discrepâncias de dificuldade, ainda não se tem o melhor ajuste de modelo.

3.2.2 Modelo Logístico de 2 Parâmetros

Agora, será iniciado o ajuste para um Modelo Logístico de 2 Parâmetros, isto é, o parâmetro Acerto ao Acaso será nulo.

Item	Dificuldade	Discriminação
1	0,54	0,54
2	1,85	0,60
3	0,51	0,57
4	3,83	0,41
5	1,80	0,44
6	1,06	0,58
7	1,59	0,37
8	1,68	0,28
9	1,72	0,30
10	0,45	0,61
11	1,07	0,38
12	-0,45	0,84
13	-0,98	0,64
14	-0,52	0,95
15	-1,29	0,55
16	-0,71	0,50
17	-0,44	0,53
18	-1,05	0,71
19	2,33	0,59
20	-0,11	0,81
21	10,21	0,19

Tabela 2: Parâmetros do ML2

A partir da análise da Tabela 2, percebe-se uma pequena diferença em relação aos valores dos parâmetros da Tabela 1, isso porque a maioria dos itens tinha o parâmetro de Acerto ao Acaso igual ou próximo a zero no ML3.

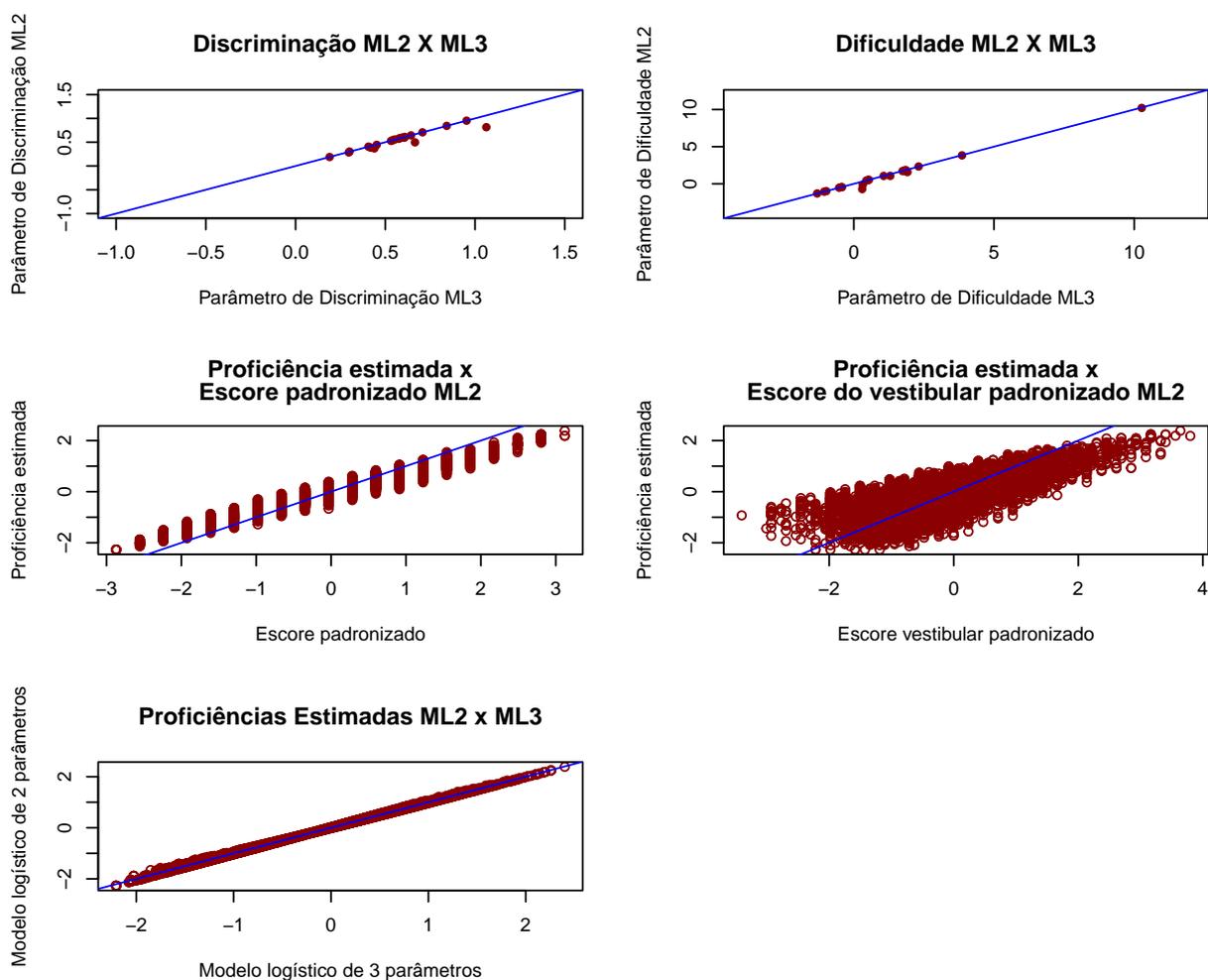


Figura 14: Comparações entre o ML2 e o ML3

