



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Um Novo Método para Avaliação da Coerência de Respostas Individuais em Provas de Múltipla Escolha

por

Thays Suelen Brito Santos

Brasília, 27 de outubro de 2021

Um Novo Método para Avaliação da Coerência de Respostas Individuais em Provas de Múltipla Escolha

por

Thays Suelen Brito Santos

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Antônio Eduardo Gomes

Brasília, 27 de outubro de 2021

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Prof. Dr. Antônio Eduardo Gomes
Orientador, EST/UnB

Prof. Dr. Heliton Ribeiro Tavares
EST/UFPA

Prof. Dr. Raul Yukihiro Matsushita
EST/UnB

Prof. Dr. André Luiz Fernandes Caçado
Suplente, EST/UnB

No meio de qualquer dificuldade encontra-se a oportunidade.

(Albert Einstein)

Para minha família amada.

Agradecimentos

A Deus, por me conceder saúde e sabedoria para seguir sempre em frente. Obrigada por ser a minha força e o meu guia em todos os momentos.

Aos meus pais, Maria e Roselito, pelo apoio e incentivo em todos os momentos da minha vida. Por acreditarem em mim, e não medirem esforços para a concretização dos meus sonhos. Sem vocês, nada seria possível.

E claro, a minha querida filha, Ana Luísa, que amo incondicionalmente e que veio dar um novo colorido à minha vida. Foi ela o meu grande estímulo nesta caminhada.

À meu irmão, Thales Santos, pelos conselhos preciosos e encorajamento naqueles momentos cruciais desta difícil jornada.

Aos meus amigos, Adolfo, Alan, Débora e Sheyla, anjos que Deus colocou em meu caminho. Mesmo com a distância, sempre se fizeram presentes na minha vida e estarão sempre em meu coração. Obrigada pelo companheirismo, apoio e amizade incondicional.

À todos que apoiaram minha decisão de concluir o mestrado.

Agradeço a todos professores do PPGEST e, particularmente, agradeço a meu orientador,

professor Antônio Eduardo, pela paciência, conselhos e auxílio neste trabalho.

Agradeço, também, ao meu orientador da graduação, professor Edson Ramos pelo incentivo na conquista de mais essa etapa.

Finalmente, gostaria de agradecer ao PPGEST e à CAPES, sem os quais essa dissertação dificilmente poderia ter sido realizada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

O trabalho propõe o uso das estatísticas de ajuste individual como diretriz para validação das pontuações em testes educacionais. Por isso, neste estudo, serão utilizadas 10 estatísticas não paramétricas de ajuste individual ajustadas com base na Teoria de Resposta ao Item para modelos ML1P, ML2P e ML3P e, uma nova estatística baseada na regressão isotônica, denotada por R_n , será proposta e comparada com as 10 estatísticas mencionadas. Foram gerados 60 conjuntos de dados a partir da simulação de dados de 3 (três) tipos de teste (17, 33 e 65 itens), 4 (quatro) porcentagens de respondentes aberrantes (5%, 10%, 25% e 50%) e 5 (cinco) tipos de respondentes (fraudadores, descuidados, criativos, adivinhos e aleatórios). O desempenho da estatística de ajuste individual será baseado nas áreas sob a curva ROC e as vantagens e desvantagens serão discutidas. Os resultados mostram que a sensibilidade das estatísticas de ajuste individual depende da quantidade de itens do teste, da porcentagem de respondentes aberrantes e do tipo de respondente. Além disso, a estatística R_n apresentará maior eficácia para respondentes fraudadores e criativos. Por fim, há uma aplicação em dados reais com a finalidade de confirmar o desempenho das estatísticas na identificação de indivíduos aberrantes em testes educacionais do SARESP (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo). Através de um teste de hipótese cuja a hipótese nula H_0 diz que as respostas dos examinados não são aberrantes, confirma-se o desempenho das estatísticas na identificação de indivíduos aberrantes nos testes educacionais. A decisão do teste de hipótese foi por meio da comparação do p -valor com um nível descritivo especificado.

