



Universidade de Brasília
Departamento de Estatística

Uso da Teoria de Resposta ao Item para a
identificação de respondentes inconsistentes

Artur Carvalho de Medeiros

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

Brasília
2023

Artur Carvalho de Medeiros

**Uso da Teoria de Resposta ao Item para a
identificação de respondentes inconsistentes**

Orientador: Prof. Dr. Antônio 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

- Gostaria de expressar minha gratidão a todos que de alguma forma me auxiliaram durante o período da minha graduação.
- Primeiramente, agradeço a Deus por me conceder forças para passar por todo esse período.
- Agradeço, também, aos meus familiares pelo apoio aos meus estudos e demais dificuldades perante a vida acadêmica. Sem o suporte de vocês, não teria sido possível alcançar este objetivo.
- Aos meus amigos, que me acompanharam desde o primeiro semestre, durante todo o período da pandemia e após o retorno das aulas presenciais. Sem a companhia de vocês, o percurso teria sido impossível.
- Agradeço a todos os demais que me incentivaram e estiveram presentes, oferecendo palavras de encorajamento nos momentos desafiadores.
- Não poderia deixar de mencionar meu Professor Orientador Antônio Eduardo, cuja orientação, conhecimento e conselhos foram fundamentais para o desenvolvimento deste trabalho. Sua dedicação e orientação foram imprescindíveis para o meu crescimento acadêmico e pessoal.
- Por fim, agradeço a todos que, de alguma forma, contribuíram para a realização deste trabalho. Estou profundamente grato pela confiança, apoio e incentivo oferecidos ao longo dessa jornada.

Resumo

O âmbito do estudo aqui presente são 10 estatísticas de ajuste individual que identificam os respondentes de exames de múltipla escolha cujas respostas não condizem com o padrão esperado nos modelos de Teoria de Resposta ao Item (TRI). Nesse sentido, buscou-se identificar quais os indivíduos poderiam ser classificados como incoerentes e, para isso, realizou-se estudos de simulação de Monte Carlo sobre os dados a fim de testar a hipótese de que o respondente apresenta padrões de respostas incoerentes. Por conseguinte, a partir de simulações compararam-se os resultados obtidos pelas 10 estatísticas onde é possível re-estimar os parâmetros do modelo proposto para os dados, obtendo dessa forma um modelo mais preciso para as possíveis respostas.

A análise foi realizada considerando duas possíveis estratégias: na primeira, foram utilizados os parâmetros gerados pelo ML3P e na segunda, considerou-se o parâmetro de acerto ao acaso igual a 0,25 para todos os itens.

Inicialmente, a primeira análise permitiu concluir que as estatísticas de quadrado médio identificavam um número superior de indivíduos incoerentes quando comparado ao valor obtido pelas estatísticas de índice de cautela. Além disso, observou-se que os parâmetros de discriminação e dificuldade do item sofriam grandes impactos pelos respondentes incoerentes.

Posteriormente, para a segunda análise, a conclusão obtida para as estatísticas de ajuste individual foram as mesmas, porém, com relação aos parâmetros, observou-se que não houve grandes mudanças com a retirada dos indivíduos incoerentes.

Palavras-chaves: ajuste individual; índice de cautela; quadrado médio ajustado.

Lista de Tabelas

1	Estatísticas paramétricas de ajuste individual	15
2	Parâmetros estimados do ML3P (Erros Padrão)	18
3	Número de indivíduos incoerentes por estatística	22
4	EAI para respondentes coerentes	23
5	EAI para respondentes incoerentes	24
6	Número de indivíduos incoerentes em comum entre as estatísticas	26
7	Percentual de indivíduos incoerentes em comum entre as estatística	26
8	Parâmetros estimados sem os respondentes incoerentes do ML3P (Erros Padrão)	28
9	Parâmetros estimados sem os respondentes incoerentes do ML3P (Erros Padrão)	30
10	Número de indivíduos incoerentes por estatística	34
11	Número de indivíduos incoerentes em comum entre as EAI	35
12	Percentual de indivíduos incoerentes em comum entre as estatística	35
13	Número de indivíduos incoerentes por estatística em comum entre os modelos	36
14	Parâmetros estimados sem os respondentes incoerentes do ML3P com parâmetro de acerto ao acaso fixado (Erros Padrão)	38

Lista de Figuras

1	Exemplo de uma Curva Característica do Item - CCI	11
2	Percentual de acertos	16
3	Proficiência Estimada	19
4	Gráfico de boxplot das EAI	20
5	Gráfico de dispersão das EAI em função das proficiências	21
6	Respostas de indivíduos coerentes	24
7	Respostas de indivíduos incoerentes	25
8	Gráficos de de dispersão dos parâmetros do ML3P com os parâmetros re-estimados	27
9	Proficiência Estimada com parâmetro de acerto ao acaso fixado	31
10	Gráfico de boxplot das EAI com parâmetro de acerto ao acaso fixado	32
11	Gráfico de dispersão das EAI em função das proficiências com parâmetro de acerto ao acaso fixado	33
12	Gráficos de de dispersão dos parâmetros do ML3P com c fixado versus os parâmetros re-estimados	37

Sumário

1 Introdução	8
1.1 Considerações iniciais	8
1.2 Objetivos	9
1.2.1 Objetivo Geral	9
1.2.2 Objetivos Específicos	9
2 Referencial Teórico	10
2.1 Teoria da Resposta ao Item(TRI)	10
2.2 Curva Característica do Item - CCI	10
2.3 Função de Informação	11
2.4 Estatísticas de Ajuste Individual	12
2.4.1 Estatísticas de quadrado médio ajustado U , ZU , W e ZW	12
2.4.2 Estatísticas índice de cautela estendido 1 a 6	13
3 Metodologia	15
3.1 Conjunto de dados	15
4 Resultados	16
4.1 Análise exploratória dos dados	16
4.2 Modelo Logístico de Três Parâmetros (ML3P)	17
4.3 Proficiência	19
4.4 Estatísticas de Ajuste Individual (EAI)	20
4.5 Simulação	22
4.6 Remodelagem	27
4.7 ML3P com parâmetro de acerto ao acaso fixado	29
4.8 Proficiência com parâmetro de acerto ao acaso fixado	31
4.9 EAI com parâmetro de acerto ao acaso fixado	32
4.10 Simulação das EAI com parâmetro de acerto ao acaso fixado	34
4.11 Remodelagem com parâmetro de acerto ao acaso fixado	36

5 Conclusão	40
5.1 Considerações finais	40
Referências	42
Apêndice	43
A Código da Programação em R	43

1 Introdução

1.1 Considerações iniciais

A avaliação do nível de conhecimento de cada indivíduo está presente na história desde 2.205 a.C (DEPRESBITERIS, 1989), onde um imperador chinês aplicava exames em seus oficiais com o intuito de demiti-los ou ascende-los de cargo. Nos dias atuais a prática de aplicação de testes se tornou natural e mostra-se presente desde a infância humana categorizando e mensurando o nível de aprendizado da população. Porém, essas análises se tornam árduas quando há concorrência por vagas, por exemplo no Exame Nacional do Ensino Médio (ENEM), que é a porta de entrada para as universidades públicas do Brasil ou, como o objeto de estudo principal desse artigo, a prova do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), que tem por objetivo produzir um diagnóstico da situação da escolaridade básica paulista.

A partir desse problema foram criados diversas metodologias que proporcionam classificações justas para os respondentes com base em seus níveis de conhecimento. Dentre esses métodos, foi desenvolvido o Modelo Logístico de 3 Parâmetros (M3PL), um modelo regressivo logístico da Teoria de Resposta ao Item (TRI) (ANDRADE; TAVARES; VALLE, 2000) que estima escores para cada indivíduo e define os parâmetros para cada um dos itens presentes no exame a partir das respostas dos participantes.

Entretanto, nem todos os respondentes fornecem respostas precisas e coerentes com seu nível de conhecimento estimado e, dessa forma, surgem as Estatísticas de Ajuste Individual (EAI), cujo cálculo permite identificar esses indivíduos de forma mais precisa. Ademais, cabe ressaltar que já foram desenvolvidas mais de vinte estatísticas similares, a exemplo do artigo "Comparing the aberrant response detection performance of thirty-six person-fit statistics" (KARABATSOS, 2003) que compara a eficiência dessas estatísticas. Outrossim, o artigo "Um Novo Método para Avaliação da Coerência de Respostas Individuais em Provas de Múltipla Escolha" (SANTOS, 2022), elenca comparativamente algumas das estatísticas não paramétricas e desenvolve um novo método utilizando regressão isotônica.

1.2 Objetivos

1.2.1 Objetivo Geral

Identificação de respondentes inconsistentes em exames via simulação de respostas aleatórias, baseado em um modelo logístico de três parâmetros a partir de estatísticas de ajuste individual.

1.2.2 Objetivos Específicos

- Comparar estatísticas paramétricas de ajuste individual;
- Verificar efeito dos respondentes incoerentes nos parâmetros estimado pelo modelo ajustado.

2 Referencial Teórico

2.1 Teoria da Resposta ao Item (TRI)

A TRI compreende um conjunto de modelos matemáticos que visam descrever a ligação entre a probabilidade de um participante responder corretamente a uma pergunta com o seu nível de proficiência, por área, em uma avaliação, utilizando atributos específicos de cada questão.

