



**Universidade de Brasília  
Departamento de Estatística**

**Aplicação da Teoria de Resposta ao Item na Avaliação das Questões da  
Prova de História do Vestibular da UnB de 2014**

**Stefany Luene Boto Martins**

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

**Brasília  
2024**

**Stefany Luene Boto Martins**

**Aplicação da Teoria de Resposta ao Item na Avaliação das Questões da  
Prova de História do Vestibular da UnB de 2014**

Orientador: Prof. Antônio Eduardo Gomes

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

**Brasília  
2024**

---

Dedico este trabalho à minha mãe, Maria de Jesus, por sempre me apoiar nos estudos e na vida.

---

# Resumo

Este projeto tem como objetivo analisar as questões da Prova de História do Vestibular da UnB de 2024, da banca examinadora CEBRASPE, utilizando técnicas e ajustes de Teoria de Resposta ao Item. O presente estudo irá abordar detalhes de cada item da prova, como seus parâmetros de dificuldade e discriminação, assim como irá se aprofundar nos Modelos Logísticos de 2 e 3 parâmetros e nos ajustes do Modelo de Resposta Gradual, buscando definir qual é a melhor ordenação, sabendo que essa prova possui penalização em caso de resposta errada, assim analisando o impacto dessa penalização.

O software R é utilizado para obter e analisar os resultados deste projeto, assim como para criar gráficos e tabelas dos dados, permitindo uma visualização clara e objetiva, auxiliando na interpretação dos resultados.

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

## **Lista de Tabelas**

1	Parâmetros do ML3 . . . . .	23
2	Parâmetros do ML2 . . . . .	27

## **Lista de Figuras**

1	Elementos da prova do vestibular 2014/2. . . . .	9
2	Gráfico da Curva Característica do Item (CCI) . . . . .	11
3	Distribuição das Respostas por Item. . . . .	14
4	Boxplots dos Escores no Vestibular. . . . .	15
5	Comparação dos Escores no Vestibular. . . . .	16
6	Distribuição dos Escores. . . . .	17
7	Proporção de Acertos por Item. . . . .	17
8	Curva Característica de cada item (CCI). . . . .	21
9	Coefficiente de Correlação Ponto-Bisserial por Item. . . . .	22
10	Proporção de respostas corretas no ML3 por parâmetro de dificuldade. . .	24
11	Curva Característica dos Itens 2, 3, 18 e 25 no ML3. . . . .	25
12	Proficiência Estimada por Escore Padronizado no ML3. . . . .	25
13	Proficiência Estimada por Escore Vestibular Padronizado no ML3. . . . .	26
14	Comparativo entre os modelos ML2 e ML3. . . . .	28
15	Medianas do Escore, Proficiência Estimada do ML3, Proficiência Estimada do ML2 e Escore do Vestibular por Item. . . . .	30
16	Proficiência Estimada MRG na ordem errado, NR, certo x Dicotômicos . .	31
17	Discriminação e Dificuldade MRG na ordem errado, NR, certo. . . . .	32
18	Proficiência Estimada MRG na ordem NR, errado e certo x Dicotômicos. .	33
19	Curva Característica de cada item para o MRG na ordem NR, errado e certo.	35
20	Discriminação e Dificuldade b1 e b2 por Item para o MRG (NR, errado, certo). . . . .	36

# Sumário

<b>1 Introdução</b> . . . . .	8
<b>2 Metodologia</b> . . . . .	9
2.1 Conjunto de dados . . . . .	9
2.2 Coeficiente de Correlação Ponto-Bisserial . . . . .	10
2.3 Teoria de Resposta ao Item . . . . .	10
2.3.1 Curva Característica do Item . . . . .	11
2.3.2 Modelos Logísticos . . . . .	11
2.3.3 Modelo de Resposta Gradual . . . . .	12
2.3.4 Estimação dos Parâmetros . . . . .	13
<b>3 Resultados</b> . . . . .	14
3.1 Análise Descritiva . . . . .	14
3.2 Ajuste dos Modelos Logísticos . . . . .	22
3.2.1 Modelo Logístico de 3 Parâmetros . . . . .	22
3.2.2 Modelo Logístico de 2 Parâmetros . . . . .	26
3.3 Ajuste do Modelo de Resposta Gradual . . . . .	30
<b>4 Conclusão</b> . . . . .	37
<b>Referências</b> . . . . .	38

# 1 Introdução

Uma prática comum em grande parte do mundo é a aplicação de provas para testar o conhecimento de uma pessoa em relação a determinado assunto. Essa é uma ação aplicada desde o ensino fundamental até o ensino superior.

As avaliações têm como objetivo medir o conhecimento, sendo essa uma característica que não é possível medir diretamente, chamada de traço latente. Uma forma de medir um traço latente é aplicando técnicas da Teoria de Resposta ao Item (TRI), sendo essa uma evolução da Teoria Clássica dos Testes (TCT).

Na TCT, é analisado o teste como um todo. Por outro lado, na TRI cada item é analisado, sendo que, de acordo Andrade, Tavares e Valle (2000), quanto maior o conhecimento ou a habilidade do respondente, maior a probabilidade do mesmo acertar o item.

Neste estudo, serão analisadas as questões da prova de história do vestibular da UnB de 2014, sendo todos esses itens do tipo A (certo ou errado). No entanto, este vestibular possui uma penalidade caso o respondente erre a questão: em caso de acerto, o indivíduo ganha 1 ponto; em caso de erro, perde 1 ponto; e se o mesmo optar por não responder o item, ele não ganha, nem perde ponto.

O intuito deste projeto é analisar as questões do vestibular, buscando identificar os itens com baixo poder de discriminação entre indivíduos com baixo e alto nível de conhecimento e assim conseguindo medir melhor as habilidades de cada indivíduo, bem como estimar a proficiência (conhecimento) dos respondentes.

Então, o objetivo geral deste projeto é realizar um estudo a partir das respostas dos itens da prova de História do vestibular da UnB de 2014 via TRI. Os objetivos específicos são: utilizar técnicas de TRI, realizar ajustes nos modelos logísticos de dois e três parâmetros, analisar a curva característica de cada questão, ajustar o Modelo de Resposta Gradual e estimar os parâmetros dos modelos de TRI.

No capítulo 2 apresentaremos a metodologia do estudo, onde será comentado também sobre o conjunto de dados utilizado, no capítulo 3 será mostrado os resultados do projeto, e, por fim, no capítulo 4 será apresentado a conclusão do estudo.



## 2 Metodologia

O desenvolvimento deste projeto começa com uma análise descritiva dos dados. Em seguida é utilizada a Teoria Clássica dos Teste (TCT), que foca no teste como um todo, e para um melhor estudo, é utilizada a Teoria de Resposta ao Item (TRI), em que é analisado cada item do teste de forma individual.

Em TRI, serão ajustados os modelos de três parâmetros (ML3) e em seguida o modelo de dois parâmetros (ML2).

Depois será utilizado o Modelo de Resposta Gradual (MRG) para analisar a melhor forma de ordenação das categorias dos itens.

Para a parte de programação deste estudo foi utilizado o software R.

Todas as técnicas e modelos citados serão detalhados a seguir.

### 2.1 Conjunto de dados

O banco de dados utilizado neste trabalho é de dados reais do vestibular de 2014 da UnB, disponibilizado pelo CEBRASPE (Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos). As questões são aplicadas em dois dias de prova como mostrado na Figura 1:

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: Elementos da prova do vestibular 2014/2.

O vestibular contém um total de 300 itens, nos dois dias de prova. No entanto, o interesse neste estudo é a prova de história que é composta por 30 questões, sendo todas do tipo A (certo/errado) e serão analisadas as respostas de 7.232 indivíduos.

## 2.2 Coeficiente de Correlação Ponto-Bisserial

O coeficiente de correlação ponto-bisserial ( $\rho_i^{PB}$ ) é calculado através do coeficiente de correlação de Pearson, obtendo a correlação linear entre os itens dicotômicos e os escores dos respondentes, sendo calculado para cada item. Prevê-se que os itens apresentem um resultado positivo para este coeficiente, assim indicando que o item é coerente para o teste; em caso de resultado negativo ou muito próximo de zero, o item deve ser excluído ou reformulado.

O valor deste coeficiente é obtido através da fórmula

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

em que,

- $U_{ij}$  é a resposta do  $j$ -ésimo respondente para o  $i$ -ésimo item, sendo 0 para resposta errada e 1 para resposta correta;
- $n$  é o número de participantes;
- $n_i$  é o número de participantes que acertaram o  $i$ -ésimo item;
- $p_i$  é a proporção de participantes que acertaram o  $i$ -ésimo item;
- $\bar{T}_j$  é o escore do  $j$ -ésimo respondente;
- $\bar{T}_A$  é o escore médio dos participantes que acertaram o  $i$ -ésimo item;
- $\bar{T}$  é o escore médio dos respondentes;
- $S_T$  é o desvio-padrão dos escores.

## 2.3 Teoria de Resposta ao Item

A Teoria de Resposta ao Item (TRI) busca medir o traço latente (habilidade) dos indivíduos através de questões. Assim, serão feitas análises considerando os itens do estudo como dicotômicos, isto é, itens com respostas binárias (certo ou errado) e no decorrer do desenvolvimento do estudo serão consideradas outras possibilidades.

### 2.3.1 Curva Característica do Item

A curva característica do item (CCI) mostra a relação entre a proficiência do respondente ( $\theta$ ) e a probabilidade do mesmo acertar o item. Na Figura abaixo é possível ver um exemplo desta curva:

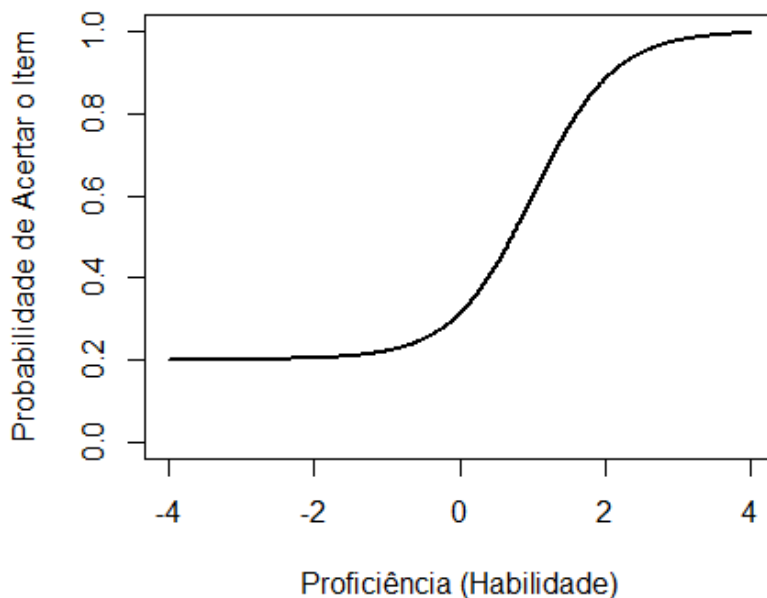


Figura 2: Gráfico da Curva Característica do Item (CCI)

### 2.3.2 Modelos Logísticos

Os modelos logísticos de dois parâmetros - ML2 (discriminação e dificuldade) e de três parâmetros - ML3 (discriminação, dificuldade e acerto ao acaso) serão utilizados neste trabalho.