Palavras-chave: ajuste individual; estatísticas não paramétricas; regressão isotônica.

Abstract

The work proposed the use of person-fit statistics as a guideline for validating educational test scores. Therefore, in this study, 10 non-parametric person-fit statistics will be used adjusted based on Item Response Theory for the ML1P, ML2P and ML3P models, and a new statistic based on isotonic regression, denoted by R_n , will be proposed and compared to the 10 statistics mentioned. Sixty data sets were generated from data simulation of 3 (three) test types (17, 33 and 65 items), 4 (four) percentages of aberrant respondents (5 %, 10 %, 25 % and 50 %) and 5 (five) types of respondents (cheaters, careless, creative, Lucky guessing and random). The performance of the person-fit statistics will be based on areas under the ROC and the advantages and disadvantages will be discussed. The results show that the sensitivity of person-fit statistics depends on the number of items in the test, the percentage of aberrant respondents, and the type of respondent. Also, a R_n stat will be more effective for cheaters and creative responders. Finally, there is an application in real data as a good way to confirm the performance of statistics in identifying aberrant information in educational tests of SARESP (School Performance Assessment System of the State of São Paulo). Through a hypothesis test whose null hypothesis H_0 says that the answers of the exams are not aberrant, the performance of the statistics in the identification of aberrant individuals in the educational tests is confirmed. The hypothesis test decision was made by comparing the p -value with a specified descriptive level.

Key-words: Item Response Theory; person-fit; non parametric statistics; isotonic regression.

Sumário

1	Introdução	1
1.1	Considerações iniciais	1
1.2	Objetivos	2
1.2.1	Objetivo Geral	2
1.2.2	Objetivos Específicos	2
1.3	Dados	2
1.4	Esboço do trabalho	3
1.5	Aspectos gerais	4
2	Revisão Bibliográfica	6
2.1	Teoria da Resposta ao Item (TRI)	6
2.2	Métodos de detecção de Coerência	8
2.2.1	Estatística G	9
2.2.2	Estatística G^*	10
2.2.3	Estatística r_{pbis}	10
2.2.4	Índice de Cautela (CI)	11
2.2.5	Índice de Cautela Modificado (MCI)	11
2.2.6	Estatística $U3$ e $ZU3$	12
2.2.7	Estatística H^T	13
2.2.8	Índice de Concordância A	14

3	Estatística R_n e Estudo de Simulação	15
3.1	Regressão isotônica	15
3.2	Uma nova estatística baseada na regressão isotônica R_n	21
3.3	Avaliando a Eficiência das estatísticas de ajuste individual	22
3.4	Simulações	23
3.5	Resultados	25
3.5.1	Simulação para modelos com 1 (um) parâmetro	25
3.5.2	Simulação para modelos com 2 (dois) parâmetros	42
3.5.3	Simulação para modelos com 3 (três) parâmetros	59
3.6	Discussão	76
4	Aplicações a Dados Reais	78
5	Conclusões	84
5.1	Considerações Finais	84
5.2	Trabalhos Futuros	85
	Referências Bibliográficas	85
A	Código da Programação em R	90
A.1	Programação para Dados Simulados	90
A.2	Programação para Dados Reais	108

Abreviações e Siglas

C	<i>Índice de Cautela</i>
DSA	<i>Diagrama da Soma Acumulada</i>
FRI	<i>Função de Resposta do Item</i>
MCI	<i>Índice de Cautela Modificado</i>
MCM	<i>Minorante Convexo Máximo</i>
ML1P	<i>Modelo Logístico de 1 (um) Parâmetro</i>
ML2P	<i>Modelo Logístico de 2 (dois) Parâmetros</i>
ML3P	<i>Modelo Logístico de 3 (três) Parâmetros</i>
PAV	<i>Algoritmo Pool-Adjacent-Violators</i>
ROC	<i>Receiver Operating Characteristic</i>
TRI	<i>Teoria da Resposta ao Item</i>