Para este trabalho será considerado o Modelo Logístico de 3 Parâmetros (ML3P), dado por:

$$P_i(\theta_j) = P(U_{ij}|\theta_j) = c_i + \frac{1 - c_i}{1 + \exp[-a_i(\theta_j - b_i)]}, \quad (2.1.1)$$

$i = 1, 2, \dots, I$ (representa os I itens);

$j = 1, 2, \dots, J$ (representa os J indivíduos que responderam ao questionário);

U_{ij} é uma variável dicotômica que assume o valor 1, quando o j -ésimo respondente acerta o i -ésimo item, ou 0 caso contrário;

θ_j é a proficiência do j -ésimo respondente;

$P_i(\theta_j) = P(U_{ij} = 1)$ é a probabilidade do respondente j responder corretamente ao i -ésimo item;

a_i é o parâmetro de discriminação do i -ésimo item, com valor proporcional à inclinação da curva característica do item no ponto b_i ;

b_i é o parâmetro de dificuldade do item i ;

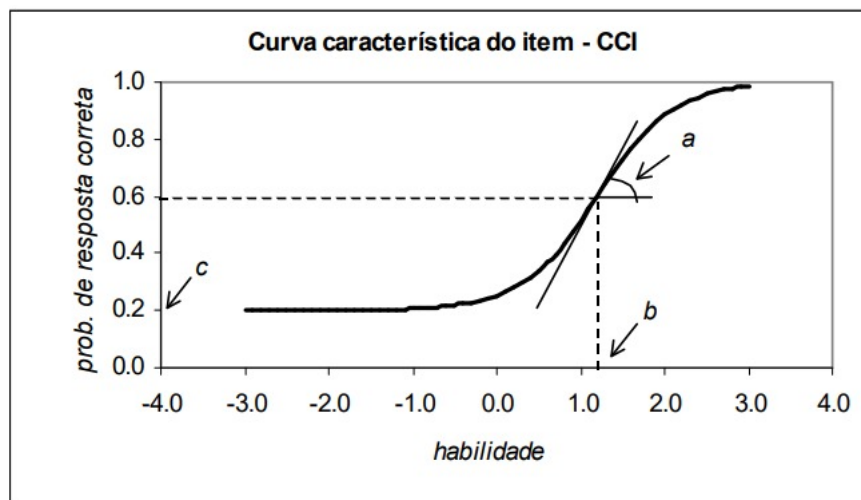
c_i é o parâmetro de acerto ao acaso.

2.2 Curva Característica do Item - CCI

É possível observar uma Curva Característica do Item (CCI) para cada item do exame e, por meio dela, os parâmetros de cada item, permitindo uma rápida visualização de seus parâmetros, além da comparação entre os itens. Outrossim, a escala da habilidade é arbitrária, de modo que o valor por si só não traz informações suficientemente relevantes, mas sim a comparação entre os itens.

O parâmetro a é proporcional à derivada da tangente da curva no ponto de inflexão. O parâmetro b é medido na mesma escala da habilidade e representa o nível de habilidade necessário para uma probabilidade de acerto igual a $\frac{1+c}{2}$. Portanto, quanto maior o valor de b , mais difícil é o item. Por fim, o parâmetro c representa a probabilidade de um aluno com baixa habilidade responder corretamente o item e é muitas vezes referido como a probabilidade de acerto ao acaso, assume sempre valores entre 0 e 1.

Figura 1: Exemplo de uma Curva Característica do Item - CCI



Fonte: Teoria da resposta ao item: conceitos e aplicações (ANDRADE; TAVARES; VALLE, 2000)

2.3 Função de Informação

A Função de Informação do Item permite verificar quanto cada item do teste apresenta de informação para a mensuração da proficiência, sendo dada por:

$$I_i(\theta) = a_i^2 \frac{[1 - P_i(\theta_j)]}{P_i(\theta_j)} \left[\frac{P_i(\theta_j) - c_i}{1 - c_i} \right]^2. \quad (2.3.1)$$

É possível calcular a Função de Informação do Teste somando as funções de informação de cada item. Que é expressa por:

$$FIT(\theta) = \sum_{i=1}^I I_i(\theta). \quad (2.3.2)$$

Esta função também pode ser representada por meio do erro-padrão de medida (erro-padrão de estimação), que é dado por:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}}. \quad (2.3.3)$$

2.4 Estatísticas de Ajuste Individual

Estatísticas de ajuste individual são utilizadas para verificar se os respondentes apresentam respostas inconsistentes com a esperada pelo modelo selecionado (ML3P).

Para o desenvolvimento das estatísticas de ajuste individual são considerados os *erros de Guttman*, que ocorrem quando um item difícil é respondido corretamente e um item fácil é respondido incorretamente pelo respondente j .

As estatísticas de ajuste individual apresentam como hipóteses para cada indivíduo:

$$\begin{cases} H_0 : \text{O padrão de resposta é congruente com o modelo de resposta especificado;} \\ H_1 : \text{O padrão de resposta não é congruente com o modelo de resposta especificado.} \end{cases}$$

2.4.1 Estatísticas de quadrado médio ajustado U , ZU , W e ZW

A estatística de ajuste quadrado médio não ponderado U (WRIGHT; STONE, 1979), também conhecida como *outfit*, é simplesmente uma média do quadrado dos resíduos padronizados de um respondente. Pode ser expressa como:

$$U = \frac{1}{I} \sum_{i=1}^I \frac{(U_{ij} - P_i(\theta_j))^2}{P_i(\theta_j)[1 - P_i(\theta_j)]} \quad (2.4.1)$$

Esta é uma estatística de ajuste sensível a outliers, e portanto é sensível a respostas inesperadas.

Ela pode ser padronizada, obtendo-se a estatística do quadrado médio não ponderado padronizado ZU (WRIGHT, 1980), a qual segue uma distribuição normal padrão e é dada por:

$$ZU = \frac{U^{1/3} - 1 + \frac{2}{9I}}{\sqrt{\frac{2}{9I}}} \quad (2.4.2)$$

Para diminuir o impacto dos outliers, a estatística de ajuste quadrado médio ponderado W (WRIGHT, 1980), também conhecida como *infit*, foi desenvolvida. Esta estatística considera cada resíduo padronizado ao quadrado e pode ser expressa por:

$$W = \frac{\sum_{i=1}^I (U_{ij} - P_i(\theta_j))^2}{\sum_{j=1}^J P_i(\theta_j)[1 - P_i(\theta_j)]} \quad (2.4.3)$$

Esta é uma estatística menos sensível a outliers do que a U . Assim como a U , ela também pode ser padronizada, ZW (WRIGHT, 1980), de forma que a nova estatística segue uma distribuição normal padrão:

$$ZW = \frac{W^{1/3} - 1 + \frac{2}{9I}}{\sqrt{\frac{2}{9I}}} \quad (2.4.4)$$

Grandes valores observados para essas estatísticas indicam que há um subajuste dos dados para o modelo observado, rejeitando-se assim H_0 .

2.4.2 Estatísticas índice de cautela estendido 1 a 6

Foram elaboradas 6 estatísticas índice de cautela estendido: ECI1, ECI2, ECI3, ECI4, ECI5 e ECI6 (TATSUOKA, 1984); que são baseadas nos parâmetros do ML3P. As estatísticas ECI1, ECI2 e ECI4 são baseadas nas covariâncias entre U_j , p , P_j e G , onde dentro do item i , G_j é a média dos $P_i(\theta_j)$ ao longo de J examinados (G não está em ECI1). ECI3 e ECI5 são análogos a ECI2 e ECI4, respectivamente, que usam correlação em vez de covariância. ECI6 é baseado em uma razão, com a correlação entre U_j e P_j no numerador e a variância dos elementos de P_j no denominador. Grandes valores dos índice de cautela estendidos, indicam a existência de uma anomalia no padrão de resposta.

$$ECI1 = 1 - \frac{Cov(U_j, p)}{Cov(P_j, p)} \quad (2.4.5)$$

$$ECI2 = 1 - \frac{Cov(U_j, G)}{Cov(P_j, G)} \quad (2.4.6)$$

$$ECI3 = 1 - \frac{Corr(U_j, G)}{Corr(P_j, G)} \quad (2.4.7)$$

$$ECI4 = 1 - \frac{Cov(U_j, P_j)}{Cov(P_j, G)} \quad (2.4.8)$$

$$ECI5 = 1 - \frac{Corr(U_j, P_j)}{Corr(P_j, G)} \quad (2.4.9)$$

$$ECI6 = 1 - \frac{Corr(U_j, p)}{Var(P_j)} \quad (2.4.10)$$

Em que:

U_j = vetor de acertos de cada item do j -ésimo respondente;

p = vetor da proporção de respostas corretas para cada item, $p = (p_1, \dots, p_i, \dots, p_I)$;

P_j = vetor com as probabilidades de acerto de cada item ($P_i(\theta_j)$) do j -ésimo respondente;

G_j = vetor com as médias de probabilidades de acerto de cada item ($P_i(\theta_j)$),
 $G = (G_1, \dots, G_i, \dots, G_I)$, com $G_i = J^{-1} \sum_{j=1}^J P_i(\theta_j)$.

3 Metodologia

Para realização da pesquisa foram selecionadas estatísticas paramétricas para detecção de resposta incoerentes por meio do pacote PerFit (TENDEIRO; MELJER; NIESSEN, 2016; TENDEIRO, 2021), além de estatísticas desenvolvidas computacionalmente. A precisão da qualificação do resposnte como incoerente ou não pelo modelo será verificada via simulação de mil respondentes incoerentes sob a hipótese nula do Modelo Logístico de 3 Parâmetros ajustado aos dados.

As estatísticas paramétricas de ajuste individual escolhidas são:

Tabela 1: Estatísticas paramétricas de ajuste individual

Estatísticas Paramétricas
U (WRIGHT; STONE, 1979)
ZU (WRIGHT, 1980)
W (WRIGHT, 1980)
ZW (WRIGHT, 1980)
$ECI1$ (TATSUOKA, 1984)
$ECI2$ (TATSUOKA, 1984)
$ECI3$ (TATSUOKA, 1984)
$ECI4$ (TATSUOKA, 1984)
$ECI5$ (TATSUOKA, 1984)
$ECI6$ (TATSUOKA, 1984)

3.1 Conjunto de dados

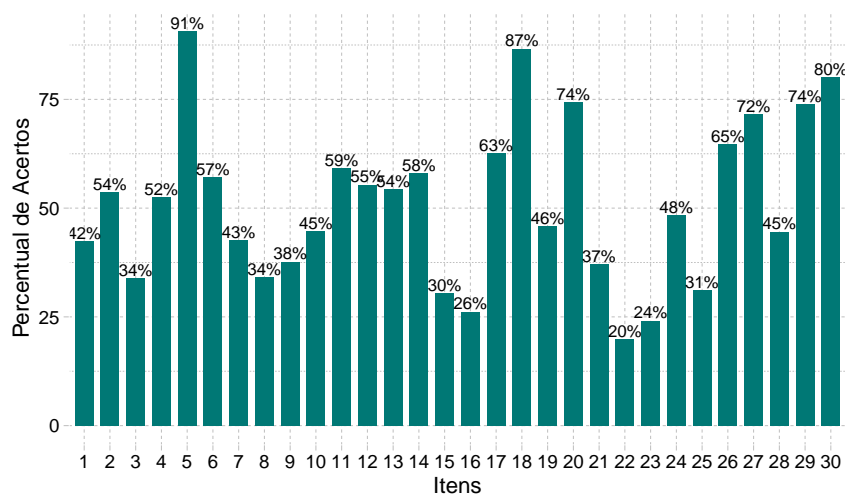
Dados da prova de Língua Portuguesa, com 30 itens, do SARESP (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo) aplicada no ano de 2007, a alunos do terceiro ano matutino de escolas públicas de São Paulo.