O modelo logístico de dois parâmetros possui o parâmetro de discriminação do item ( $a_i$ ), que corresponde à inclinação da curva característica do item (CCI) e é este parâmetro que avalia o poder do item em separar os respondentes com proficiência alta dos com proficiência baixa em relação ao parâmetro de dificuldade ( $b_i$ ), sendo este o outro parâmetro deste modelo. O parâmetro de dificuldade avalia o quão “difícil” é um item.

A sua Função de Resposta do Item (FRI), que corresponde a probabilidade do  $j$ -ésimo respondente acertar o  $i$ -ésimo item, dado um valor para o traço latente  $\theta_j$ , é

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, \quad (2.3.1)$$

com  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, n$ , onde,

- $P(U_{ij} = 1|\theta_j)$  é a probabilidade do  $j$ -ésimo respondente acertar o item  $i$ , dado sua proficiência  $\theta_j$
- $U_{ij} = 1$  se o  $j$ -ésimo respondente acerta o item  $i$ ;
- $\theta_j$  é a proficiência do  $j$ -ésimo respondente;
- $a_i$  é o parâmetro de discriminação do item  $i$ ;
- $b_i$  é o parâmetro de dificuldade do item  $i$ .

No modelo logístico de três parâmetros é acrescentado o parâmetro de acerto ao acaso ( $c_i$ ), que é a probabilidade de um candidato responder corretamente o item  $i$  ao acaso mesmo tendo uma habilidade ( $\theta_j$ ) baixa.

A sua Função de Resposta ao Item (FRI) é dada por

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, \quad (2.3.2)$$

com  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, n$ .

Os dois modelos serão testados, porque, em outras provas, a probabilidade de acerto ao acaso ficou próxima de zero em função da penalização.

### 2.3.3 Modelo de Resposta Gradual

O modelo de resposta gradual (MRG) foi proposto por Samejima (1969), com o objetivo de analisar as respostas dos indivíduos além de se acertaram ou não o item, podendo classificar os itens através do seu grau de importância. Este modelo é utilizado quando os itens não são dicotômicos, como no presente estudo em que as questões possuem três opções (certo, errado e não resposta). Como visto anteriormente que o vestibular da UnB possui penalidade para resposta errada, o indivíduo pode optar por não responder o item por não saber ou por receio, então para este modelo, uma das opções de ordenação das categorias seria “não resposta” como uma opção intermediária.

A fórmula do MRG é dada por

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k})}}, \quad (2.3.3)$$

com  $i = 1, 2, \dots, I$ ;  $j = 1, 2, \dots, n$  e  $k = 1, 2, \dots, m$ , onde,

- $P_{i,k}^+(\theta_j)$  é a probabilidade do indivíduo escolher a categoria  $k + 1$  ou superior, dado sua proficiência  $\theta_j$
- $b_{i,k}$  é o parâmetro de dificuldade da categoria  $k$  do item  $i$ .

#### 2.3.4 Estimação dos Parâmetros

Nos modelos de TRI, a estimação dos parâmetros das questões e das proficiências dos respondentes é obtida através do Método Da Máxima Verossimilhança, realizando processos iterativos, como pelo método Newton-Raphson ou “Scoring” de Fisher, como mostra Anjos e Andrade (2012) e Bastos (2024).

### 3 Resultados

Os resultados apresentados a seguir foram obtidos a partir do Software R.

#### 3.1 Análise Descritiva

Para começar a análise descritiva, a figura 3 mostra a distribuição das respostas por item.

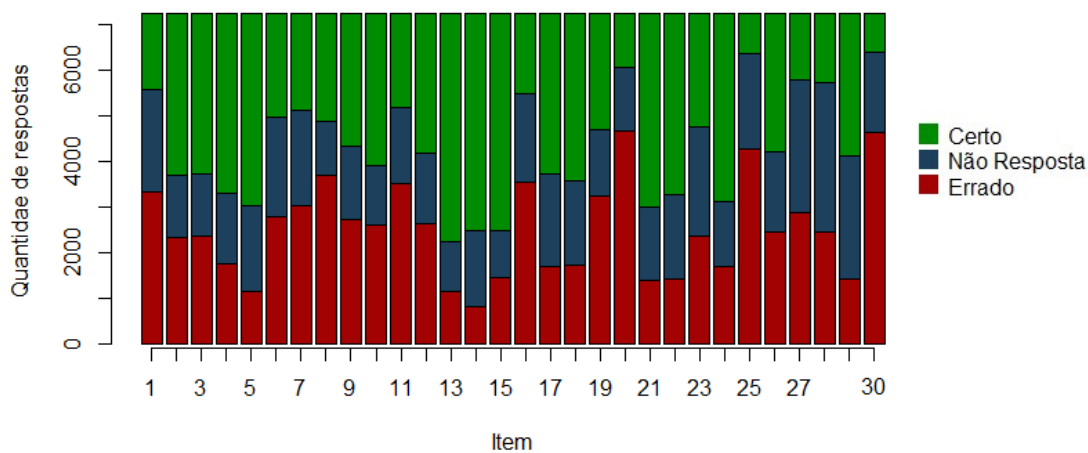


Figura 3: Distribuição das Respostas por Item.

Observa-se pela Figura 3, que os itens em que mais os respondentes erraram foram o 20, com 4679 respostas erradas, seguido dos itens 30 e 25. Já o item com mais respostas corretas foi o 13 (4992 acertos), logo após vem as questões 14 e 15. Por fim, os itens em que os respondentes mais optaram por deixar em branco foram o 28, sendo essa a questão com maior quantidade de não respostas (3257), seguido dos itens 27 e 29.

O esperado é que as questões com maior número de respostas erradas ou em branco apresentem um maior nível de dificuldade em comparação aos itens em que os respondentes acertaram mais.

Como já foi mencionado anteriormente, o vestibular da UnB possui uma penalidade em caso de erro, então o score é calculado da seguinte maneira: no caso da prova de história todas as questões são do tipo A, portanto resposta errada = -1, resposta em branco = 0 e resposta correta = 1. Dessa forma, a Figura 4 apresenta os boxplots dos itens.

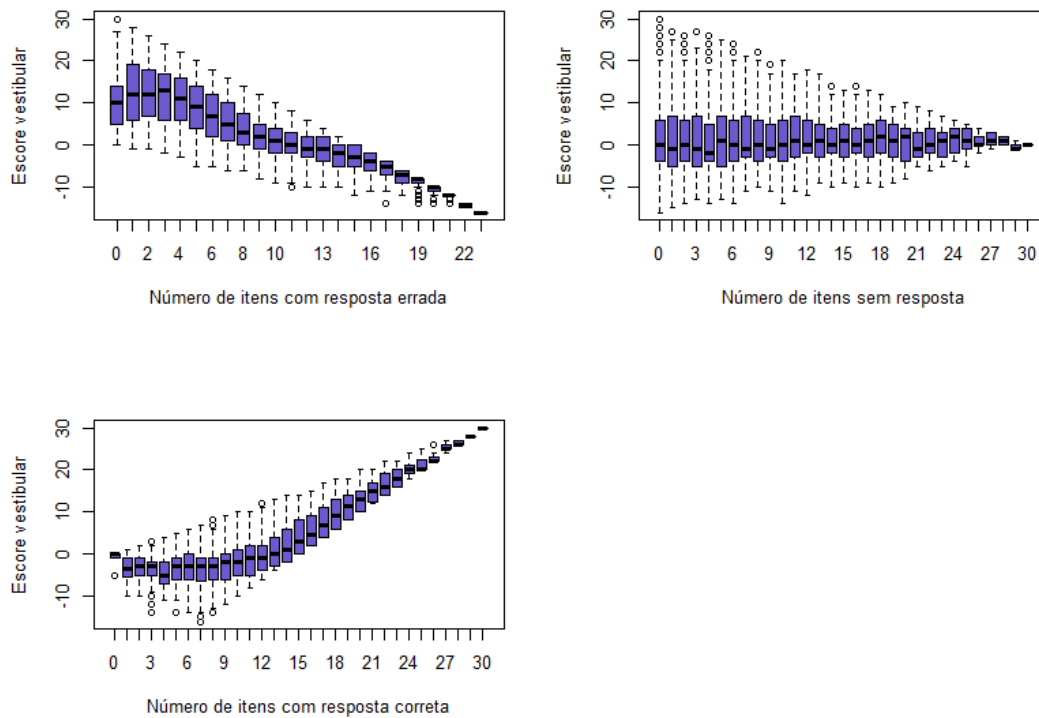


Figura 4: Boxplots dos Escores no Vestibular.

Pelos boxplots da Figura 4, é possível observar que existe uma relação entre os escores dos respondentes no vestibular da UnB e se os mesmos acertaram, erraram ou deixaram em branco os itens. Nota-se que conforme o número de itens com resposta correta aumenta, o escore também aumenta; percebe-se também que, quando o número de questões erradas aumenta, o escore diminui.

Observa-se também pela Figura 4, que há uma maior variabilidade do escore do vestibular para respondentes com menor número de respostas corretas. O mesmo vale para respondentes com menor número de itens errados e para os respondentes com menor número de não respostas.

Em seguida, uma codificação dicotômica será feita, em que os itens errados e os itens não respondidos serão iguais a 0 e os itens corretos serão iguais a 1.

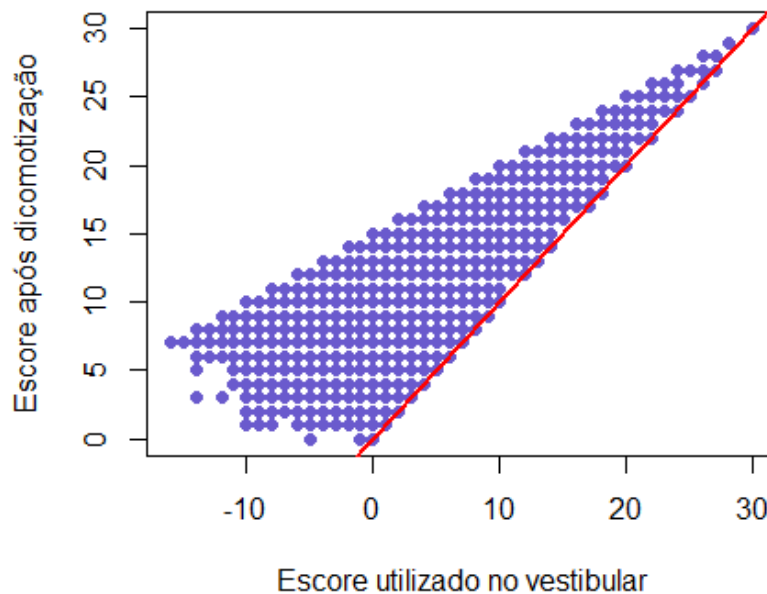


Figura 5: Comparação dos Escores no Vestibular.