Na Figura 14, há gráficos comparando os parâmetros dos Modelos Logísticos de 2 e de 3 parâmetros. Nos dois primeiros gráficos, é possível notar que tanto o índice de discriminação quanto o parâmetro de dificuldade são bastante semelhantes nos dois ajustes de modelo. Logo abaixo, são apresentados os gráficos que comparam a proficiência estimada com o escore padronizado e com o escore do vestibular padronizado e, assim como no ML3, no ML2 o escore padronizado usando TRI é melhor, com maior correlação. Por fim, no último gráfico da figura, é possível notar uma grande correlação e semelhança entre as proficiências estimadas nos dois modelos.

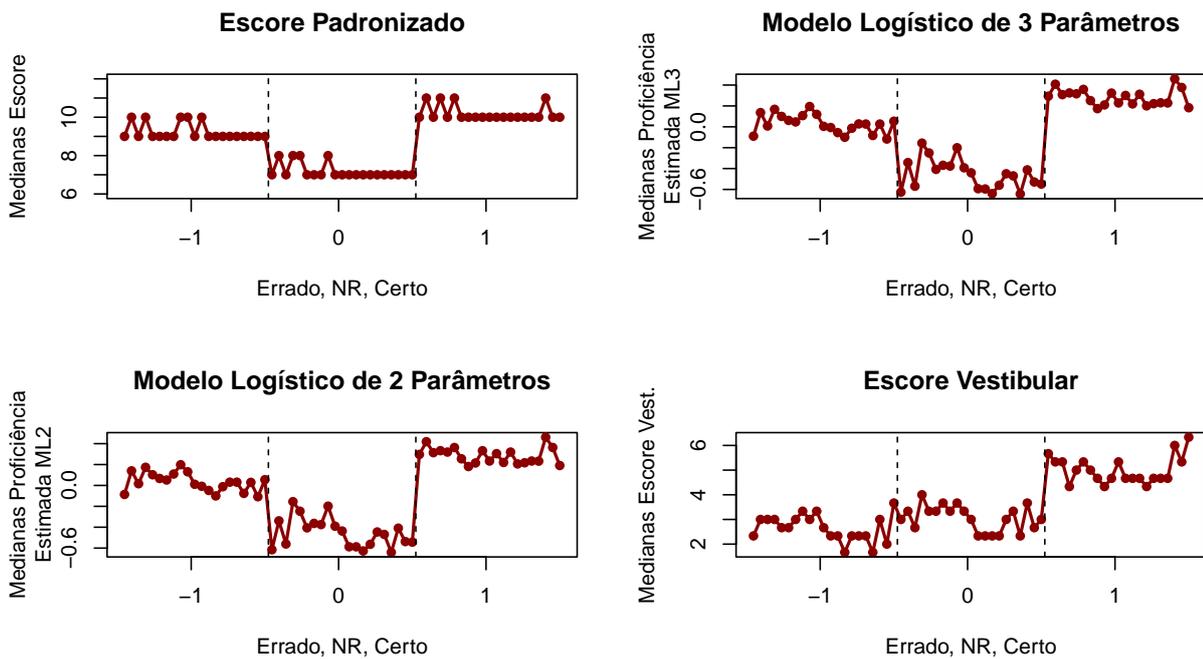


Figura 15: Comparação das Medianas

A Figura 15 mostra as medianas dos escores e das proficiências estimadas dos respondentes analisadas por três grupos: errado (-1), não resposta (0) e certo (1) para todos os itens da prova de geografia. Os três primeiros gráficos (escore padronizado, proficiência estimada no ML3 e no ML2) são gerados a partir de ajustes de TRI, enquanto o quarto gráfico é relacionado ao escore do Vestibular.

