

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM MODELAGEM COMPUTACIONAL DO
CONHECIMENTO

STEFAN CAVALCANTE GOMES

**DESEMPENHO DE COTISTAS E NÃO-COTISTAS EM PROVA DE VESTIBULAR:
UMA ANÁLISE PSICOMÉTRICA**

Maceió-AL

Novembro de 2020

STEFAN CAVALCANTE GOMES

**DESEMPENHO DE COTISTAS E NÃO-COTISTAS EM PROVA DE VESTIBULAR:
UMA ANÁLISE PSICOMÉTRICA**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Modelagem Computacional do Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador: Jorge Artur Peçanha de Miranda Coelho

Maceió-AL

Novembro de 2020

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

G633d Gomes, Stefan Cavalcante.

Desempenho de cotistas e não-cotistas em prova de vestibular: uma análise psicométrica / Stefan Cavalcante Gomes. – 2020.
69 f. : il., figs. e tabs. color.

Orientador: Jorge Arthur Peçanha de Miranda Coelho.
Dissertação (Mestrado em Modelagem Computacional do Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2020.

Bibliografia: f. 47-49.
Apêndices: f. 50-63.

1. Cotas. 2. Vestibular. 3. Funcionamento Diferencial dos Itens (DIF).
4. Teoria de Resposta ao Item (TRI). I. Título.

CDU: 378

Folha de Aprovação

STEFAN CAVALCANTE GOMES
DESEMPENHO DE COTISTAS E NÃO-COTISTAS EM PROVA DE VESTIBULAR: UMA
ANÁLISE PSICOMÉTRICA

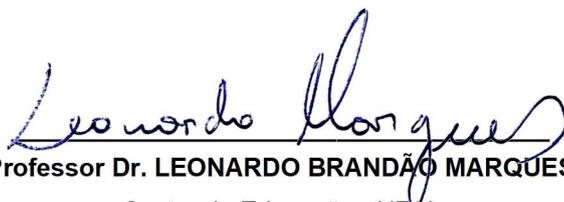
Dissertação submetida ao corpo docente do
Programa de Pós-Graduação em Modelagem
Computacional de Conhecimento da
Universidade Federal de Alagoas e aprovada
em 30 de NOVEMBRO de 2020.



Prof. Dr. JORGE ARTUR PEÇANHA DE MIRANDA COELHO

Faculdade de Medicina - UFAL
Orientador

Banca Examinadora:



Professor Dr. LEONARDO BRANDÃO MARQUES

Centro de Educação - UFAL
Examinador interno



Professor Dr. LEOGILDO ALVES FREIRES

Instituto de Psicologia - UFAL
Examinador externo

*Aos meus filhos, Lara e Vitor,
as maiores alegrias da minha vida.*

AGRADECIMENTOS

Agradeço,

Ao meu amigo e orientador Prof. Dr. Jorge Artur, que não mediu esforços para contribuir com o sucesso deste trabalho. Seu conhecimento e disponibilidade foram fundamentais nesta jornada.

À minha esposa Fabianna, por todo companheirismo de vida e incentivo incondicional neste curso de mestrado.

À minha mãe Nilza, por todo incentivo à educação desde sempre.

À Deus por tudo isso.

RESUMO

Dentre as mais importantes políticas de ações afirmativas da sociedade brasileira, está o sistema de cotas para ingresso no ensino superior. Desta forma, este mecanismo precisa estar em constante avaliação, em relação a sua efetividade na questão da justiça social e quanto à eficiência na seleção dos candidatos que ingressam nas universidades. Trata-se de um estudo transversal em que buscou-se avaliar o desempenho de estudantes cotistas e de ampla concorrência do ensino médio em um processo de avaliação (vestibular) de uma universidade pública do Estado de Alagoas. Considerou-se os dados do ano 2015 de todos os avaliandos (N = 3807), sendo 2858 de ampla concorrência e 949 cotistas. Optou-se por executar as análises apenas com os candidatos do curso de Bacharelado em Medicina por possuir a maior população (N = 2722), sendo 535 (20%) cotistas e 2187 (80%) não cotistas, e por apresentar o menor erro padrão além de ser o curso com os maiores escores. O instrumento de avaliação foi a prova de vestibular nos moldes do ENEM. Utilizou-se o R e por meio do pacote mirt verificou-se os três parâmetros para cada competência e item da prova, bem como o *DIF (Differential Item Functioning)* para checar a diferença de probabilidade de acerto de um item entre os grupos - cotistas e não cotistas. O Modelo TRI de três parâmetros (3PL) se mostrou adequado para a análise das provas, bem como o método Expected a Posteriori para a proficiência dos candidatos. A prova de Inglês apresentou o menor escore médio de dificuldade dos itens ($b = -0,98$) e a de Matemática o maior ($b = 1,25$). A prova de Ciências da Natureza apresentou o melhor ajuste entre dificuldade da prova e proficiência dos candidatos, enquanto a de Matemática apresentou a maior discrepância. Através do método Mantel-Haenszel para DIF identificou-se 23 itens que discriminam cotistas de não cotistas, sendo a maior concentração (10 itens) na prova de Ciências Humanas. Os resultados demonstram que há DIF para alguns itens e que esse aspecto precisa ser considerado para composição do escore de rendimento na prova. Além disso, os itens que são utilizados em provas de vestibular demandam ao menos algum sistema prévio de classificação e validação do nível de dificuldade - como o método Angoff. O presente estudo salienta a relevância do funcionamento diferencial dos itens para compreensão do desempenho de estudantes cotistas e não cotistas em prova de vestibular nos moldes do ENEM.

Palavras-chaves: Cotas, Vestibular, Funcionamento Diferencial dos Itens (DIF), Teoria de Resposta ao Item (TRI).

ABSTRACT

Among the most important affirmative action policies of Brazilian society, there is the quota system for entering higher education. Thus, this mechanism needs to be constantly evaluated, in relation to its effectiveness in the issue of social justice and in terms of efficiency in the selection of candidates who enter universities. This is a cross-sectional study that sought to evaluate the performance of quota students and broad competition from high school in an evaluation process (vestibular) of a public university in the State of Alagoas. The data for the year 2015 of all appraisers was considered ($N = 3807$), 2858 of which were highly competitive and 949 quota holders. It was decided to perform the analyzes only with the candidates of the Bachelor of Medicine course because it has the largest population ($N = 2722$), being 535 (20%) quota holders and 2187 (80%) non-quota holders, and for presenting the lowest standard error in addition to being the course with the highest scores. The assessment instrument was the entrance exam according to the ENEM model. The R was used and through the mirt package, the three parameters were verified for each competence and item of the test, as well as the DIF (Differential Item Functioning) to check the difference in the probability of an item's success between the quota holders and non-quota holders. The three-parameter IRT Model (3LP) proved to be adequate for the analysis of the tests, as well as the Expected a Posteriori method for the proficiency of the candidates. The English test had the lowest average difficulty score of the items ($b = -0.98$) and the Mathematics test the highest ($b = 1.25$). The Natural Sciences test presented the best fit between the difficulty of the test and the proficiency of the candidates, while the Mathematics test presented the greatest discrepancy. Through the Mantel-Haenszel method for DIF, 23 items were identified that discriminate between quota holders and non-quota holders, with the highest concentration (10 items) in the Humanities test. The results demonstrate that there is DIF for some items and that this aspect needs to be considered for the composition of the performance score in the test. In addition, the items that are used in entrance exams require at least some previous system of classification and validation of the level of difficulty - such as the Angoff method. The present study highlights the relevance of the differential functioning of the items for understanding the performance of quota students and non-quota students in the entrance exam in the manner of ENEM.

Keywords: Quotas, University Entrance Exam, Differential Item Functioning (DIF), Item Response Theory (IRT).

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 – Disposição dos parâmetros na Curva Característica do Item - CCI. Fonte: SINAPE 2000 | 29 |
| Figura 2 – Exemplo de item com <i>DIF</i> uniforme. Fonte: Psicologia: Reflexão e Crítica, 2000 | 30 |
| Figura 3 – Exemplo de item com <i>DIF</i> não-uniforme. Fonte: Psicologia: Reflexão e Crítica, 2000 | 31 |
| Figura 4 – Diferença de médias entre grupos de cotistas e não cotistas | 34 |
| Figura 5 – Diferença de médias entre grupos de cotistas e não cotistas por curso | 36 |
| Figura 6 – Diferença de médias entre grupos de cotistas e não cotistas por disciplina, para os candidatos do curso de Bacharelado em Medicina | 37 |
| Figura 7 – Curva Característica dos Itens da prova de Ciências da Natureza | 38 |
| Figura 8 – Análise de proficiência e dificuldade da prova de Ciências Humanas | 39 |
| Figura 9 – Análise de proficiência e dificuldade da prova de Ciências da Natureza | 40 |
| Figura 10 – Análise de proficiência e dificuldade da prova de Matemática | 40 |
| Figura 11 – Análise de proficiência e dificuldade da prova de Linguagens | 40 |
| Figura 12 – Análise de proficiência e dificuldade da prova de Inglês | 41 |
| Figura 13 – Análise de proficiência e dificuldade da prova de Espanhol | 41 |
| Figura 14 – Gráficos do <i>DIF</i> das provas de Ciências Humanas, Ciências da Natureza, Matemática, Linguagens, Inglês e Espanhol, respectivamente | 42 |
| Figura 15 – Curva Característica dos Itens da prova de Ciências Humanas | 52 |
| Figura 16 – Curva Característica dos Itens da prova de Matemática | 53 |
| Figura 17 – Curva Característica dos Itens da prova de Linguagens | 53 |
| Figura 18 – Curva Característica dos Itens da prova de Inglês | 54 |
| Figura 19 – Curva Característica dos Itens da prova de Espanhol | 54 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Classificação do parâmetro a | 28 |
| Tabela 2 – Classificação do parâmetro b | 29 |
| Tabela 3 – Inscritos por curso | 33 |
| Tabela 4 – Análise de variância de notas por curso | 35 |
| Tabela 5 – Marcadores dos escores das provas | 38 |
| Tabela 6 – Parâmetros de todas as provas | 39 |
| Tabela 7 – Estatística <i>DIF</i> da prova de Ciências Humanas com o método Mantel-Haenszel | 43 |
| Tabela 8 – Estatística <i>DIF</i> da prova de Ciências da Natureza com o método Mantel-Haenszel | 55 |
| Tabela 9 – Estatística <i>DIF</i> da prova de Matemática com o método Mantel-Haenszel . . | 56 |
| Tabela 10 – Estatística <i>DIF</i> da prova de Linguagens com o método Mantel-Haenszel . . | 57 |
| Tabela 11 – Estatística <i>DIF</i> da prova de Inglês com o método Mantel-Haenszel | 57 |
| Tabela 12 – Estatística <i>DIF</i> da prova de Espanhol com o método Mantel-Haenszel . . . | 57 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|------|---|
| PPI | Pretos, Pardos ou Indígenas |
| IBGE | Instituto Brasileiro de Geografia e Estatística |
| ENEM | Exame Nacional do Ensino Médio |
| TCT | Teoria Clássica dos Testes |
| TRI | Teoria de Resposta ao Item |
| DIF | Differential Item Functioning |
| CCI | Curva Característica do Item |

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO | 21 |
| 1.1 | Contextualização | 21 |
| 1.2 | Cotas no Ensino Superior | 22 |
| 1.3 | Problemática | 23 |
| 1.4 | Escopo | 23 |
| 1.5 | Objetivos | 24 |
| 1.5.1 | Objetivo Geral | 24 |
| 1.5.2 | Objetivos Específicos | 24 |
| 1.6 | Organização do texto | 24 |
| 2 | REVISÃO DA BIBLIOGRAFIA | 25 |
| 2.1 | Teoria Clássica dos Testes - TCT | 25 |
| 2.1.1 | Postulados | 25 |
| 2.1.2 | Análise de Itens | 26 |
| 2.2 | Teoria de Resposta ao Item - TRI | 26 |
| 2.2.1 | Postulados | 26 |
| 2.2.2 | Parâmetros dos Itens | 27 |
| 2.2.2.1 | Parâmetro <i>a</i> : discriminação | 27 |
| 2.2.2.2 | Parâmetro <i>b</i> : dificuldade | 28 |
| 2.2.2.3 | Parâmetro <i>c</i> : acerto ao acaso | 28 |
| 2.2.3 | Curva Característica do Item - CCI | 29 |
| 2.3 | <i>Differential Item Functioning</i> - <i>DIF</i> | 29 |
| 2.3.1 | Tipos de <i>DIF</i> | 30 |
| 2.3.2 | Métodos de Detecção de <i>DIF</i> | 30 |
| 2.3.2.1 | Baseados em modelos da TRI | 31 |
| 2.3.2.2 | Clássicos | 31 |
| 3 | MÉTODO | 33 |
| 3.1 | Fonte de Dados | 33 |
| 3.2 | Normalização dos dados | 34 |
| 3.3 | Diferença entre grupos | 34 |
| 3.4 | Análise TCT | 36 |
| 3.5 | Análise TRI | 37 |
| 3.6 | Análise <i>DIF</i> | 41 |
| 4 | DISCUSSÃO | 45 |
| 4.1 | Desempenho de Cotistas e Não Cotistas | 45 |

| | | |
|--------------|--|-----------|
| 4.2 | Instrumento de Avaliação | 45 |
| 4.2.1 | Qualidade dos Itens | 46 |
| 4.2.2 | Capacidade de avaliação em grupos distintos | 46 |
| 5 | CONCLUSÃO | 47 |
| | REFERÊNCIAS | 49 |
| | APÊNDICE A – CURVA CARACTERÍSTICA DOS ITENS DO INSTRUMENTO POR COMPETÊNCIAS | 52 |
| | APÊNDICE B – ESTATÍSTICA DIF DOS ITENS DO INSTRUMENTO POR COMPETÊNCIAS | 55 |
| | APÊNDICE C – CÓDIGO R PARA EXECUÇÃO DAS ANÁLISES . . | 58 |

1 INTRODUÇÃO

Neste capítulo, apresentaremos a contextualização da pesquisa, a problemática abordada, o escopo do estudo, os objetivos a serem alcançados e a organização do texto.

1.1 Contextualização

O ensino superior no Brasil ainda é um privilégio de poucos. De acordo com o IBGE (2019), apenas 12,8% da população possui nível superior. E esta situação é ainda mais crítica no estado de Alagoas, onde apenas 7,9% possui esta condição. Os números deixam claro que é um tema que carece de atenção. Neste contexto, mais especificamente o método de acesso ao ensino superior.

Desde o final da década de 1960, o acesso ao ensino superior é realizado por meio de concurso vestibular. À época, o exame era composto apenas por itens objetivos nas áreas de conhecimento consideradas essenciais (ANDRIOLA, 2011). Hoje em dia, o formato do método de acesso pode variar de acordo com as necessidades e intenções das instituições que selecionam seus candidatos (EDUCAÇÃO, 1999), desde que combinado à uma prova de redação, que é exigência da legislação brasileira (Ministério da Educação do Brasil, 2002).

Em 2012 o governo federal brasileiro instituiu um sistema de cotas, em que definiu-se que metade das vagas ofertadas nas instituições federais de ensino superior (Universidades e Institutos Federais), deveriam ser para alunos oriundos de escolas públicas. Estes 50% ainda apresentam outras subdivisões considerando renda e etnia (Brasil, 2012b).

Frequentemente observa-se debates sobre a eficiência do sistema de cotas, principalmente quando se compara aspectos como a Excelência Acadêmica, que é a necessidade das instituições selecionarem os melhores alunos para o ingresso, Reparação Histórica para candidatos pretos e pardos, além da Justiça Social, que trata da viabilização de melhor equilíbrio de oportunidades aos alunos de maior vulnerabilidade social.

As ações afirmativas podem ser analisadas observando diversos aspectos. No âmbito do desempenho dos candidatos no exame de seleção, Vilela et al. (2017) observou que a divisão em cotas não tem impactos significativos na nota média dos aprovados em cada cota, com relação ao cenário de não haver divisão das vagas. Esta análise é interessante em discussões sobre a qualidade dos alunos ingressantes na instituição.

Outro aspecto relevante é o tipo de cota mais efetivo dentre os ofertados à sociedade. Vilela et al. (2017) mostra que a cota para alunos egressos de escola pública é mais relevante que as direcionadas aos alunos de baixa renda ou mesmo os denominados PPIs (pretos, pardos e indígenas). Neste estudo foi verificado que esta cota chega a proporcionar o ingresso de aproximadamente o dobro de alunos egressos da escola pública comparado à situação de não

haver esta demanda.

No entanto, Santos, Souza e Sasaki (2013) destacam a preferência das instituições pelas cotas sociais, que observa principalmente a condição de renda familiar do candidato. O autor destaca que muitas vezes a cota social consegue abranger elementos de outras ações afirmativas como cota para deficientes ou a étnico-racial.

Os estudos sobre o tema nem sempre apresentam resultados convergentes quanto à efetividade do uso de ações afirmativas em seleções para vagas no ensino superior. Com relação ao desempenho acadêmico, trabalhos como os de Alvarenga et al. (2012) e Queiroz et al. (2015) reportam que não há diferença estatisticamente relevante entre o desempenho de alunos cotistas e de ampla concorrência durante a graduação. Ao contrário, Cavalcanti (2015) e Júnior (2018) demonstram que alunos cotistas obtiveram desempenho inferior durante a graduação em cursos específicos.

Neste contexto de análises de efetividade das ações afirmativas, este trabalho procura abordar o tema analisando não apenas o desempenho dos candidatos que concorrem à uma vaga na universidade, mas também pelo prisma dos instrumentos de avaliação aos quais eles são submetidos.

1.2 Cotas no Ensino Superior

A desigualdade étnico-racial e social no Brasil remontam à época da escravidão. A sociedade brasileira possui em sua história mais de três séculos de relações sociais baseadas na exploração do trabalho escravo de homens e mulheres negras (THEODORO, 2008).

O período escravista teve fim no ano de 1888 e deixou marcas profundas na sociedade brasileira. Após a abolição o Brasil viveu uma época de crescimento econômico importante com o novo modelo das relações de trabalho assalariado (THEODORO, 2008).

Mesmo com a ascensão econômica da sociedade, os negros recém libertos não conseguiram aproveitar essas oportunidades. À época, acontecia no Brasil uma intensa imigração de europeus que acabou por alterar o perfil de mão-de-obra disponível no Brasil. Embora os imigrantes em sua maioria não dispusessem de qualificação profissional especializada, tal qual os negros brasileiros, o preconceito vigente acabou por proporcionar melhores oportunidades aos europeus brancos e marginalizar a população negra atribuindo-lhes funções menores ou a informalidade (THEODORO, 2008).

Neste contexto de evolução social, a educação superior que já era estabelecida como sendo acessível apenas para a elite, tornava-se praticamente inalcançável para os indivíduos socialmente vulneráveis.

Em 2001, o Brasil participou da III Conferência Mundial contra o Racismo, Discriminação Racial, Xenofobia e Intolerâncias Correlatas, ocorrido na África do Sul. Em decorrência deste

evento, em 2002 foi criado o Programa de Ações Afirmativas do Brasil, onde estão dispostas diversas providências que devem ser tomadas pelo Estado brasileiro, dentre elas, a instituição de política de cotas sociais e étnico-raciais para o acesso ao ensino superior público e para o ingresso de servidores no serviço público (MOCELIN, 2020).

Embora a lei de cotas para acesso ao ensino superior tenha sido sancionada em 2012, diversas universidades já implementavam suas próprias políticas de ações afirmativas para o ingresso. O Rio de Janeiro foi o primeiro estado a aderir a um sistema de cotas, com reserva de 50% das vagas das universidades estaduais para alunos oriundos de escolas públicas estaduais ou municipais daquele estado. Já a Universidade de Brasília foi a pioneira em um sistema de cotas de cunho étnico-racial (MARQUES; RODRIGUES, 2020).

Este pioneirismo enfrentou muitos questionamentos, no entanto, a necessidade de prover meios para democratização de acesso ao ensino superior mostrou-se maior e culminou com a aprovação da lei de cotas nº 12.711/2012.