Através da Figura 5, é possível observar uma comparação entre o escore utilizado no vestibular e o escore após a dicomotização, e nota-se que devido à penalidade presente no vestibular, existe uma grande variabilidade entre o escore utilizado no vestibular e o escore depois de realizar a dicomotização. Esta variabilidade é maior para os respondentes com escore baixo. Isto ocorre porque um respondente que acertou muitas questões, mas também errou muitas questões (deixando poucas questões sem resposta) e outro que acertou poucas questões, mas também errou poucas (deixando muitas questões sem resposta) terão o escore do vestibular com valores próximos, ao passo que o escore dado apenas pelo número de respostas corretas será maior para aquele respondente que acertou muitas questões, mesmo que tenham errado muitas questões.

Na Figura 6, observa-se pelos histogramas que os escores do vestibular não parecem seguir uma distribuição Normal, apresentando assimetria considerável, enquanto os escores padronizados (após dicotomização) têm uma distribuição mais próxima da Normal.



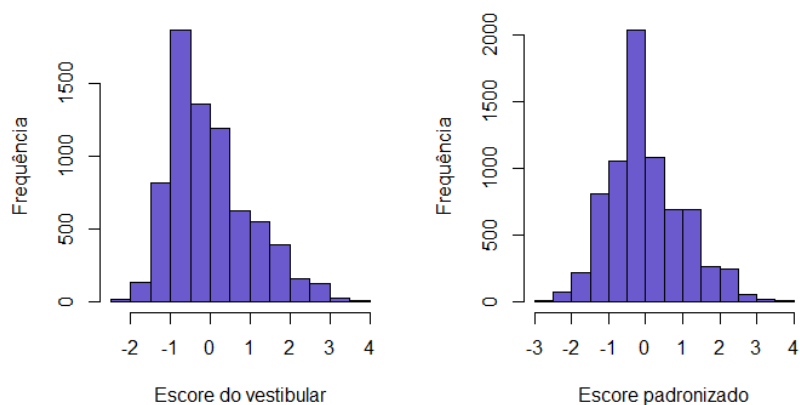


Figura 6: Distribuição dos Escores.

A Figura 7 mostra a proporção de acertos por item da prova de história. Nota-se que a maior parte dos itens obteve menos de 60% de acertos. Os itens que alcançaram mais de 60% de acerto foram: 13, 14 e 15; e o item 30 é o que obteve a menor proporção de acertos.

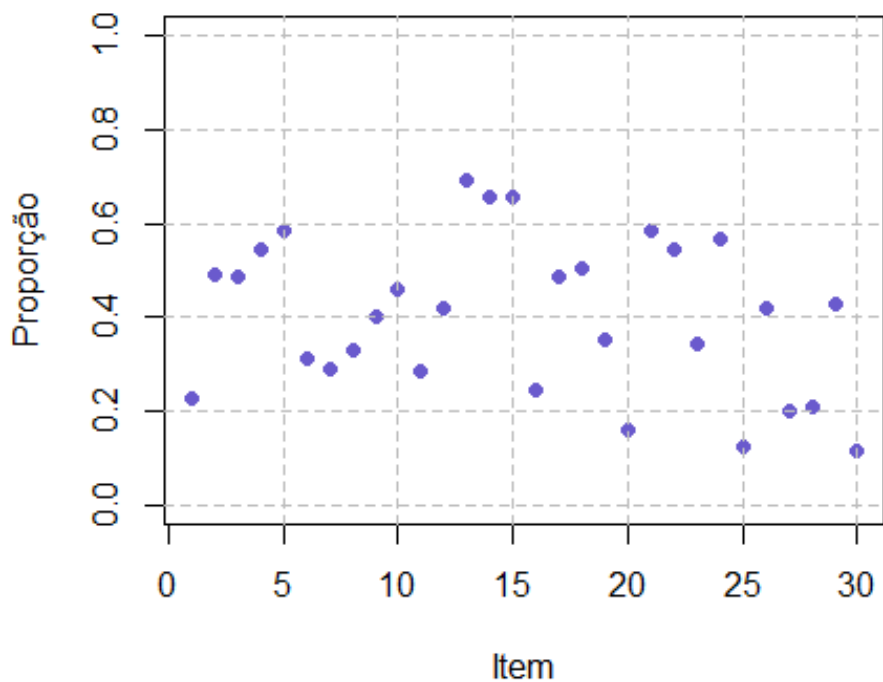
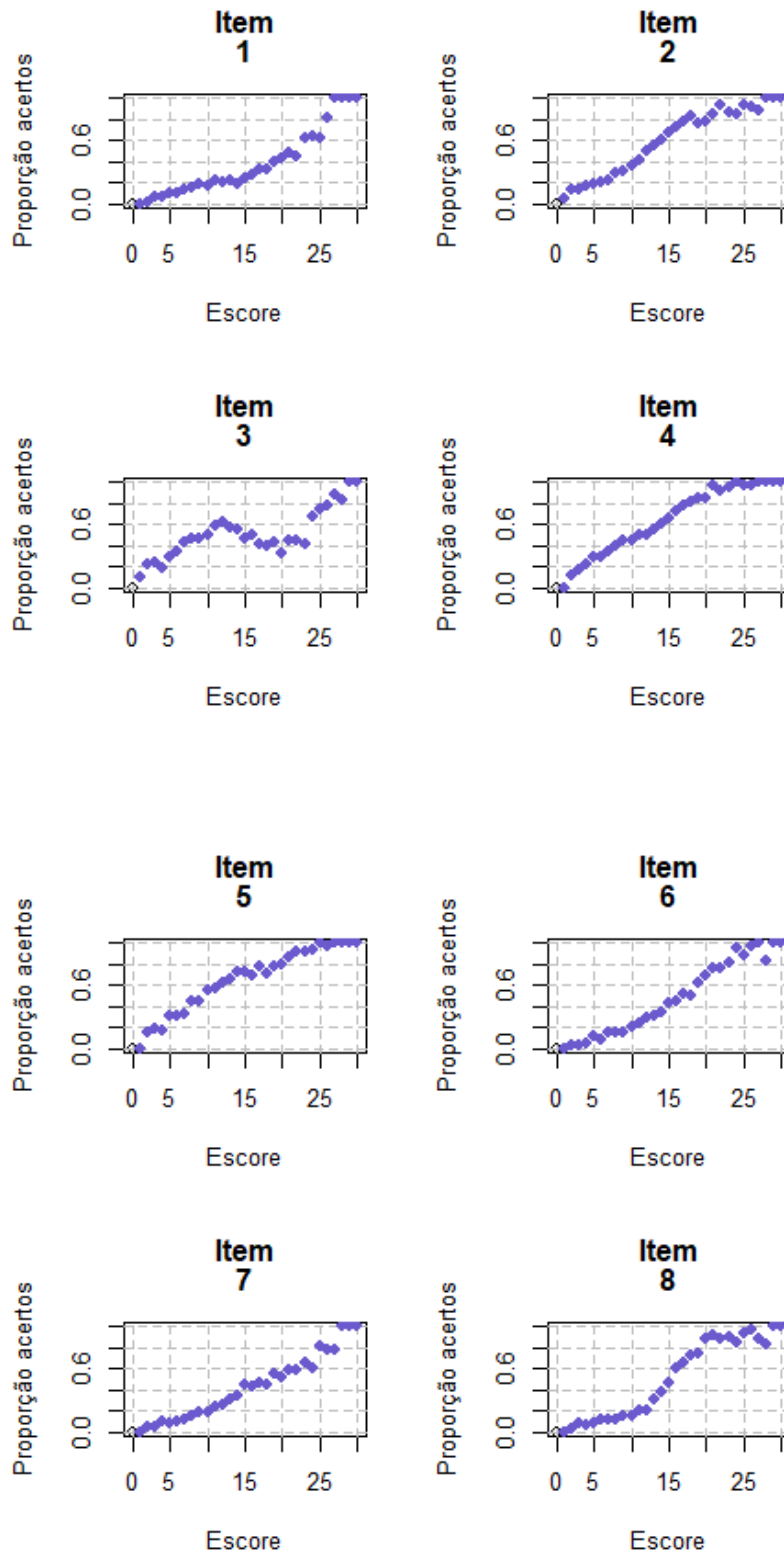
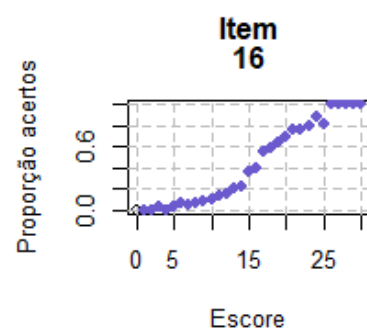
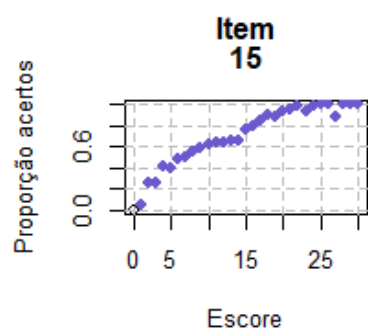
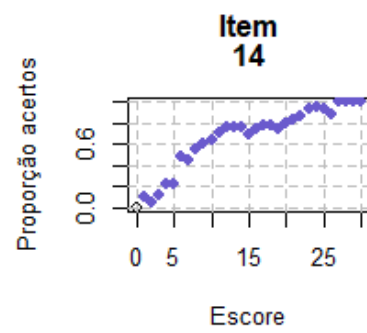
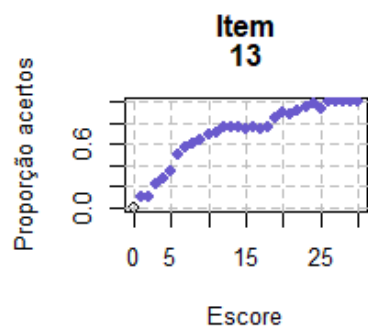
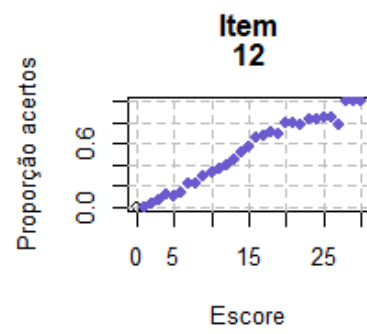
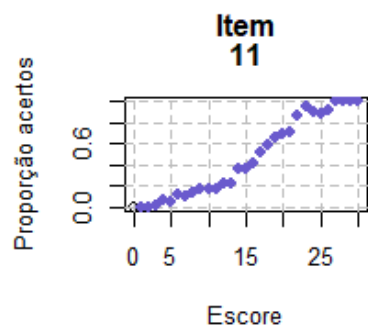
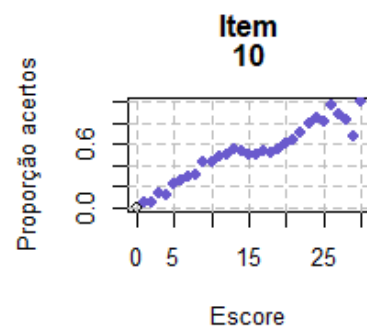
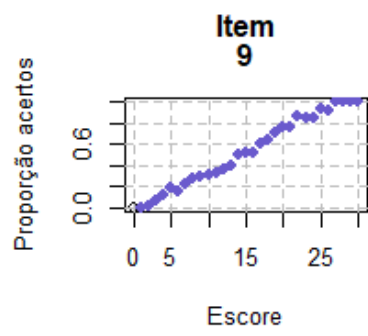
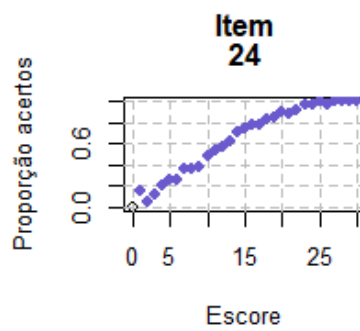
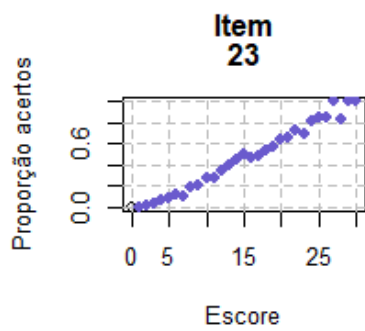
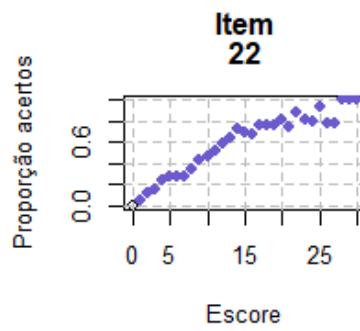
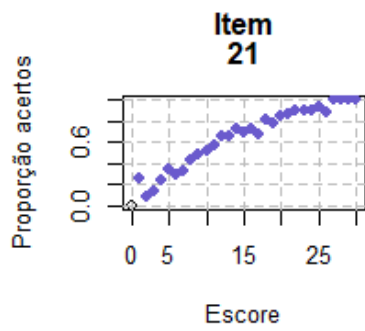
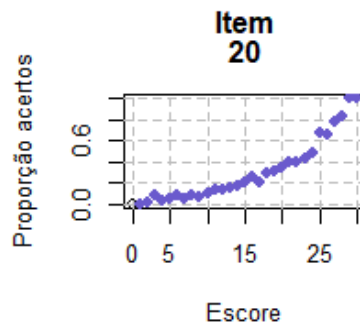
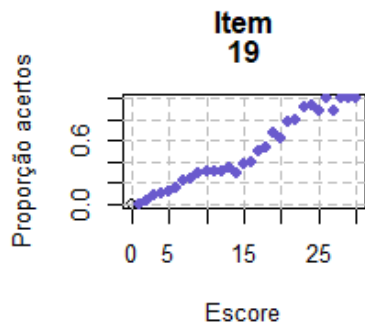
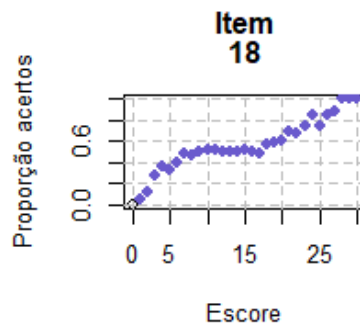
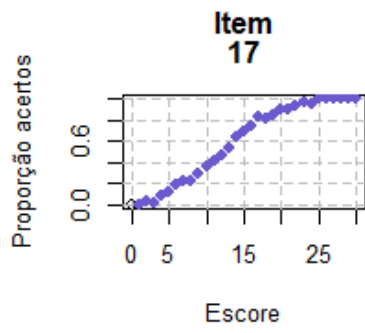