É perceptível, nos três primeiros gráficos, que respondentes que erram a questão tendem a ter maior número de respostas corretas na prova do que os que não respondem, enquanto o escore do vestibular esconde este fato, o que mostra que uma possibilidade para considerar respostas categorizadas seria com a ordenação das categorias como: “não resposta”, “errado”, “certo”.

3.3 Ajuste do Modelo de Resposta Gradual

Para iniciar o ajuste do Modelo de Resposta Gradual, será considerada a ordem “errado”, “não resposta” e “certo”.

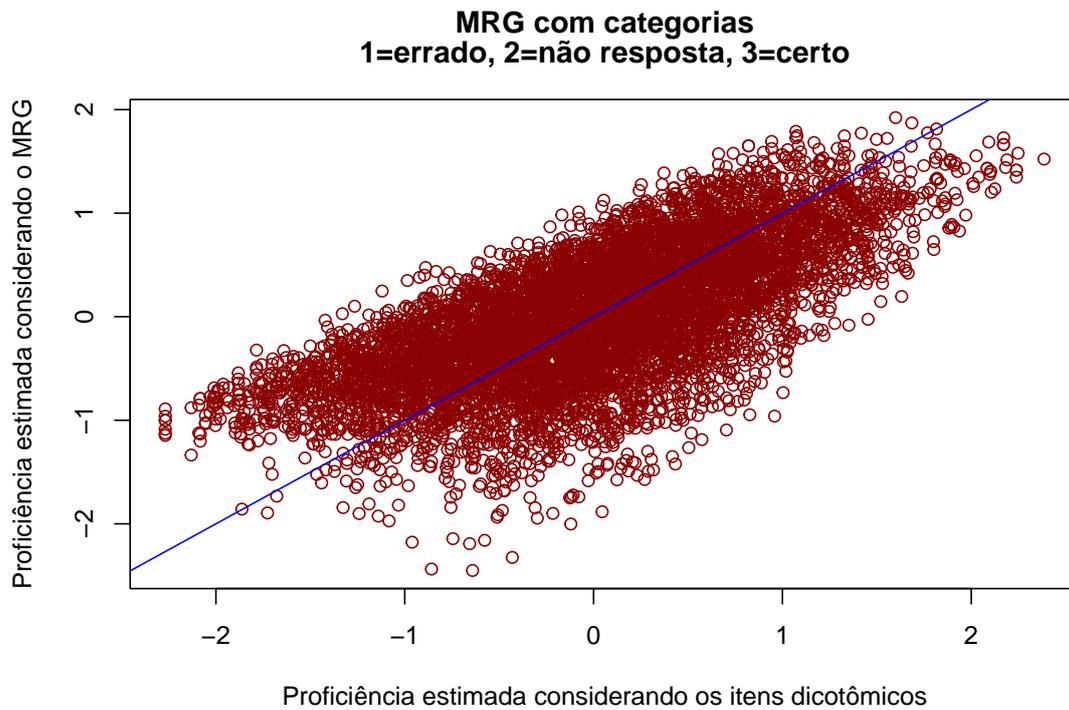


Figura 16: Proficiência Estimada MRG x Dicotômicos (errado, NR, certo)

Com a ajuda da Figura 16, pode-se perceber que a proficiência estimada considerando o Modelo de Resposta Gradual na ordem “errado”, “não resposta” e “certo” não está muito correlacionada com a proficiência estimada considerando os itens dicotômicos.