1.3 Problemática

É, de certa forma, intuitivo fazermos um paralelo entre o desempenho do candidato em uma avaliação e o grupo social ao qual ele pertence. Candidatos de grupos distintos, sejam socialmente, etnicamente ou regionalmente, podem apresentar desempenhos diferentes em uma prova, no entanto, o instrumento de avaliação deve ser elaborado a fim de garantir a proficiência de todos os aprovados, independentemente do grupo ao qual pertencem.

A habilidade de um indivíduo, assim como a dificuldade de um item, não deveriam possuir qualquer dependência em relação a um grupo dentro de uma população (SISTO, 2006).

Assim, um bom instrumento de avaliação deve garantir a equivalência de habilidades e competências dos candidatos, cotistas e de ampla concorrência, em processo seletivo vestibular.

1.4 Escopo

O contexto de cotas e ações afirmativas é amplo e, no Brasil, é aplicado em diversas áreas como contratação de funcionários com deficiência para o setor privado (Brasi, 1991) e contratação de pessoas com deficiência e negros no serviço público (Brasil, 2014).

Este trabalho pretende abordar as cotas de ações afirmativas aplicadas ao ingresso no ensino superior, analisando um processo vestibular com instrumento de seleção nos moldes do Exame Nacional do Ensino Médio - ENEM.

1.5 Objetivos

1.5.1 Objetivo Geral

Avaliar o desempenho de estudantes do ensino médio, cotistas e ampla concorrência, em prova de vestibular de uma universidade pública.

1.5.2 Objetivos Específicos

1. Analisar a qualidade psicométrica da prova de vestibular por meio da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI);
2. Identificar o Funcionamento Diferencial dos Itens (*DIF*), por grupos - cotistas e ampla concorrência.

1.6 Organização do texto

A Revisão Bibliográfica no capítulo 2 traz conceitos importantes para a realização deste trabalho, como a Teoria Clássica dos Testes - TCT (seção 2.1), Teoria de Resposta ao Item - TRI (seção 2.2) e *Differential Item Functioning* - *DIF* (seção 2.3).

No capítulo 3 são executados todos os modelos necessários para a obtenção dos objetivos propostos na seção 1.5.1, iniciando com o detalhamento do instrumento de avaliação (seção 3.1) e normatização de dados necessária para a execução dos modelos, seguindo com a aplicação das técnicas de TCT (seção 3.4), TRI (seção 3.5) e *DIF* (seção 3.6).

A exposição dos resultados e análises dos modelos executados são apresentadas no capítulo 4. A análise está dividida em duas seções: uma analisando o desempenho dos grupos avaliados - cotistas e não cotistas (seção 4.1), e outra discorrendo sobre o instrumento de avaliação (seção 4.2).

Por fim, o capítulo 5 apresenta as conclusões desta pesquisa e desafios para trabalhos futuros.

2 REVISÃO DA BIBLIOGRAFIA

Neste capítulo veremos conceitos importantes utilizados para a realização deste trabalho.

2.1 Teoria Clássica dos Testes - TCT

Criada inicialmente a partir dos trabalhos dos psicólogos Charles Spearman e Louis Thurstone, a Teoria Clássica dos Testes apresenta-se como uma das primeiras medidas formais de medida em Psicologia, e consiste no cálculo do escore de um sujeito ao responder um determinado instrumento ou teste (HUTZ; BANDEIRA; TRENTINI, 2015)

2.1.1 Postulados

Para entendermos a TCT, precisamos primeiramente apresentar três conceitos que são necessários para o funcionamento deste modelo. Pasquali (2011) os relaciona da seguinte forma:

T é o **escore bruto** ou **escore empírico** do sujeito, que consiste no somatório dos pontos obtidos pelo respondente em um teste. V o **escore verdadeiro**, que seria a magnitude real daquilo que o teste quer medir no sujeito. E por fim, E o **erro** cometido nesta medida.

A TCT conta com diversos postulados na literatura, no entanto, dois deles explicam a essência do seu funcionamento.

O primeiro postulado versa sobre o escore bruto T , quando diz que (a) seu valor equivale à soma do escore verdadeiro V e do erro E (PASQUALI, 2011). Na fórmula:

$$T = V + E$$

E, conseqüentemente:

$$V = T - E$$

Podemos observar que V seria o próprio T se não houvesse o erro de medida, e este é um conceito importante para o entendimento do próximo postulado.

Ao rótulo de “erro” podem-se atribuir diversos fatores na aplicação e análise de um teste, como defeito no instrumento de testagem, estereótipos ou vieses dos sujeitos testados, além de fatores históricos ou ambientais aleatórios (PASQUALI, 2011).

O erro é um evento casual que pode ter valor positivo ou negativo. Em uma situação hipotética em que um teste seja respondido infinitas vezes e de forma independente por um sujeito (ou seja, todas elas sem qualquer lembrança das aplicações anteriores), a média dos valores dos erros dos testes seria igual à zero (HUTZ; BANDEIRA; TRENTINI, 2015).

Assim, temos o segundo postulado estipulando que, (b) o valor esperado para o resultado do teste é o **escore verdadeiro**, uma vez que precisão do resultado aumenta de acordo com o número de observações (SARTES; SOUZA-FORMIGONI, 2013).

2.1.2 Análise de Itens

De forma muito mais simples que na TRI (seção 2.2.2.2), na TCT a dificuldade do item é dada pelo cálculo do coeficiente de dificuldade D , dividindo o número A da quantidade de sujeitos que um item de forma correta, pelo número N da quantidade sujeitos submetidos ao item (PASQUALI, 2011). Na fórmula:

$$D = \frac{A}{N}$$

Vale ressaltar que o conceito de “dificuldade” de um item, só faz sentido no contexto de instrumentos que medem aptidão, como testes e provas.

Também é importante observar que o cálculo da dificuldade do item depende diretamente da amostra em que o teste foi aplicado. Um item pode ter seu coeficiente de dificuldade calculado como baixo se aplicado à uma amostra de sujeitos com muita aptidão e coeficiente alto se aplicado à outra amostra com sujeitos de pouca aptidão (PASQUALI, 2011) . Ou seja, o item pode ser fácil para um grupo de sujeitos habilidosos e difícil para outro grupo menos habilidoso.

2.2 Teoria de Resposta ao Item - TRI

A TRI ou teoria do traço latente corresponde a um conjunto de modelos matemáticos que procura descrever a probabilidade de uma pessoa responder corretamente à um item, dada a sua habilidade latente e dos parâmetros do mesmo (LINDEN, 2016). A teoria do traço latente afirma que o rendimento observável, como a habilidade de responder corretamente aos itens de um teste de inteligência, são reflexos de traços não observáveis mais básicos.

Embora a TRI venha sendo considerada como um dos maiores avanços em termos de psicometria, o Modelo da Psicometria Clássica continua sendo utilizado, unicamente ou em combinação com a TRI, com o objetivo de oferecer evidências adicionais sobre a qualidade do teste, na validação de instrumentos. Bem como, auxiliando na análise exploratória dos itens e possibilitando apontar discrepâncias nos dados e itens ambíguos ou problemáticos (BECHGER et al., 2003).

2.2.1 Postulados

A TRI assume dois postulados: (a) o desempenho de um dado indivíduo em um item pode ser explicado por meio de um conjunto de variáveis hipotéticas (traços latentes); (b) a relação entre o desempenho e os traços latentes pode ser descrita por uma função matemática

monotônica crescente, cujo gráfico é denominado de Curva Característica do Item - CCI (PASQUALI, 2017). Comumente, se faz uso de termos como: “estimação de habilidade” e “estimação ou calibração dos parâmetros dos itens”. Estimar habilidade significa determinar o valor do traço latente (ou teta - θ) para cada um dos indivíduos avaliados em um teste (HAMBLETON HARIHARAN SWAMINATHAN, 1991). O θ pode ser, por exemplo, o grau de habilidade na disciplina da prova realizada.

Postula-se que para se fazer uso da TRI, o teste deve avaliar apenas um único traço latente, isto é, supõe-se que existe apenas uma aptidão responsável pela realização de um conjunto de itens. Em geral, para satisfazer tal postulado, é suficiente que haja um fator dominante responsável pelas respostas dos indivíduos a um conjunto de itens (HAMBLETON HARIHARAN SWAMINATHAN, 1991) (PASQUALI, 2017).

Outro pressuposto é o da independência local. Diz respeito ao fato de que, mantidas constantes as aptidões que afetam o teste (exceto a aptidão dominante que está sendo avaliada), as respostas das pessoas aos itens são estatisticamente independentes. Isso quer dizer que o desempenho do indivíduo em um item não afeta o desempenho nos demais itens, pois cada item é respondido exclusivamente em função do tamanho da aptidão dominante (HAMBLETON HARIHARAN SWAMINATHAN, 1991) (PASQUALI, 2009).

O postulado da independência local é importante e útil porque, sendo ele verdadeiro, a probabilidade para uma sequência de respostas do indivíduo a uma série de itens, será o produto das probabilidades de cada item individual (PASQUALI, 2017). Na maioria dos casos, quando o pressuposto da unidimensionalidade é satisfeito, igualmente é o da independência local (EMBRETSON, 2013).

Os modelos de TRI propostos na literatura dependem basicamente de três fatores: (a) do número de atributos ou dimensões consideradas (uma ou mais); (b) do formato dos itens (múltipla escolha; dicotômico/politômico); e (c) do número de parâmetros dos itens a serem estimados (NUNNALLY, 1967) (LINDEN, 2016).

2.2.2 Parâmetros dos Itens

São três os parâmetros dos itens, como seguem:

2.2.2.1 Parâmetro a : discriminação

Também conhecido como *slope*, *inclination* ou *dispersion*, é a inclinação da CCI (ver Seção 2.2.3), no momento da inflexão. Ou seja, quando corta a linha que corresponde à probabilidade 0,5.

O valor deste parâmetro é definido como o poder do item para diferenciar indivíduos com magnitudes próximas do traço latente que está sendo aferido. O parâmetro a refere-se à inclinação da curva, sendo itens com curvas características dos itens mais inclinadas, mais úteis

para diferenciar indivíduos com habilidades diferentes do que itens com curvas menos inclinadas. A métrica do parâmetro a , teoricamente, vai de $-\infty$ a $+\infty$. Na prática, varia entre 0,0 e 2,0, de acordo com Baker (2001), que também sugere ser 0,65 um valor moderado.

Itens com parâmetro a negativo não são esperados, uma vez que indicariam que a probabilidade de responder corretamente o item diminui com o aumento da habilidade. Baixos valores de a indicam que o item tem pouco poder de discriminação, isto é, tanto examinandos com baixa habilidade quanto examinandos com alta habilidade têm praticamente a mesma probabilidade de responder corretamente o item. Valores muito altos do parâmetro a , por sua vez, indicam itens com curvas características muito “íngremes”, que discriminam os examinandos basicamente em dois grupos: os que possuem habilidade abaixo do valor do parâmetro b e os que possuem habilidades acima do parâmetro b (ANDRADE; TAVARES; VALLE, 2000).

Baker (2001) apresenta a classificação do parâmetro de discriminação por faixa de valores, observado na Tabela 1:

Tabela 1 – Classificação do parâmetro a

| Discriminação | Faixa de valores |
|---------------|------------------|
| Nenhuma | 0,0 |
| Muito baixa | 0,01 a 0,34 |
| Baixa | 0,35 a 0,64 |
| Moderada | 0,65 a 1,34 |
| Alta | 1,35 a 1,69 |
| Muito alta | a partir de 1,70 |

2.2.2.2 Parâmetro b : dificuldade

O parâmetro b quantifica a dificuldade do item analisado. Esse parâmetro é medido na mesma escala da habilidade do respondente e corresponde ao valor de θ para o qual a probabilidade de acerto é 0,50. Quanto maior o valor do parâmetro b do item, maior o nível de θ requerido para um indivíduo ter 50% de chance de acertá-lo (HAMBLETON HARIHARAN SWAMINATHAN, 1991). O parâmetro b indica ponto na escala de aptidão no qual a probabilidade de uma resposta correta é 0,5. A métrica teoricamente vai de $-\infty$ a $+\infty$, mas na prática vai de -3 a +3. Também é conhecido como *location*, *threshold* ou *position*.

Pasquali (2017) recomenda critérios de classificação e distribuição quanto ao parâmetro b , tal como descrito na Tabela 2:

2.2.2.3 Parâmetro c : acerto ao acaso

Esse parâmetro avalia a resposta correta dada ao acaso, isto é, a probabilidade de um examinando com habilidade muito baixa de acertar o item. Os valores de c podem variar de 0,0 a 1,0. Em geral, são aceitáveis probabilidades de acerto iguais ou inferiores a 0,2 para itens com cinco opções de marcação, 0,25 para itens com quatro opções e 0,50 para itens com duas opções

Tabela 2 – Classificação do parâmetro b

| Dificuldade | Intervalo de valores | Porcentagem ideal |
|----------------|----------------------|-------------------|
| Muito fáceis | Menor que -1,28 | 10% |
| Fáceis | -1,28 a -0,52 | 20% |
| Medianos | -0,52 a 0,52 | 40% |
| Difíceis | 0,52 a 1,28 | 20% |
| Muito difíceis | Maior que 1,28 | 10% |

(ANDRADE; TAVARES; VALLE, 2000) (HAMBLETON HARIHARAN SWAMINATHAN, 1991).

2.2.3 Curva Característica do Item - CCI

A Curva Característica do Item é um gráfico logístico em forma de S (ogiva), que representa a probabilidade de acerto de um item por um sujeito, dado um θ (habilidade) do mesmo (ANDRADE; TAVARES; VALLE, 2000).

Klein (2009) afirma que esta curva expressa tudo que a TRI pode descobrir sobre o item. De fato, nela estão sinalizados os três parâmetros do item dispostos em uma curva traçada em função de θ , como podemos observar na Figura 1 (NUNES; OLIVEIRA; NAMETALA, 2020).

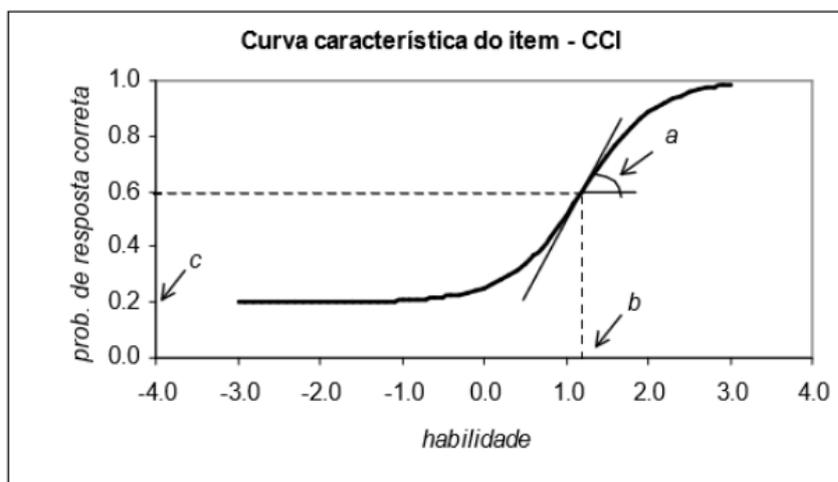


Figura 1 – Disposição dos parâmetros na Curva Característica do Item - CCI. Fonte: SINAPE 2000

2.3 Differential Item Functioning - DIF

Ao submeter uma população à um instrumento de avaliação, é desejável que os itens que compõem este instrumento sejam capazes de aferir a habilidade de todos os sujeitos da mesma forma, sem que haja distinção entre grupos de características semelhantes, como gênero, credo ou classe social.

A existência de *DIF* em um item, significa que este tem desempenho diferente, ou mede de forma diferente, indivíduos em subgrupos distintos de uma população com mesmo nível de habilidade (traço latente). Desta forma, o item que apresenta *DIF* é, necessariamente, menos válido para pelo menos um dos subgrupos da população (STEINBERG; THISSEN, 2006).

Para fins de análise do *DIF*, os subgrupos da população com mesma habilidade e alguma característica que os diferencie, são chamados de **Grupo de Referência** e **Grupos Focal** (SISTO, 2006).

2.3.1 Tipos de *DIF*

São dois os tipos de *DIF*, como seguem:

- **Uniforme ou Consistente:** Ocorre quando as CCI's do item estudado para o Grupo de Referência e para o Grupo Focal são diferentes e não se cruzam em nenhum ponto do gráfico de proficiência (Figura 2), indicando que o item favorece uniformemente um dos grupos, em relação ao outro (AGUIAR, 2010).
- **Não-uniforme ou Inconsistente:** Ocorre quando as CCI's são diferentes e se cruzam em algum ponto da curva de proficiência (Figura 3). Neste tipo, o favorecimento a um dos grupos não é uniforme (AGUIAR, 2010).

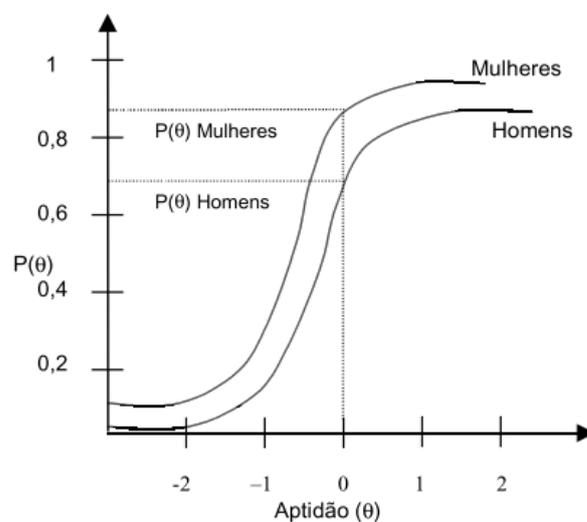


Figura 2 – Exemplo de item com *DIF* uniforme. Fonte: Psicologia: Reflexão e Crítica, 2000

2.3.2 Métodos de Detecção de *DIF*

Existem vários métodos para detecção de itens com *DIF* e podem ser divididos em dois grupos, como seguem:

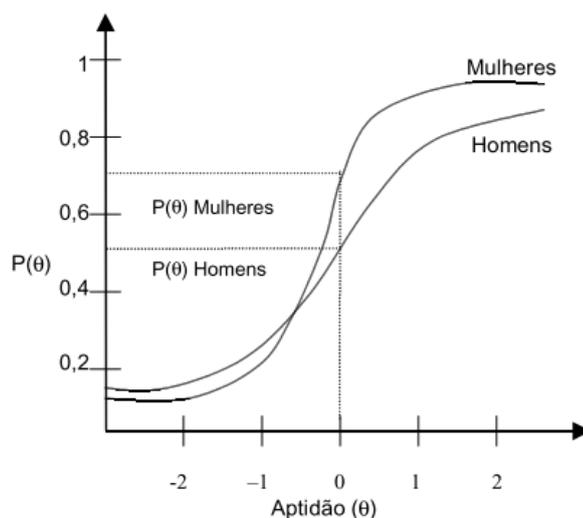


Figura 3 – Exemplo de item com *DIF* não-uniforme. Fonte: Psicologia: Reflexão e Crítica, 2000

2.3.2.1 Baseados em modelos da TRI

São métodos que se utilizam dos parâmetros dos itens, conceituados e estabelecidos pela TRI, para detectar o *DIF* em um item. Estes métodos não utilizam técnicas derivadas da TRI na detecção do *DIF*, ou seja, são métodos não-paramétricos.

Entre os métodos que utilizam a TRI, podemos citar o *Método das Áreas*, o *Método das Probabilidades* o *Método de Comparação dos Parâmetros Métricos dos Itens* e o *Método do Qui-quadrado de Lord*

2.3.2.2 Clássicos

Estes métodos não utilizam técnicas derivadas da TRI na detecção do *DIF*, ou seja, são métodos não-paramétricos. No entanto, necessitam de uma proficiência já conhecida entre os respondentes.