Capítulo 1

Introdução

1.1 Considerações iniciais

A avaliação de desempenho é algo comum. Algumas variáveis de interesse não podem ser medidas, diretamente, tal como o comportamento de um examinado ao fazer um teste, de modo que a pontuação no teste pode ou não refletir suas verdadeiras habilidades. Os testes são para identificar a coerência de respostas individuais em provas de múltipla escolha como por exemplo, um indivíduo com alta habilidade erra itens fáceis ou um examinado com baixa habilidade responder corretamente os itens difíceis e um candidato que responde aleatoriamente ao longo do teste.

A discrepância entre a verdadeira capacidade do examinando e o resultado do teste pode ser medido em uma grande variedade de maneiras. O interesse por análise individual surgiu com as pesquisas (Spearman, 1910; Thurstone, 1927; Guttman, 1944) e se intensificou com o uso da Teoria de Resposta ao Item (TRI) como método de psicometria (Rasch, 1960). Além disso, o estudo de Meijer e Sijtsma (2001) e Karabatsos (2003) mostrou que existem mais de quarenta estatísticas disponíveis para testar o ajuste individual.

1.2 Objetivos

1.2.1 Objetivo Geral

Discutir o uso de estatísticas não paramétricas ajustadas à pessoa para verificação da capacidade de detecção de examinados com padrões aberrantes de respostas aos itens, a fim de obter um consenso sobre quais são as melhores estatísticas para detectar respostas coerentes do examinado em um teste.

1.2.2 Objetivos Específicos

- Apresentar estatísticas não paramétricas de ajuste individual com base em dados de estudos anteriores, mas apresentando uma análise de desempenho mais detalhada e uma comparação relativa à literatura existente;
- Utilizar a regressão isotônica como um novo método de ajuste individual;
- Realizar simulação e análise ROC (Receiver Operating Characteristic) para avaliação do desempenho das estatísticas.

1.3 Dados

O conjunto de dados do SARESP (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo) corresponde a uma amostra de 3054 respondentes da prova de Língua Portuguesa aplicada, com 30 itens, no ano de 2007, a alunos do terceiro ano dos turnos matutino, vespertino e noturno de escolas públicas de São Paulo.

- **Descrição:** Uma amostra de 3 testes (manhã, tarde e noite) de Língua Portuguesa aplicados para alunos do terceiro ano do ensino médio em 2007.
- **Utilização:** Aplicação a dados reais com Modelo logístico de 3 parâmetros.

- **Disponível em:** <http://www.ufpr.br/~aanjos/TRI/sinape/dados/saresp.dat>, Fonte: Secretaria Estadual da Educação de São Paulo.

Para esse trabalho será utilizado apenas o turno da manhã, que possui 1001 respondentes.

1.4 Esboço do trabalho

Este trabalho encontra-se dividido em 4 capítulos, a saber:

- No Capítulo 1 é feita uma introdução a avaliação de desempenho do examinado ao fazer um teste, e a importância de métodos para identificar a coerência da resposta individuais em prova de múltipla escolha, são abordados as considerações iniciais, o objetivo geral e os específicos, justificativa, e os aspectos gerais.
- No Capítulo 2 são apresentados os métodos utilizados neste trabalho na análise de ajuste individual.
- No Capítulo 3 são apresentados a nova estatística proposta e os resultados da simulação de dados das estimativas das proficiências e dos parâmetros para ML1P, ML2P e ML3P para calcular o desempenho das estatísticas de ajuste individual e comparar com a nova estatística proposta.
- No Capítulo 4 são apresentados resultados da aplicação de dados reais com a finalidade de confirmar o desempenho das estatísticas de ajuste individual, com as estimativas das proficiências e dos parâmetros para ML3P.
- No Capítulo 5 são apresentadas as considerações finais e recomendações para trabalhos futuros.