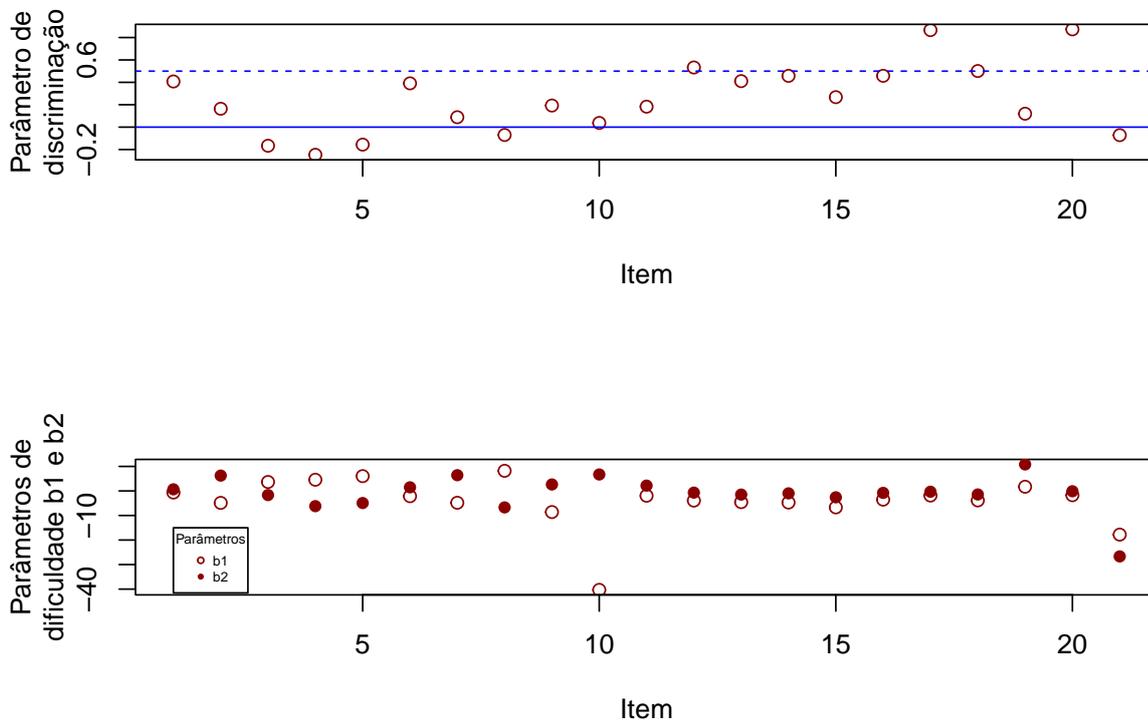


Figura 17: Discriminação e Dificuldade MRG (errado, NR, certo)

Os gráficos da Figura 17 mostram valores absurdos para o parâmetro de dificuldade e valores negativos para o parâmetro de discriminação, além de valores também absurdos para b_1 e b_2 (b_1 não deveria ser superior ao b_2), mostrando que o MRG com as categorias “1 = errado, 2 = não resposta, 3 = certo” não está bem ajustado. Então, será realizado um novo ajuste com a ordem “não resposta”, “errado” e “certo”.

A possibilidade dessa ordenação diferente se ajustar melhor é uma descoberta muito importante, pois pode mudar a forma que se analisam as respostas das provas do Vestibular da UnB e de outros processos seletivos, o que mostra o quão importante é a Teoria da Resposta ao Item e sua análise com o MRG.

Então, para iniciar essa análise com nova ordenação, é preciso fazer os gráficos que comparam a proficiência estimada neste Modelo de Resposta Gradual com a proficiência estimada nos Modelos Logísticos de 2 e 3 parâmetros.