Figura 7: Proporção de Acertos por Item.

A seguir são apresentadas, na Figura 8, as Curvas Características dos Itens (CCI), de forma não paramétrica.







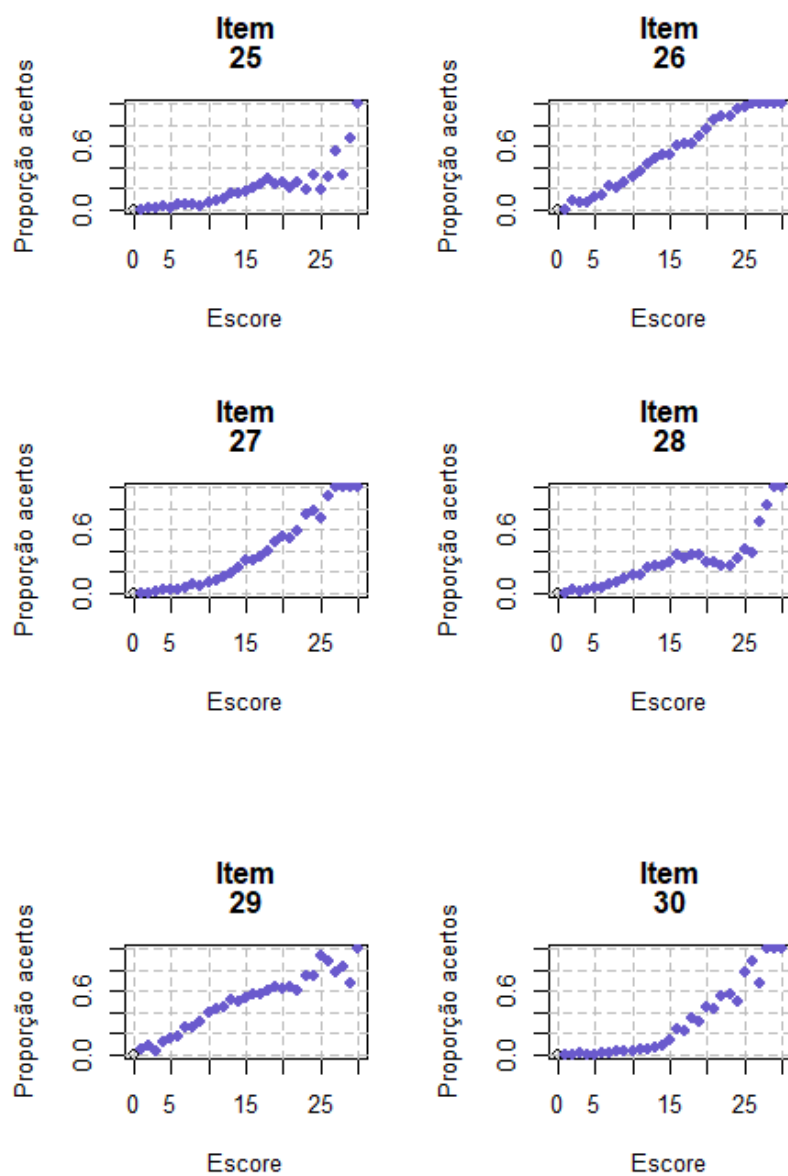


Figura 8: Curva Característica de cada item (CCI).

Na Figura 8, observa-se que na maioria dos itens, quanto maior o escore do respondente, maior sua proporção de acertos. No entanto, nos itens 3, 10, 25 e 28 não fica muito claro tal fato. Este comportamento irregular pode ser causado por uma frequência muito baixa de respondentes com escores na faixa de valores em que o comportamento irregular ocorre. Tal comportamento será estudado posteriormente após o ajuste dos modelos de TRI.

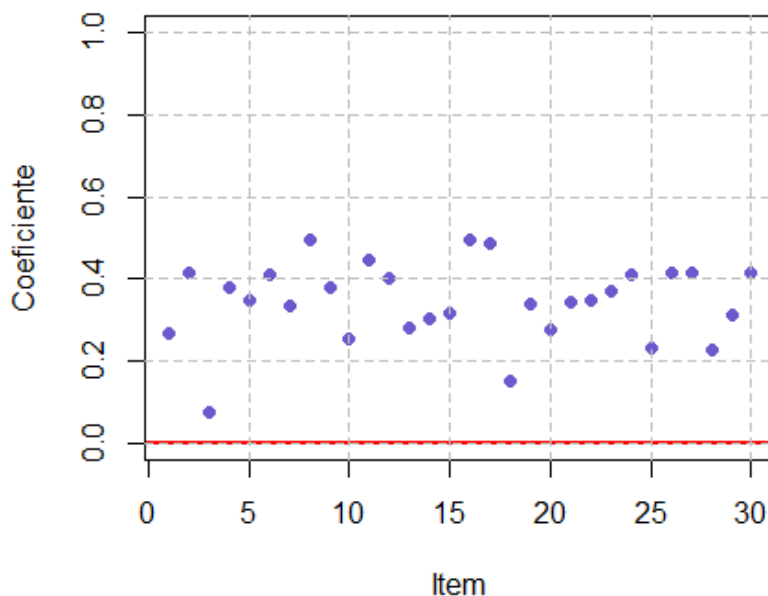


Figura 9: Coeficiente de Correlação Ponto-Biserial por Item.

A Figura 9 apresenta o Coeficiente de Correlação Ponto-Biserial por item. Nota-se que nenhum dos itens apresenta coeficiente menor do que zero, porém o item 3 apresenta um valor muito próximo de zero para o coeficiente, aproximadamente 0,07, o que pode ser uma indicação de que o item 3 talvez não deva ser utilizado em provas futuras ou que deva ser remodelado para ser mais coerente com a prova de história do vestibular da UnB.

## 3.2 Ajuste dos Modelos Logísticos

### 3.2.1 Modelo Logístico de 3 Parâmetros

Abaixo, na Tabela 1, são apresentados os 3 parâmetros (Acerto ao Acaso, Dificuldade e Discriminação) deste modelo.