4 Resultados

4.1 Análise exploratória dos dados

Foi utilizado o banco de dados do SARESP que apresenta uma amostra de 1001 respondentes de um exame de 30 itens de Língua Portuguesa aplicada a alunos do terceiro ano do turno matutino de escolas públicas de São Paulo em 2007.

Figura 2: Percentual de acertos



A Figura 2 reafirma as especificidades da prova do SARESP, visto que é utilizado na construção dessa avaliação a TRI, de modo que, os itens já foram previamente planejados para a adequação a esse modelo. Nesse sentido, o gráfico demonstra o percentual de respostas positivas para cada um dos 30 itens presentes no banco, visto que cada um dos itens apresenta 4 alternativas sendo somente uma a correta, percebe-se também que os itens estão condizentes com os respondentes uma vez que a maioria dos itens apresentam frequência de acerto distantes das extremidades (0% ou 100%), ou seja, nenhum item foi respondido corretamente ou incorretamente por todos os estudantes.

Apesar disso, alguns itens ainda apresentam um percentual de acertos relativamente altos, em especial o item 5, o que pode ser um indicativo de que estes itens sejam mais fáceis e ao mesmo tempo não possuam um poder de discriminação eficiente para os respondentes. Dessa forma, são itens que podem ser reavaliados para as provas posteriores.

4.2 Modelo Logístico de Três Parâmetros (ML3P)

O ajuste do modelo ajuda a ter uma avaliação mais precisa a respeito dos itens e dos respondentes. Diante disso, os resultados para os parâmetros estimados de discriminação (a), dificuldade do item (b) e acerto ao acaso (c) podem ser verificados na Tabela 2.

Tabela 2: Parâmetros estimados do ML3P (Erros Padrão)

Item	Discriminação (<i>a</i>)	Dificuldade (<i>b</i>)	Acerto ao acaso (<i>c</i>)
1	0,63 (0,16)	1,98 (0,07)	0,24 (0,04)
2	1,89 (0,13)	0,25 (1,09)	0,19 (0,56)
3	1,72 (0,16)	1,11 (0,03)	0,16 (0,04)
4	1,08 (0,05)	0,25 (0,25)	0,14 (0,15)
5	0,94 (0,18)	-2,73 (0,12)	0,04 (0,03)
6	0,57 (0,03)	-0,40 (0,04)	0,04 (0,04)
7	0,81 (0,15)	0,78 (0,08)	0,10 (0,03)
8	0,52 (0,02)	1,35 (0,02)	0,00 (0,07)
9	1,52 (0,06)	1,04 (0,14)	0,17 (0,05)
10	1,47 (0,07)	1,29 (0,04)	0,31 (0,18)
11	0,60 (1,09)	-0,64 (0,4)	0,01 (0,53)
12	0,77 (0,45)	0,97 (1,28)	0,32 (1,11)
13	1,17 (0,61)	-0,11 (0,11)	0,04 (0,42)
14	1,23 (0,71)	0,23 (0,45)	0,24 (0,81)
15	1,75 (0,46)	2,30 (0,49)	0,26 (1,79)
16	1,61 (0,77)	1,71 (0,87)	0,16 (0,64)
17	2,31 (0,51)	0,80 (0,75)	0,49 (0,13)
18	3,14 (2,74)	0,00 (1,86)	0,73 (1,16)
19	0,96 (0,29)	0,43 (0,39)	0,08 (0,80)
20	1,32 (0,31)	0,53 (0,12)	0,59 (0,57)
21	0,58 (0,44)	0,98 (0,42)	0,00 (0,42)
22	2,83 (0,28)	2,20 (0,24)	0,17 (0,30)
23	2,25 (0,30)	2,35 (0,09)	0,21 (0,33)
24	0,80 (0,49)	2,59 (0,14)	0,40 (0,37)
25	0,96 (0,30)	1,00 (0,33)	0,00 (0,98)
26	1,36 (0,51)	-0,25 (0,81)	0,19 (1,15)
27	1,99 (29)	0,61 (0,53)	0,58 (0,09)
28	1,52 (1,42)	0,36 (0,97)	0,07 (0,56)
29	1,22 (0,19)	-1,08 (0,30)	0,00 (0,72)
30	1,36 (0,28)	-0,62 (0,13)	0,43 (0,39)

Ao avaliar o poder de discriminação dos itens, percebe-se que os itens tem valores de discriminação condizentes, visto que, nenhum possui o parâmetro *a* com valores negativos, já os itens 18 e 22 são os que apresentam maiores valores de discriminação, portanto, são os melhores para dividir os alunos em dois grupos: os que possuem proficiência abaixo do parâmetro *b* e os que possuem proficiência acima do parâmetro *b* respectivo a

cada item.

Ao observar os valores de b em comparação com os valores obtidos na Figura 2, verifica-se que não necessariamente os itens mais acertados serão considerados os mais fáceis, devido ao percentual de acerto do item não definir por si só o nível de dificuldade do mesmo, uma vez que outros parâmetros influenciam no cálculo deste, como por exemplo a proficiência dos respondentes que acertaram ou não o item.

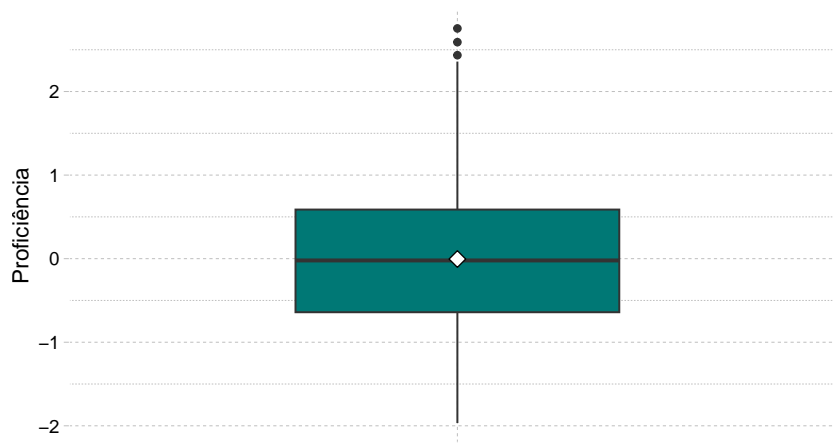
Já para o parâmetro c , ou seja, de indivíduos com baixos níveis de proficiência acertarem os itens, essa probabilidade é estimada com base nos acertos dos itens por pessoas com baixa proficiência, de modo que as estimativas das probabilidades podem estar enviesadas e portanto existe a possibilidade de haver presença de valores mais extremos, próximos a 0 ou a 1.

Desse modo, como os valores observados divergem muito do que era esperado para o acerto ao acaso (0,25), visto que, existem 4 alternativas para cada item, pode-se realizar a modelagem fixando o parâmetro c em 0,25 para todos os itens.

Isto posto, primeiramente serão calculadas as estatísticas de ajuste individual considerando o parâmetro c obtido pelo ML3P e em seguida fixá-lo em 0,25.

4.3 Proficiência

Figura 3: Proficiência Estimada



Quadro 1: Medidas resumo da proficiência estimada para os respondentes

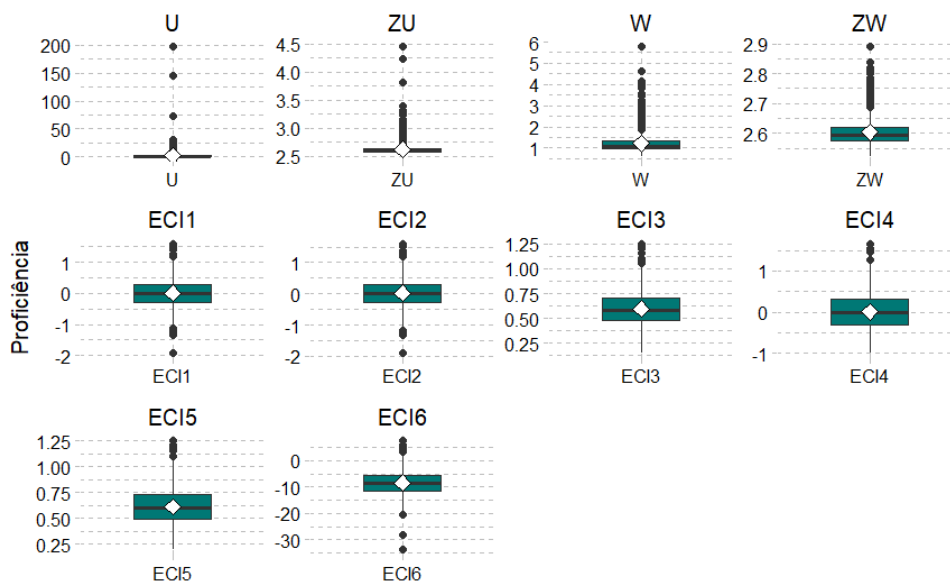
Estatística	Valor
Média	0,06
Desvio Padrão	0,85
Mínimo	-1,87
1º Quartil	-0,55
Mediana	0,07
3º Quartil	0,62
Máximo	2,73

A Figura 3 permite visualizar que as proficiências estimadas demonstram uma distribuição simétrica e observam-se três pontos cujo nível de proficiência foram outliers superiores, ou seja, se destacaram dos demais.