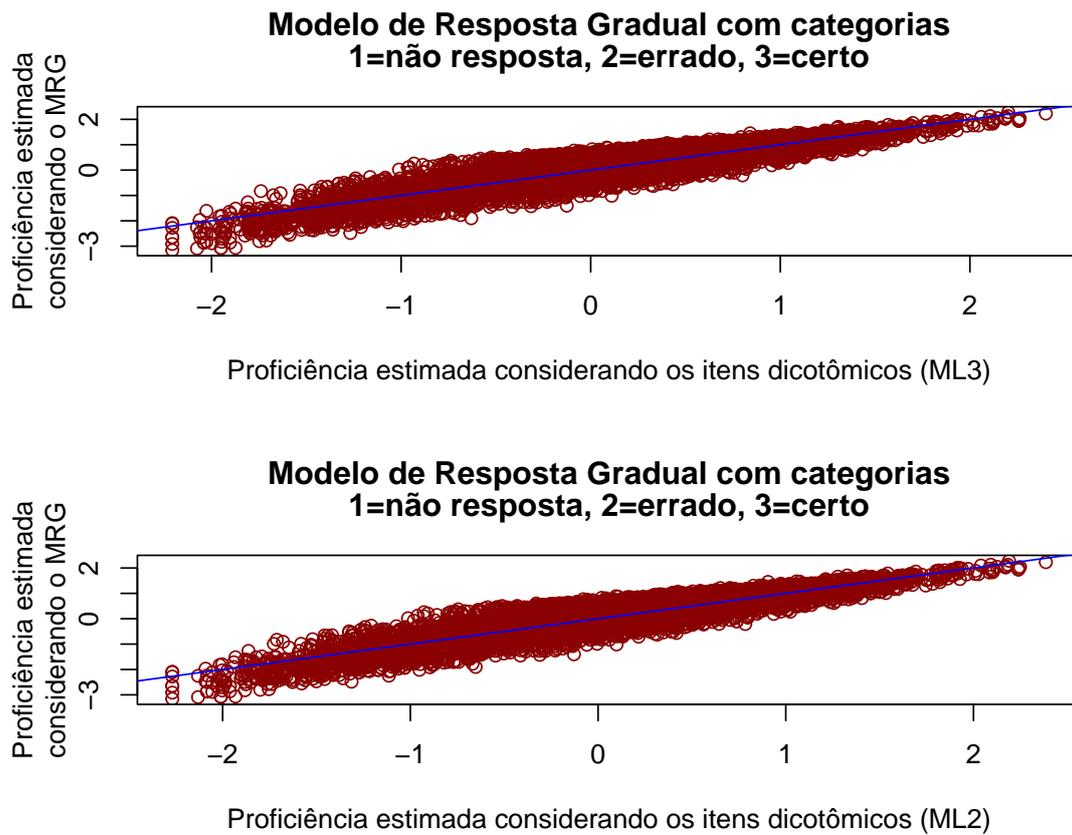
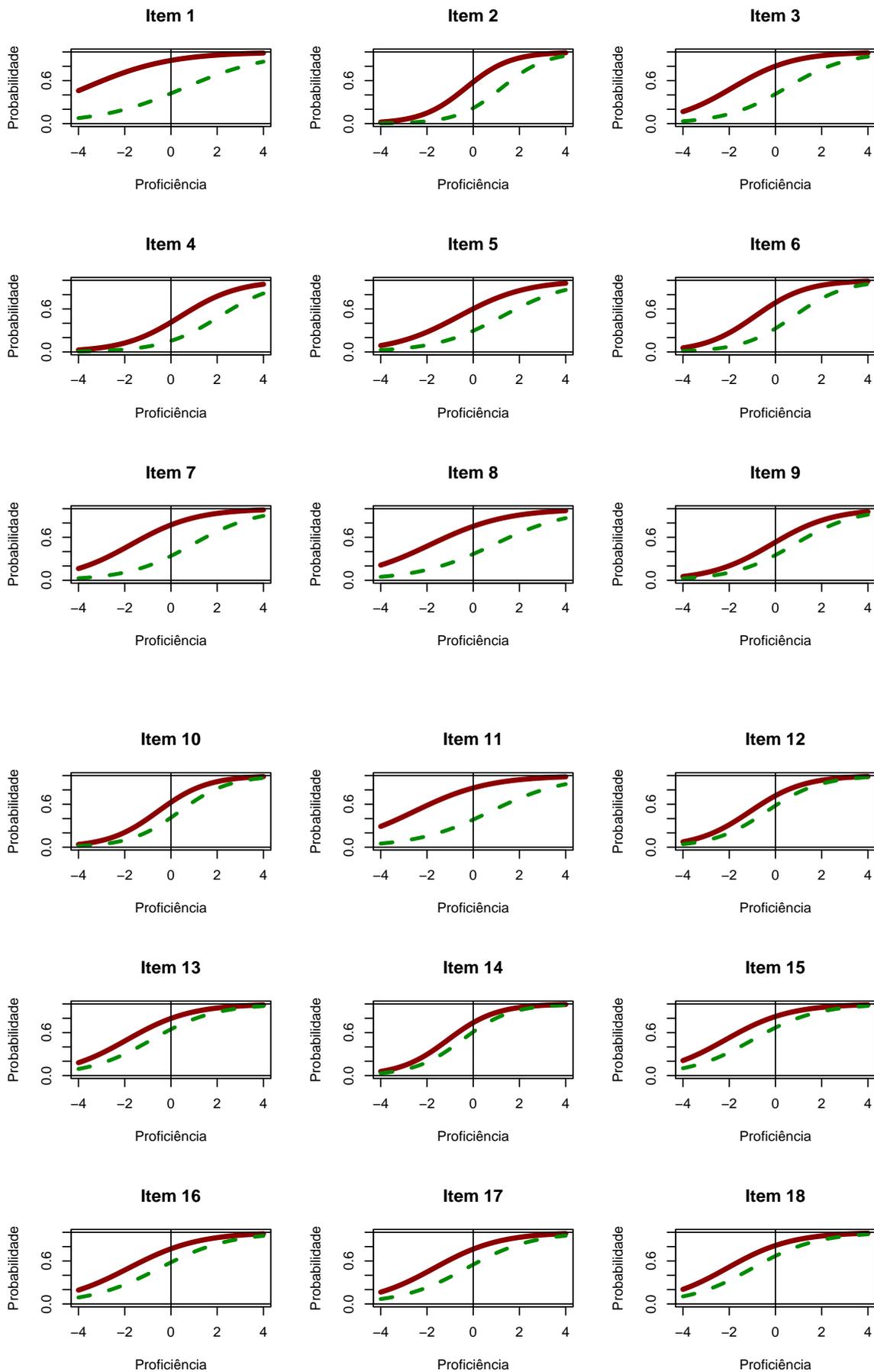