Item	Acerto ao Acaso	Dificuldade	Discriminação
1	0,20	2,02	3,02
2	0,00	0,10	1,06
3	0,00	-0,30	-0,17
4	0,47	1,29	22,10
5	0,25	0,36	0,83
6	0,20	1,37	2,35
7	0,00	1,30	0,81
8	0,12	0,91	2,85
9	0,00	0,62	0,78
10	0,42	2,09	1,75
11	0,11	1,24	1,85
12	0,25	1,14	1,67
13	0,65	1,45	2,67
14	0,59	1,52	1,51
15	0,60	1,26	19,43
16	0,10	1,20	3,04
17	0,09	0,30	1,85
18	0,42	-1,13	-4,75
19	0,29	1,37	39,08
20	0,00	2,49	0,75
21	0,00	-0,47	0,78
22	0,00	-0,20	0,89
23	0,00	0,97	0,81
24	0,00	-0,20	1,59
25	0,00	3,02	0,73
26	0,00	0,41	1,20
27	0,06	1,65	1,51
28	0,19	3,30	1,43
29	0,38	1,99	1,94
30	0,00	1,84	1,54

Tabela 1: Parâmetros do ML3

Pode-se observar valores atípicos para o parâmetro de discriminação ( $a_i$ ) dos itens 4 ( $a_i = 22,10$ ), 15 ( $a_i = 19,43$ ) e 19 ( $a_i = 39,08$ ), assim como os itens 3 e 18 que apresentam valores negativos. Além disso, o parâmetro de acerto ao acaso não apresenta valores tão baixos para uma boa parte dos itens (1, 4 a 6, 8, 10 a 16, 18, 19, 28 a 30), o que indica que o modelo ML3 não é adequado. Nota-se também que o item mais difícil é

o 28, seguido do 25 e 20, já o item mais fácil é o 18, seguido do 21 e 3.

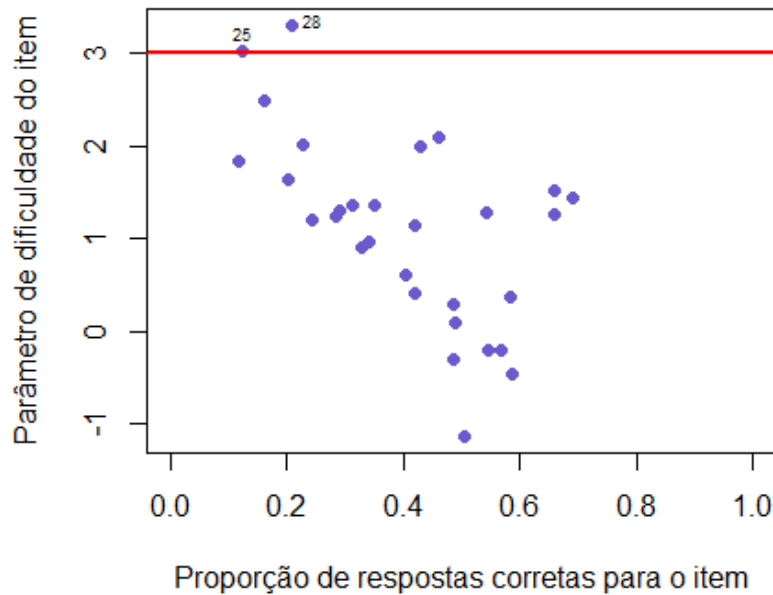


Figura 10: Proporção de respostas corretas no ML3 por parâmetro de dificuldade.

Na Figura 10, é possível observar através do gráfico a proporção de respostas corretas pelo parâmetro de dificuldade de cada item. Assim como foi reparado na Tabela 1, nota-se que os itens 28 e 25 são os mais difíceis. Também, observa-se que conforme o nível de dificuldade do item diminui, a proporção de acertos aumenta, o que é o esperado.



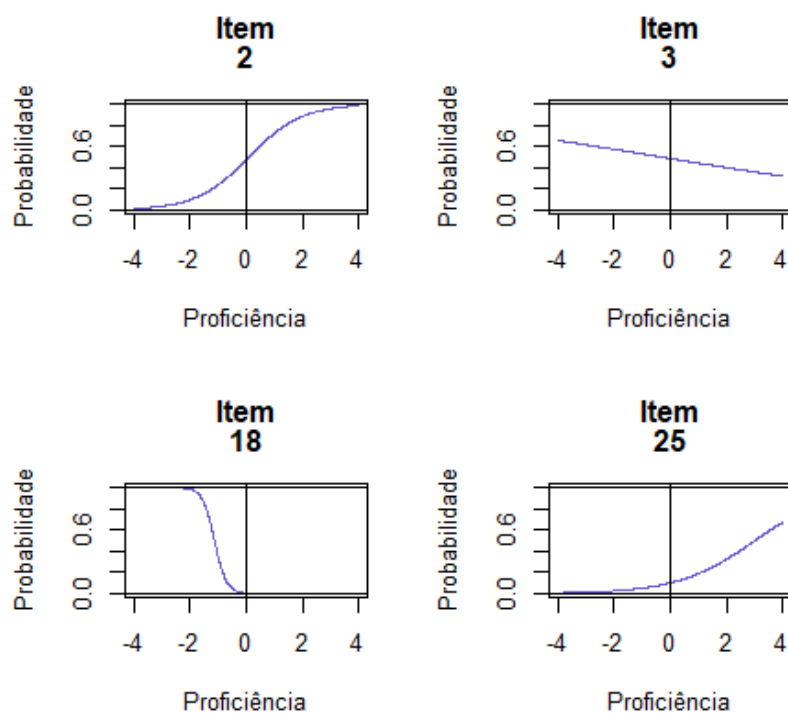


Figura 11: Curva Característica dos Itens 2, 3, 18 e 25 no ML3.

A Figura 11 mostra a curva característica de alguns itens no modelo ML3. Através dos gráficos, é possível observar que o item 2 apresenta o resultado esperado. Já nos itens 3 e 18 nota-se que, conforme a proficiência aumenta, a probabilidade de acerto diminui, o que é o oposto do esperado. O item 25 apresenta baixo poder de discriminação e um nível de dificuldade alto.

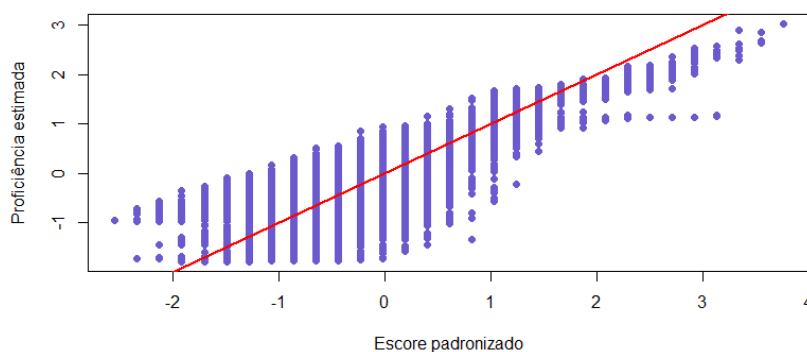


Figura 12: Proficiência Estimada por Escore Padronizado no ML3.

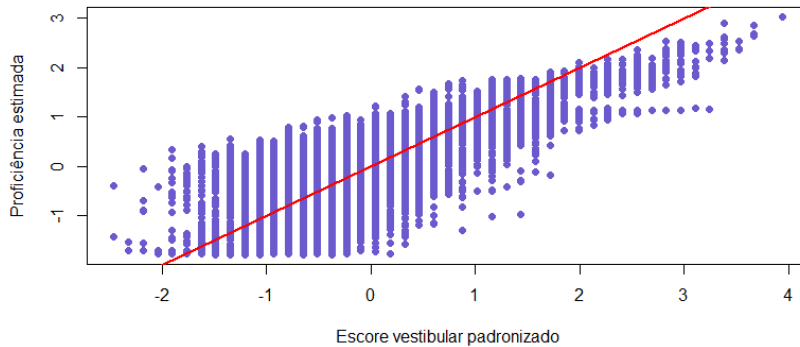


Figura 13: Proficiência Estimada por Escore Vestibular Padronizado no ML3.

Na Figura 12 é exibida a proficiência estimada por escore padronizado, já na Figura 13 é a proficiência estimada por escore do vestibular padronizado. Nota-se que a correlação entre a proficiência estimada e o escore é maior no gráfico da Figura 12. Ao calcular o coeficiente de correlação de Pearson, obteve-se para o primeiro 0.81, um valor considerado alto, para o segundo o resultado foi 0.76. Apesar desses resultados indicarem que o ajuste do Modelo é coerente, esse ainda não é o melhor ajuste, devido a algumas razões citadas anteriormente, como o parâmetro de discriminação que apresenta valores aberrantes e o parâmetro de acerto ao acaso que não apresenta valores baixos (ou próximos de zero) para todos os itens.

### 3.2.2 Modelo Logístico de 2 Parâmetros

Neste modelo, o parâmetro de acerto ao acaso será nulo, sobrando os parâmetros de dificuldade e de discriminação. Assim, se dará início ao ajuste para um Modelo Logístico de 2 Parâmetros.

---

Item	Dificuldade	Discriminação
1	4.10	0.30
2	0.04	1.05
3	-0.40	-0.14
4	-0.33	0.56
5	-0.57	0.63
6	1.09	0.82
7	1.34	0.74
8	0.65	1.57
9	0.56	0.79
10	0.61	0.28
11	1.04	1.08
12	0.44	0.83
13	-2.19	0.38
14	-1.61	0.42
15	-1.55	0.44
16	1.06	1.46
17	0.04	1.58
18	0.21	-0.10
19	1.68	0.38
20	2.57	0.71
21	-0.53	0.73
22	-0.26	0.82
23	0.95	0.78
24	-0.28	1.28
25	3.22	0.66
26	0.38	1.05
27	1.59	1.05
28	5.19	0.26
29	0.84	0.35
30	1.81	1.53

---

Tabela 2: Parâmetros do ML2

Na Tabela 2, nota-se que houve uma diferença em relação aos parâmetros de dificuldade e discriminação dos apresentados na Tabela 1. Isso se deve ao parâmetro Acerto ao Acaso que apresentou valores não muito baixos no ML3.