4.4 Estatísticas de Ajuste Individual (EAI)

Após o ajuste do modelo calculou-se as 10 estatísticas para os 1001 respondentes. A Figura 4 apresenta o comportamento de cada estatística por meio de seus boxplots.

Figura 4: Gráfico de boxplot das EAI



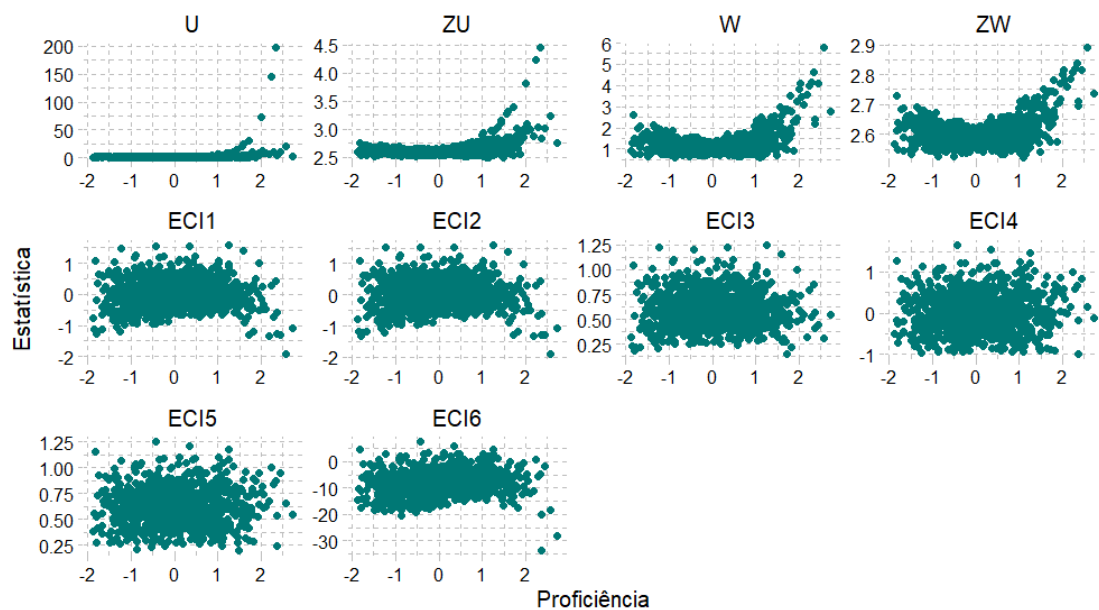
Quadro 2: Medidas resumo das EAI

Estatística	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
Média	1,91	2,62	1,22	2,60	0,01	0,01	0,60	0,01	0,061	-8,46
Desvio Padrão	8,20	0,12	0,48	0,04	0,45	0,45	0,17	0,46	0,18	4,45
Mínimo	0,48	2,50	0,61	2,52	-1,92	-1,91	0,15	-0,99	0,19	-33,87
1º Quartil	0,91	2,57	0,95	2,58	-0,28	-0,28	0,47	-0,31	0,49	-11,45
Mediana	1,11	2,60	1,10	2,59	-0,01	-0,01	0,58	-0,02	0,60	-8,64
3º Quartil	1,45	2,63	1,32	2,62	0,29	0,29	0,70	0,32	0,73	-5,59
Máximo	197,02	4,45	5,77	2,89	1,60	1,60	1,25	1,66	1,25	7,63

Ao analisar a Figura 4, é relevante observar os outliers superiores das estatísticas, visto que, a elevação desses valores são indicativos de que há incoerência no padrão de respostas do indivíduo. Percebe-se também que as estatísticas de quadrado médio ajustado apresentam um maior número de outliers do que as de índice de cautela, portanto, espera-se de forma geral que essas estatísticas identifiquem o maior número de respondentes incoerentes.

A Figura 5, por outro lado, apresenta as estatísticas em função da proficiência que permite verificar quais níveis de proficiências estes respondentes estão alcançando.

Figura 5: Gráfico de dispersão das EAI em função das proficiências



Para as estatísticas de quadrado médio ajustado U, ZU, W e ZW, altos valores observados indicam insistência no padrão de respostas do respondente, de modo que ao analisar a Figura 5 é possível verificar que os indivíduos incoerentes apresentem pro-

ficiências entre 1 e 2. Isso ocorre pois para essas estatísticas são consideradas as diferenças entre respostas certas para itens difíceis e erradas para itens fáceis, ou seja, os alunos que acertaram itens difíceis tiveram suas proficiências alavancadas. Porém, a inconsistência ainda pode ser verificada.

Da mesma forma, as estatísticas de índice de cautela (ECI1 a ECI6) com valores altos indicam respondentes incoerentes, pois ao observar essas estatísticas, as quais consideram a covariância, correlação e variância das probabilidades e proporções de acertos dos itens, verifica-se que aqueles alunos com maiores valores para essas estatísticas apresentam proficiências entre -2 e 2. Dessa forma, espera-se que os respondentes tidos como incoerentes para essas estatísticas não sejam necessariamente os mesmos indicados pelas estatísticas de quadrado médio ajustado.

4.5 Simulação

Para poder inferir quais indivíduos são considerados incoerentes para cada estatística foram realizadas simulações de Monte Carlo. Primeiramente foram geradas 1000 amostras sob H_0 para cada respondente, ou seja, respondentes simulados cujo padrão de resposta é congruente com o modelo de resposta especificado. Na sequência, verificou-se quantos dos valores simulados para as estatísticas eram menores do que as dos respondentes reais e, por fim, foram considerados como incoerentes aqueles indivíduos que apresentaram 5% ou menos valores maiores do que o real.

Tabela 3: Número de indivíduos incoerentes por estatística

Estatística	N(%)
U	236 (24%)
ZU	247 (25%)
W	295 (29%)
ZW	297 (30%)
ECI1	81 (8%)
ECI2	74 (7%)
ECI3	81 (8%)
ECI4	99 (10%)
ECI5	91 (9%)
ECI6	97 (10%)
Todas	41 (4%)

Diante disso, é interessante observar o número total de indivíduos identificados para cada estatística de ajuste individual, de modo que o valor de indivíduos tidos como incoerentes corresponde a cerca de 27% dos respondentes para as estatísticas de quadrado médio ajustado e aproximadamente 8% para as de índice de cautela. Apesar disso, apenas 4% foram tidos como incoerentes por todas as estatísticas em estudo.

Analisando a Tabela 3, verifica-se que as estatísticas de índice de cautela foram as que apresentaram menor número de indivíduos incoerentes. As demais estatísticas classificaram mais de 20% dos respondentes como incoerentes.

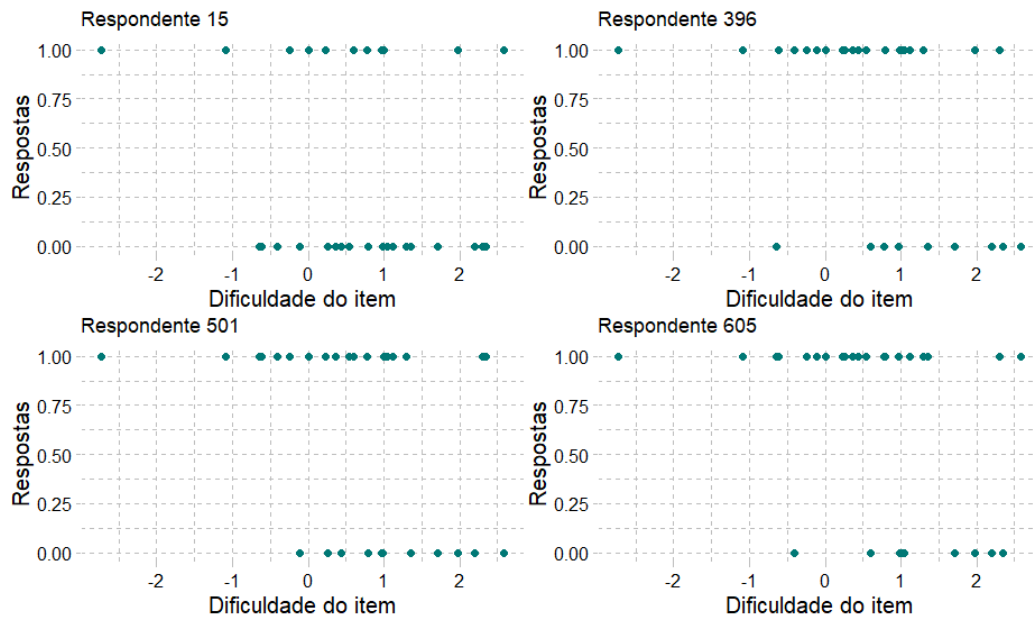
Assim, pode-se definir como menos rigorosas as estatísticas de quadrado médio ajustado, visto que, estas apresentam menos critérios para identificação do respondente, rejeitando H_0 mais facilmente. Por outro lado, as estatísticas de índice de cautela, apresentam maior rigor, já que, estas apresentam maior dificuldade em rejeitar a hipótese nula.

Na Tabela 4 serão apresentados tabelas com os resultados das estatísticas para alguns dos alunos coerentes e em seguida o gráfico de suas respostas em relação ao nível de dificuldade do item estimado pelo ML3P.