Figura 18: Proficiência Estimada MRG x Dicotômicos (NR, errado, certo)

Ao analisar a Figura 18, nota-se uma grande correlação entre as proficiências estimadas no MRG e com os itens dicotômicos, o que mostra que essa ordenação pode se ajustar melhor, todavia é necessário estudar mais este modelo.



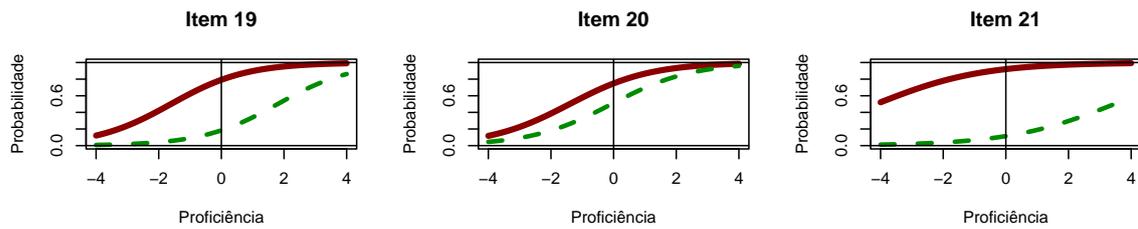


Figura 19: Curvas Características do MRG (NR, errado, certo)

Ao analisar a curva característica de cada item no novo ajuste (apresentadas na Figura 19), em que as curvas vermelha e verde estão relacionadas, respectivamente, aos parâmetros de dificuldade b_1 e b_2 , é notório como este modelo se ajustou melhor do que a ordenação anterior, porém também é importante frisar que os itens 1 e 21 aparentam ter um menor poder de discriminação em relação aos outros.