Neste grupo, podemos destacar o *Método de Regressão Logística* e o *Método de Mantel-Haenszel*.

3 MÉTODO

3.1 Fonte de Dados

Utilizou-se para o estudo, dados oriundos do concurso vestibular de uma universidade pública de Alagoas. Os dados correspondem à prova aplicada no processo seletivo do ano de 2015.

Embora seja uma instituição Estadual, a universidade oferta metade de suas vagas para alunos oriundos de escolas públicas. Contamos com uma amostra de 3807 candidatos, sendo 2858 (75,1%) inscritos de ampla concorrência e 949 (24,9%) inscritos cotistas. Neste certame as demandas possuem concorrência exclusiva, ou seja, os candidatos cotistas não concorrem concomitantemente às vagas de cota e ampla. A distribuição de candidatos por curso e demanda pode ser verificada na Tabela 3.

A prova tem seus itens distribuídos nas competências de Ciências Humanas (30 itens), Ciências da Natureza (30 itens), Matemática (30 itens), Linguagens (20 itens) e Língua Estrangeira (10 itens), além de uma prova de Redação. Na prova de Língua Estrangeira o candidato pôde escolher sua preferência pelos idiomas Inglês ou Espanhol.

Ao dividir a prova de Língua Estrangeira em dois idiomas, tornou-se inviável calcular apenas uma nota para esta competência. Com itens diferentes, foi necessário calcular as notas separadamente para as provas de Língua Estrangeira - Inglês e Língua Estrangeira - Espanhol.

O instrumento foi aplicado utilizando a Teoria Clássica dos Testes, em que o escore dos indivíduos é calculado pelo somatório dos acertos em cada uma das competências.

Tabela 3 – Inscritos por curso

| Curso | Ampla | Cota | Total |
|---|-------|------|-------|
| 001 - Bacharelado em Enfermagem | 178 | 117 | 295 |
| 002 - Bacharelado em Fisioterapia | 178 | 109 | 287 |
| 003 - Bacharelado em Fonoaudiologia | 92 | 39 | 131 |
| 004 - Bacharelado em Medicina | 2185 | 537 | 2722 |
| 005 - Bacharelado em Terapia Ocupacional | 65 | 37 | 102 |
| 006 - Tecnologia em Análise e Desenvolvimento de Sistemas | 49 | 31 | 80 |
| 007 - Tecnologia em Processos Gerenciais | 20 | 8 | 28 |
| 008 - Tecnologia em Radiologia | 67 | 55 | 122 |
| 009 - Tecnologia em Sistemas Biomédicos | 24 | 16 | 40 |
| Total | 2858 | 949 | 3807 |

3.2 Normatização dos dados

O instrumento aplicado contou com quatro modelos de gabaritos diferentes, com variação de posição dos itens e ordem das alternativas, a fim de dificultar qualquer tipo de fraude no momento da aplicação.

Para utilizar os registros dos candidatos que realizaram a prova com qualquer dos quatro gabaritos, os dados foram organizados para o formato adequado à utilização das ferramentas de análise, ajustando para que os itens ficassem na mesma posição.

A prova foi realizada em dois dias, sendo Ciências Humanas e Ciências da Natureza no primeiro dia e Linguagens, Língua Estrangeira e Matemática no segundo. Este fato possibilitou que alguns candidatos participassem de apenas um dia de prova, levando à situação de registros com respostas consistentes em apenas metade dos itens. Por este motivo, decidiu-se por excluir observações de candidatos faltosos em qualquer dos dias de prova.

Por fim, para os experimentos relacionados à TRI e *DIF*, os itens que foram anulados pela banca de realização da prova foram removidos. Para estes itens, foram atribuídas as pontuações a todos os candidatos e, desta forma, não interferem no modelo.

3.3 Diferença entre grupos

Na sequência, foi investigada a existência de diferença estatisticamente relevante entre os grupos de alunos cotistas e não cotistas, com um intervalo de confiança de 95%. A análise de variância entre os grupos apresentou valor de $p < 0,05$, indicando que existe diferença entre as notas das duas populações. O gráfico das médias com intervalo de confiança pode ser observado na Figura 4.

A fim de aprimorar a análise, investigou-se a variância das notas entre os cursos. Para

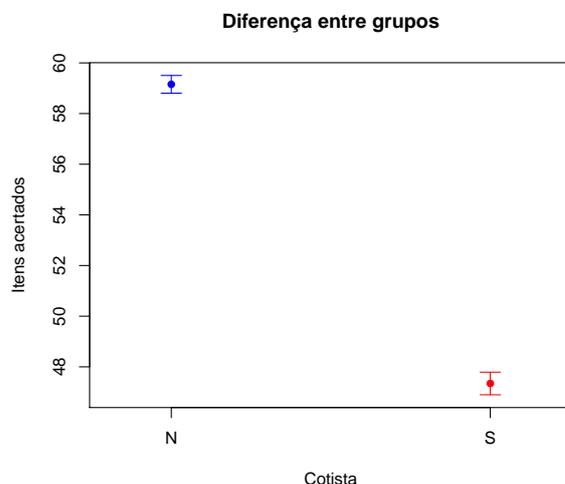


Figura 4 – Diferença de médias entre grupos de cotistas e não cotistas

tanto, utilizou-se o teste ANOVA para detectar se existe algum curso com médias significativamente diferentes e um teste *posthoc* para comparar as médias grupo a grupo. O teste de variância apresentou o valor de $p < 0,05$ e, portanto, evidenciando a diferença entre os grupos.

O teste *posthoc* de Tukey produziu os resultados apresentados na Tabela 4, onde podemos perceber que o curso 004 - Bacharelado em Medicina, se destaca de todos os demais (os códigos dos cursos podem ser observados na Tabela 3).

Tabela 4 – Análise de variância de notas por curso

| Curso-Curso | diff | lwr | upr | p adj |
|-------------|----------|----------|----------|---------|
| 002-001 | 0.1172 | -3.9936 | 4.2279 | 1.0000 |
| 003-001 | 2.3096 | -2.8343 | 7.4536 | 0.9006 |
| 004-001 | 19.5241 | 16.4886 | 22.5596 | 0.0000* |
| 005-001 | 0.8587 | -4.8206 | 6.5379 | 0.9999 |
| 006-001 | 1.5058 | -4.7847 | 7.7964 | 0.9982 |
| 007-001 | -1.3275 | -11.4320 | 8.7770 | 1.0000 |
| 008-001 | -0.6960 | -6.3097 | 4.9176 | 1.0000 |
| 009-001 | 0.3009 | -8.0145 | 8.6163 | 1.0000 |
| 003-002 | 2.1924 | -2.9792 | 7.3641 | 0.9269 |
| 004-002 | 19.4070 | 16.3248 | 22.4891 | 0.0000* |
| 005-002 | 0.7415 | -4.9628 | 6.4458 | 1.0000 |
| 006-002 | 1.3887 | -4.9245 | 7.7019 | 0.9990 |
| 007-002 | -1.4447 | -11.5633 | 8.6739 | 1.0000 |
| 008-002 | -0.8132 | -6.4522 | 4.8258 | 1.0000 |
| 009-002 | 0.1837 | -8.1488 | 8.5163 | 1.0000 |
| 004-003 | 17.2145 | 12.8486 | 21.5805 | 0.0000* |
| 005-003 | -1.4509 | -7.9395 | 5.0376 | 0.9989 |
| 006-003 | -0.8038 | -7.8336 | 6.2261 | 1.0000 |
| 007-003 | -3.6371 | -14.2176 | 6.9435 | 0.9788 |
| 008-003 | -3.0057 | -9.4369 | 3.4256 | 0.8779 |
| 009-003 | -2.0087 | -10.8965 | 6.8791 | 0.9988 |
| 005-004 | -18.6654 | -23.6509 | -13.6799 | 0.0000* |
| 006-004 | -18.0183 | -23.6904 | -12.3462 | 0.0000* |
| 007-004 | -20.8516 | -30.5831 | -11.1201 | 0.0000* |
| 008-004 | -20.2202 | -25.1308 | -15.3095 | 0.0000* |
| 009-004 | -19.2232 | -27.0812 | -11.3653 | 0.0000* |
| 006-005 | 0.6472 | -6.7833 | 8.0777 | 1.0000 |
| 007-005 | -2.1862 | -13.0371 | 8.6647 | 0.9995 |
| 008-005 | -1.5547 | -8.4216 | 5.3122 | 0.9988 |
| 009-005 | -0.5578 | -9.7658 | 8.6502 | 1.0000 |
| 007-006 | -2.8333 | -14.0163 | 8.3496 | 0.9972 |
| 008-006 | -2.2019 | -9.5824 | 5.1786 | 0.9915 |
| 009-006 | -1.2050 | -10.8021 | 8.3921 | 1.0000 |
| 008-007 | 0.6314 | -10.1853 | 11.4481 | 1.0000 |
| 009-007 | 1.6284 | -10.8068 | 14.0636 | 1.0000 |
| 009-008 | 0.9969 | -8.1707 | 10.1646 | 1.0000 |

Para melhor compreensão das características da diferença entre o curso 004 - Bacharelado

em Medicina e os demais, observemos o gráfico da Figura 5, onde fica evidente o quão discrepante é a média deste curso. Ambas as demandas possuem média superior às médias de todos os outros cursos.

Outro grupo que merece destaque é o curso 009 - Tecnologia em Sistemas Biomédicos, onde a nota dos cotistas é maior que a nota dos candidatos não cotistas, o que vai de encontro à análise apresentada na Figura 4.

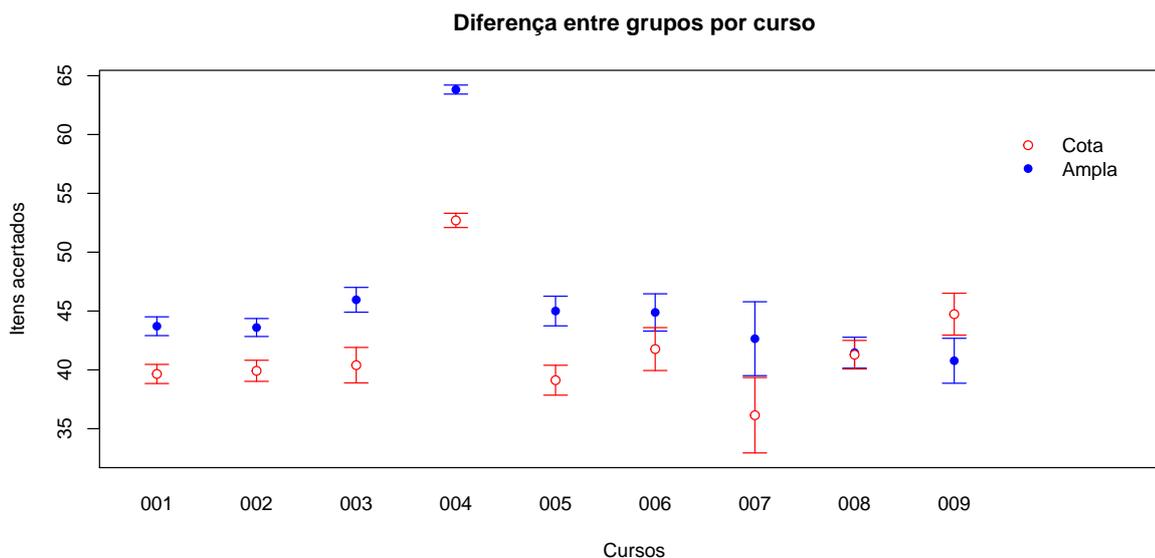


Figura 5 – Diferença de médias entre grupos de cotistas e não cotistas por curso

Diante destes fatos, decidiu-se que não seria uma boa opção executar as análises com as observações de todos os candidatos que realizaram a prova, mas sim de um curso específico. Neste particular pelo curso de Bacharelado em Medicina, por possuir a maior população ($N = 2722$), sendo 535 (20%) cotistas e 2187 (80%) não cotistas, menor erro padrão, além de ser o curso com os maiores escores. A partir daqui, os dados mencionados no texto sempre farão referência ao curso de Bacharelado em Medicina.

Após a remoção de registros de candidatos faltosos, como descrito na seção 3.2, contou-se com 2480 observações para as competências de Ciências Humanas, Ciências da Natureza, Linguagens e Matemática, 1042 observações para Inglês e 1438 para a prova de Espanhol.

3.4 Análise TCT

Utilizando a TCT, foram calculadas as notas dos candidatos em cada uma das competências e em seguida analisamos as notas médias de cotistas e não cotistas em cada uma delas. A análise de variância para cada uma das seis competências apresentou valor de $p < 0,01$ e, portanto, existe diferença estatisticamente relevante entre as médias de cotistas e não cotistas em cada uma das competências da prova. A representação gráfica destas diferenças como pode ser observada no gráfico da Figura 6.

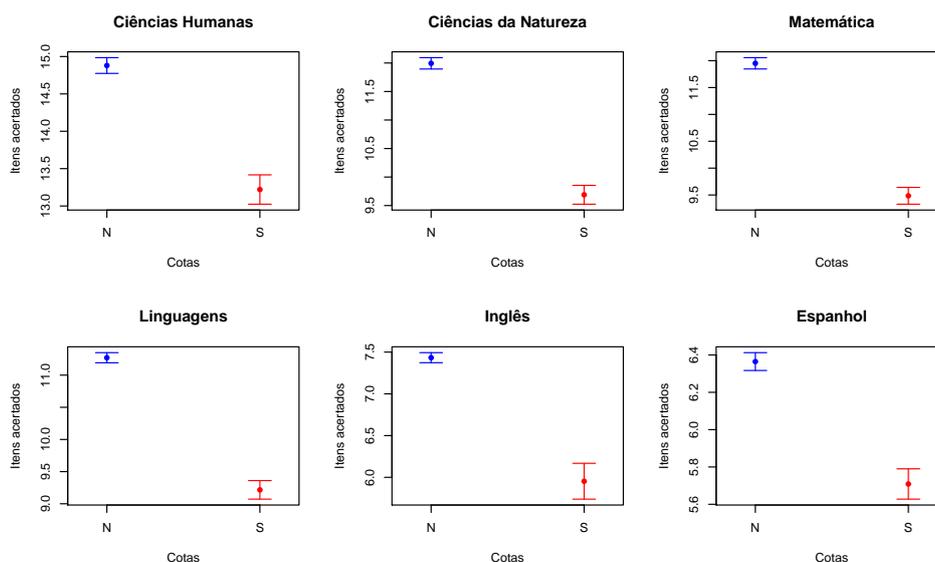


Figura 6 – Diferença de médias entre grupos de cotistas e não cotistas por disciplina, para os candidatos do curso de Bacharelado em Medicina

Os escores maiores para os grupos não cotistas, evidencia a maior quantidade de acertos no grupo, mas não pode fazer uma análise criteriosa sobre quais itens foram marcados de forma correta, tão pouco sobre a qualidade destes itens. Para uma análise focada não apenas nos escores totais de cada competência, mas sim em cada item individualmente, fez-se uso de modelos da TRI, a seguir.

3.5 Análise TRI

Conforme já abordado na seção 2.2, a TRI dispõe de diversos modelos, em que os mais conhecidos são os com 1 (Rasch), 2 (2PL) ou 3 (3PL) parâmetros. Aqui, faz-se uso do modelo de 3 parâmetros, onde além dos parâmetros de *discriminação* (a) e *dificuldade* (b), o modelo também utiliza o parâmetro *acerto ao acaso* (c) para penalizar as respostas não consistentes aos itens (KLEIN, 2013). Com este recurso, podemos quantificar não apenas o desempenho do respondente, mas também a qualidade do instrumento de avaliação.

Utilizando o pacote *mirt* (v1.32.1) do *R* (v4.0.3), foi calculado o modelo de 3PL pra cada uma das competências e analisados os parâmetros de cada item. Verificou-se que alguns itens possuíam o parâmetro b demasiadamente baixo ou alto demais e, conseqüentemente, parâmetro a muito baixo. Ou seja, itens tão fáceis ou tão difíceis que não discriminam bem a proficiência do respondente.

Para ilustrar, pode-se observar na Figura 7, o gráfico da curva característica dos itens *CN_7*, *CN_20* e *CN_27* da prova de Ciências da Natureza. Com 3 itens de parâmetro a muito baixo esta foi a competência com maior número de itens que precisaram ser reconsiderados. O gráfico da curva característica dos itens das outras competências encontram-se no apêndice A.

Para não enviesar o modelo, decidiu-se por retirar os itens discrepantes desta análise. Também foram descartados os itens *CH_17* da prova de Ciências Humanas, item *MA_13* da prova de Matemática e o item *LE_4* da prova de Espanhol.

A configuração dos novos conjuntos de itens de cada prova, juntamente com os marcadores dos seus escores, pode ser observado na Tabela 5.

Tabela 5 – Marcadores dos escores das provas

| Prova | n | mean | sd | median | min | max | itens |
|----------------------|------|-------|------|--------|-----|-----|-------|
| Ciências Humanas | 2480 | 14.25 | 4.67 | 14 | 3 | 27 | 28 |
| Ciências da Natureza | 2480 | 10.70 | 4.32 | 10 | 1 | 24 | 25 |
| Matemática | 2480 | 11.22 | 4.59 | 10 | 1 | 27 | 29 |
| Linguagens | 2480 | 10.86 | 3.55 | 11 | 0 | 20 | 20 |
| Inglês | 1042 | 7.28 | 1.94 | 8 | 0 | 9 | 9 |
| Espanhol | 1438 | 5.99 | 1.54 | 6 | 1 | 9 | 9 |

Com o novo conjunto de itens, os modelos foram recalculados e obteve-se os parâmetros descritos na Tabela 6.

Existem diversos métodos para estimar os escores de proficiência dos sujeitos observados, como o de *Máxima Verossimilhança* ou o *Maximum a Posteriori*. Neste estudo, utilizou-se o método *Expected a Posteriori*, que é o mesmo método utilizado para calcular as notas dos candidatos do ENEM (Brasil, 2012a).