Tabela 4: EAI para respondentes coerentes

Respondente	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
Resp.15	0,96	2,58	0,96	2,58	-0,08	-0,09	0,59	-0,11	0,57	-10,80
Resp.396	0,96	2,58	1,05	2,59	0,28	0,27	0,67	0,16	0,66	-6,85
Resp.501	1,20	2,61	1,28	2,62	0,24	0,24	0,67	0,22	0,70	-5,10
Resp.605	1,07	2,59	1,12	2,60	-0,10	-0,11	0,53	-0,04	0,58	-9,94

Figura 6: Respostas de indivíduos coerentes

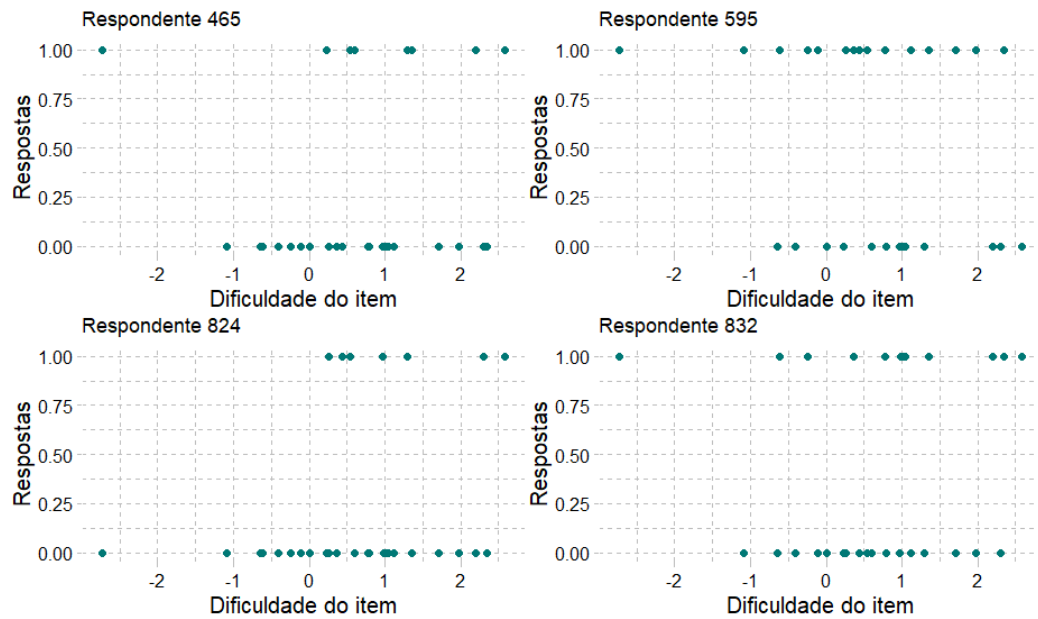


A Figura 6 apresenta os gráficos de acerto e erro em relação ao nível de dificuldade estimado dos itens, onde atribui-se 1 ao item acertado e 0 ao item errado, possibilitando dessa forma visualizar que, de modo geral, esses alunos identificados como coerentes erram itens considerados mais difíceis e acertam aqueles considerados mais fáceis, já os considerados incoerentes erraram os itens fáceis e acertaram os itens difíceis o que explicita valores maiores para suas EAI calculadas. Assim, denota-se que esse comportamento condiz com os valores obtidos para cada uma das estatísticas apresentadas nas Tabela 6.

Tabela 5: EAI para respondentes incoerentes

Respondente	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
Resp.465	2,05	2,69	1,54	2,64	0,72	0,74	0,88	0,86	0,94	-0,32
Resp.595	1,57	2,64	1,50	2,64	0,82	0,78	0,92	1,06	1,02	1,65
Resp.824	6,96	2,93	2,37	2,71	1,07	1,08	1,04	1,23	1,10	3,12
Resp.832	2,04	2,69	1,70	2,66	1,54	1,52	1,22	1,55	1,21	5,69

Figura 7: Respostas de indivíduos incoerentes



Retomando ao contexto dos respondentes identificados como incoerentes, percebe-se que suas respostas não seguem uma ordem lógica de acerto e erro em função do nível de dificuldade do item. Em alguns casos, as respostas apresentam um comportamento inverso ao esperado, no qual o indivíduo apresenta maior número de respostas positivas para itens mais difíceis e menor para mais fáceis, como o indivíduo 595. Outro indicador é a aleatoriedade nas respostas do indivíduo, ou seja, acertos e erros distribuídos igualmente ao longo de todo o parâmetro de dificuldade dos itens, como pode ser observado para os demais indivíduos exemplificados, além disso, verifica-se a proeminência das estatísticas quando comparados aos valores obtidos para os respondentes coerentes.

As Tabelas 6 e 7 apresentam o número absoluto e o percentual de indivíduos incoerentes em comum entre as estatísticas, para que seja possível verificar possíveis relações entre elas:

Tabela 6: Número de indivíduos incoerentes em comum entre as estatísticas

Estatística	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
U	236	232	213	214	53	53	53	72	68	70
ZU	232	247	221	221	55	53	55	75	71	73
W	213	221	295	291	61	57	63	84	80	84
ZW	214	221	291	297	59	56	61	82	79	82
ECI1	53	55	61	59	81	71	74	72	64	69
ECI2	53	53	57	56	71	74	72	67	61	66
ECI3	53	55	63	61	74	72	81	71	67	72
ECI4	72	75	84	82	72	67	71	99	84	89
ECI5	68	71	80	79	64	61	67	84	91	85
ECI6	70	73	84	82	69	66	72	89	85	97

Tabela 7: Percentual de indivíduos incoerentes em comum entre as estatísticas

Estatística	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
U	100%	98%	90%	91%	22%	22%	22%	31%	29%	30%
ZU	94%	100%	89%	89%	22%	21%	22%	30%	29%	30%
W	72%	75%	100%	99%	21%	19%	21%	28%	27%	28%
ZW	72%	74%	98%	100%	20%	19%	21%	28%	27%	28%
ECI1	65%	68%	75%	73%	100%	88%	91%	89%	79%	85%
ECI2	72%	72%	77%	76%	96%	100%	97%	91%	82%	89%
ECI3	65%	68%	78%	75%	91%	89%	100%	88%	83%	89%
ECI4	73%	76%	85%	83%	73%	68%	72%	100%	85%	90%
ECI5	75%	78%	88%	87%	70%	67%	74%	92%	100%	93%
ECI6	72%	75%	87%	85%	71%	68%	74%	92%	88%	100%

Como esperado as estatísticas de quadrado médio ajustado apresentaram maior intersecção entre si, na mesma medida em que, as de índice de cautela trouxeram resultados parecidos.

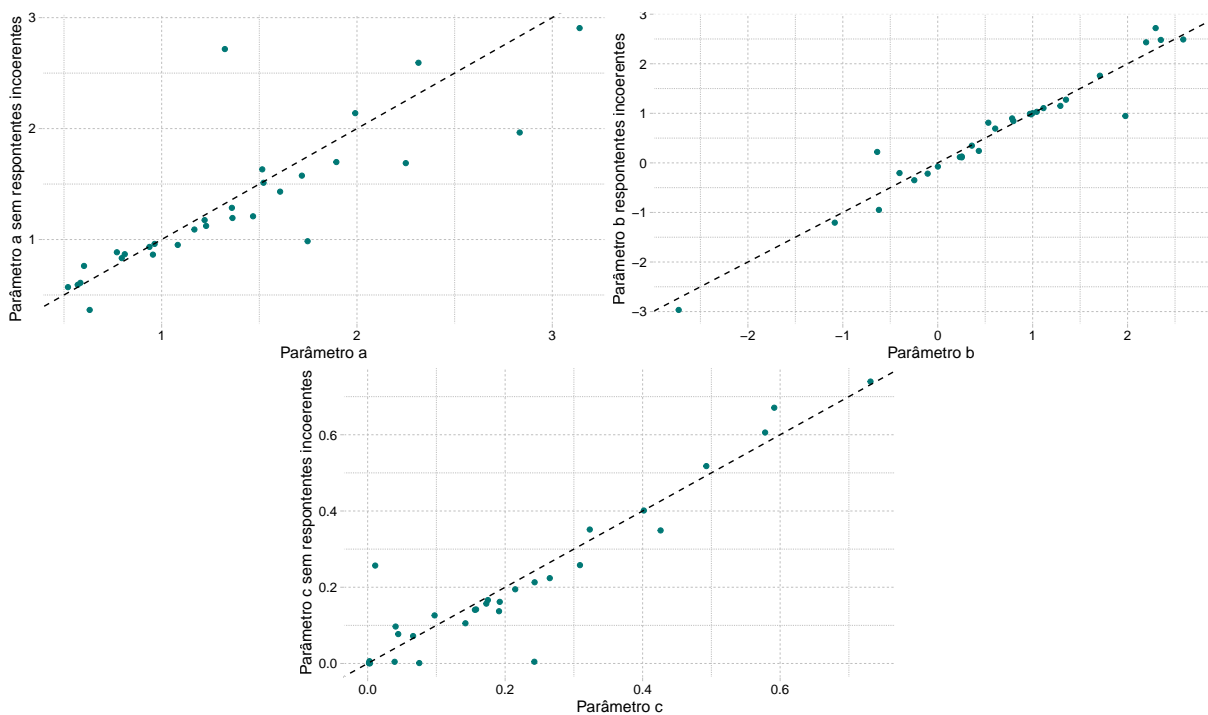
Outro ponto a ser observado são as estatísticas ZU, ZW, ECI4 e ECI6, as quais apresentaram maiores percentuais de intersecção. Essas estatísticas apresentam coeficientes mais precisos para a identificação dos respondentes incoerentes.

Apesar disso, também é interessante ressaltar que nenhuma estatística divergiu muito das demais, ou seja, todas elas identificaram respondentes em comum o que ajuda para a validação dos critérios de cada estatística.

4.6 Remodelagem

Após o estudo das estatísticas de ajuste individual, foram verificados que 41 respondentes foram tidos como incoerentes para todas as estatísticas, sendo assim, retirados da base de dados e submetendo o restante a uma nova modelagem, obtendo os seguintes parâmetros para o ML3P expostos na Tabela 8. Além disso, a Figura 8 apresenta gráficos comparando os parâmetros do modelo inicial com o modelo sem os respondentes incoerentes.

Figura 8: Gráficos de de dispersão dos parâmetros do ML3P com os parâmetros re-estimados



A partir dos gráficos da Figura 8, verifica-se que houve uma diferença principalmente nos parâmetros de discriminação e de chance de acerto ao acaso, visto que os pontos não apresentam um comportamento retilíneo como era esperado, ilustrado pela linha tracejada, representando a igualdade entre os parâmetros. Por outro lado, o parâmetro de dificuldade reverbera esse comportamento, exceto para os valores superiores de dificuldade que apresentaram distâncias maiores da linha central.