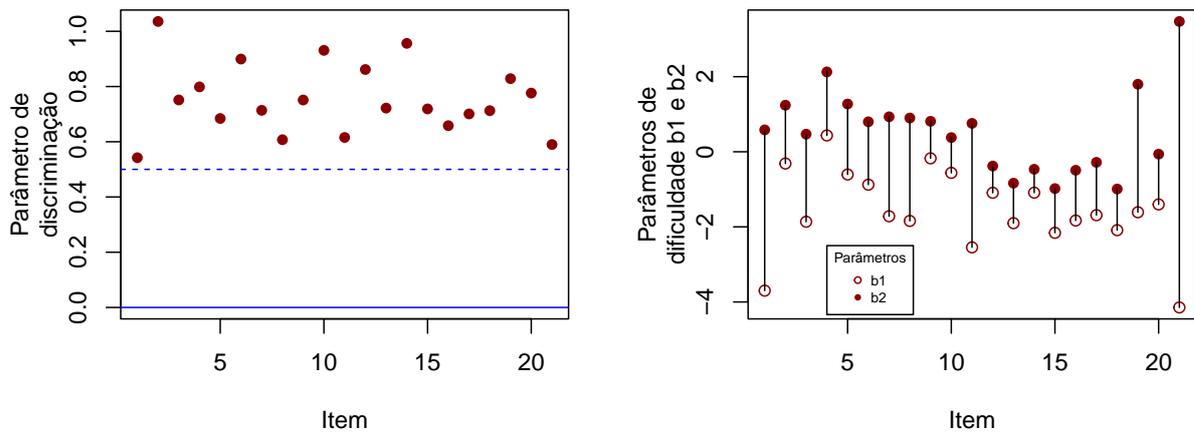


Figura 20: Discriminação e Dificuldade (NR, errado, certo)

A Figura 20 mostra que todos os itens apresentam valores acima de 0,5 para o parâmetro de discriminação e estimativas razoáveis para os parâmetros b_1 e b_2 , exceto o item 21, que apresenta uma diferença grande entre as estimativas de b_1 (menor que -3) e b_2 (maior que 3), o que indica que a probabilidade de escolha da categoria intermediária é alta, que seria errar a questão. Nota-se também que, com b_1 abaixo de -3, há uma baixa probabilidade de escolha da categoria “não resposta” no item 1.

4 Conclusão

Este estudo analisou os itens da prova de geografia do vestibular da UnB de 2014 (banca CEBRASPE) e, ao total, foram analisadas 21 questões e seus parâmetros de discriminação, dificuldade e probabilidade de acerto ao acaso.

Assim como fora esperado ao começo do estudo, os Modelos Logísticos para Itens Dicotômicos não obtiveram os melhores resultados, então, foi realizado o ajuste para Modelo de Resposta Gradual com a ordenação “errado”, “não resposta” e “certo”. Todavia, por causa da penalização dada pelo CEBRASPE às respostas erradas, o melhor modelo foi o MRG com ordenação “não resposta”, “errado” e “certo”.

Ao analisar este ajuste de modelo, percebeu-se que os respondentes com maior nível de proficiência (conhecimento) tendem a tentar responder a maior parte dos itens, o que aumenta a probabilidade de errarem mais questões. Enquanto os respondentes com menor nível de proficiência, quando estão com dúvidas em uma questão, tendem a não respondê-la, pois caso arrisquem, podem perder pontos por causa da penalização da banca.

Por fim, com o último ajuste do MRG, todos os itens apresentaram valores razoáveis de discriminação, o que é favorável, pois diferencia os respondentes com diferentes níveis de habilidade. Além disso, as estimativas dos parâmetros de dificuldade também foram satisfatórias, porém o item 21, que já apresentara outros problemas ao longo do projeto, apresentou uma grande diferença entre as estimativas de b_1 e b_2 e se mostrou um item mais difícil, pois há uma alta probabilidade de errá-lo.

Todavia, o item 21 é um dos dois itens de tipo C da prova de Geografia do Vestibular, então pode-se esperar que tenha maior probabilidade do respondente errá-lo em comparação a itens do tipo A, mas mesmo assim a dificuldade ainda está acima do esperado.

Portanto, mesmo que um item tenha uma probabilidade de erro alta, mas tenha um bom poder de discriminação, ele ainda pode ser útil para diferenciar entre os alunos de alto e baixo desempenho, o que é o caso do item 21, então, não é necessária a retirada de nenhum item da prova.

Referências

ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. *ABE, Sao Paulo*, 2000.

ANJOS, A.; ANDRADE, D. F. Teoria da resposta ao item com uso do r. *João Pessoa: ABE*, 2012.

ARAUJO, E. A. C. d.; ANDRADE, D. F. d.; BORTOLOTTI, S. L. V. Teoria da resposta ao item. *Revista da Escola de Enfermagem da USP, SciELO Brasil*, v. 43, p. 1000–1008, 2009.

SAMEJIMA, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*, 1969.