1.5 Aspectos gerais

A medição precisa do desempenho do examinado é fundamental para tomada de decisão na educação, como a avaliação do desempenho escolar dos alunos no Brasil, além de permitir educadores a adaptar o planejamento curricular e ajudar nas decisões políticas sobre questões educacionais. Por outro lado, medições imprecisas podem levar a consequências negativas e positivas ao examinado, como indivíduo qualificado obter pontuação baixa e ser injustamente excluído de um processo seletivo ou indivíduo não qualificado obter pontuação alta e, consequentemente, a vaga (Karabatsos, 2003).

Deve-se notar que a medição imprecisa pode ser classificada como aleatória ou sistemática (Smith, 1991). Há fatores que podem fazer com que a pontuação de um examinado em um teste seja considerada incoerente: resposta descuidada, adivinhação, resposta criativa, resposta aleatória e fraudadores (Meijer, 1996). A resposta descuidada ocorre quando o examinado, de maneira descuidada, responde a certos itens incorretamente, para os quais ele é capaz de responder corretamente. Adivinhação de sorte acontece quando o examinado adivinha as respostas corretas de alguns itens de teste para os quais ele não sabe a resposta correta. Alguns examinados com alta capacidade envolvem-se em respostas criativas, onde obtêm respostas incorretas para certos itens fáceis, porque eles interpretam esses itens de uma maneira única e criativa. Enquanto distúrbios sistemáticos podem ocorrer, por exemplo, com deficiência em um conteúdo de certa área, tendência de teste ou diferentes instruções (Karabatsos, 2003).

Já os fraudadores referem-se ao comportamento em que o examinando obtém injustamente as respostas corretas nos itens de teste que ele não consegue responder corretamente. Dentre as condutas de fraude de um teste, o uso da cola que consiste em conseguir resposta de três formas: lembretes, ou por intermédio de comunicação eletrônica, que possui característica mais danosa devido ao envolvimento de um grupo de candidato obtendo resposta de uma única fonte. Os métodos não paramétricos de ajuste individual podem comparar as respostas entre os pares de indivíduos com objetivos de detectar esses tipos de incorrências entre os examinados (Wollack

e Maynes, 2017).

As estatísticas de ajuste individual são usadas para identificar padrões de resposta considerados inadequados de serem observados com base no modelo. Os estudos de Trabin e Weiss (1983) simularam a eficiência das estatísticas de ajuste individual com base no modelo logístico de 3 (três) parâmetros. Já Klauer e Rettig (1990) e Dimitrov e Smith (2006) estudaram as estatísticas de ajuste individual no contexto do modelo Rasch e, Meijer (1994) utilizou o modelo logístico de 2 (dois) parâmetros a fim de compará-los.

Para detectar padrões aberrantes, pesquisas anteriores desenvolveram inúmeras estatísticas de ajuste individual. O estudo de Li e Olejnik (1997) compararam a eficiência de cinco estatísticas de ajuste individual, usando dados simulados de 30 a 60 itens dentro da estrutura do modelo de Rasch e, que a identificação de respostas aberrantes aumenta com o comprimento do teste. Dimitrov e Smith (2006) realizaram um estudo com testes curtos (10, 20 e 30 itens) e comparou a estatística t (transformada da raiz cúbica do quadrado médio) e a $ZU3$ (estatística normalizada de $U3$) com a estatística H^T (mais eficiente na literatura), e observou que para teste com 20 a 30 itens, a estatística H^T superou as estatísticas t e $ZU3$, mas, para teste com 10 itens, não apresentou diferença significativa.