Tabela 8: Parâmetros estimados sem os respondentes incoerentes do ML3P (Erros Padrão)

Item	Discriminação (a)	Dificuldade (b)	Acerto ao acaso (c)
1	0,37 (0,07)	0,95 (0,09)	0,00 (0,04)
2	1,70 (0,2)	0,13 (2,70)	0,14 (0,55)
3	1,58 (0,12)	1,10 (0,01)	0,14 (0,04)
4	0,95 (0,08)	0,11 (0,28)	0,11 (0,13)
5	0,93 (0,14)	-2,97 (0,15)	0,08 (0,05)
6	0,59 (0,04)	-0,21 (0,04)	0,10 (0,05)
7	0,87 (0,03)	0,90 (0,04)	0,13 (0,02)
8	0,57 (0,02)	1,27 (0,02)	0,00 (0,07)
9	1,51 (0,00)	1,03 (0,16)	0,17 (0,04)
10	1,21 (0,06)	1,15 (0,13)	0,26 (0,30)
11	0,76 (0,20)	0,22 (0,36)	0,26 (0,49)
12	0,89 (0,56)	0,99 (3,17)	0,35 (1,24)
13	1,09 (0,53)	-0,22 (0,09)	0,00 (0,40)
14	1,12 (0,71)	0,12 (0,94)	0,21 (0,83)
15	0,98 (0,33)	2,72 (0,53)	0,22 (1,17)
16	1,43 (0,69)	1,76 (1,05)	0,14 (0,61)
17	2,59 (0,10)	0,84 (1,82)	0,52 (0,09)
18	2,91 (1,79)	-0,07 (1,33)	0,74 (1,15)
19	0,86 (0,09)	0,24 (0,40)	0,00 (0,91)
20	2,72 (0,30)	0,81 (0,21)	0,67 (0,71)
21	0,61 (0,09)	0,98 (0,37)	0,00 (0,39)
22	1,96 (0,30)	2,43 (0,41)	0,16 (0,37)
23	1,69 (0,28)	2,48 (0,09)	0,19 (0,32)
24	0,83 (0,46)	2,49 (0,39)	0,40 (0,43)
25	0,96 (0,21)	1,00 (0,33)	0,00 (0,61)
26	1,29 (0,46)	-0,35 (0,99)	0,16 (1,06)
27	2,14 (0,10)	0,69 (1,74)	0,61 (0,09)
28	1,63 (0,95)	0,35 (0,73)	0,07 (0,57)
29	1,17 (0,11)	-1,21 (0,29)	0,01 (0,83)
30	1,19 (0,30)	-0,95 (0,15)	0,35 (0,37)

Após o reajuste foram calculadas novamente as estatísticas de ajuste individual para os respondentes e ao comparar os parâmetros do novo modelo com os da Tabela 2 verifica-se que o parâmetro de discriminação não apresentou grandes mudanças em seus valores, porém todos foram reduzidos. Já com relação ao nível de dificuldade dos itens, percebe-se que em geral aumentaram e, em alguns como o item 11, essa diferença foi

grande, pois anteriormente esse item apresentava $b = -0,64$ e passou para $b = 0,22$. Já o parâmetro de acerto ao acaso não variou tanto quanto o esperado já que, para alguns itens, se manteve o mesmo valor, como para o item 4, mas para outros, houve uma diferença significativa, a exemplo do item 1.

Portanto esse reajuste auxilia no real entendimento dos parâmetros para o modelo e colabora fundamentalmente para obtenção de um parâmetro mais preciso.

4.7 ML3P com parâmetro de acerto ao acaso fixado

Como visto na Tabela 2, o parâmetro c não apresentou valores próximos ao esperado de 0,25. Visto isso, foi realizado o ajuste do modelo fixando o parâmetro $c = 0,25$ de forma que possa ser realizada uma comparação com os valores estimados pelo modelo. A Tabela 9 apresenta a estimação dos parâmetros a e b , com o critério adotado para o acerto ao acaso.

Tabela 9: Parâmetros estimados sem os respondentes incoerentes do ML3P (Erros Padrão)

Item	Discriminação (<i>a</i>)	Dificuldade (<i>b</i>)	Acerto ao acaso (<i>c</i>)
1	0,65 (0,14)	2,02 (0,20)	0,25
2	2,31 (0,67)	0,40 (0,14)	0,25
3	2,85 (0,17)	1,27 (0,10)	0,25
4	1,35 (0,19)	0,55 (0,24)	0,25
5	1,04 (0,35)	-2,23 (0,17)	0,25
6	0,75 (0,10)	0,44 (0,10)	0,25
7	1,23 (0,14)	1,22 (0,11)	0,25
8	1,00 (0,65)	2,18 (1,17)	0,25
9	2,02 (0,10)	1,21 (0,15)	0,25
10	1,06 (0,17)	1,23 (0,09)	0,25
11	0,76 (0,22)	0,26 (40,53)	0,25
12	0,67 (0,00)	0,64 (0,10)	0,25
13	1,66 (0,46)	0,39 (0,11)	0,25
14	1,21 (0,09)	0,26 (0,26)	0,25
15	1,34 (0,12)	2,45 (0,12)	0,25
16	2,90 (0,15)	1,81 (0,33)	0,25
17	0,92 (0,57)	0,01 (0,18)	0,25
18	1,15 (0,18)	-1,64 (0,12)	0,25
19	1,41 (0,19)	0,91 (0,20)	0,25
20	0,64 (0,33)	-1,10 (0,19)	0,25
21	1,08 (0,12)	1,79 (0,12)	0,25
22	11,25 (0,21)	2,18 (0,16)	0,25
23	8,37 (0,45)	2,18 (0,76)	0,25
24	0,33 (0,13)	2,49 (0,18)	0,25
25	2,12 (0,20)	1,51 (0,11)	0,25
26	1,48 (0,20)	-0,10 (19,73)	0,25
27	0,78 (0,00)	-0,70 (0,12)	0,25
28	2,37 (0,37)	0,73 (0,18)	0,25
29	1,49 (0,12)	-0,59 (0,33)	0,25
30	1,16 (0,18)	-1,09 (0,16)	0,25

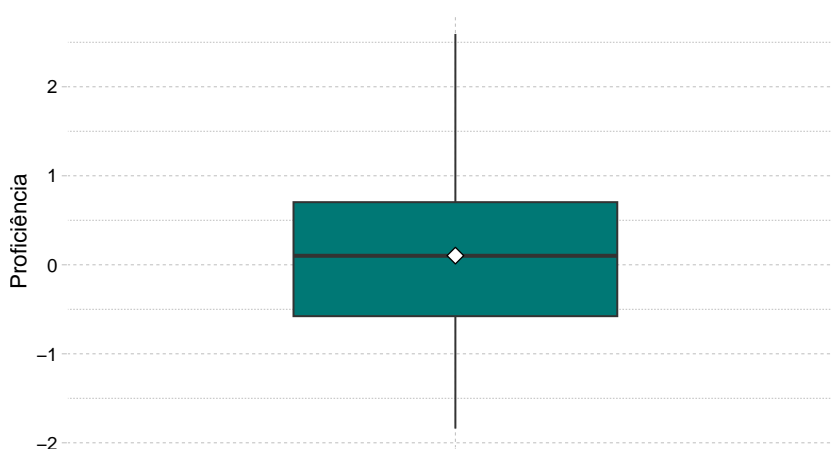
Percebe-se que os parâmetros de discriminação e de dificuldade divergem daqueles estimados na Tabela 2. Para o parâmetro de discriminação *a*, o item 22 continua sendo aquele que melhor discrimina, apesar disso o item 18 reduziu significativamente sua capacidade discriminativa, visto que, anteriormente este apresentava $c = 0,74$ e passou para $c = 0,25$, afetando diretamente na capacidade discriminativa. Por outro lado outros

itens passaram a ter sua capacidade discriminativa maior, como é o caso do item 23 que passou a ser o segundo melhor item para discriminação.

Quando se compara a dificuldade dos itens também houveram mudanças, porém estas foram menos significativas. Já para a Tabela 2, os itens com maior dificuldade eram os itens 24 e 23. Da mesma forma, na Tabela 9 estes itens ainda são os mais difíceis.

4.8 Proficiência com parâmetro de acerto ao acaso fixado

Figura 9: Proficiência Estimada com parâmetro de acerto ao acaso fixado



Quadro 3: Medidas resumo da proficiência estimada para os respondentes

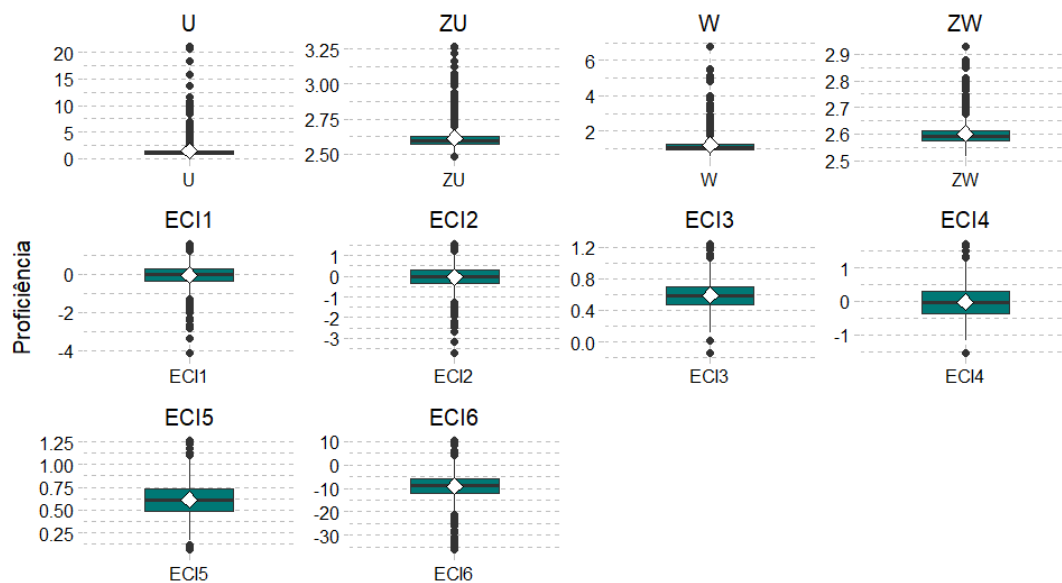
Estatística	Valor
Média	0,10
Desvio Padrão	0,86
Mínimo	-1,84
1º Quartil	-0,58
Mediana	0,10
3º Quartil	0,70
Máximo	2,59

Ao comparar as proficiências estimadas verifica-se que não houve diferença significativa na distribuição da variável. Apesar disso, percebe-se que agora não há a presença dos outliers verificados na Figura 3.