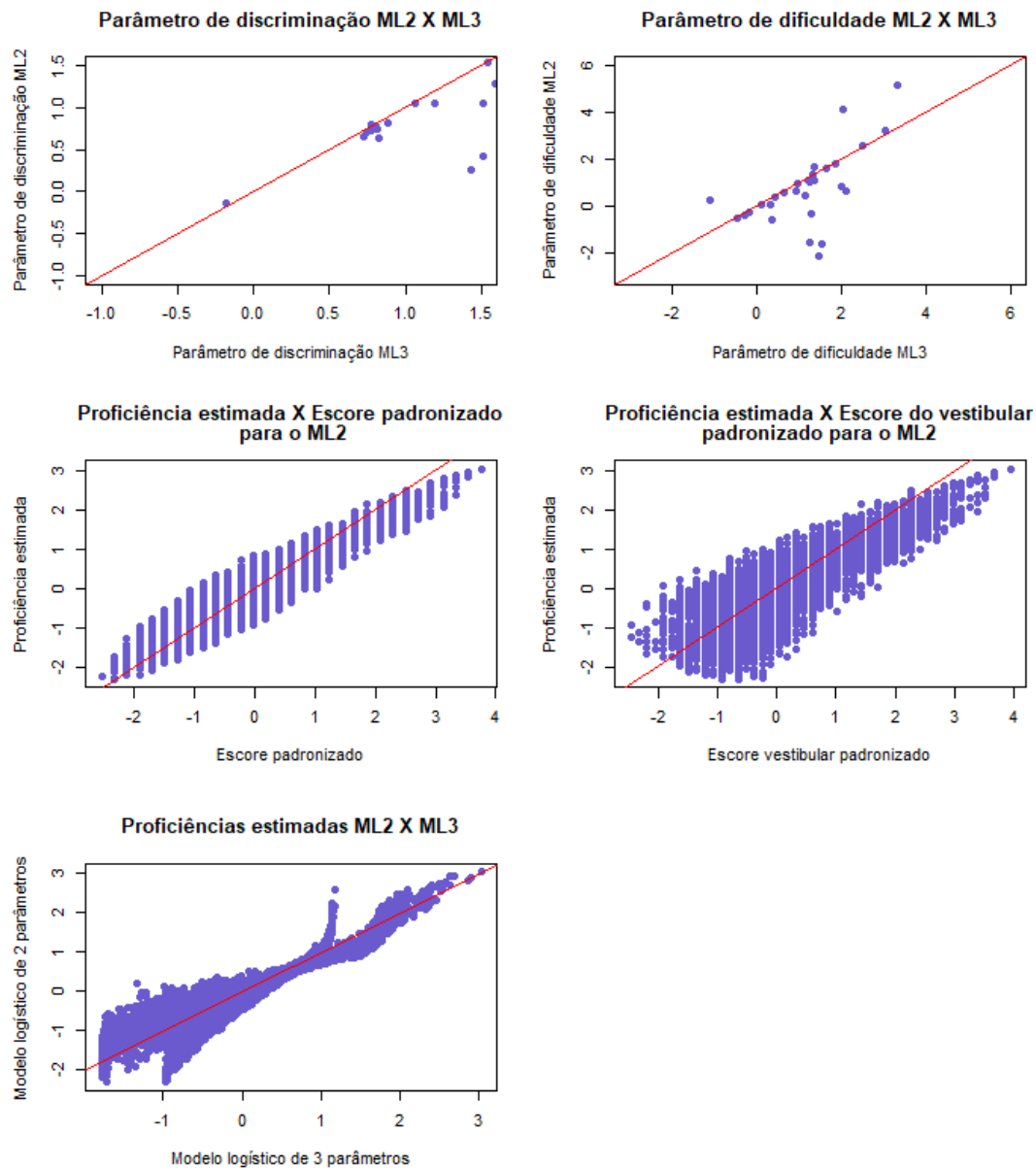


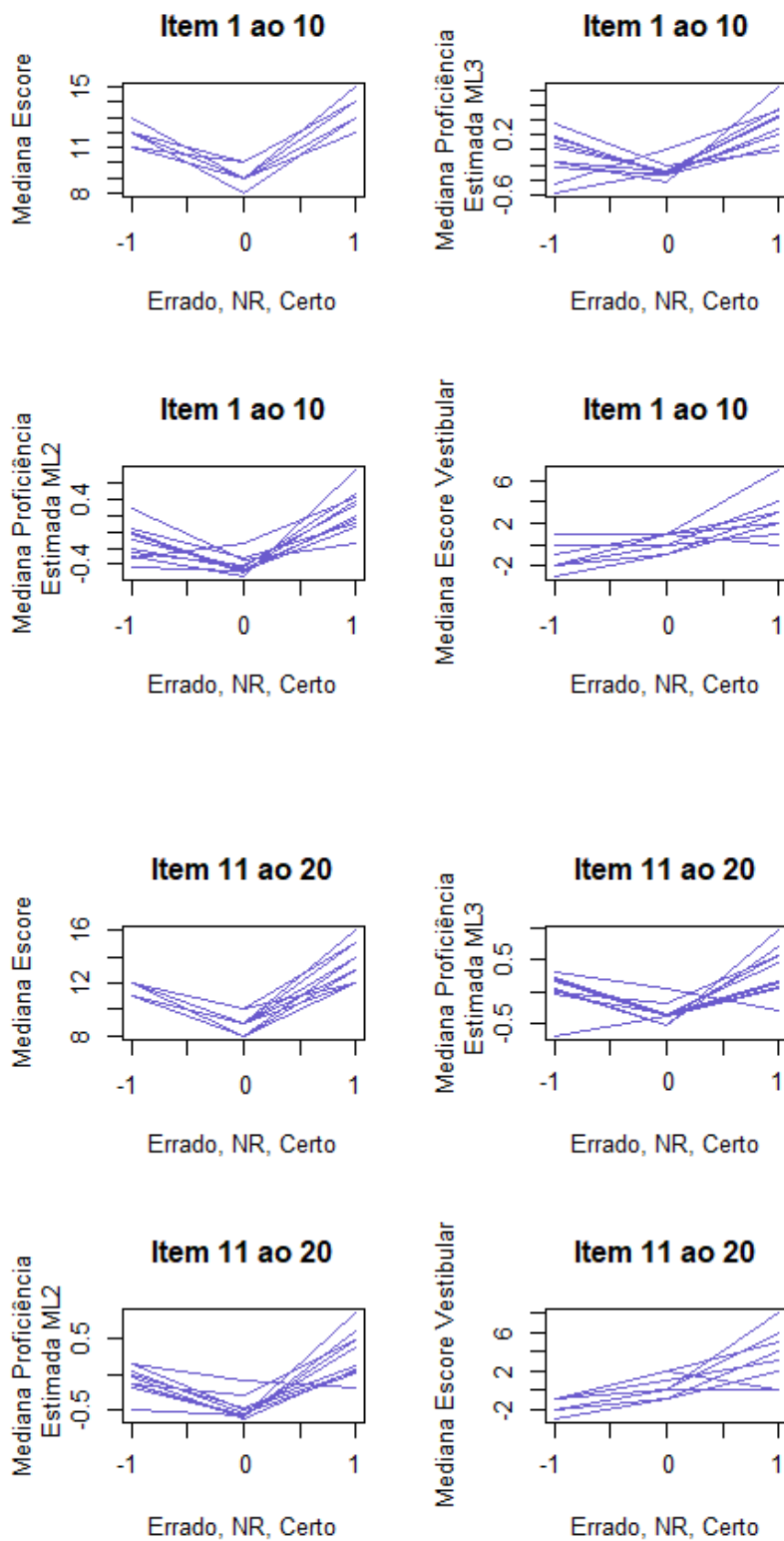
Figura 14: Comparativo entre os modelos ML2 e ML3.

A Figura 14 é composta por 5 gráficos que fazem comparação entre os parâmetros dos Modelos Logísticos de 2 e de 3 parâmetros. No primeiro e no segundo gráfico, observa-se que os parâmetros de discriminação e de dificuldade são parecidos para os dois modelos.

Os gráficos 3 e 4 mostram as proficiências estimadas pelo escore padronizado e pelo escore do vestibular padronizado, ambos para o Modelo ML2. Nota-se que, assim como no ML3, o ajuste utilizando TRI é melhor, apresenta uma maior correlação, tendo como resultados para o coeficiente de correlação de Pearson para o terceiro gráfico 0.93 e para o quarto gráfico 0.80.

Enfim, o último gráfico da Figura 14 representa a comparação entre as pro-

ficiências estimadas no ML2 e no ML3. Observa-se uma grande correlação entre as proficiências, com coeficiente de correlação de Pearson igual a 0.93.



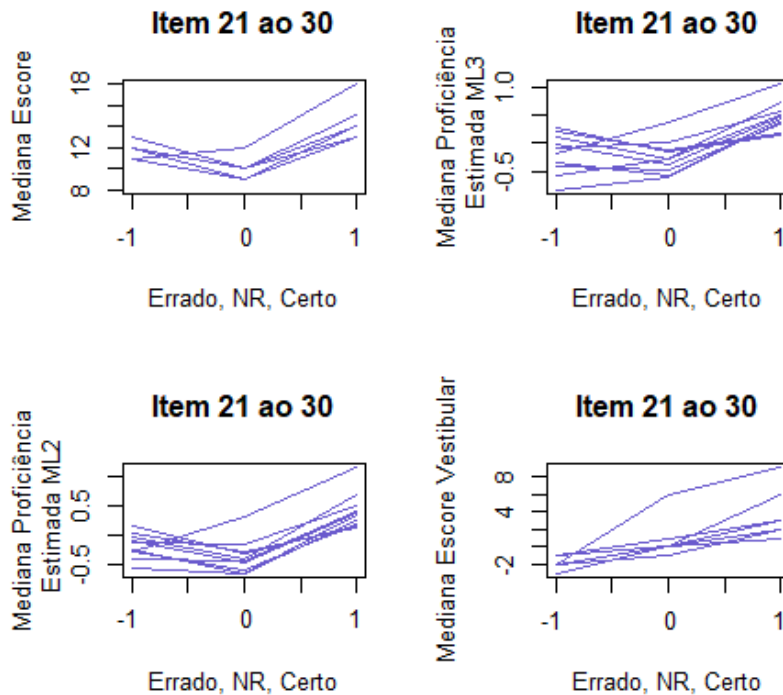


Figura 15: Medianas do Escore, Proficiência Estimada do ML3, Proficiência Estimada do ML2 e Escore do Vestibular por Item.

A Figura 15 apresenta as medianas do escore (prova de história), das proficiências estimadas pelos modelos ML3 e ML2 e do escore do vestibular, sendo os três primeiros com ajustes de TRI, para as três opções: errado, não resposta (NR) e certo. Pelos gráficos, nota-se que a maioria dos itens apresentam o mesmo comportamento, as medianas do escore e das proficiências estimadas são maiores para os respondentes que erraram o item do que para os que não responderam, com exceção do item 30.

Já os gráficos das medianas do escore do vestibular mostram um comportamento diferente do resto, onde os respondentes que não respondem apresentam um escore maior do que os que erram. Portanto, levando a considerar a ordem “não resposta”, “errado” e “certo”.