Karabatsos (2003) comparou 36 estatísticas de ajuste individual paramétricas e não paramétricas com testes de 17, 33 e 65 itens, e constatou que a estatística H^T (Sijtsma, 1986) é a melhor na detecção de padrões de respostas aberrantes, seguida pelas estatísticas paramétricas $D(\theta)$ (Trabin e Weiss, 1983) e não paramétrica CI (Sato, 1975).

As estatísticas de ajuste individual podem ter vários usos, entre eles, a detecção de outliers (Meijer, 2002; Felt et al., 2017). Assim, tem ocorrido um crescente interesse em identificar valores discrepantes, assim como entender o motivo dessa resposta (Meijer et al., 2008; Meijer e Sijtsma, 2001; Reise, 1990)

Nesta dissertação, serão utilizados métodos não paramétricos que detectam padrões individuais inadequados tratados no trabalho de Karabatsos (2003) e a implementação de uma nova estatística baseada na regressão isotônica.

Capítulo 2

Revisão Bibliográfica

2.1 Teoria da Resposta ao Item (TRI)

A TRI é um conjunto de modelos estatísticos nos quais a probabilidade de resposta a um item é modelada como função da habilidade (proficiência) do aluno (variável não observável) e de parâmetros que expressam certas características dos itens, como a característica de que quanto maior a proficiência do aluno, maior a probabilidade dele acertar o item (Andrade, Tavares e Valle, 2000).

O conjunto dos modelos de TRI depende do tipo de função matemática adotada e dos parâmetros dos itens que se queira investigar. Há dois tipos de função que podem ser encontrados na literatura: logística e normal acumulada. Neste trabalho, será abordado apenas o modelo de função logística, pois permite melhor tratamento matemático e também é mais frequente na literatura especializada (Muniz, 1990).

Dentre os modelos da TRI para itens com resposta dicotômica ou itens de múltipla escolha (corrigido como certo/errado) acumulativos, destacam-se: o Modelo Logístico de 1 parâmetro, o Modelo Logístico de 2 parâmetros e o Modelo Logístico de 3 parâmetros. Todos os modelos citados serão utilizados neste trabalho e são dado por:

ML1P

$$P_{jn} = P(U_{jn} = 1|\theta_n) = \frac{1}{1 + e^{-(\theta_n - b_j)}},$$

ML2P

$$P_{jn} = P(U_{jn} = 1|\theta_n) = \frac{1}{1 + e^{-Da_j(\theta_n - b_j)}}, \quad (2.1)$$

ML3P

$$P_{jn} = P(U_{jn} = 1|\theta_n) = c_j + (1 - c_j) \frac{1}{1 + e^{-Da_j(\theta_n - b_j)}},$$

em que,

$j = 1, 2, \dots, J$ (representa os J itens propostos para avaliar o traço latente considerado) e $n = 1, 2, \dots, N$ (representa os N elementos que compõem a amostra, que podem ser indivíduos, empresas, etc.);

U_{jn} é uma variável dicotômica que assume os valores 1, quando o respondente n responde corretamente, concorda ou satisfaz as condições do item j , ou 0 caso contrário;

θ_n pode representar o traço latente do respondente n ;

$P(U_{jn} = 1|\theta_n)$ é a probabilidade de o respondente n , condicionado no seu traço latente, responder corretamente, ou concordar ou satisfazer as condições do item j e é denominada de Função de Resposta do Item FRI;

b_j é o parâmetro de dificuldade (ou de posição) do item j , medido na mesma escala do traço latente;

a_j é o parâmetro de discriminação (ou de inclinação) do item j , com valor proporcional à inclinação da curva característica do item no ponto b_j . Itens com maiores valores de a_j fornecem melhores discriminações;

c_j é o parâmetro do acerto casual;

D é uma constante de escala igual a 1, mas se utiliza o valor 1, 7, quando se quer que os valores da função logística se aproximem da função ogiva normal.