4.9 EAI com parâmetro de acerto ao acaso fixado

Observa-se o comportamento dos novos valores obtidos para as estatísticas de ajuste individual por meio de seus boxplots na Figura 10.

Figura 10: Gráfico de boxplot das EAI com parâmetro de acerto ao acaso fixado



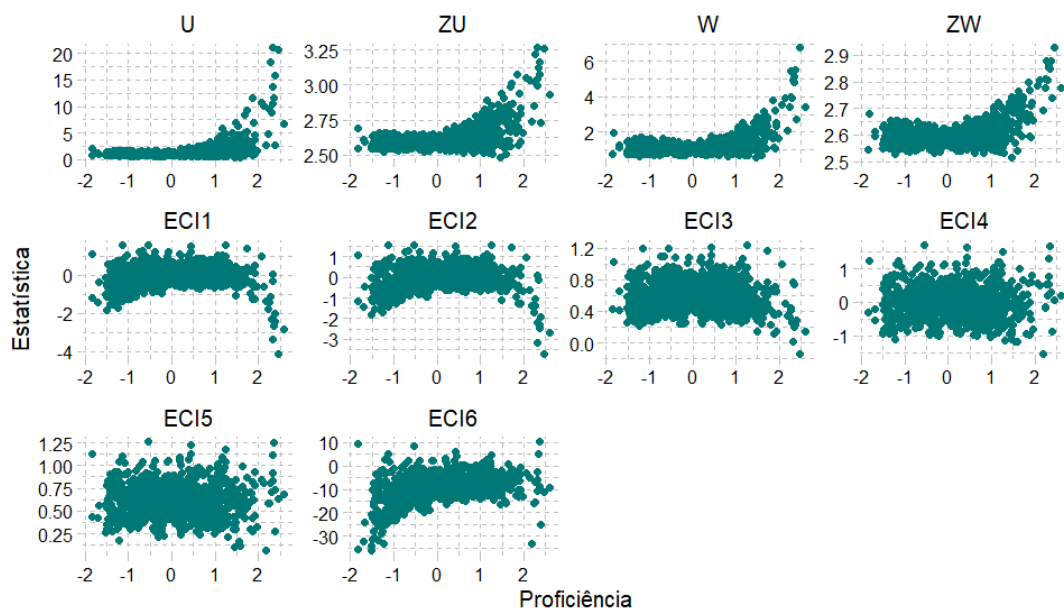
Quadro 4: Medidas resumo das EAI com parâmetro de acerto ao acaso fixado

Estatística	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
Média	1,45	2,61	1,21	2,60	-0,05	-0,05	0,59	-0,02	0,61	-9,19
Desvio Padrão	1,63	0,08	0,54	0,05	0,55	0,54	0,18	0,48	0,18	5,67
Mínimo	0,41	2,48	0,57	2,52	-4,15	-3,73	-0,14	-1,54	0,06	-36,24
1º Quartil	0,91	2,57	0,94	2,57	-0,34	-0,34	0,47	-0,35	0,49	-11,94
Mediana	1,09	2,59	1,08	2,59	-0,03	-0,03	0,58	-0,04	0,60	-8,78
3º Quartil	1,35	2,62	1,27	2,61	0,28	0,29	0,70	0,29	0,73	-5,79
Máximo	21,03	3,26	6,76	2,93	1,60	1,58	1,24	1,71	1,26	10,37

Em geral as estatísticas aparentam ter menores variâncias e médias. Isso é impacto dos valores estimados estarem menores que os vistos no Quadro 4. Ademais, ainda é perceptível, pela Figura 10, um grande número de outliers, de forma que ainda são esperados muitos respondentes incoerentes. O comportamento é similar ao observado anteriormente na Figura 4 e no Quadro 2, ou seja, maior número de outliers para as estatísticas de quadrado médio ajustado e um valor menor para as de índice de cautela.

Além disso, pode-se observar também os valores em função das novas proficiências estimadas pela Figura 11.

Figura 11: Gráfico de dispersão das EAI em função das proficiências com parâmetro de acerto ao acaso fixado



A partir da comparação entre as Figuras 5 e 11, observa-se que o comportamento das estatísticas, ao longo da proficiência estimada, variou bastante do ML3P para o modelo com parâmetro c fixado, visto que, a mudança da chance de acerto ao acaso implica diretamente no cálculo de cada estatística, além do cálculo da proficiência. Para as estatísticas U e ZU , percebe-se uma redução nos valores do eixo Y , isso permite uma melhor visualização dos postos extremos pelo gráfico, sendo assim mais fácil verificar os níveis de proficiência nos quais os respondentes incoerentes estão (de 1 a 2).

Por conseguinte, as estatísticas W e ZW foram as que apresentaram menor diferença de comportamento quando comparadas entre os modelos, já em relação as estatísticas de índice de cautela, percebe-se que muitas estatísticas tiveram valores menores para o modelo com parâmetro c fixado, principalmente para altas proficiências nas estatísticas $ECI1$ e $ECI2$ e $ECI6$. Além disso, a estatística $ECI6$ também apresentou essa redução nos valores para proficiências mais baixas, entre -2 e -1 .

4.10 Simulação das EAI com parâmetro de acerto ao acaso fixado

Assim como feito para o modelo inicial, foram simulados valores para as estatísticas de ajuste individual sob o modelo H_0 , que para este caso é no ML3P com o parâmetro c igual a 0,25.

A Tabela 10 apresenta o número de indivíduos identificados como incoerentes para cada uma das 10 estatísticas em estudo.

Tabela 10: Número de indivíduos incoerentes por estatística

Estatística	N(%)
U	202 (20%)
ZU	240 (24%)
W	266 (27%)
ZW	310 (31%)
ECI1	74 (7%)
ECI2	77 (8%)
ECI3	68 (7%)
ECI4	89 (9%)
ECI5	89 (9%)
ECI6	83 (8%)
Todas	31 (3%)

Quando comparado o número de respondentes identificados como incoerentes para cada estatística presente na Tabela 10 com os valores da Tabela 3, verifica-se que a diferença percentual não foi muito elevada para nenhuma estatística e isso se afirma na diferença percentual máxima de apenas 4%, como explicitado anteriormente. Contudo, nas estatísticas do novo modelo em ZW e ECI2 notou-se unicidade no que tange à identificação de mais respondentes incoerentes e isso decorre pois, houve aumento de um ponto percentual no número desses indivíduos.

Por outro lado, no âmbito geral, o número de indivíduos identificados como incoerentes foi menor, passando de 41 indivíduos identificados por todas as estatísticas do modelo inicial para apenas 31, uma redução de 1% na base de dados.

Adiante, as Tabelas 11 e 12 apresentam as intersecções absolutas e percentuais de indivíduos incoerentes para as estatísticas entre si, respectivamente.

Tabela 11: Número de indivíduos incoerentes em comum entre as EAI

Estatística	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
U	202	176	180	178	48	51	47	62	63	59
ZU	176	240	179	217	47	49	45	61	61	57
W	180	179	266	246	56	60	53	75	77	70
ZW	178	217	246	310	57	59	54	75	76	71
ECI1	48	47	56	57	74	67	59	59	56	57
ECI2	51	49	60	59	67	77	64	62	60	59
ECI3	47	45	53	54	59	64	68	57	57	57
ECI4	62	61	75	75	59	62	57	89	78	76
ECI5	63	61	77	76	56	60	57	78	89	80
ECI6	59	57	70	71	57	59	57	76	80	83

Tabela 12: Percentual de indivíduos incoerentes em comum entre as estatística

Estatística	U	ZU	W	ZW	ECI1	ECI2	ECI3	ECI4	ECI5	ECI6
U	100%	87%	89%	88%	24%	25%	23%	31%	31%	29%
ZU	73%	100%	75%	90%	20%	20%	19%	25%	25%	24%
W	68%	67%	100%	92%	21%	23%	20%	28%	29%	26%
ZW	57%	70%	79%	100%	18%	19%	17%	24%	25%	23%
ECI1	65%	64%	76%	77%	100%	91%	80%	80%	76%	77%
ECI2	66%	64%	78%	77%	87%	100%	83%	81%	78%	77%
ECI3	69%	66%	78%	79%	87%	94%	100%	84%	84%	84%
ECI4	70%	69%	84%	84%	66%	70%	64%	100%	88%	85%
ECI5	71%	69%	87%	85%	63%	67%	64%	88%	100%	90%
ECI6	71%	69%	84%	86%	69%	71%	69%	92%	96%	100%

Nesse sentido, ao comparar os valores presentes nas Tabelas 6 e 7, verifica-se que em geral, o valor absoluto de intersecções foi menor, justamente pelo fato do número de indivíduos identificados ser inferior de forma geral. Mas apesar disso, nota-se que, para as intersecções das estatísticas de quadrado médio ajustado entre si, se anteriormente apresentavam percentuais variando de 72% a 98%, passaram a variar de 57% a 89%, portanto uma expressiva redução na intersecção.

Por outro lado, as intersecções das estatísticas de índice de cautela passaram de uma variação de 65% a 92% para uma variação de 63% a 96%, ou seja, não houve grande mudança. O mesmo ocorre quando se compara a variação no percentual de intersecção entre os grupos das estatísticas (quadrado médio ajustado e índice de cautela), observa-se que esta passou de 19% a 30% para 17% a 31%.

A Tabela 13 apresenta o número de indivíduos em comum identificados para cada estatística entre os modelos, ou seja, os indivíduos identificados pelo modelo inicial que também foram identificados pelo modelo com parâmetro c fixado.