### 3.3 Ajuste do Modelo de Resposta Gradual

Para começar o ajuste do Modelo de Resposta Gradual, primeiro será aplicada a ordem “errado” (-1 ponto), “não resposta” (0 pontos) e “certo” (1 ponto).

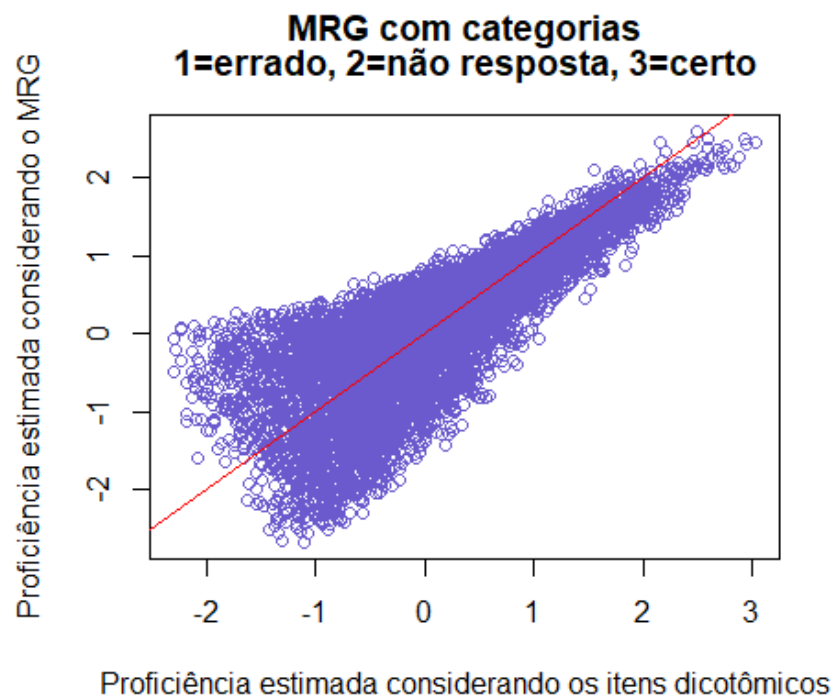


Figura 16: Proficiência Estimada MRG na ordem errado, NR, certo x Dicotômicos

Pela Figura 16, nota-se que a proficiência estimada levando em consideração o Modelo de Resposta Gradual (MRG) com a ordem “errado”, “não resposta” e “certo” e a proficiência estimada tendo em conta os itens dicotômicos não possuem muita correlação.

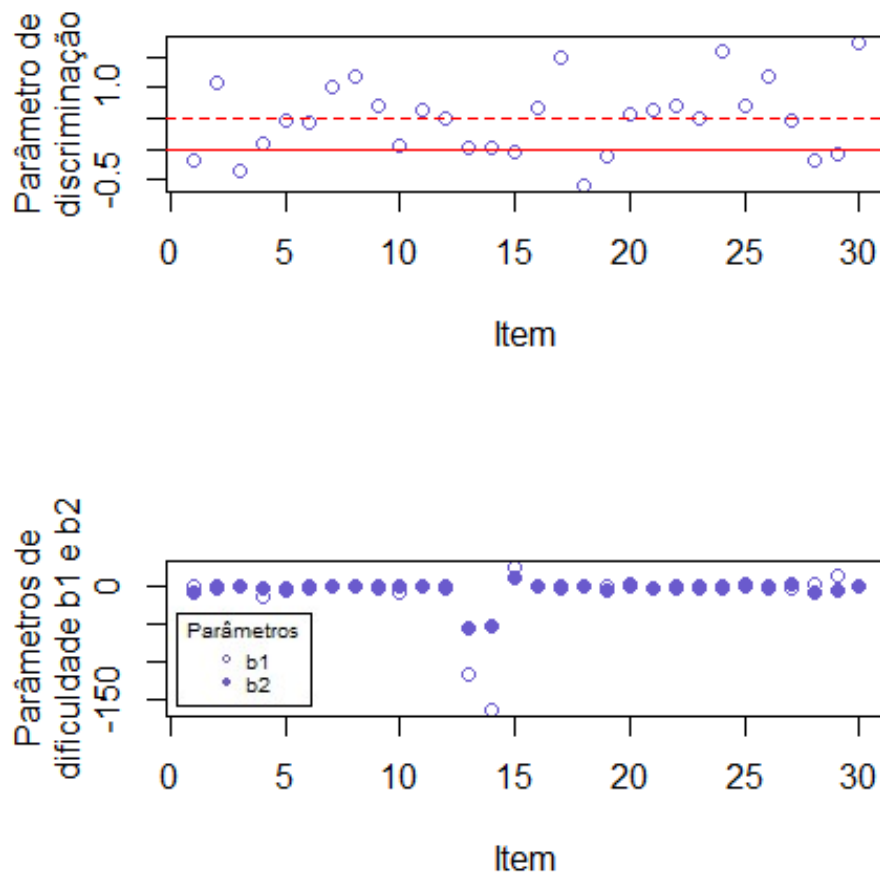


Figura 17: Discriminação e Dificuldade MRG na ordem errado, NR, certo.

A Figura 17 mostra no primeiro gráfico valores negativos para o parâmetro de discriminação para alguns itens e no segundo gráfico valores atípicos para o parâmetro de dificuldade de determinados itens, um deles com um valor próximo de -150. Assim como alguns valores do  $b_1$  são superiores ao  $b_2$ , o que não poderia acontecer. Esses comportamentos mostram que o modelo com a ordem “errado”, “não resposta” e “certo” não está bem ajustado.

Portanto, agora será testado o Modelo de Resposta Gradual com a ordem “não resposta”, “errado” e “certo”, ou seja, os respondentes que erraram em um nível superior aos que deixaram o item em branco. O que seria uma inovação para a correção da prova do Vestibular da UnB e até mesmo de outras provas, caso o MRG com essa ordem apresente um bom ajuste.

Logo, a seguir serão iniciadas as análises com esta nova ordem para o MRG.



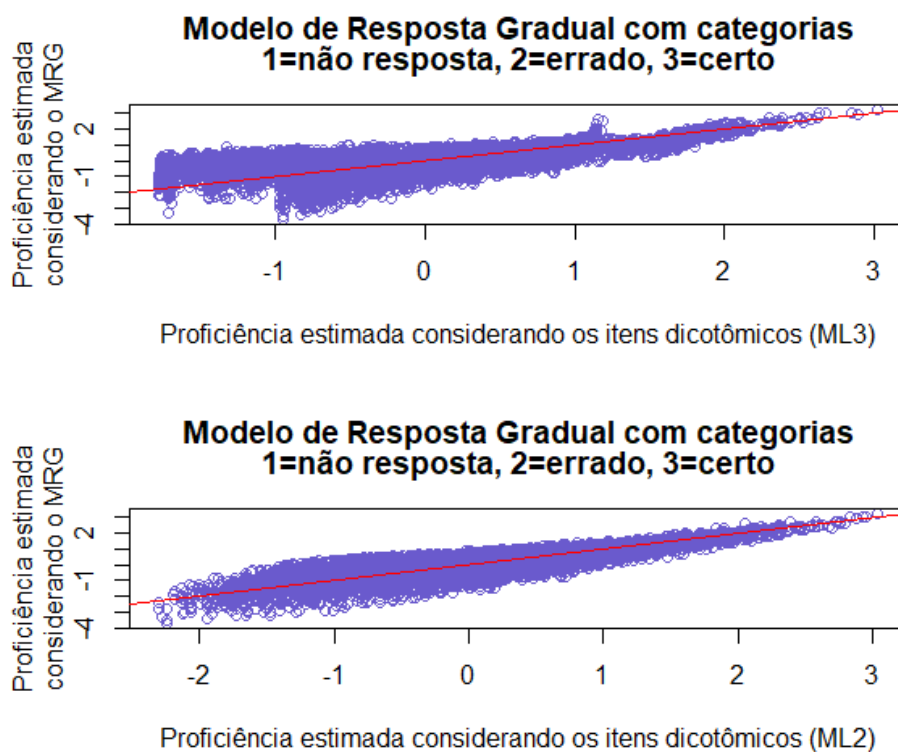
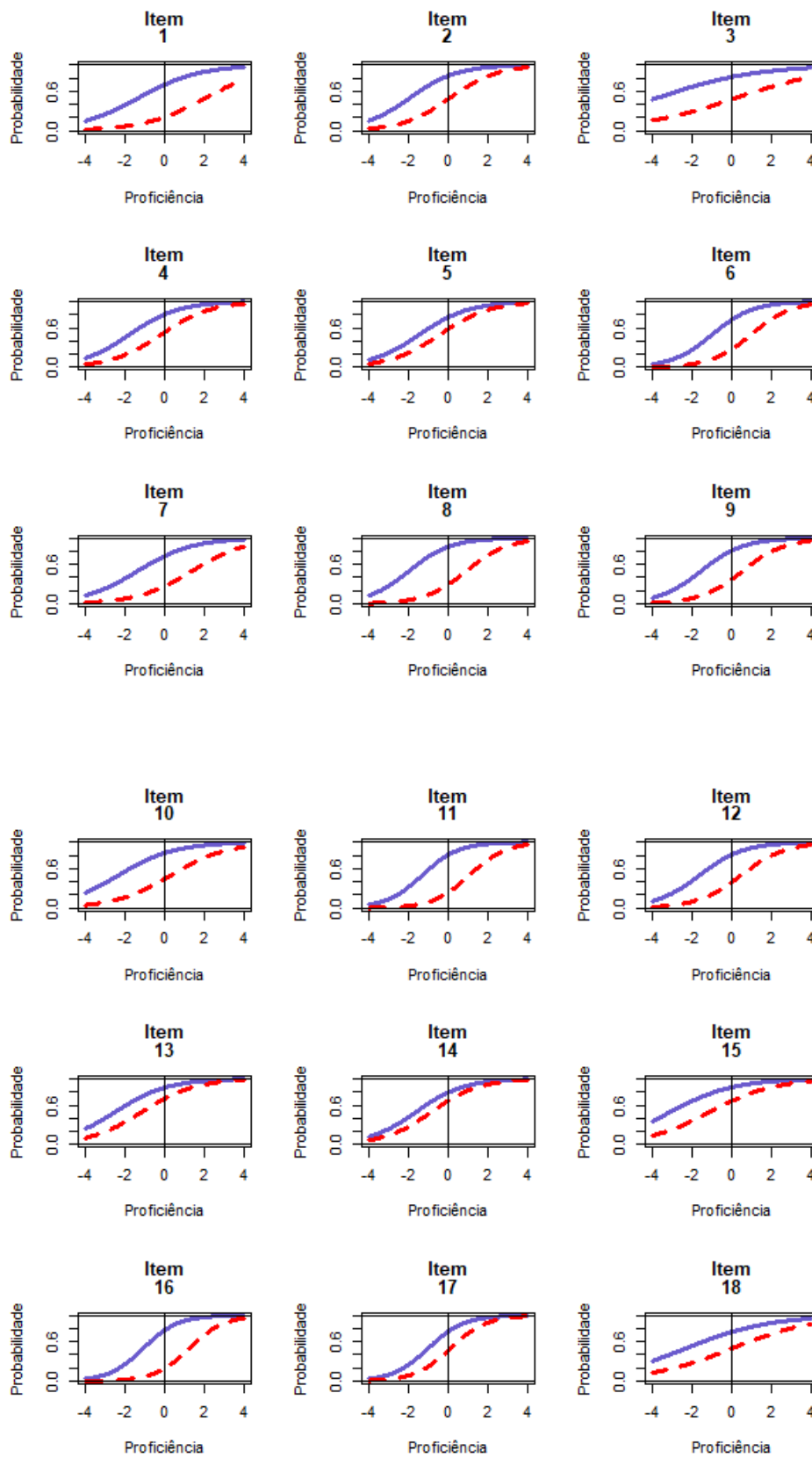


Figura 18: Proficiência Estimada MRG na ordem NR, errado e certo x Dicotômicos.