2.2 Métodos de detecção de Coerência

Não selecionamos todos os índices propostos na literatura, mas utilizaremos índices não paramétricos (Tabela 2.1) com base no estudo de Karabatsos (2003). O foco é comparar os índices que apresentaram um bom desempenho com uma nova estatística de ajuste individual baseada na regressão isotônica.

Uma estatística adaptada ao indivíduo mede o grau de razoabilidade das respostas de um examinado para um conjunto de itens em teste. Para formular esta ideia, matematicamente, denota-se $X_{nj} = 1$ como uma resposta correta feita pelo examinado n no item de teste j , $X_{nj} = 0$ uma resposta incorreta, $\{n = 1, \dots, N\}$ denota uma amostra aleatória de examinados que respondem para o conjunto de itens de teste $\{j = 1, \dots, J\}$ e a pontuação correta do examinado (NC) é dado pela soma das pontuações dos itens $r_n = \sum_{j=1}^J X_{nj}$.

Quase todas as estatísticas de ajuste individual são sensíveis aos erros de Guttman. Seja uma particular escolha de pesos (w) definindo uma estatística de ajuste individual. Suponha que N examinados façam um teste com J itens. Considerando um teste com itens pontuados dicotomicamente e ordenados do mais fácil para o mais difícil, de acordo com a proporção correta do item j (aumentando a dificuldade), espera-se que o examinado acerte os itens fáceis e erre os itens difíceis. Quando um item difícil é respondido corretamente e, um item fácil, incorretamente, considera-se um erro de Guttman.

Estatísticas	Literatura Referência	Tipo de Item
G	(Guttman, 1944, 1950)	Dicotômico
G^*	(van der Flier, 1977)	Dicotômico
r_{pbis}	(Donlon e Fischer, 1968)	Dicotômico
C	(Sato, 1975)	Dicotômico
MCI	(Harnisch e Linn, 1981)	Dicotômico
$U3$	(van der Flier, 1980)	Dicotômico
$ZU3$	(van der Flier, 1982)	Dicotômico
H^T	(Sijtsma, 1986)	Dicotômico
A	(Kane e Brennan, 1980)	Dicotômico
R_n	(proposta neste trabalho)	Dicotômico

Tabela 2.1: As 10 (dez) estatísticas de ajuste individual

2.2.1 Estatística G

Assim, a estatística de Guttman é um teste baseado na habilidade do indivíduo, onde todos os itens apresentam sempre a mesma resposta quando um examinado tem um determinado nível de habilidade. Desse modo, a probabilidade assume valores 0 ou 1 e, a curva característica do item de Guttman é uma função degrau (Figura 2.1).

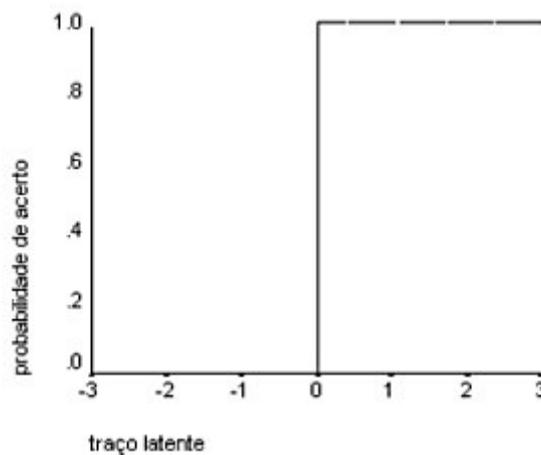


Figura 2.1: Curva característica de um item de Guttman.