Curva característica dos itens de Ciências da Natureza

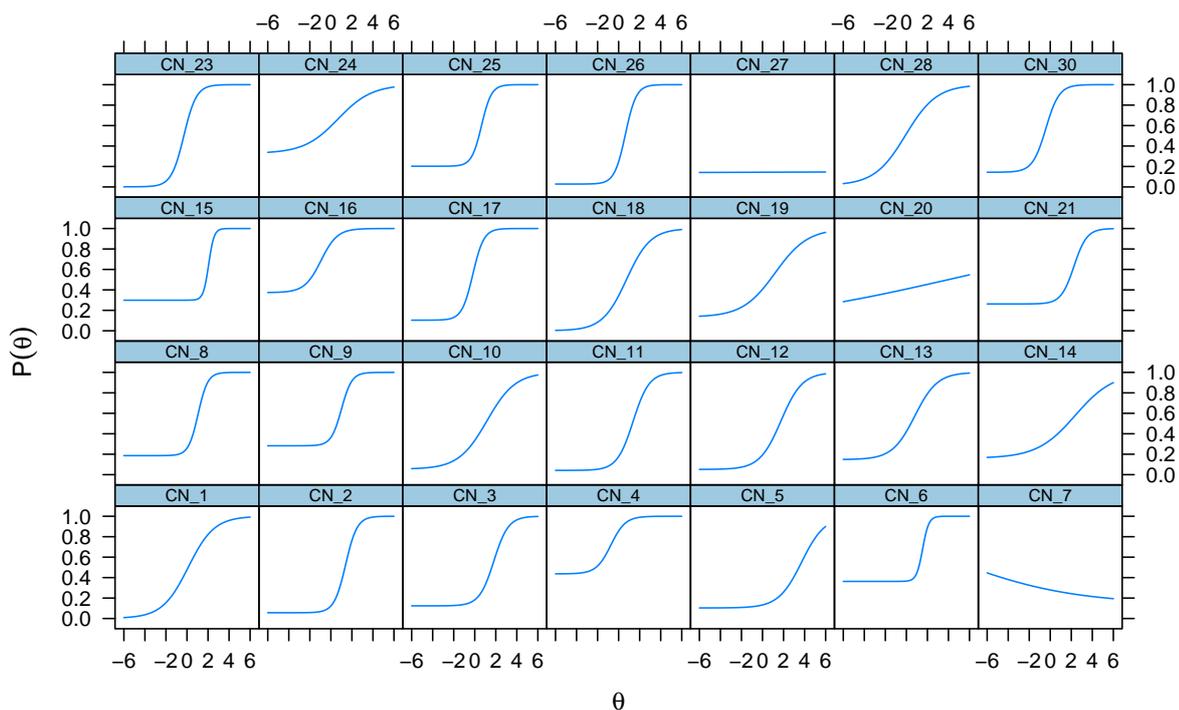


Figura 7 – Curva Característica dos Itens da prova de Ciências da Natureza

Tabela 6 – Parâmetros de todas as provas

| Prova | Parâmetro | n | mean | sd | median | min | max |
|----------------------|-----------|----|-------|------|--------|-------|-------|
| Ciências Humanas | a | 28 | 1.00 | 0.58 | 0.92 | -0.59 | 2.11 |
| | b | 28 | 0.10 | 1.24 | 0.47 | -3.80 | 1.70 |
| | c | 28 | 0.13 | 0.11 | 0.10 | 0.00 | 0.38 |
| Ciências da Natureza | a | 25 | 1.46 | 0.78 | 1.32 | 0.58 | 3.59 |
| | b | 25 | 0.96 | 1.07 | 1.00 | -0.85 | 3.74 |
| | c | 25 | 0.16 | 0.14 | 0.13 | 0.00 | 0.45 |
| Matemática | a | 29 | 1.49 | 0.56 | 1.44 | 0.39 | 2.77 |
| | b | 29 | 1.25 | 0.79 | 1.31 | -0.53 | 2.57 |
| | c | 29 | 0.18 | 0.15 | 0.16 | 0.00 | 0.67 |
| Linguagens | a | 20 | 1.10 | 0.42 | 1.14 | 0.32 | 1.77 |
| | b | 20 | 0.20 | 1.19 | 0.23 | -2.01 | 2.29 |
| | c | 20 | 0.12 | 0.10 | 0.12 | 0.00 | 0.29 |
| Inglês | a | 9 | 2.55 | 1.36 | 1.94 | 1.43 | 5.54 |
| | b | 9 | -0.98 | 0.68 | -0.69 | -2.37 | -0.40 |
| | c | 9 | 0.22 | 0.21 | 0.27 | 0.00 | 0.54 |
| Espanhol | a | 9 | 0.76 | 0.44 | 0.92 | 0.25 | 1.58 |
| | b | 9 | -0.46 | 1.85 | -0.80 | -3.34 | 2.01 |
| | c | 9 | 0.16 | 0.25 | 0.04 | 0.00 | 0.73 |

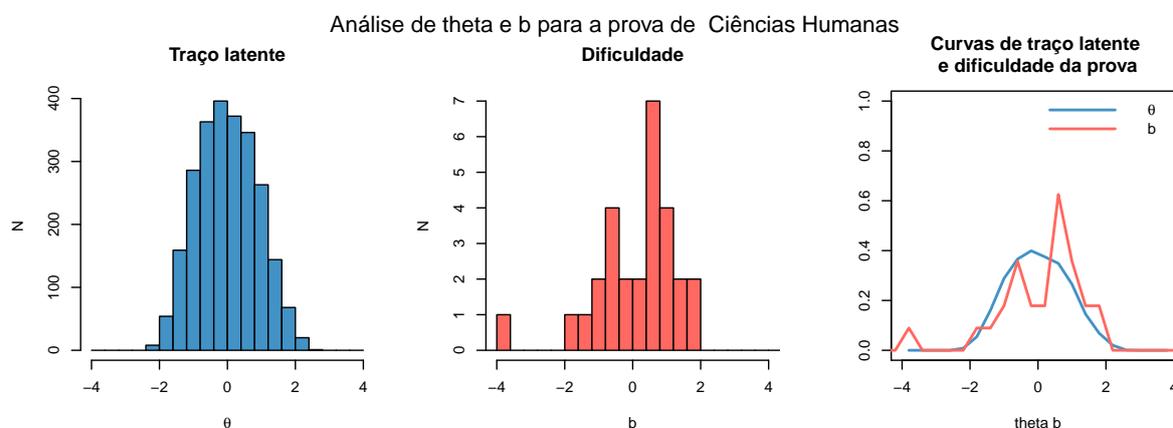


Figura 8 – Análise de proficiência e dificuldade da prova de Ciências Humanas

Com os 3 parâmetros da prova calculados bem como conhecendo os escores dos respondentes, o desempenho dos candidatos foi comparado com a dificuldade da prova.

Tomando como exemplo a prova de Ciências Humanas e considerando apenas o parâmetro b (dificuldade), em média os itens apresentam score de 0,10. Esse parâmetro é medido na mesma escala da habilidade do respondente, que corresponde ao valor do traço latente (ou teta - θ). Quanto maior o valor do parâmetro b do item, maior o nível de θ requerido para um indivíduo ter a chance de acertar o item (HAMBLETON HARIHARAN SWAMINATHAN, 1991).

Observa-se que, de acordo com os critérios estabelecidos por Pasquali (2017), a prova de Ciências Humanas tem um escore médio de dificuldade que pode ser classificado como

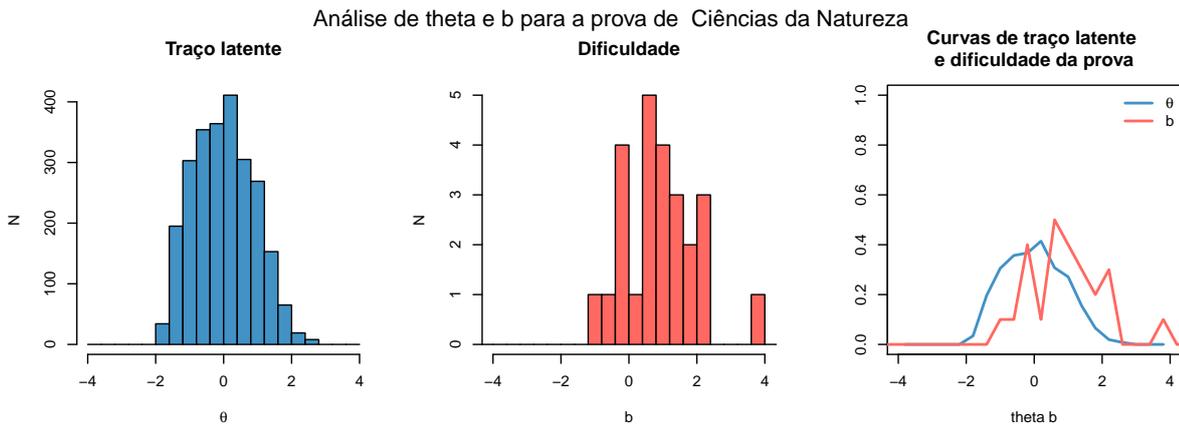


Figura 9 – Análise de proficiência e dificuldade da prova de Ciências da Natureza

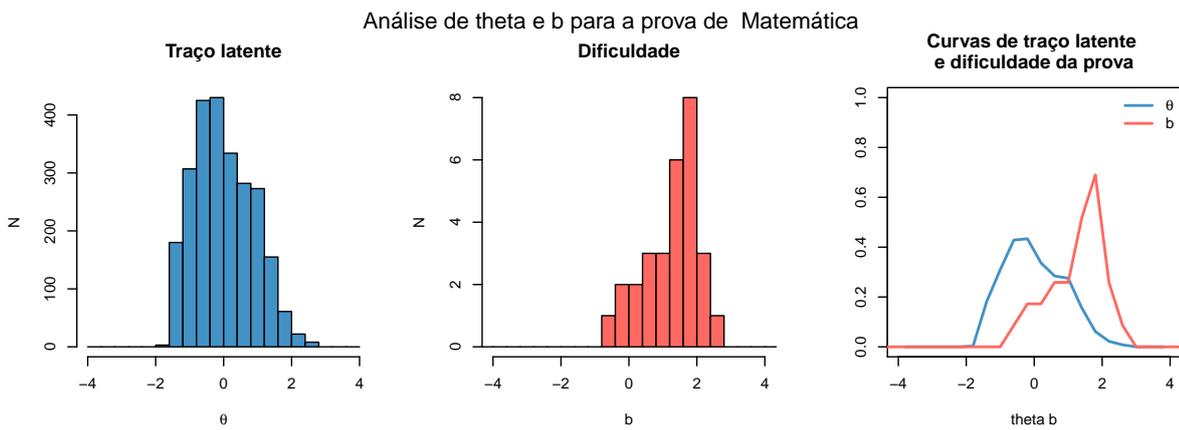


Figura 10 – Análise de proficiência e dificuldade da prova de Matemática

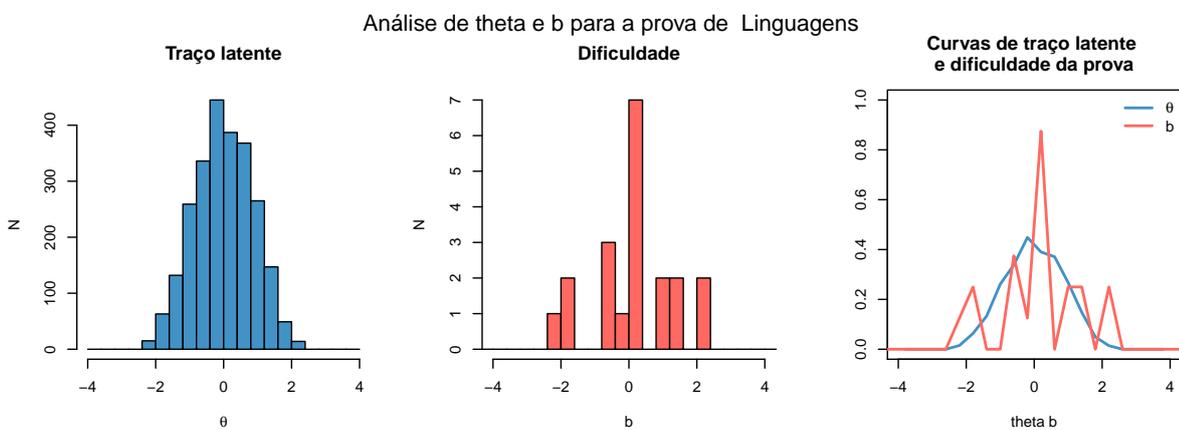


Figura 11 – Análise de proficiência e dificuldade da prova de Linguagens

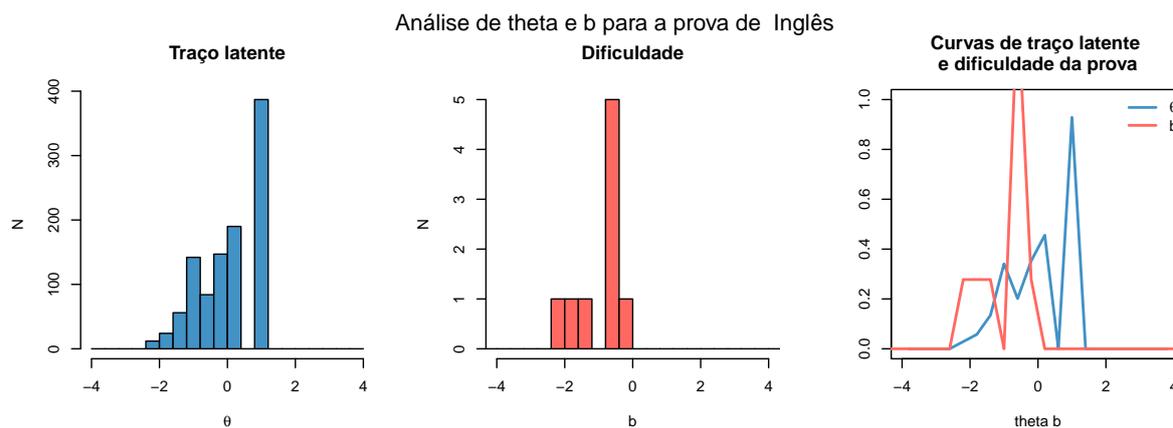


Figura 12 – Análise de proficiência e dificuldade da prova de Inglês

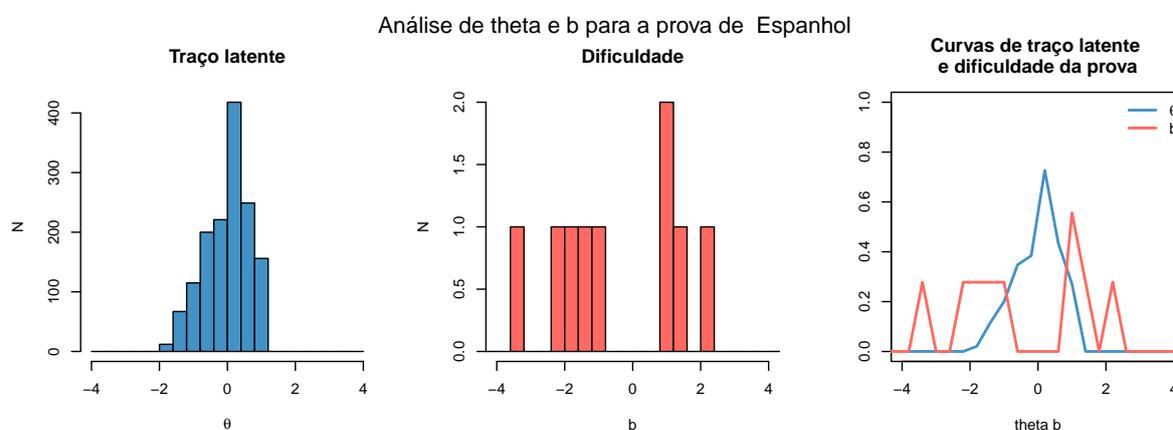


Figura 13 – Análise de proficiência e dificuldade da prova de Espanhol

“intermediário”. No entanto, constata-se na Figura 8 que as curvas de b e θ são semelhantes, demonstrando que a dificuldade da prova é adequada à habilidade dos candidatos.

Situação diferente é encontrada na prova de Matemática, onde curvas de b e θ são muito diferentes. O parâmetro b apresenta curva mais a direita que a curva de θ , caracterizando uma prova com nível de dificuldade mais alto que o nível de proficiência dos respondentes.

3.6 Análise DIF

Conforme já mencionado na Seção 2.3, o *Differential Item Functioning - DIF* trata da análise de diferença de probabilidade de acerto de um item entre grupos de indivíduos com o mesmo nível de habilidade. De acordo com Sisto (2006), um item enviesado será aquele cujas probabilidades de êxito são diferentes, apesar da igualdade da capacidade das pessoas que responderam a ele.

Existem diversos métodos de cálculo do *DIF*. Para este estudo, adotou-se o método Mantel-Haenszel por ser “amplamente utilizado pelo *Educational Testing Service (ETS)*” (ANDRIOLA, 2001).

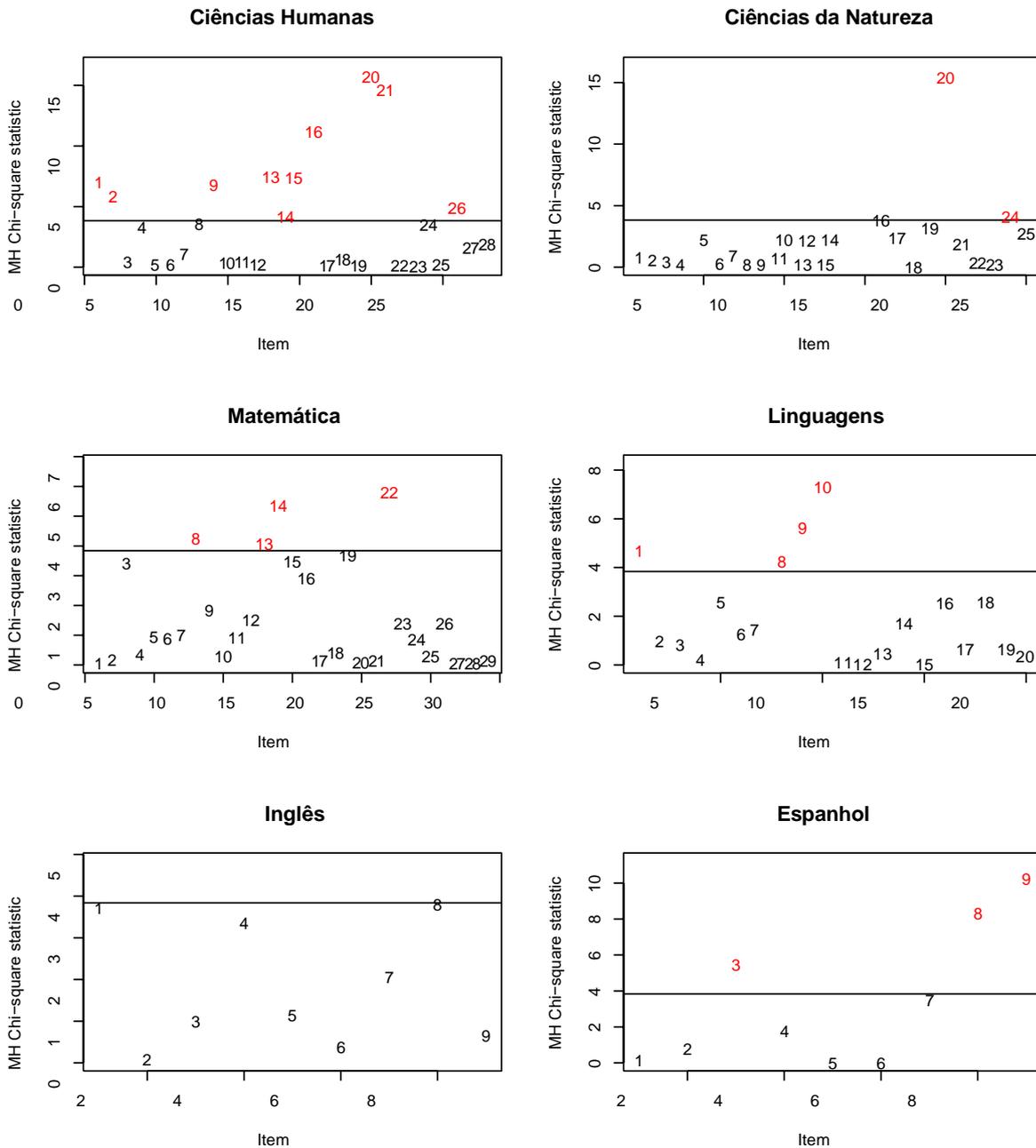


Figura 14 – Gráficos do *DIF* das provas de Ciências Humanas, Ciências da Natureza, Matemática, Linguagens, Inglês e Espanhol, respectivamente

De forma simplificada, este método infere que, se o item não apresenta *DIF*, a razão entre o número de pessoas que acertam o item e aquelas que o erram deve ser a mesma nos dois grupos comparados (referência e focal), em todos os níveis de pontuação (FIDALGO; SCALON, 2012).

O *DIF* das seis competências do instrumento foi calculado por meio do pacote *difR* (v5.1) da linguagem *R*.

Na Figura 14, pode-se verificar a representação gráfica do *DIF* nos itens, aplicados em

grupos distintos, em cada uma das disciplinas. Os números em vermelho representam os itens que apresentaram diferença estatisticamente relevante entre os respondentes cotistas e não cotistas.

A prova de Ciências Humanas apresentou a maior quantidade de itens com *DIF*, como mostra a estatística descritiva desta prova, na Tabela 7. As estatísticas das outras disciplinas encontram-se no Apêndice B.