Tabela 13: Número de indivíduos incoerentes por estatística em comum entre os modelos

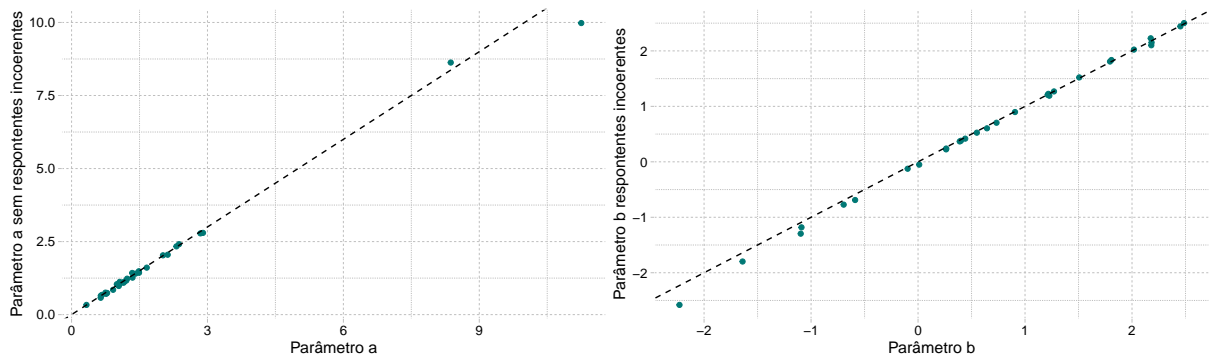
Estatística	N
U	184
ZU	223
W	246
ZW	290
ECI1	70
ECI2	65
ECI3	63
ECI4	75
ECI5	72
ECI6	74
Todas	30

Ao observar o número de respondentes incoerentes em comum para cada estatística entre os modelos, verifica-se que os indivíduos encontrados não foram todos os mesmos, porém uma proporção considerável foi igual, portanto, espera-se que haja um impacto na estimação dos parâmetros sem esses respondentes, da mesma forma que ocorreu para o ML3P. Dos 37 indivíduos identificados por todas as estatísticas no segundo modelo, 30 foram identificados por todas as estatísticas no modelo inicial, ou seja, esses indivíduos de fato apresentam respostas incoerentes. Portanto, há um impacto significativo no ajuste do modelo, visto que, cerca de 25% dos respondentes que eram classificados como incoerentes no modelo inicial não foram identificados pelo modelo com parâmetro c fixado.

4.11 Remodelagem com parâmetro de acerto ao acaso fixado

Para a remodelagem, novamente excluem-se os indivíduos incoerentes que, para este modelo, totalizam 30. Os resultados do reajuste são observados na Tabela 14, além da comparação com os parâmetros do modelo contendo os respondentes incoerentes pela Figura 12.

Figura 12: Gráficos de de dispersão dos parâmetros do ML3P com c fixado versus os parâmetros re-estimados



Observa-se que foram poucos os itens em que houve divergências na estimativa dos parâmetros, visto que, a maioria dos pontos permanece na reta que representa a igualdade dos parâmetros.

Tabela 14: Parâmetros estimados sem os respondentes incoerentes do ML3P com parâmetro de acerto ao acaso fixado (Erros Padrão)

Item	Discriminação (<i>a</i>)	Dificuldade (<i>b</i>)	Acerto ao acaso (<i>c</i>)
1	0,65 (0,15)	2,02 (0,20)	0,25
2	2,34 (0,65)	0,38 (0,13)	0,25
3	2,78 (0,18)	1,27 (0,10)	0,25
4	1,27 (0,19)	0,53 (0,24)	0,25
5	0,98 (0,35)	-2,58 (0,18)	0,25
6	0,75 (0,09)	0,42 (0,10)	0,25
7	1,23 (0,14)	1,23 (0,11)	0,25
8	1,04 (0,70)	2,10 (1,12)	0,25
9	2,03 (0,10)	1,21 (0,15)	0,25
10	1,12 (0,17)	1,19 (0,09)	0,25
11	0,71 (0,22)	0,22 (22,80)	0,25
12	0,68 (0,00)	0,60 (0,11)	0,25
13	1,61 (0,44)	0,37 (0,11)	0,25
14	1,17 (0,09)	0,24 (0,26)	0,25
15	1,43 (0,12)	2,44 (0,12)	0,25
16	2,80 (0,15)	1,84 (0,33)	0,25
17	0,85 (0,56)	-0,05 (0,17)	0,25
18	1,09 (0,19)	-1,80 (0,12)	0,25
19	1,40 (0,20)	0,90 (0,20)	0,25
20	0,58 (0,33)	-1,30 (0,20)	0,25
21	1,07 (0,12)	1,81 (0,12)	0,25
22	9,98 (0,20)	2,23 (0,15)	0,25
23	8,63 (0,46)	2,15 (0,73)	0,25
24	0,33 (0,13)	2,50 (0,18)	0,25
25	2,05 (0,20)	1,52 (0,11)	0,25
26	1,48 (0,20)	-0,12 (10,98)	0,25
27	0,74 (0,00)	-0,77 (0,12)	0,25
28	2,41 (0,36)	0,70 (0,18)	0,25
29	1,43 (0,12)	-0,69 (0,34)	0,25
30	1,13 (0,18)	-1,18 (0,16)	0,25

Ao comparar os resultados da Tabela 14 com os da Tabela 9, percebe-se que quase não há impacto nos parâmetros *a* e *b*, dado que os valores obtidos no ML3P são praticamente os mesmos dos obtidos no modelo com *c* fixado.

Por conseguinte, a remoção dos respondentes incoerentes não trouxe o impacto

esperado no modelo de forma que, não é necessária a retirada desses indivíduos quando o parâmetro c for fixado. Apesar disso, entende-se que não são todos os casos em que a chance de acerto ao acaso é 0,25, visto que, diversos outros fatores influenciam na hora da marcação de uma alternativa cuja o respondente não conhece a resposta. Portanto, apesar do modelo com o parâmetro c fixado anular o efeito dos respondentes incoerentes, é importante uma avaliação dos itens para poder definir se essa fixação é válida.

5 Conclusão

5.1 Considerações finais

Diante do exposto, a análise realizada revelou uma série de descobertas significativas. De antemão, foi ajustado o Modelo Logístico de 3 Parâmetros e apurou-se os parâmetros obtidos para o modelo, ao analisar que o parâmetro de acerto ao acaso c divergia muito do esperado de 0,25 para a maioria dos itens. Dessa forma, optou-se por duas análises sobre os respondentes incoerentes: a primeira considerando o modelo com parâmetros já estimados e a segunda, o modelo com o parâmetro c fixado em 0,25 para todos os itens.

Para a primeira análise, foram calculadas as estatísticas de ajuste individual e observou-se que as de quadrado médio ajustado consideravam um maior número de respondentes como incoerentes, cerca de 27%. Entretanto, as de índice de cautela apresentaram um percentual relativamente menor, correspondendo a cerca de 9% e ambas em conjunto apontaram para 4% dos dados como incoerentes. Logo, as estatísticas de índice de cautela levam a testes de hipótese menos poderosos do que as de quadrado médio ajustado, devido a apresentar maior dificuldade para rejeitar a hipótese nula de que as respostas são coerentes.

Dentro dessa perspectiva, após essa identificação, os parâmetros do modelo foram re-estimados sem os 41 respondentes incoerentes, para que fosse possível comparar o efeito desses respondentes nos parâmetros estimados pelo ML3P e com isso verificou-se um severo impacto nos índices de dificuldade dos itens além de um impacto menor na capacidade de discriminação. Em contrapartida, não houve mudanças significativas nas chances de acerto ao acaso.

Na esfera da segunda análise, realizou-se o mesmo procedimento para o Modelo Logístico de 3 Parâmetros com $c = 0,25$, de modo a promover uma sutil redução (1%) no índice geral de respondentes tidos como incoerentes. Ademais, apenas 3% dos dados foram considerados incoerentes por todas as estatísticas, portanto, a retirada destes indivíduos quase não impactou em mudanças para a estimação de novos parâmetros. Sendo assim, comprova-se que a fixação do parâmetro c reduz o impacto dos respondentes incoerentes na estimação dos outros dois parâmetros (a e b).

Outrossim, quando se observa a intersecção dos indivíduos incoerentes entre as estatísticas de ajuste individual, nota-se que para ambos os modelos há uma maior in-

tersecção das estatísticas em seus respectivos grupos e um menor percentual quando são realizadas entre os grupos. Além disso, verificou-se que essa convergência entre grupos juntamente com as estatísticas de índice de cautela, não apresentaram grandes variações em seus percentuais de intersecção. Em contrapartida, as intersecções das estatísticas de quadrado médio ajustado reverberam uma redução expressiva em sua variação (passando de 72% a 98% para 57% a 89%).

Referências

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

DEPRESBITERIS, L. *O desafio da avaliação da aprendizagem: dos fundamentos a uma proposta inovadora*. [S.l.]: EPU, 1989.

KARABATSOS, G. Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, Taylor & Francis, v. 16, n. 4, p. 277–298, 2003.

SANTOS, T. S. B. Um novo método para avaliação da coerência de respostas individuais em provas de múltipla escolha. 2022.

TATSUOKA, K. K. Caution indices based on item response theory. *Psychometrika*, Springer, v. 49, p. 95–110, 1984.

TENDEIRO, J. N. Package ‘perfit’. *CRAN*, 2021.

TENDEIRO, J. N.; MEIJER, R. R.; NIESSEN, A. S. M. Perfit: An r package for person-fit analysis in irt. *Journal of Statistical Software*, v. 74, p. 1–27, 2016.

WRIGHT, B. D.; STONE, M. H. Best test design. Mesa press, 1979.

WRIGHT, J. Monika thiel-horstmann: Sadani-lieder: Studien zu einer nordindischen volksliedliteratur.(neuindische studien, bd. 5.) 276 pp. wiesbaden: Otto harrassowitz, 1978. dm 106. *Bulletin of the School of Oriental and African Studies*, Cambridge University Press, v. 43, n. 2, p. 421–422, 1980.

Apêndice

A Código da Programação em R

Código em R utilizado para a confecção do trabalho.