Sejam $X_{nj} \in \{0, 1\}$ a resposta do respondente n para o j -ésimo item e $r_n = \sum_{j=1}^J X_{nj}$ o total de respostas corretas do indivíduo n . Na situação ideal, este indivíduo deve fornecer as respostas corretas para os r_n itens mais fáceis e as respostas incorretas para $J - r_n$ itens mais difíceis. De modo geral, quando ordenar os itens pelo número de acertos, de forma que, comece pelos itens mais fáceis para os itens mais difíceis, espera-se que o padrão de resposta do examinado acerte as r_n questões, ou seja, 1 's para os primeiros r_n itens e erre os últimos, ou seja, 0 's nas últimas $J - r_n$ itens. A equação geral é dada por:

$$G = \sum_{h=1}^{J-1} \sum_{e=h+1}^J X_{nh}(1 - X_{ne}) \quad (2.2)$$

onde, X_{nh} é a resposta do indivíduo n para o item mais difícil no par de itens $[h, e]$ e X_{ne} é a

resposta do indivíduo n para o item mais fácil no par de itens $[h, e]$.

Quanto à estatística $G = 0$, o padrão de resposta do examinado está de acordo com o padrão perfeito de Guttman. Quanto maior for o valor da estatística, maior será o desajuste do padrão de resposta do examinado.

2.2.2 Estatística G^*

van der Flier (1977) definiu sua estatística G^* como o número de erros de Guttman padronizados (G) em relação ao número máximo de erros de Guttman, onde este máximo é igual a $r_n(J - r_n)$, ou seja, normalizada G para o intervalo $[0, 1]$.

$$G^* = \frac{\sum_{h=1}^{J-1} \sum_{e=h+1}^J X_{nh}(1 - X_{ne})}{r_n(J - r_n)} \quad (2.3)$$

Quando o valor da estatística G^* é nulo, o padrão de resposta do examinado está de acordo com o padrão perfeito de Guttman e, quando G^* assume o valor 1, o padrão de resposta do examinado está de acordo com o padrão inverso de Guttman.

2.2.3 Estatística r_{pbis}

Donlon e Fischer (1968) consideraram que a correlação pessoal ponto-bisserial (r_{pbis}) é a correlação entre o padrão de pontuação do item do candidato X_n e o vetor da proporção da pontuação correta do item na amostra $p = (p_1, \dots, p_j, \dots, p_J)$, onde esse vetor inclui a pontuação do examinado. Na situação ideal, espera-se que o candidato considerado erre os itens com pouco acerto e acerte os itens com muitos acertos apresentando valor de r_{pbis} mais próximo de 1. Se o valor da estatística r_{pbis} estiver próximo de -1 , o padrão de resposta do examinado está de acordo com o padrão inverso de Guttman e apresenta inconsistência. A estatística r_{pbis} é dada por:

$$r_{pbis} = Corr(X_n, p), \quad (2.4)$$

sendo X_n o vetor de respostas do n -ésimo indivíduo e p o vetor de proporções de respostas positivas para cada item.

2.2.4 Índice de Cautela (CI)

O Índice de Cautela (CI) é definido como o complemento da razão de dois termos de covariância. Suponha $p = (p_1, \dots, p_j, \dots, p_J)$, onde p_j é a proporção de respostas corretas obtidas pelos N examinados no item j , de modo que $p_1 > p_2 > \dots > p_j > \dots > p_J$. Seja X um vetor contendo pontuações individuais (itens binários) de um examinado, e seja X^* um vetor com 1's nas primeiras r_n posições e 0's nas últimas $j - r_n$ posições (este vetor é chamado de Guttman). Além disso, se $\sigma(\cdot, \cdot)$ denota a covariância entre os elementos de dois vetores, então Sato (1975) propôs o Índice de Cautela CI dado por

$$CI = 1 - \frac{Cov(X, p)}{Cov(X^*, p)}. \quad (2.5)$$

O Índice de Cautela CI é igual a 0, se o vetor de pontuação dos itens de um indivíduo for um vetor de Guttman; CI é igual a 1, se a covariância entre o vetor de pontuação do item e o vetor total dos itens é nulo. Se a covariância entre o vetor de pontuação do item e o vetor total de itens é negativo, CI excede 1. Assim, valores relativamente grandes de CI indicam que o comportamento da resposta é atípico em comparação com o comportamento das respostas das outras pessoas do grupo. Para o Índice de Cautela CI , um valor de corte de 0,5 foi sugerido (Sato, 1975).