Tabela 7 – Estatística *DIF* da prova de Ciências Humanas com o método Mantel-Haenszel

| item | MH | p.value | Signif |
|-------|---------|---------|--------|
| CH_1 | 7.0262 | 0.0080 | * |
| CH_2 | 5.8302 | 0.0158 | * |
| CH_4 | 0.2459 | 0.6200 | |
| CH_5 | 3.3080 | 0.0689 | |
| CH_6 | 0.0020 | 0.9644 | |
| CH_7 | 0.0258 | 0.8723 | |
| CH_8 | 0.8716 | 0.3505 | |
| CH_9 | 3.5518 | 0.0595 | |
| CH_10 | 6.7952 | 0.0091 | * |
| CH_11 | 0.1409 | 0.7074 | |
| CH_12 | 0.2584 | 0.6112 | |
| CH_13 | 0.0369 | 0.8477 | |
| CH_14 | 7.3953 | 0.0065 | * |
| CH_15 | 4.1285 | 0.0422 | * |
| CH_16 | 7.2936 | 0.0069 | * |
| CH_18 | 11.1589 | 0.0008 | * |
| CH_19 | 0.0080 | 0.9286 | |
| CH_20 | 0.5061 | 0.4768 | |
| CH_21 | 0.0007 | 0.9788 | |
| CH_22 | 15.6549 | 0.0001 | * |
| CH_23 | 14.6402 | 0.0001 | * |
| CH_24 | 0.1206 | 0.7284 | |
| CH_25 | 0.0078 | 0.9297 | |
| CH_26 | 3.5224 | 0.0605 | |
| CH_27 | 0.1551 | 0.6937 | |
| CH_28 | 4.8827 | 0.0271 | * |
| CH_29 | 1.5317 | 0.2159 | |
| CH_30 | 1.9130 | 0.1666 | |

Considerando o conjunto de itens com *DIF* calculado, o instrumento apresentou 23 itens com baixo desempenho quando aplicado a grupos distintos. Para fins de análise de qualidade dos itens, se este instrumento fosse uma prova de calibração, os itens destacados deveriam ser substituídos para a versão final de aplicação ou desconsiderados na composição da nota do candidato. Como a concorrência é intra-grupo, não há prejuízo para os candidatos.

4 DISCUSSÃO

Os dados foram analisados da seguinte forma:

4.1 Desempenho de Cotistas e Não Cotistas

A análise de desempenho entre cotistas e não cotistas foi realizada inicialmente em todos os candidatos do certame, sem distinção de curso. O valor de p calculado na análise de variância, juntamente com o gráfico (4) torna evidente o melhor desempenho dos candidatos não cotistas.

O modelo de cotas adotado neste certame, está alinhado com o estudo de Vilela et al. (2017), que aponta maior relevância das cotas para alunos egressos de escola pública do que as direcionadas aos alunos de baixa renda ou baseadas em etnia.

Uma vez que as vagas de cotas foram destinadas a candidatos que cursaram, exclusiva e integralmente, os anos finais do Ensino Fundamental (6º ao 9º ano) e todo o Ensino Médio (1º ao 3º ano), em escolas públicas ou Cenevistas do Estado de Alagoas, pode-se inferir que um dos fatores determinantes para o pior desempenho deste grupo no exame de seleção é a qualidade no ensino público do Estado.

Ao realizar a análise segmentando os candidatos por curso, observamos que o comportamento de melhor desempenho de alunos não cotistas, se repete para os cursos de Bacharelado em Enfermagem, Bacharelado em Fisioterapia, Bacharelado em Fonoaudiologia, Bacharelado em Medicina, Bacharelado em Terapia Ocupacional e Tecnologia em Processos Gerenciais (gráfico 5). Os cursos de Tecnologia em Análise e Desenvolvimento de Sistemas e Tecnologia em Radiologia não apresentam diferença estatisticamente relevante entre suas médias. Porém, o curso de Tecnologia em Sistemas Biomédicos apresentou diferença entre as médias dos grupos, sendo o melhor desempenho dos cotistas.

O curso de Tecnologia em Sistemas Biomédicos contou com número baixo de inscritos para a realização da prova (24 não cotistas e 16 cotistas). Este N reduzido pode contribuir com resultados discrepantes, no entanto, é um resultado relevante e poderia ter suas causas rastreadas. Informações como a escola de origem dos candidatos, composição familiar e situação de vulnerabilidade social seriam interessantes para uma análise mais apurada.

4.2 Instrumento de Avaliação

A prova contou com 120 itens de 6 competências, que foram aplicados em dois dias. A banca de elaboração revisou e anulou 5 destes itens, sendo 1 de Ciências Humanas, 2 de Ciências da Natureza, 1 de Língua Estrangeira (tanto na prova de Inglês quanto na prova de Espanhol) e 1 de Matemática. Embora o ideal seja que todos os itens da prova sejam válidos, esta quantidade de anulações (4%), não representa um problema para a qualidade da prova.

4.2.1 Qualidade dos Itens

Utilizando a TRI para análise dos itens, foram identificados 6 deles com parâmetro a inadequado, ou seja, que não discriminam bem o traço latente dos seus respondentes. Foram 3 itens na prova de Ciências de Natureza, 1 na de Ciências Humanas, 1 na de Matemática e 1 na de Espanhol. Aqui, optou-se por remover estes itens das análises.

De acordo com os critérios estabelecidos por Pasquali (2017), considerando-se o parâmetro b (dificuldade), classifica-se como **fácil** a prova de Inglês, **mediana** as provas de Ciências Humanas, Linguagens e Espanhol, e **difícil** as provas de Ciências da Natureza e Matemática.

Com relação ao parâmetro a (discriminação), a escala de Baker (2001) classifica as provas de Ciências da Humanas, Linguagens e Espanhol como discriminação **moderada**, Ciências da Natureza e Matemática como de **alta** discriminação e a prova de Inglês como sendo de discriminação **muito alta**.

Desta forma, pode-se afirmar que a instrumento aplicado possui qualidade adequada para a seleção de candidatos em concurso vestibular.

4.2.2 Capacidade de avaliação em grupos distintos

Com a análise do *DIF* foram identificados 23 itens com desempenho diferente para grupos de cotistas e não cotistas, sendo 10 itens na prova de Ciências Humanas, 2 na de Ciências da Natureza, 4 na de Matemática, 4 na de Linguagens e 3 na de Espanhol.

Estes resultados, com atenção especial para a prova de Ciências Humanas, demonstram que esse aspecto precisa ser considerado para composição do escore de rendimento na prova. Os itens que são utilizados em provas de vestibular demandam ao menos algum sistema prévio de classificação e validação do nível de dificuldade - como o método Angoff.

5 CONCLUSÃO

O ingresso de um aluno no ensino superior é um marco importante, não só para aquele que inicia sua jornada na graduação, mas também para a instituição que o acolhe e para a sociedade como um todo. Proporcionar seleções justas e adequadas ao que a sociedade demanda é de fundamental importância neste processo.

Contudo, esta não é uma tarefa fácil. Produzir instrumentos de avaliação de alta qualidade é bastante complexo. Criar itens relevantes para compor uma prova e garantir que estes possuam dificuldade adequada e boa discriminação, demanda recursos humanos e materiais. Além disso, o uso de ferramentas como a TRI e *DIF* é complexo e sua difusão ainda é difícil.

Mesmo assim, foi possível constatar com precisão que há sim diferença entre desempenho de cotistas e não cotistas; sendo o melhor desempenho dos alunos não cotistas. Exceto para o curso de Tecnologia em Sistemas Biomédicos. Como já foi dito anteriormente, este grupo precisa ser investigado e variáveis como renda e escolaridade dos pais precisam ser levados em consideração. Além da possibilidade de ser um efeito coerente com a lei dos grandes números. Note que é o menor *slot* da população estudada (N = 24 - não cotistas; N = 16 - cotistas).

No geral, a prova atende aos critérios de avaliação. Mas, para um uso mais acurado e para a construção da régua de proficiência seria demandado eliminar alguns itens para composição das notas dos estudantes. Trata-se de um procedimento recorrente, inclusive no ENEM. Não por acaso, não se pode chegar a própria nota somando o número de acertos na prova. Portanto, o instrumento apresenta qualidade psicométrica, mas a composição da nota no certame em estudo é feita com base apenas na TCT. Versões futuras precisam considerar o uso e aplicação da TRI para estabelecer a proficiência dos alunos.

Também, identificou-se o funcionamento diferencial de alguns itens entre os dois grupos estudados. Tal achado, é importante para destacar a manutenção e sistemática de avaliação considerando o aspecto de justiça social. A variável que impacta nesta diferença é a formação prévia do estudante. Apesar do *DIF* evidenciado, os candidatos não foram prejudicados, pois a composição da nota implicou em ranqueamento entre os cotista e os não cotistas em separado. Isto é, a nota dos não cotista foi comparada com os indivíduos do mesmo grupo e, de igual forma, a nota dos cotistas foi comparada com os indivíduos do mesmo grupo.

Como direcionamento para estudos futuros é importante considerar, no caso da medicina, o Teste do Progresso (TP). O TP é um instrumento de monitoramento da performance do estudante de medicina ao longo dos anos de formação e constitui uma importante ferramenta avaliativa de caráter formativo. Deste modo, ter-se-á um estudo longitudinal com os ingressos cotistas e não cotistas, sendo de suma importância para avaliação do aluno ao longo dos anos de formação.

Por fim, o mais importante é utilizar as melhores práticas conhecidas pela ciência para

tomada de decisão com base em dados. Não se pode fazer política pública e justiça social baseada em evidências, sem as evidências. É neste particular que o presente estudo se insere, evidenciado “como?” e “onde?” as diferenças de desempenho dos alunos cotista e não cotistas aparecem em um processo seletivo vestibular nos moldes do ENEM. Para proporcionar aos candidatos e universidades um processo em que sejam selecionados os melhores candidatos de acordo com a necessidade das instituições, não há outro caminho que não seja por meio do uso e aplicação da TRI e construção da régua de proficiência.

REFERÊNCIAS

- AGUIAR, G. d. S. O Funcionamento Diferencial do Item (DIF) como estratégia para captar ênfases curriculares diferenciadas em matemática. *Estudos em Avaliação Educacional*, v. 21, n. 45, p. 169, 4 2010. ISSN 1984-932X. Disponível em: <<http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2032>>. Citado na página 30.
- ALVARENGA, C. F. et al. Desafios do Ensino Superior para Estudantes de Escola Pública: um Estudo na UFLA. *Revista Pensamento Contemporâneo em Administração*, v. 6, n. 1, p. 55, 2012. ISSN 1982-2596. Citado na página 22.
- ANDRADE, D. F. d.; TAVARES, H. R.; VALLE, R. d. C. *Teoria da Resposta ao Item: Conceitos e Aplicações*. 1. ed. [s.n.], 2000. 151 p. Disponível em: <https://docs.ufpr.br/~aanjos/CE095/LivroTRI_DALTON.pdf>. Citado 2 vezes nas páginas 28 e 29.
- ANDRIOLA, W. B. Descrição dos Principais Métodos para Detectar o Funcionamento Diferencial dos Itens (DIF). *Psicologia: Reflexão e Crítica*, v. 14, n. 3, p. 643–652, 2001. ISSN 0102-7972. Citado na página 41.
- ANDRIOLA, W. B. Doze motivos favoráveis à adoção do exame nacional do ensino médio (ENEM) pelas instituições federais de ensino superior (IFES). *Ensaio*, v. 19, n. 70, p. 107–126, 2011. ISSN 01044036. Citado na página 21.
- BAKER, F. *Baker (2001) Item Response Theory.pdf*. Second. ERIC Clearinghouse on Assessment and Evaluation, 2001. 186 p. ISBN 1-886047-03-0. Disponível em: <<http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED458219%5Cnpapers2://publication/uuid/53C840DD-C92B-4719-8EC3-AF2076EDCAB3>>. Citado 2 vezes nas páginas 28 e 46.
- BECHGER, T. M. et al. Using Classical Test Theory in Combination with Item Response Theory. *Applied Psychological Measurement*, v. 27, n. 5, p. 319–334, 9 2003. ISSN 0146-6216. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0146621603257518>>. Citado na página 26.
- Brasi. *LEI Nº 8.213*. 1991. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L8213compilado.htm>. Citado na página 23.
- Brasil. Entenda a sua Nota no ENEM: Guia do Participante. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*, p. 38, 2012. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/guia_participante/2012/guia_do_participante_notas_2012.pdf>. Citado na página 38.
- Brasil. *LEI Nº 12.711*. 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112711.htm>. Citado na página 21.
- Brasil. *LEI Nº 12.990*. 2014. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112990.htm>. Citado na página 23.
- CAVALCANTI, I. T. d. N. *Análise dos diferenciais de desempenho dos estudantes cotistas e não cotistas da UFBA pelo propensity score matching*. Tese (Doutorado) — Universidade Federal da Bahia, 2015. Disponível em: <<http://repositorio.ufba.br/ri/handle/ri/18125>>. Citado na página 22.

EDUCAÇÃO, C. N. d. *Cp 98/99*. 1999. Disponível em: <<http://portal.mec.gov.br/cne/arquivos/pdf/PNCP098.pdf>>. Citado na página 21.

EMBRETSON, S. P. R. S. E. *Item Response Theory for Psychologists*. [S.l.]: Psychology Press, 2013. ISBN 0-8058-2818-4. Citado na página 27.

FIDALGO, A. M.; SCALON, J. D. Uso dos métodos Mantel-Haenszel para a detecção do funcionamento diferencial dos itens e software relacionado. *Psicologia: Reflexão e Crítica*, v. 25, n. 1, p. 60–68, 2012. ISSN 0102-7972. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-79722012000100008&lng=pt&nrm=iso&tlng=en>. Citado na página 42.

HAMBLETON HARIHARAN SWAMINATHAN, H. J. R. R. K. *Fundamentals of Item Response Theory*. 1. ed. [S.l.]: Sage, 1991. ISBN 0-8039-3646-X. Citado 4 vezes nas páginas 27, 28, 29 e 39.

HUTZ, C. S.; BANDEIRA, D. R.; TRENTINI, C. M. *Psicometria*. Artmed Editora, 2015. (Avaliação Psicológica). ISBN 9788582712368. Disponível em: <<https://books.google.com.br/books?id=cVIICgAAQBAJ>>. Citado na página 25.

IBGE. *IBGE - Instituto Brasileiro de Geografia e Estatística*. 2019. 1–5 p. Disponível em: <<https://www.ibge.gov.br/>>. Citado na página 21.

JÚNIOR, F. J. d. S. *Diferencial de desempenho entre alunos cotistas e não cotistas nos cursos do Centro de Tecnologia da UFC*. Tese (Doutorado) — Universidade Federal do Ceará, 2018. Disponível em: <<http://www.repositorio.ufc.br/handle/riufc/34529>>. Citado na página 22.

KLEIN, R. Fundamentos da Teoria da Resposta ao Item –TRI. *Meta: Avaliação*, v. 1, n. 2, p. 283–296, 2009. ISSN 21752753. Citado na página 29.

KLEIN, R. Alguns aspectos da teoria de resposta ao item relativos à estimação das proficiências. *Ensaio*, v. 21, n. 78, p. 35–56, 2013. ISSN 01044036. Citado na página 37.

LINDEN, W. J. van der. *Handbook of Item Response Theory, Volume One*. Boca Raton, FL : CRC Press, 2015-: Chapman and Hall/CRC, 2016. ISBN 9781315374512. Disponível em: <<https://www.taylorfrancis.com/books/9781315374512>>. Citado 2 vezes nas páginas 26 e 27.

MARQUES, T. M.; RODRIGUES, R. a Política De Cotas Como Forma De Acesso Democrático Ao Ensino Superior Público. *Revista Ciências Humanas*, v. 13, n. 2, p. 74–83, 2020. ISSN 2179-1120. Citado na página 23.

Ministério da Educação do Brasil. *PORTARIA Nº 391*. 2002. Disponível em: <<https://pesquisa.in.gov.br/imprensa/servlet/INPDFViewer?jornal=1&pagina=4&data=13/02/2002&captchafield=firstAccess>>. Citado na página 21.

MOCELIN, C. E. Uma análise marxiana da política de cotas no ensino superior público brasileiro. *Revista Katálysis*, v. 23, n. 1, p. 101–110, 2020. ISSN 1982-0259. Citado na página 23.

NUNES, G. H.; OLIVEIRA, B. A. S.; NAMETALA, C. A. L. A Computational Strategy for Classification of Enem Issues Based on Item Response Theory. *Revista de Informática Teórica e Aplicada*, v. 27, n. 1, p. 92–111, 2020. ISSN 01034308. Citado na página 29.

NUNNALLY, I. B. J. *Psychometric Theory*. New York: McGraw-Hill, 1967. 736 p. ISBN 978-0070478497. Citado na página 27.

PASQUALI, L. *Psicometria - Teoria dos Testes na Psicologia e na Educação*. [s.n.], 2009. v. 43. 992–999 p. ISSN 0080-6234. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-62342009000500002&lng=pt&tlng=pt>. Citado na página 27.

PASQUALI, L. *Psicometria - Teoria dos Testes na Psicologia e na Educação*. 4. ed. [S.l.]: Editora Vozes, 2011. ISBN 9788532656124. Citado 2 vezes nas páginas 25 e 26.

PASQUALI, L. *Psicometria - Teoria dos Testes na Psicologia e na Educação*. [S.l.]: Editora Vozes, 2017. Citado 4 vezes nas páginas 27, 28, 39 e 46.

QUEIROZ, Z. C. L. S. et al. A lei de cotas na perspectiva do desempenho acadêmico na Universidade Federal de Uberlândia. *Revista Brasileira de Estudos Pedagógicos*, v. 96, n. 243, p. 299–320, 8 2015. ISSN 2176-6681. Disponível em: <http://emaberto.inep.gov.br/index.php/rbep/article/view/242http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2176-66812015000200299&lng=pt&tlng=pt>. Citado na página 22.

SANTOS, H.; SOUZA, M. G. d.; SASAKI, K. O subproduto social advindo das cotas raciais na educação superior do Brasil. *Revista Brasileira de Estudos Pedagógicos*, v. 94, n. 237, p. 542–563, 2013. ISSN 2176-6681. Citado na página 22.

SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. d. Avanços na psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica*, v. 26, n. 2, p. 241–250, 2013. ISSN 0102-7972. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-79722013000200004&lng=pt&nrm=iso&tlng=en>. Citado na página 26.

SISTO, F. F. O funcionamento diferencial dos itens. *Psico-USF*, v. 11, n. 1, p. 35–43, 6 2006. ISSN 1413-8271. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-82712006000100005&lng=pt&tlng=pt>. Citado 3 vezes nas páginas 23, 30 e 41.

STEINBERG, L.; THISSEN, D. Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*, v. 11, n. 4, p. 402–415, 12 2006. ISSN 1939-1463. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/1082-989X.11.4.402>>. Citado na página 30.

THEODORO, M. L. As políticas públicas e a desigualdade racial no Brasil: 120 anos após a abolição: in Capítulo 1, A FORMAÇÃO DO MERCADO DE TRABALHO E A QUESTÃO RACIAL NO BRASIL. *Ipea*, p. 15–44, 2008. Disponível em: <http://www.ipea.gov.br/portal/images/stories/Livro_desigualdadesraciais.pdf>. Citado na página 22.

VILELA, L. et al. As cotas nas universidades públicas diminuem a qualidade dos ingressantes? *Estudos em Avaliação Educacional*, v. 28, n. 69, p. 652, 12 2017. ISSN 1984-932X. Disponível em: <<http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/4427>>. Citado 2 vezes nas páginas 21 e 45.