Através da Figura 18, é possível observar uma comparação entre a proficiência estimada considerando o MRG com a ordem “não resposta”, “errado” e “certo” e a proficiência estimada considerando os itens dicotômicos dos modelos ML3 e ML2.

Nota-se que as proficiências estimadas no MRG e com os itens dicotômicos nos Modelos Logísticos de 2 e 3 parâmetros apresentam uma forte correlação, então a princípio essa ordem parece ser melhor para o modelo, mas é importante fazer uma análise mais aprofundada.

Abaixo são apresentados os gráficos da curva característica de cada item para o MRG com a nova ordem, onde as curvas azul e vermelha estão associadas, respectivamente, aos parâmetros de dificuldade  $b_1$  e  $b_2$ .



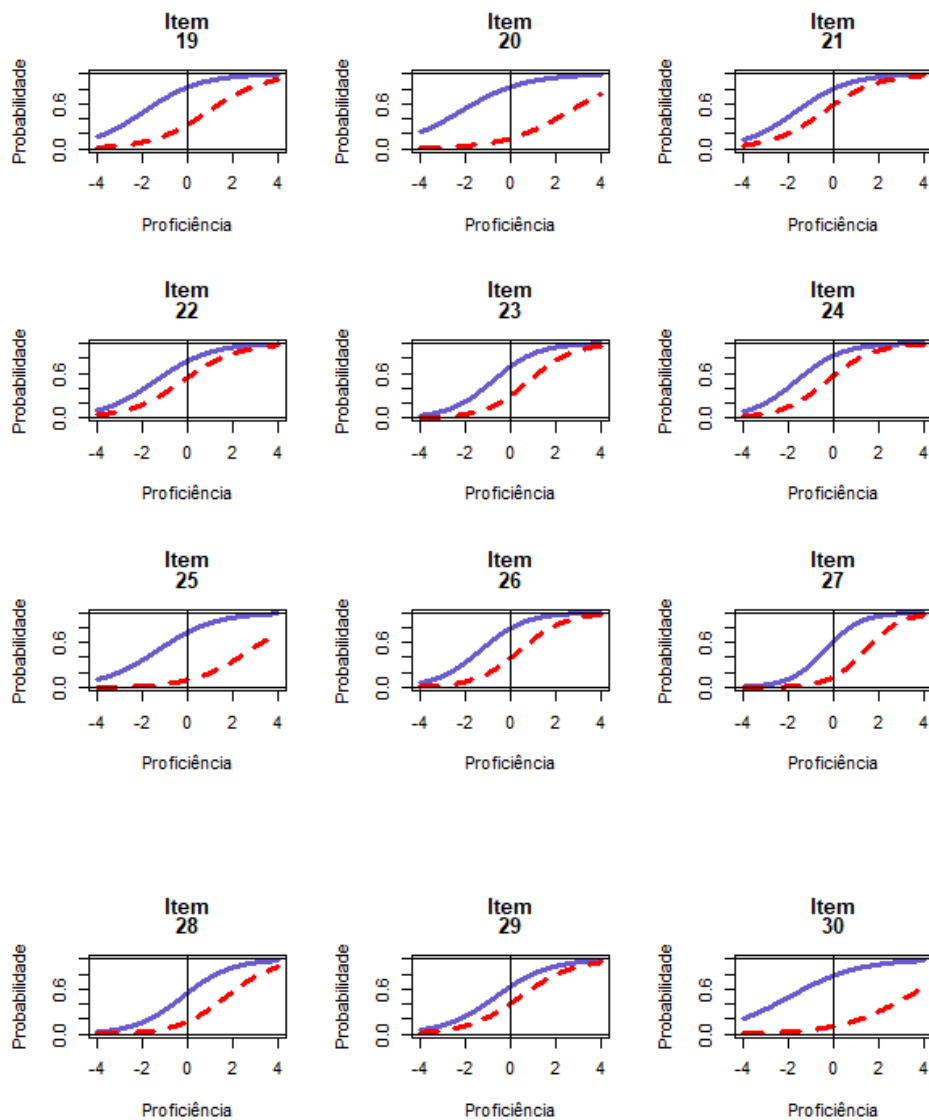


Figura 19: Curva Característica de cada item para o MRG na ordem NR, errado e certo.

Pela Figura 19, é possível notar que essa nova ordenação (NR, errado, certo) se ajusta melhor para o MRG do que a ordem testada anteriormente (errado, NR, certo). No entanto, observa-se que os itens 20, 25 e 30 apresentam um menor valor para o parâmetro de discriminação.

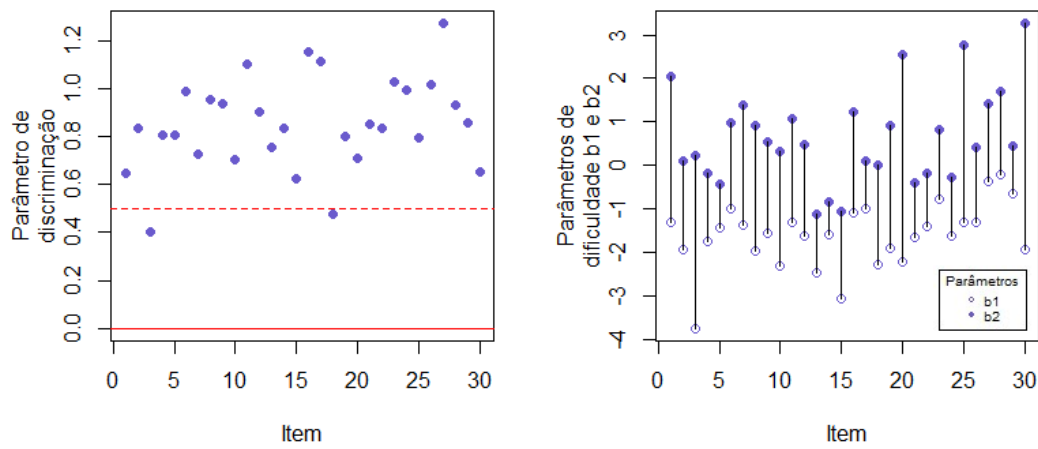


Figura 20: Discriminação e Dificuldade  $b_1$  e  $b_2$  por Item para o MRG (NR, errado, certo).

A Figura 20 mostra que a ordenação “não resposta”, “errado” e “certo” realmente apresenta um melhor ajuste para o modelo. Todos os itens, com exceção do 3 e 18, possuem um valor maior do que 0,5 para o parâmetro de discriminação. Analisando o segundo gráfico, todos os itens apresentaram o parâmetro  $b_2$  superior ao  $b_1$ , como é o esperado. O item 3 possui o  $b_1$  próximo de -4, o que significa que é um item menos difícil e tem uma probabilidade menor do candidato não responder.

## 4 Conclusão

Neste projeto analisamos a Prova de História do Vestibular da UnB de 2014, da banca examinadora CEBRASPE, composta por 30 questões, sendo todas do tipo A. Realizamos um estudo do parâmetro de dificuldade, discriminação e a probabilidade de acerto ao acaso, sendo que essa prova possui penalidade caso o respondente erre o item.

Observamos no começo do projeto que os Modelos Logísticos para Itens Dicotômicos não seria a melhor opção. Portanto, ajustamos o Modelo de Resposta Gradual considerando-se inicialmente a ordem “errado”, “não resposta” e “certo”.

Porém, observamos que os respondentes que mais optavam por não responder um item, devido a penalização de respostas erradas, apresentavam um conhecimento (proficiência) menor do que aqueles que erravam a questão, porque as pessoas com proficiência mais alta respondiam mais questões, tendo maior probabilidade de errar a questão. Concluímos que o melhor ajuste para o MRG foi com a ordenação “não resposta”, “errado” e “certo”. Este resultado também foi observado para outras provas do vestibular da UnB de 2014, como mostra Souza (2023) e Veiga (2022).

Enfim, com o devido ajuste do MRG, todos os itens, exceto o 3 e 18, apresentaram um bom poder de discriminação, ou seja, possuem o poder de separar os respondentes com baixa e alta proficiência. O item 3 mostrou ser mais problemático desde o início do estudo, sendo um item menos difícil e com parâmetro de discriminação baixo, além de apresentar um valor próximo de zero para o coeficiente de correlação ponto-bisserial. Portanto, os itens 3 e 18 da prova de História devem ser revisados antes de uma utilização futura.

## 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. dos; ANDRADE, D. F. de. Teoria da resposta ao item com uso do r. João Pessoa, 2012.

BAKER, F. B. *Item Response Theory*. New York: Marcel Dekker, 1992.

BAKER, F. B.; KIM, S.-H. *Item Response Theory: parameter estimation techniques*. New York: Marcel Dekker, 2004.

BASTOS, L. M. *Deep IRT: an application of deep learning methods to Item Response Theory*. 51 p. Dissertação (Mestrado em Estatística) — Universidade de Brasília, Brasília, 2024.

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

SOUZA, G. S. d. *Teoria da Resposta ao Item aplicada à avaliação das questões da prova de geografia do vestibular da UnB de 2014*. 34 p. Trabalho de Conclusão de Curso — Universidade de Brasília, 2023.

VEIGA, G. P. *Aplicação de modelos de Teoria de Resposta ao Item na avaliação das questões da prova de língua portuguesa do vestibular da UnB*. 50 p. Trabalho de Conclusão de Curso — Universidade de Brasília, Brasília, 2022.