APÊNDICE A – CURVA CARACTERÍSTICA DOS ITENS DO INSTRUMENTO POR COMPETÊNCIAS

Curva característica dos itens de Ciências Humanas

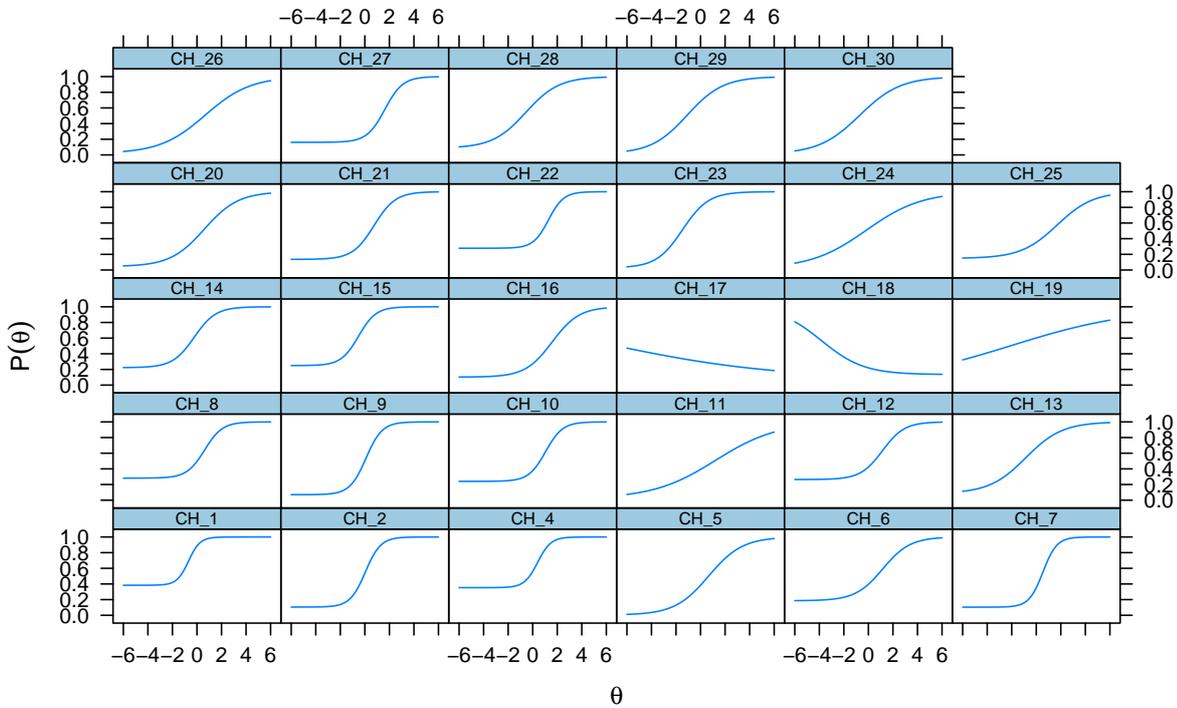


Figura 15 – Curva Característica dos Itens da prova de Ciências Humanas

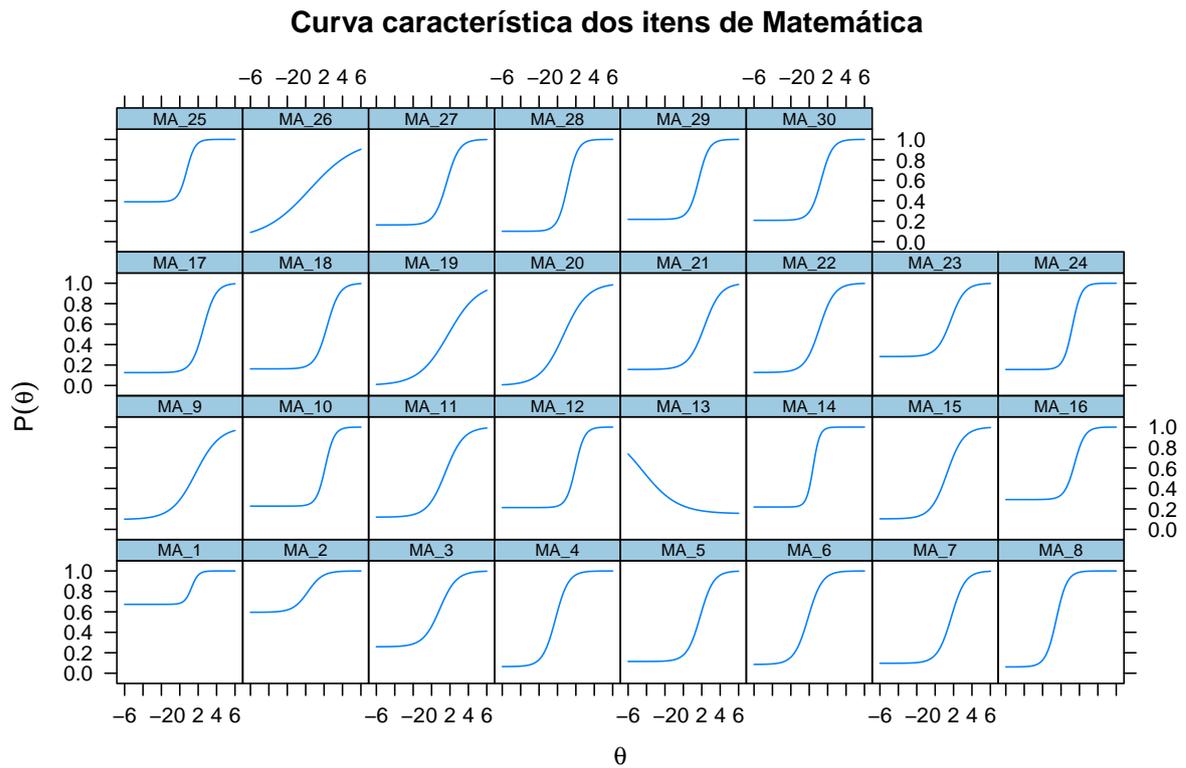


Figura 16 – Curva Característica dos Itens da prova de Matemática

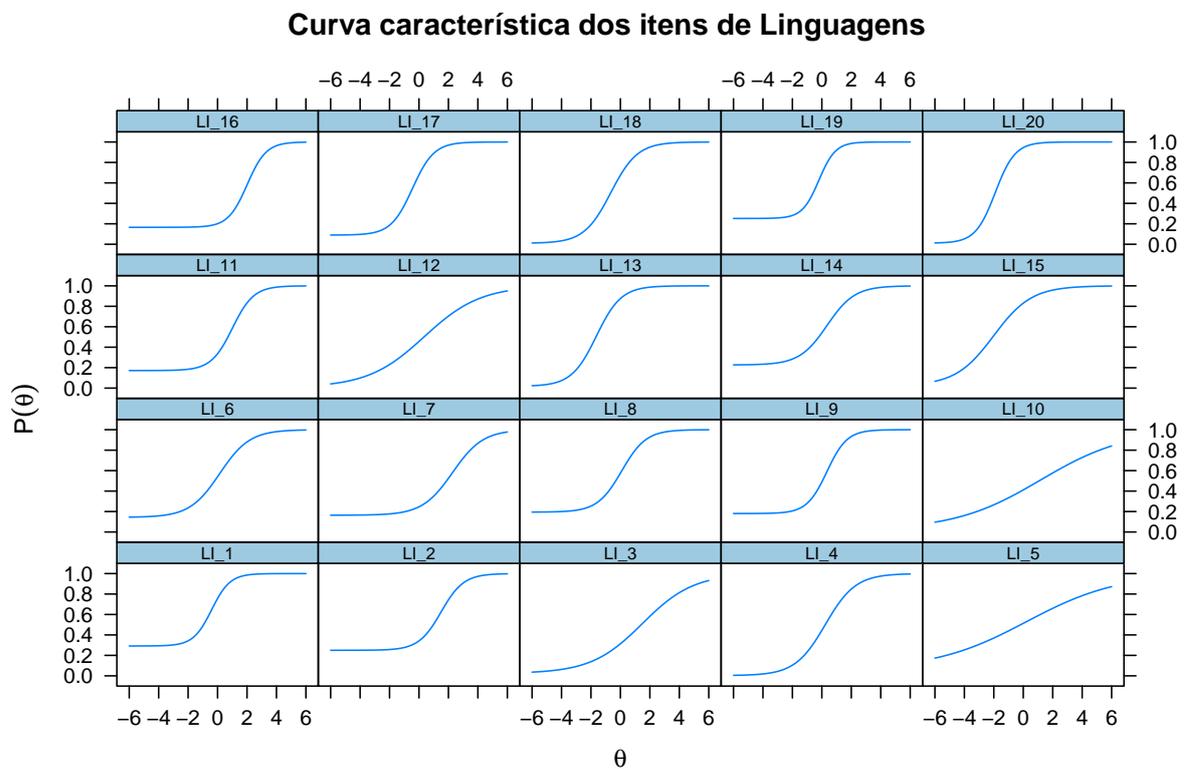


Figura 17 – Curva Característica dos Itens da prova de Linguagens

Curva característica dos itens de Inglês

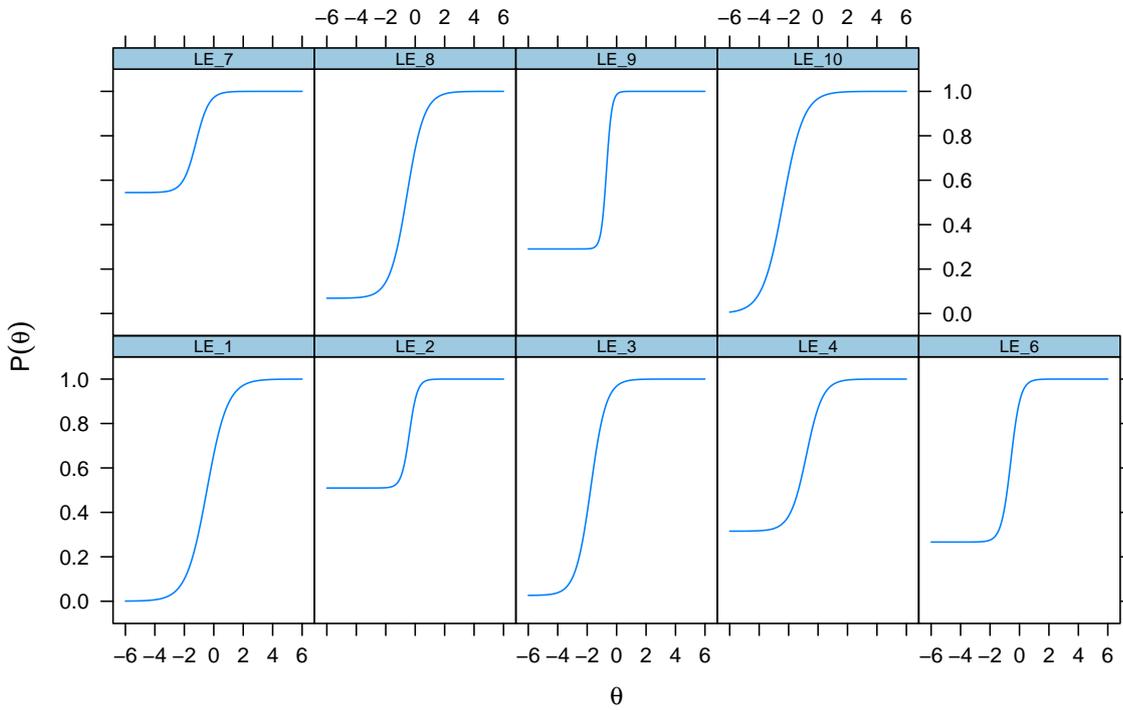


Figura 18 – Curva Característica dos Itens da prova de Inglês

Curva característica dos itens de Espanhol

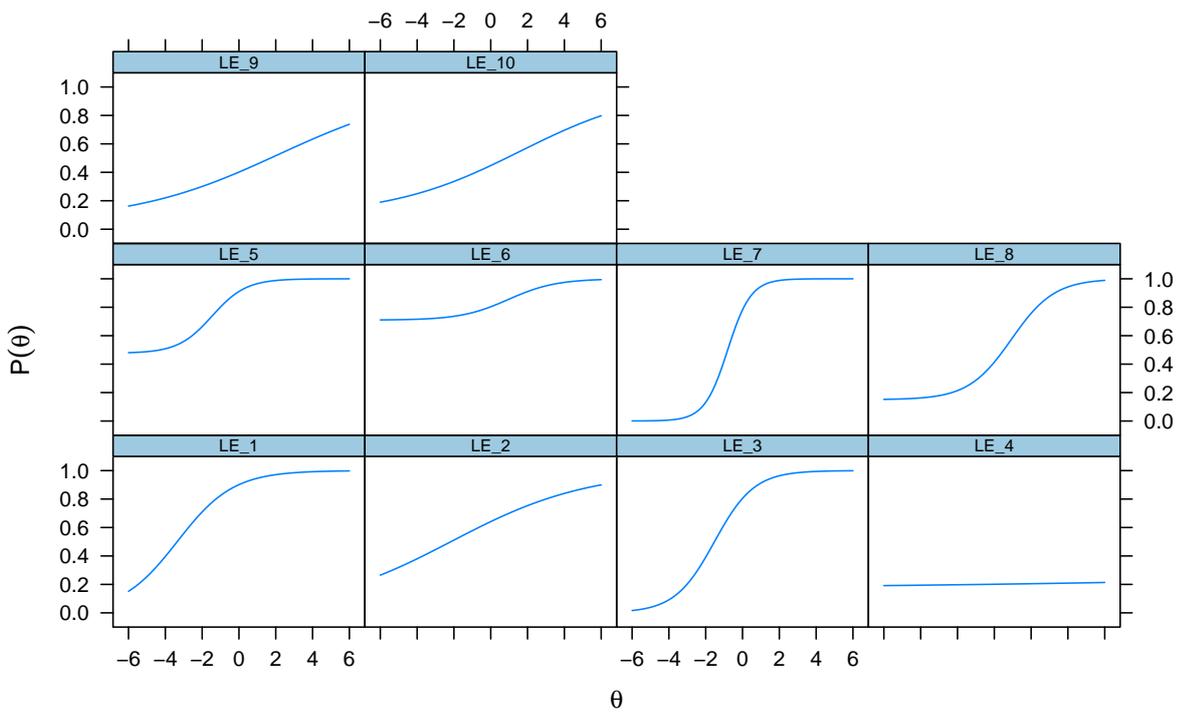


Figura 19 – Curva Característica dos Itens da prova de Espanhol

**APÊNDICE B – ESTATÍSTICA DIF DOS ITENS DO INSTRUMENTO POR
COMPETÊNCIAS**

Tabela 8 – Estatística *DIF* da prova de Ciências da Natureza com o método Mantel-Haenszel

| item | MH | p.value | Signif |
|-------|---------|---------|--------|
| CN_1 | 0.8300 | 0.3623 | |
| CN_2 | 0.5117 | 0.4744 | |
| CN_3 | 0.3089 | 0.5784 | |
| CN_4 | 0.1249 | 0.7237 | |
| CN_5 | 2.1698 | 0.1407 | |
| CN_6 | 0.1175 | 0.7318 | |
| CN_8 | 0.7472 | 0.3874 | |
| CN_9 | 0.0007 | 0.9787 | |
| CN_10 | 0.0326 | 0.8567 | |
| CN_11 | 2.0107 | 0.1562 | |
| CN_12 | 0.4843 | 0.4865 | |
| CN_13 | 1.9322 | 0.1645 | |
| CN_14 | 0.0404 | 0.8408 | |
| CN_15 | 2.0176 | 0.1555 | |
| CN_16 | 0.0289 | 0.8651 | |
| CN_17 | 3.7560 | 0.0526 | |
| CN_18 | 2.3156 | 0.1281 | |
| CN_19 | 0.0006 | 0.9806 | |
| CN_21 | 3.1707 | 0.0750 | |
| CN_23 | 15.4026 | 0.0001 | * |
| CN_24 | 1.8423 | 0.1747 | |
| CN_25 | 0.4758 | 0.4903 | |
| CN_26 | 0.2151 | 0.6428 | |
| CN_28 | 3.9231 | 0.0476 | * |
| CN_30 | 2.7312 | 0.0984 | |

Tabela 9 – Estatística *DIF* da prova de Matemática com o método Mantel-Haenszel

| item | MH | p.value | Signif |
|-------|--------|---------|--------|
| MA_1 | 0.0186 | 0.8914 | |
| MA_2 | 0.1616 | 0.6876 | |
| MA_3 | 3.4030 | 0.0651 | |
| MA_4 | 0.3568 | 0.5503 | |
| MA_5 | 1.0209 | 0.3123 | |
| MA_6 | 0.9239 | 0.3365 | |
| MA_7 | 1.0580 | 0.3037 | |
| MA_8 | 4.2314 | 0.0397 | * |
| MA_9 | 1.8407 | 0.1749 | |
| MA_10 | 0.2862 | 0.5926 | |
| MA_11 | 0.9240 | 0.3364 | |
| MA_12 | 1.4678 | 0.2257 | |
| MA_14 | 4.0538 | 0.0441 | * |
| MA_15 | 5.3434 | 0.0208 | * |
| MA_16 | 3.3939 | 0.0654 | |
| MA_17 | 2.8985 | 0.0887 | |
| MA_18 | 0.1037 | 0.7475 | |
| MA_19 | 0.3641 | 0.5462 | |
| MA_20 | 3.6830 | 0.0550 | |
| MA_21 | 0.0660 | 0.7973 | |
| MA_22 | 0.1023 | 0.7491 | |
| MA_23 | 5.7825 | 0.0162 | * |
| MA_24 | 1.3797 | 0.2401 | |
| MA_25 | 0.7346 | 0.3914 | |
| MA_26 | 0.2125 | 0.6448 | |
| MA_27 | 1.3799 | 0.2401 | |
| MA_28 | 0.0262 | 0.8714 | |
| MA_29 | 0.0375 | 0.8465 | |
| MA_30 | 0.0789 | 0.7787 | |

Tabela 10 – Estatística *DIF* da prova de Linguagens com o método Mantel-Haenszel

| item | MH | p.value | Signif |
|-------|--------|---------|--------|
| LI_1 | 4.6775 | 0.0306 | * |
| LI_2 | 0.9960 | 0.3183 | |
| LI_3 | 0.7414 | 0.3892 | |
| LI_4 | 0.2139 | 0.6437 | |
| LI_5 | 2.5752 | 0.1085 | |
| LI_6 | 1.2549 | 0.2626 | |
| LI_7 | 1.4490 | 0.2287 | |
| LI_8 | 4.2298 | 0.0397 | * |
| LI_9 | 5.6281 | 0.0177 | * |
| LI_10 | 7.2845 | 0.0070 | * |
| LI_11 | 0.0912 | 0.7627 | |
| LI_12 | 0.0205 | 0.8862 | |
| LI_13 | 0.4845 | 0.4864 | |
| LI_14 | 1.7292 | 0.1885 | |
| LI_15 | 0.0195 | 0.8888 | |
| LI_16 | 2.5413 | 0.1109 | |
| LI_17 | 0.6475 | 0.4210 | |
| LI_18 | 2.6296 | 0.1049 | |
| LI_19 | 0.6611 | 0.4162 | |
| LI_20 | 0.2920 | 0.5889 | |

Tabela 11 – Estatística *DIF* da prova de Inglês com o método Mantel-Haenszel

| item | MH | p.value | Signif |
|-------|--------|---------|--------|
| LE_1 | 3.7183 | 0.0538 | |
| LE_2 | 0.0980 | 0.7543 | |
| LE_3 | 1.0038 | 0.3164 | |
| LE_4 | 3.3628 | 0.0667 | |
| LE_6 | 1.1396 | 0.2857 | |
| LE_7 | 0.3921 | 0.5312 | |
| LE_8 | 2.0662 | 0.1506 | |
| LE_9 | 3.8017 | 0.0512 | |
| LE_10 | 0.6563 | 0.4179 | |

Tabela 12 – Estatística *DIF* da prova de Espanhol com o método Mantel-Haenszel

| item | MH | p.value | Signif |
|-------|---------|---------|--------|
| LE_1 | 0.1592 | 0.6899 | |
| LE_2 | 0.8056 | 0.3694 | |
| LE_3 | 5.4503 | 0.0196 | * |
| LE_5 | 1.7493 | 0.1860 | |
| LE_6 | 0.0002 | 0.9883 | |
| LE_7 | 0.0214 | 0.8836 | |
| LE_8 | 3.4790 | 0.0622 | |
| LE_9 | 8.3338 | 0.0039 | * |
| LE_10 | 10.2254 | 0.0014 | * |

APÊNDICE C – CÓDIGO R PARA EXECUÇÃO DAS ANÁLISES

```
1 #CONFIGURAÇÕES INICIAIS
2
3 #Atualizando R
4 install.packages("installr")
5 library(installr)
6 updateR()
7
8 #Instalando pacotes necessários
9 install.packages("readr")
10 install.packages("ggplot2")
11 install.packages("Rcmdr")
12 install.packages("psych")
13 install.packages("mirt")
14 install.packages("latticeExtra")
15 install.packages("dplyr")
16 install.packages("stringr")
17 install.packages("gridExtra")
18 install.packages("difR")
19 install.packages("ltm")
20 install.packages("lme4")
21 install.packages("dplyr")
22 install.packages("tidyr")
23
24 #Carregando pacotes
25 library(readr)
26 library(ggplot2)
27 library(Rcmdr)
28 library(psych)
29 library(mirt)
30 library(latticeExtra)
31 library(dplyr)
32 library(stringr)
33 library(gridExtra)
34 library(difR)
35 library(ltm)
36 library(lme4)
37 library(dplyr)
38 library(tidyr)
39
40
41 #Tratamento dos dados
42 #Importando base de dados
43
44 todos_2015 <- read_delim("C:/Dados/R/todos_2015.txt",
45                         "\t", escape_double = FALSE, locale = locale(decimal_mark = ",",
46                         grouping_mark = "."),
47                         encoding = "WINDOWS-1252"),
48                         trim_ws = TRUE)
49 #Selecionando respostas de Medicina - Candidatos não faltosos
50 med_2015 <- subset(todos_2015, CURSO=="004 - BACHARELADO EM MEDICINA" & ((STATUS=="Convocado")
51 | (STATUS=="Habilitado") | (STATUS=="Eliminado na redação")))
52
53 #Separando itens por prova
54 campos_itens_ch <- c("CH_1", "CH_2", "CH_4", "CH_5", "CH_6", "CH_7", "CH_8", "CH_9", "
55 CH_10", "CH_11", "CH_12", "CH_13", "CH_14", "CH_15", "CH_16", "CH_17", "CH_18", "CH_19", "
56 CH_20", "CH_21", "CH_22", "CH_23", "CH_24", "CH_25", "CH_26", "CH_27", "CH_28", "CH_29", "
```

```

    CH_30")
53 campos_itens_cn <- c("CN_1", "CN_2", "CN_3", "CN_4", "CN_5", "CN_6", "CN_7", "CN_8", "CN_9", "
    CN_10", "CN_11", "CN_12", "CN_13", "CN_14", "CN_15", "CN_16", "CN_17", "CN_18", "CN_19", "
    CN_20", "CN_21", "CN_23", "CN_24", "CN_25", "CN_26", "CN_27", "CN_28", "
    CN_30")
54 campos_itens_ma <- c("MA_1", "MA_2", "MA_3", "MA_4", "MA_5", "MA_6", "MA_7", "MA_8", "MA_9", "
    MA_10", "MA_11", "MA_12", "MA_13", "MA_14", "MA_15", "MA_16", "MA_17", "MA_18", "MA_19", "
    MA_20", "MA_21", "MA_22", "MA_23", "MA_24", "MA_25", "MA_26", "MA_27", "MA_28", "MA_29", "
    MA_30")
55 campos_itens_li <- c("LI_1", "LI_2", "LI_3", "LI_4", "LI_5", "LI_6", "LI_7", "LI_8", "LI_9", "
    LI_10", "LI_11", "LI_12", "LI_13", "LI_14", "LI_15", "LI_16", "LI_17", "LI_18", "LI_19", "
    LI_20")
56 campos_itens_in <- c("LE_1", "LE_2", "LE_3", "LE_4", "LE_6", "LE_7", "LE_8", "LE_9", "
    LE_10")
57 campos_itens_es <- c("LE_1", "LE_2", "LE_3", "LE_4", "LE_5", "LE_6", "LE_7", "LE_8", "LE_9", "
    LE_10")
58
59 itens_ch <- med_2015[campos_itens_ch]
60 itens_cn <- med_2015[campos_itens_cn]
61 itens_ma <- med_2015[campos_itens_ma]
62 itens_li <- med_2015[campos_itens_li]
63 itens_in <- subset(med_2015, IDIOMA=="I", select=campos_itens_in)
64 itens_es <- subset(med_2015, IDIOMA=="E", select=campos_itens_es)
65
66
67 #Removendo variáveis com resposta única
68 #Este procedimento foi realizado removendo os nomes das colunas nas variáveis de campos a
    partir das mensagens de retorno na construção do modelo TRI
69
70 #ANÁLISES INICIAIS
71
72 #Preparando conjunto de dados
73 campos_candidato <- c("COTA", "SEXO", "IDIOMA", "STATUS", "CLASSIFICACAO", "MEDIA_FINAL")
74 dados_todos <- todos_2015[c(campos_candidato, "CURSO", campos_itens_ch, campos_itens_cn,
    campos_itens_ma, campos_itens_li, campos_itens_in, campos_itens_es)]
75 dados_todos <- subset(dados_todos, ((STATUS=="Convocado") | (STATUS=="Habilitado") | (STATUS=="
    Eliminado na redação")))
76 dados_todos <- transform(dados_todos, TOTAL = rowSums(dados_todos[c(campos_itens_ch, campos_
    itens_cn, campos_itens_ma, campos_itens_li, campos_itens_in, campos_itens_es)]))
77 dados_todos["COD_CURSO"] <- substring(dados_todos$CURSO, 1, 3)
78
79 #Teste de normalidade (Shapiro-Wilk)
80 normalityTest(~TOTAL, test="lillie.test", data=dados_todos)
81
82 #Análise gráfica
83 plot.new()
84 par(mfcol = c(1,2), oma = c(0,0,0,0))
85 qqnorm(dados_todos$TOTAL, pch = 1, frame = TRUE, xlab = "", ylab = "Acertos", main = "Todos")
86 qqline(dados_todos$TOTAL, lwd = 2)
87 hist(dados_todos$TOTAL, main = "", xlab = "Acertos", ylab = "Frequência")
88
89 #Teste de variância - análise entre grupos (todos os cursos)
90 aov_todos <- lm(TOTAL ~ COTA, data = dados_todos)
91 summary(aov_todos)
92
93 #Gráficos de intervalo de confiança (todos os cursos)
94 library(sciplot)
95 par(mfcol = c(1,1), oma = c(0,0,0,0))
96 lineplot.CI(COTA, TOTAL, data = dados_todos,
97             main="Diferença entre grupos",

```

```

98     xlab = "Cotista",
99     ylab = "Itens acertados",
100    legend = TRUE,
101    leg.lab = c("Cota", "Ampla"),
102    type = "p",
103    col = c("black", "black"))
104
105 #Análise de variância por curso
106 aov_todos_curso <- aov(TOTAL ~ COD_CURSO, data=dados_todos)
107 anova(aov_todos_curso) #*** indica que existe variância
108
109 #Teste Posthoc pra verificar quais grupos se destacam
110 analyse_tukey <- TukeyHSD(aov_todos_curso)
111 analyse_tukey_df <- as.data.frame(analyse_tukey$COD_CURSO)
112
113 #Exportando para LaTeX
114 xtable(analyse_tukey_df, type = "latex", digits = c(0,4,4,4,4), align = c("lrrrr"),
115        caption = c("tbl:Análise de variância de notas por curso", "table"), label = "tbl:Aná
        lise de variância de notas por curso", )
116
117 #TCT
118
119 #Separando totalizadores por prova
120 campos_candidato <- c("COTA", "SEXO", "IDIOMA", "STATUS", "CLASSIFICACAO", "MEDIA_FINAL")
121
122 dados_ch <- med_2015[c(campos_candidato, campos_itens_ch)]
123 dados_ch <- transform(dados_ch, TOTAL = rowSums(dados_ch[campos_itens_ch]))
124
125 dados_cn <- med_2015[c(campos_candidato, campos_itens_cn)]
126 dados_cn <- transform(dados_cn, TOTAL = rowSums(dados_cn[campos_itens_cn]))
127
128 dados_ma <- med_2015[c(campos_candidato, campos_itens_ma)]
129 dados_ma <- transform(dados_ma, TOTAL = rowSums(dados_ma[campos_itens_ma]))
130
131 dados_li <- med_2015[c(campos_candidato, campos_itens_li)]
132 dados_li <- transform(dados_li, TOTAL = rowSums(dados_li[campos_itens_li]))
133
134 dados_in <- subset(med_2015, IDIOMA=="I", select=c(campos_candidato, campos_itens_in))
135 dados_in <- transform(dados_in, TOTAL = rowSums(dados_in[campos_itens_in]))
136
137 dados_es <- subset(med_2015, IDIOMA=="E", select=c(campos_candidato, campos_itens_es))
138 dados_es <- transform(dados_es, TOTAL = rowSums(dados_es[campos_itens_es]))
139
140
141 #Teste de normalidade (Shapiro-Wilk) em cada competência normal
142 normalityTest(~TOTAL, test="lillie.test", data=dados_ch)
143 normalityTest(~TOTAL, test="lillie.test", data=dados_cn)
144 normalityTest(~TOTAL, test="lillie.test", data=dados_ma)
145 normalityTest(~TOTAL, test="lillie.test", data=dados_li)
146 normalityTest(~TOTAL, test="lillie.test", data=dados_in)
147 normalityTest(~TOTAL, test="lillie.test", data=dados_es)
148
149 #Análise gráfica
150 par(mfcol = c(2,6), oma = c(4,1,1,1))
151 qqnorm(dados_ch$TOTAL, pch = 1, frame = TRUE, xlab = "", ylab = "Acertos", main = "Ciências
        Humanas")
152 qqline(dados_ch$TOTAL, lwd = 2)
153 hist(dados_ch$TOTAL, main = "", xlab = "Acertos", ylab = "Frequência")
154
155 qqnorm(dados_cn$TOTAL, pch = 1, frame = TRUE, xlab = "", ylab = "Acertos", main = "Ciências da

```

```

    Natureza")
156 qqline(dados_cn$TOTAL, lwd = 2)
157 hist(dados_cn$TOTAL, main = "", xlab = "Acertos", ylab = "Frequência")
158
159 qqnorm(dados_ma$TOTAL, pch = 1, frame = TRUE, xlab = "", ylab = "Acertos", main = "Matemática"
)
160 qqline(dados_ma$TOTAL, lwd = 2)
161 hist(dados_ma$TOTAL, main = "", xlab = "Acertos", ylab = "Frequência")
162
163 qqnorm(dados_li$TOTAL, pch = 1, frame = TRUE, xlab = "", ylab = "Acertos", main = "Linguagens"
)
164 qqline(dados_li$TOTAL, lwd = 2)
165 hist(dados_li$TOTAL, main = "", xlab = "Acertos", ylab = "Frequência")
166
167 qqnorm(dados_in$TOTAL, pch = 1, frame = TRUE, xlab = "", ylab = "Acertos", main = "Inglês")
168 qqline(dados_in$TOTAL, lwd = 2)
169 hist(dados_in$TOTAL, main = "", xlab = "Acertos", ylab = "Frequência")
170
171 qqnorm(dados_es$TOTAL, pch = 1, frame = TRUE, xlab = "", ylab = "Acertos", main = "Espanhol")
172 qqline(dados_es$TOTAL, lwd = 2)
173 hist(dados_es$TOTAL, main = "", xlab = "Acertos", ylab = "Frequência")
174
175
176 #Análise de variância entre grupos
177 aov_ch <- lm(TOTAL ~ COTA, data = dados_ch)
178 summary(aov_ch)
179
180 aov_cn <- lm(TOTAL ~ COTA, data = dados_cn)
181 summary(aov_cn)
182
183 aov_ma <- lm(TOTAL ~ COTA, data = dados_ma)
184 summary(aov_ma)
185
186 aov_li <- lm(TOTAL ~ COTA, data = dados_li)
187 summary(aov_li)
188
189 aov_in <- lm(TOTAL ~ COTA, data = dados_in)
190 summary(aov_in)
191
192 aov_es <- lm(TOTAL ~ COTA, data = dados_es)
193 summary(aov_es)
194
195
196 #Gráficos de intervalo de confiança
197 library(sciplot)
198 par(mfrow = c(2,3), oma = c(0,0,0,0))
199 lineplot.CI(COTA, TOTAL, data = dados_ch, xlab = "Cotas", ylab = "Itens acertados", type = "p"
, col = c("black","black"), main="Ciências Humanas")
200 lineplot.CI(COTA, TOTAL, data = dados_cn, xlab = "Cotas", ylab = "Itens acertados", type = "p"
, col = c("black","black"), main="Ciências da Natureza")
201 lineplot.CI(COTA, TOTAL, data = dados_ma, xlab = "Cotas", ylab = "Itens acertados", type = "p"
, col = c("black","black"), main="Matemática")
202 lineplot.CI(COTA, TOTAL, data = dados_li, xlab = "Cotas", ylab = "Itens acertados", type = "p"
, col = c("black","black"), main="Linguagens")
203 lineplot.CI(COTA, TOTAL, data = dados_in, xlab = "Cotas", ylab = "Itens acertados", type = "p"
, col = c("black","black"), main="Inglês")
204 lineplot.CI(COTA, TOTAL, data = dados_es, xlab = "Cotas", ylab = "Itens acertados", type = "p"
, col = c("black","black"), main="Espanhol")
205
206 #TRI

```

```

207
208 #Calibra itens
209 modTRI3_ini_ch <- mirt(itens_ch, 1, itemtype='3PL');
210 modTRI3_ini_cn <- mirt(itens_cn, 1, itemtype='3PL');
211 modTRI3_ini_ma <- mirt(itens_ma, 1, itemtype='3PL');
212 modTRI3_ini_li <- mirt(itens_li, 1, itemtype='3PL');
213 modTRI3_ini_in <- mirt(itens_in, 1, itemtype='3PL');
214 modTRI3_ini_es <- mirt(itens_es, 1, itemtype='3PL');
215
216
217 #Parâmetros dos itens (parametrização intercept e slope)
218 coef(modTRI3_ini_ch, simplify=TRUE)
219 coef(modTRI3_ini_cn, simplify=TRUE)
220 coef(modTRI3_ini_ma, simplify=TRUE)
221 coef(modTRI3_ini_li, simplify=TRUE)
222 coef(modTRI3_ini_in, simplify=TRUE)
223 coef(modTRI3_ini_es, simplify=TRUE)
224
225 #Parâmetros dos itens (parametrização a e b TRI)
226 coef(modTRI3_ini_ch, simplify=TRUE, IRTpars=TRUE)
227 coef(modTRI3_ini_cn, simplify=TRUE, IRTpars=TRUE)
228 coef(modTRI3_ini_ma, simplify=TRUE, IRTpars=TRUE)
229 coef(modTRI3_ini_li, simplify=TRUE, IRTpars=TRUE)
230 coef(modTRI3_ini_in, simplify=TRUE, IRTpars=TRUE)
231 coef(modTRI3_ini_es, simplify=TRUE, IRTpars=TRUE)
232
233 #Gráficos
234 #todos separados
235 plot(modTRI3_ini_cn,
236       type = 'trace',
237       main="Curva característica dos itens de Ciências da Natureza",
238       color = "red",
239       par.settings = list(strip.background = list(col = c("#9ECAE1")), strip.border = list(col
240 = "black")),
241 )
242 #theta p(theta)
243 #por item
244 itemplot(modTRI3_ini_ch, item = 16, main="Item CH_16")
245 itemplot(modTRI3_ini_cn, item = 7, main="Item CN_7")
246 itemplot(modTRI3_ini_cn, item = 20, main="Item CN_20")
247 itemplot(modTRI3_ini_cn, item = 26, main="Item CN_27")
248 itemplot(modTRI3_ini_ma, item = 13, main="Item MA_13")
249 itemplot(modTRI3_ini_es, item = 4, main="Item ES_4")
250
251
252 #Remoção dos itens muito fáceis ou muito difíceis (com dificuldade fora do intervalo [-4, 4])
253 #Ciências Humanas
254 itens_normat_ch <- itens_ch
255 itens_normat_ch$CH_17 <- NULL
256 #Ciências da Natureza
257 itens_normat_cn <- itens_cn
258 itens_normat_cn$CN_7 <- NULL
259 itens_normat_cn$CN_20 <- NULL
260 itens_normat_cn$CN_27 <- NULL
261 #Matemática
262 itens_normat_ma <- itens_ma
263 itens_normat_ma$MA_13 <- NULL
264 #Linguagens
265 itens_normat_li <- itens_li

```

```

266 #Inglês
267 itens_normat_in <- itens_in
268 #Espanhol
269 itens_normat_es <- itens_es
270 itens_normat_es$LE_4 <- NULL
271
272 #Calcula modelos novamente com itens normatizados
273 modTRI3_ch <- mirt(itens_normat_ch, 1, itemtype='3PL');
274 modTRI3_cn <- mirt(itens_normat_cn, 1, itemtype='3PL');
275 modTRI3_ma <- mirt(itens_normat_ma, 1, itemtype='3PL');
276 modTRI3_li <- mirt(itens_normat_li, 1, itemtype='3PL');
277 modTRI3_in <- mirt(itens_normat_in, 1, itemtype='3PL');
278 modTRI3_es <- mirt(itens_normat_es, 1, itemtype='3PL');
279
280
281 #Parâmetros dos itens (parametrização intercept e slope)
282 coef(modTRI3_ch, simplify=TRUE)
283 coef(modTRI3_cn, simplify=TRUE)
284 coef(modTRI3_ma, simplify=TRUE)
285 coef(modTRI3_li, simplify=TRUE)
286 coef(modTRI3_in, simplify=TRUE)
287 coef(modTRI3_es, simplify=TRUE)
288
289 #Parâmetros dos itens (parametrização a e b TRI)
290 coef(modTRI3_ch, simplify=TRUE, IRTPars=TRUE)
291 coef(modTRI3_cn, simplify=TRUE, IRTPars=TRUE)
292 coef(modTRI3_ma, simplify=TRUE, IRTPars=TRUE)
293 coef(modTRI3_li, simplify=TRUE, IRTPars=TRUE)
294 coef(modTRI3_in, simplify=TRUE, IRTPars=TRUE)
295 coef(modTRI3_es, simplify=TRUE, IRTPars=TRUE)
296
297 #Estima habilidades - método "EAP"for the expected a-posteriori (default)
298 theta_ch = fscores(modTRI3_ch, method = "EAP", full.scores.SE = TRUE)
299 theta_cn = fscores(modTRI3_cn, method = "EAP", full.scores.SE = TRUE)
300 theta_ma = fscores(modTRI3_ma, method = "EAP", full.scores.SE = TRUE)
301 theta_li = fscores(modTRI3_li, method = "EAP", full.scores.SE = TRUE)
302 theta_in = fscores(modTRI3_in, method = "EAP", full.scores.SE = TRUE)
303 theta_es = fscores(modTRI3_es, method = "EAP", full.scores.SE = TRUE)
304
305 #Estima habilidades - método "ML"for maximum likelihood
306 theta_ch_ml = fscores(modTRI3_ch, method = "ML", full.scores.SE = TRUE)
307
308 teste <- as.data.frame(theta_ch)
309
310 str(theta_ch)
311 str(theta_cn)
312 str(theta_ma)
313 str(theta_li)
314 str(theta_in)
315 str(theta_es)
316
317 #Gráficos
318 #theta_p(theta)
319 #separados
320 plot.new()
321 par(mfrow = c(2,3), oma = c(0,0,0,0)) #apenas para conferência. deve ser gerado
    individualmente
322 plot(modTRI3_ch, type = 'trace')
323 plot(modTRI3_cn, type = 'trace')
324 plot(modTRI3_ma, type = 'trace')

```

```

325 plot(modTRI3_li, type = 'trace')
326 plot(modTRI3_in, type = 'trace')
327 plot(modTRI3_es, type = 'trace')
328
329 #juntos
330 plot.new()
331 par(mfrow = c(2,3), oma = c(1,1,1,1)) #apenas para conferência. deve ser gerado
      individualmente
332 plot(modTRI3_ch, type = 'trace', facet_items=FALSE)
333 plot(modTRI3_cn, type = 'trace', facet_items=FALSE)
334 plot(modTRI3_ma, type = 'trace', facet_items=FALSE)
335 plot(modTRI3_li, type = 'trace', facet_items=FALSE)
336 plot(modTRI3_in, type = 'trace', facet_items=FALSE)
337 plot(modTRI3_es, type = 'trace', facet_items=FALSE)
338
339 #por item
340 itemplot(modTRI3_cn, type = "infotrace", item = 26)
341
342 #Qual a relação entre escore total e theta
343 #Primeiramente calculamos o escore total.
344 #Abaixo usei a função apply para calcular os escores
345 tot_ch <- apply(itens_normat_ch, MARGIN = 1, FUN = sum, na.rm=TRUE)
346 tot_cn <- apply(itens_normat_cn, MARGIN = 1, FUN = sum, na.rm=TRUE)
347 tot_ma <- apply(itens_normat_ma, MARGIN = 1, FUN = sum, na.rm=TRUE)
348 tot_li <- apply(itens_normat_li, MARGIN = 1, FUN = sum, na.rm=TRUE)
349 tot_in <- apply(itens_normat_in, MARGIN = 1, FUN = sum, na.rm=TRUE)
350 tot_es <- apply(itens_normat_es, MARGIN = 1, FUN = sum, na.rm=TRUE)
351
352 #cria pequena tabela com escore total e theta (F1 e SE_F1)
353 scrs_ch <- as.data.frame(cbind(tot = tot_ch, eap = theta_ch))
354 scrs_cn <- as.data.frame(cbind(tot = tot_cn, eap = theta_cn))
355 scrs_ma <- as.data.frame(cbind(tot = tot_ma, eap = theta_ma))
356 scrs_li <- as.data.frame(cbind(tot = tot_li, eap = theta_li))
357 scrs_in <- as.data.frame(cbind(tot = tot_in, eap = theta_in))
358 scrs_es <- as.data.frame(cbind(tot = tot_es, eap = theta_es))
359
360 dados_ch$tot <- NULL; dados_ch$F1 <- NULL; dados_ch$SE_F1 <- NULL;
361 dados_cn$tot <- NULL; dados_cn$F1 <- NULL; dados_cn$SE_F1 <- NULL;
362 dados_ma$tot <- NULL; dados_ma$F1 <- NULL; dados_ma$SE_F1 <- NULL;
363 dados_li$tot <- NULL; dados_li$F1 <- NULL; dados_li$SE_F1 <- NULL;
364 dados_in$tot <- NULL; dados_in$F1 <- NULL; dados_in$SE_F1 <- NULL;
365 dados_es$tot <- NULL; dados_es$F1 <- NULL; dados_es$SE_F1 <- NULL;
366
367 #adiciona colunas ao final do dataset dados e altera nome das últimas colunas
368 dados_ch <- cbind(dados_ch, scrs_ch)
369 dados_cn <- cbind(dados_cn, scrs_cn)
370 dados_ma <- cbind(dados_ma, scrs_ma)
371 dados_li <- cbind(dados_li, scrs_li)
372 dados_in <- cbind(dados_in, scrs_in)
373 dados_es <- cbind(dados_es, scrs_es)
374
375 #Cria dataframe com os parâmetros dos itens
376 item_par_ch <- as.data.frame(coef(modTRI3_ch, IRTpars=TRUE, simplify=TRUE))
377 item_par_cn <- as.data.frame(coef(modTRI3_cn, IRTpars=TRUE, simplify=TRUE))
378 item_par_ma <- as.data.frame(coef(modTRI3_ma, IRTpars=TRUE, simplify=TRUE))
379 item_par_li <- as.data.frame(coef(modTRI3_li, IRTpars=TRUE, simplify=TRUE))
380 item_par_in <- as.data.frame(coef(modTRI3_in, IRTpars=TRUE, simplify=TRUE))
381 item_par_es <- as.data.frame(coef(modTRI3_es, IRTpars=TRUE, simplify=TRUE))
382
383 describe(item_par_ch)

```

```

384 describe(item_par_cn)
385 describe(item_par_ma)
386 describe(item_par_li)
387 describe(item_par_in)
388 describe(item_par_es)
389
390 #Exportando tabela de parâmetros (a, b, c), para LaTeX
391 library(xtable)
392 #separados
393 d <- describe(item_par_ch)
394 rownames(d) <- c("a", "b", "c", "u", "means", "f1")
395 xtable(subset(d, d$var<=3 , select=c(2, 3, 4, 5, 8, 9)), type = "latex", digits = c
      (0,0,4,4,4,4,4,4), align = c("lrrrrrr"),
396       caption = c("Parâmetros da prova de Ciências Humanas", "table"), label = "tbl:Parâ
      metros da prova de Ciências Humanas", )
397
398 d <- describe(item_par_cn)
399 rownames(d) <- c("a", "b", "c", "u", "means", "f1")
400 xtable(subset(d, d$var<=3 , select=c(2, 3, 4, 5, 8, 9)), type = "latex", digits = c
      (0,0,4,4,4,4,4,4), align = c("lrrrrrr"),
401       caption = c("Parâmetros da prova de Ciências da Natureza", "table"), label = "tbl:Parâ
      metros da prova de Ciências da Natureza", )
402
403 d <- describe(item_par_ma)
404 rownames(d) <- c("a", "b", "c", "u", "means", "f1")
405 xtable(subset(d, d$var<=3 , select=c(2, 3, 4, 5, 8, 9)), type = "latex", digits = c
      (0,0,4,4,4,4,4,4), align = c("lrrrrrr"),
406       caption = c("Parâmetros da prova de Matemática", "table"), label = "tbl:Parâmetros da
      prova de Matemática", )
407
408 d <- describe(item_par_li)
409 rownames(d) <- c("a", "b", "c", "u", "means", "f1")
410 xtable(subset(d, d$var<=3 , select=c(2, 3, 4, 5, 8, 9)), type = "latex", digits = c
      (0,0,4,4,4,4,4,4), align = c("lrrrrrr"),
411       caption = c("Parâmetros da prova de Linguagens", "table"), label = "tbl:Parâmetros da
      prova de Linguagens", )
412
413 d <- describe(item_par_in)
414 rownames(d) <- c("a", "b", "c", "u", "means", "f1")
415 xtable(subset(d, d$var<=3 , select=c(2, 3, 4, 5, 8, 9)), type = "latex", digits = c
      (0,0,4,4,4,4,4,4), align = c("lrrrrrr"),
416       caption = c("Parâmetros da prova de Inglês", "table"), label = "tbl:Parâmetros da prova
      de Inglês", )
417
418 d <- describe(item_par_es)
419 rownames(d) <- c("a", "b", "c", "u", "means", "f1")
420 xtable(subset(d, d$var<=3 , select=c(2, 3, 4, 5, 8, 9)), type = "latex", digits = c
      (0,0,4,4,4,4,4,4), align = c("lrrrrrr"),
421       caption = c("Parâmetros da prova de Espanhol", "table"), label = "tbl:Parâmetros da
      prova de Espanhol", )
422
423 rm(d)
424
425 #separados
426 Parametro <- c("a", "b", "c")
427 Prova <- c(rep("Ciências Humanas", 1, 3))
428 d <- describe(item_par_ch)
429 d <- subset(d, d$var<=3)
430 d <- cbind(Prova, Parametro, d)
431 marcadores <- d

```

```

432
433 Prova <- c(rep("Ciências da Natureza", 1, 3))
434 d <- describe(item_par_cn)
435 d <- subset(d, d$var<=3)
436 d <- cbind(Prova, Parametro, d)
437 marcadores <- rbind(marcadores, d)
438
439 Prova <- c(rep("Matemática", 1, 3))
440 d <- describe(item_par_ma)
441 d <- subset(d, d$var<=3)
442 d <- cbind(Prova, Parametro, d)
443 marcadores <- rbind(marcadores, d)
444
445 Prova <- c(rep("Linguagens", 1, 3))
446 d <- describe(item_par_li)
447 d <- subset(d, d$var<=3)
448 d <- cbind(Prova, Parametro, d)
449 marcadores <- rbind(marcadores, d)
450
451 Prova <- c(rep("Inglês", 1, 3))
452 d <- describe(item_par_in)
453 d <- subset(d, d$var<=3)
454 d <- cbind(Prova, Parametro, d)
455 marcadores <- rbind(marcadores, d)
456
457 Prova <- c(rep("Espanhol", 1, 3))
458 d <- describe(item_par_es)
459 d <- subset(d, d$var<=3)
460 d <- cbind(Prova, Parametro, d)
461 marcadores <- rbind(marcadores, d)
462
463 tabela_latex <- xtable(subset(marcadores, select=c(1, 2, 4, 5, 6, 7, 10, 11)), type = "latex",
464                       digits = c(0,0,0,0,2,2,2,2,2,2), align = c("lllrrrrrrr"),
465                       caption = c("Parâmetros de todas as provas", "table"), label = "tbl:Parâ
466                               metros de todas as provas", )
467
468 print(tabela_latex, include.rownames=FALSE)
469
470 #Exportando parâmetros da prova para LaTeX
471 library(xtable)
472 d <- describe(tot_ch)
473 rownames(d) <- c("Ciências Humanas")
474 d["itens"] <- dim(itens_normat_ch)[2]
475 marcadores <- d
476
477 d <- describe(tot_cn)
478 rownames(d) <- c("Ciências da Natureza")
479 d["itens"] <- dim(itens_normat_cn)[2]
480 marcadores <- rbind(marcadores, d)
481
482 d <- describe(tot_ma)
483 rownames(d) <- c("Matemática")
484 d["itens"] <- dim(itens_normat_ma)[2]
485 marcadores <- rbind(marcadores, d)
486
487 d <- describe(tot_li)
488 rownames(d) <- c("Linguagens")
489 d["itens"] <- dim(itens_normat_li)[2]

```

```

490 marcadores <- rbind(marcadores, d)
491
492 d <- describe(tot_in)
493 rownames(d) <- c("Inglês")
494 d["itens"] <- dim(itens_normat_in)[2]
495 marcadores <- rbind(marcadores, d)
496
497 d <- describe(tot_es)
498 rownames(d) <- c("Espanhol")
499 d["itens"] <- dim(itens_normat_es)[2]
500 marcadores <- rbind(marcadores, d)
501
502 xtable(subset(marcadores, select=c(2, 3, 4, 5, 8, 9, 14)), type = "latex", digits = c
      (0,0,2,2,0,0,0,0), align = c("lrrrrrrr"),
503       caption = c("Marcadores dos escores das provas", "table"), label = "tbl:Marcadores dos
      escores das provas", )
504
505
506 #Escolher cada uma das 6 linhas a seguir para rodar os gráficos de theta, b e theta/b de cada
      uma das disciplinas
507 plot_theta <- theta_ch; plot_item_par <- item_par_ch; nome_prova <- "Ciências Humanas";
508 plot_theta <- theta_cn; plot_item_par <- item_par_cn; nome_prova <- "Ciências da Natureza";
509 plot_theta <- theta_ma; plot_item_par <- item_par_ma; nome_prova <- "Matemática";
510 plot_theta <- theta_li; plot_item_par <- item_par_li; nome_prova <- "Linguagens";
511 plot_theta <- theta_in; plot_item_par <- item_par_in; nome_prova <- "Inglês";
512 plot_theta <- theta_es; plot_item_par <- item_par_es; nome_prova <- "Espanhol";
513
514 #INÍCIO - Gráficos de theta, b e theta/b de cada uma das disciplinas
515 plot.new()
516 par(mfrow = c(1,3), oma = c(0,0,1,0))
517 #Histograma de Theta
518 histogram_theta <- hist(plot_theta[,1], xlim = c(-4,4), main="Traço latente", col="
      gray60", freq = TRUE, ylab="N", xlab=expression(theta), breaks = seq(-4, 4, 0.4))
519 #Histograma de b
520 histogram_b <- hist(plot_item_par$items.b, xlim = c(-4,4), main="Dificuldade", col="
      gray80", freq = TRUE, ylab="N", xlab="b", breaks = seq(-10, 10, 0.4)) #
      nclass =
      20
521 #Linhas de theta e b
522 plot(
523   x=histogram_theta$mids,
524   y = histogram_theta$density,
525   main="Curvas de traço latente \ne dificuldade da prova",
526   xlab=paste(expression(theta), "b"),
527   ylab="",
528   xlim = c(-4,4), ylim = c(0,1),
529   col="gray60",
530   type="l",
531   lwd=2,
532 )
533
534 lines(
535   x=histogram_b$mids,
536   y = histogram_b$density,
537   col="gray80",
538   lwd=2,
539   lty=1
540 )
541
542 lines(

```

```

543 legend("topright", legend=c(expression(theta), "b"), lty=c(1, 1), col=c("gray60", "gray80"),
      lwd=2, bty="n", y = 50)
544 )
545 mtext(paste("Análise de theta e b para a prova de ", nome_prova), side = 3, line = -1, outer =
      TRUE)
546
547 #FIM - Gráficos de theta, b e theta/b de cada uma das disciplinas
548
549 describe(tot_ch)
550
551 #DIF
552
553 #Selecionando campos
554 dadosDIF_ch <- dados_ch[c("COTA", campos_itens_ch)]
555 dadosDIF_cn <- dados_cn[c("COTA", campos_itens_cn)]
556 dadosDIF_ma <- dados_ma[c("COTA", campos_itens_ma)]
557 dadosDIF_li <- dados_li[c("COTA", campos_itens_li)]
558 dadosDIF_in <- dados_in[c("COTA", campos_itens_in)]
559 dadosDIF_es <- dados_es[c("COTA", campos_itens_es)]
560
561 #Removendo itens não utilizados (definido na análise TRI)
562 dadosDIF_ch$CH_17 <- NULL
563 dadosDIF_cn$CN_7 <- NULL
564 dadosDIF_cn$CN_20 <- NULL
565 dadosDIF_cn$CN_27 <- NULL
566 dadosDIF_ma$MA_13 <- NULL
567 dadosDIF_es$LE_4 <- NULL
568
569 #Calculando modelos
570 modDIF_MH_ch <- difMH(Data=dadosDIF_ch, group="COTA", focal.name="S", purify=TRUE, nrIter=20)
571 modDIF_MH_cn <- difMH(Data=dadosDIF_cn, group="COTA", focal.name="S", purify=TRUE, nrIter=20)
572 modDIF_MH_ma <- difMH(Data=dadosDIF_ma, group="COTA", focal.name="S", purify=TRUE, nrIter=20)
573 modDIF_MH_li <- difMH(Data=dadosDIF_li, group="COTA", focal.name="S", purify=TRUE, nrIter=20)
574 modDIF_MH_in <- difMH(Data=dadosDIF_in, group="COTA", focal.name="S", purify=TRUE, nrIter=20)
575 modDIF_MH_es <- difMH(Data=dadosDIF_es, group="COTA", focal.name="S", purify=TRUE, nrIter=20)
576
577 #Criando datafrme com estatística e p-valor
578 modDif_ch_df <- cbind(as.data.frame(modDIF_MH_ch$names), modDIF_MH_ch$MH, modDIF_MH_ch$p.value
      )
579 colnames(modDif_ch_df) <- c("item", "MH", "p.value")
580 modDif_ch_df <- mutate(modDif_ch_df, Sig.cod = if_else(p.value < 0.05, "*", ""))
581
582 modDif_cn_df <- cbind(as.data.frame(modDIF_MH_cn$names), modDIF_MH_cn$MH, modDIF_MH_cn$p.value
      )
583 colnames(modDif_cn_df) <- c("item", "MH", "p.value")
584 modDif_cn_df <- mutate(modDif_cn_df, Sig.cod = if_else(p.value < 0.05, "*", ""))
585
586 modDif_ma_df <- cbind(as.data.frame(modDIF_MH_ma$names), modDIF_MH_ma$MH, modDIF_MH_ma$p.value
      )
587 colnames(modDif_ma_df) <- c("item", "MH", "p.value")
588 modDif_ma_df <- mutate(modDif_ma_df, Sig.cod = if_else(p.value < 0.05, "*", ""))
589
590 modDif_li_df <- cbind(as.data.frame(modDIF_MH_li$names), modDIF_MH_li$MH, modDIF_MH_li$p.value
      )
591 colnames(modDif_li_df) <- c("item", "MH", "p.value")
592 modDif_li_df <- mutate(modDif_li_df, Sig.cod = if_else(p.value < 0.05, "*", ""))
593
594 modDif_in_df <- cbind(as.data.frame(modDIF_MH_in$names), modDIF_MH_in$MH, modDIF_MH_in$p.value
      )
595 colnames(modDif_in_df) <- c("item", "MH", "p.value")

```

```

596 modDif_in_df <- mutate(modDif_in_df, Sig.cod = if_else(p.value < 0.05, "*", ""))
597
598 modDif_es_df <- cbind(as.data.frame(modDIF_MH_es$names), modDIF_MH_es$MH, modDIF_MH_es$p.value
)
599 colnames(modDif_es_df) <- c("item", "MH", "p.value")
600 modDif_es_df <- mutate(modDif_es_df, Sig.cod = if_else(p.value < 0.05, "*", ""))
601
602
603 #Exportando tabelas
604 library(xtable)
605 tabela_latex <- xtable(modDif_ch_df, select=c(1, 2, 3, 4), type = "latex", digits = c
(0,0,4,4,0), align = c("llrll"),
606 caption = "Estatística DIF da prova de Ciências Humanas com o método
Mantel-Haenszel", label = "tbl:Estatística DIF Ciências Humanas", )
607 print(tabela_latex, include.rownames=FALSE)
608
609 tabela_latex <- xtable(modDif_cn_df, select=c(1, 2, 3, 4), type = "latex", digits = c
(0,0,4,4,0), align = c("llrll"),
610 caption = "Estatística DIF da prova de Ciências da Natureza com o mé
todo Mantel-Haenszel", label = "tbl:Estatística DIF Ciências da Natureza", )
611 print(tabela_latex, include.rownames=FALSE)
612
613 tabela_latex <- xtable(modDif_ma_df, select=c(1, 2, 3, 4), type = "latex", digits = c
(0,0,4,4,0), align = c("llrll"),
614 caption = "Estatística DIF da prova de Matemática com o método Mantel-
Haenszel", label = "tbl:Estatística DIF Matemática", )
615 print(tabela_latex, include.rownames=FALSE)
616
617 tabela_latex <- xtable(modDif_li_df, select=c(1, 2, 3, 4), type = "latex", digits = c
(0,0,4,4,0), align = c("llrll"),
618 caption = "Estatística DIF da prova de Linguagens com o método Mantel-
Haenszel", label = "tbl:Estatística DIF Linguagens", )
619 print(tabela_latex, include.rownames=FALSE)
620
621 tabela_latex <- xtable(modDif_in_df, select=c(1, 2, 3, 4), type = "latex", digits = c
(0,0,4,4,0), align = c("llrll"),
622 caption = "Estatística DIF da prova de Inglês com o método Mantel-
Haenszel", label = "tbl:Estatística DIF Inglês", )
623 print(tabela_latex, include.rownames=FALSE)
624
625 tabela_latex <- xtable(modDif_es_df, select=c(1, 2, 3, 4), type = "latex", digits = c
(0,0,4,4,0), align = c("llrll"),
626 caption = "Estatística DIF da prova de Espanhol com o método Mantel-
Haenszel", label = "tbl:Estatística DIF Espanhol", )
627 print(tabela_latex, include.rownames=FALSE)
628
629 #Gráficos de DIF
630 plot.new()
631 par(mfrow = c(3,2), oma = c(0,0,0,0))
632 plot(modDIF_MH_ch)
633 plot(modDIF_MH_cn)
634 plot(modDIF_MH_ma)
635 plot(modDIF_MH_li)
636 plot(modDIF_MH_in)
637 plot(modDIF_MH_es)

```