2.2.5 Índice de Cautela Modificado (MCI)

Como o Índice de Cautela CI não tem limite superior fixo, a interpretação dos valores de CI pode ser problemática, então Harnisch e Linn (1981) propuseram o Índice de Cautela Modificado (MCI) que é uma versão modificada do Índice de Cautela que limita o Índice de

Cautela entre 0 e 1. Então, MCI pode ser escrito como:

$$MCI = \frac{Cov(X^*, p) - Cov(X, p)}{Cov(X^*, p) - Cov(X', p)}, \quad (2.6)$$

onde X é o vetor de resposta do examinado n , X^* um vetor com 1's nas primeiras r_n posições e 0's nas últimas $j - r_n$ posições (este vetor é chamado de Guttman) e X' denota o vetor de pontuação do item com 0's nas primeiras $j - r_n$ posições e 1's nas últimas r_n posições. O vetor e é o número máximo de erros Guttman e, portanto, é chamado de Vetor Guttman invertido.

De modo geral, o MCI é a razão entre duas covariâncias, ou seja, a covariância de X_n com o padrão perfeito e a covariância de X_n com o padrão perfeitamente inconsistente entre os itens de teste J , o padrão perfeitamente inconsistente contém respostas corretas apenas para o r_n itens mais difícil. Ou seja, quando $MCI = 0$, um examinando responde os r_n itens mais fáceis corretamente e os $J - r_n$ itens mais difíceis incorretamente. Assim, as pontuações dos itens do examinando estão de acordo com o vetor de Guttman e, quando $MCI = 1$, o padrão de pontuação do item é igual ao Guttman invertido padrão. Foi sugerido por Harnisch e Linn (1981) um valor de corte de 0,3.

2.2.6 Estatística $U3$ e $ZU3$

A estatística $U3$ (van der Flier, 1980) é uma estatística baseada em grupo com uma distribuição nula. Esta distribuição de amostragem pode ser usada para obter valores críticos para classificar padrões de pontuação de item como adequados ou não adequados. Então, o índice $U3$ é definido como:

$$U3 = \frac{\log(X_n^*) - \log(X_n)}{\log(X_n^*) - \log(X_n')} \quad (2.7)$$

sendo $\log(X_n) = \sum_{j=1}^J X_{nj} \log(p_j/(1-p_j))$, $\log(X_n^*) = \sum_{j=1}^J X_{nj}^* \log(p_j/(1-p_j))$ e $\log(X_n') = \sum_{j=1}^J X_{nj}' \log(p_j/(1-p_j))$ em que, p_j designando a proporção de respostas positivas para o j -

ésimo item, X_n é o vetor de pontuação do item de um indivíduo n com pontuações 0 (zero) e 1 (um) de itens dicotômicos, onde 1 indica uma resposta correta e 0 indica uma resposta incorreta. Além disso, X_n^* é o vetor que satisfaz o modelo de Guttman (1950), ou seja, todos os 1's estão à esquerda de todos os 0's e, por fim, X_n' é o vetor Guttman invertido com todos 1's à direita e todos 0's à esquerda.

$U3$ varia de 0 (zero) a 1 (um), sendo que o valor 0 indica que o item observado padrão de pontuação é um vetor Guttman e o valor 1 indica que o padrão observado é um vetor invertido de Guttman. Os valores crescentes indicam que os padrões são mais distantes dos padrões perfeitos de Guttman. As pontuações $U3$ dependem da pontuação correta, portanto van der Flier (1982) propôs $ZU3$ como uma padronização. As fórmulas para calcular $E(U3)$ e $Var(U3)$ podem ser encontradas em van der Flier (1982). $ZU3$ tem uma distribuição normal padrão assintótica. A estatística padronizada é

$$ZU3 = \frac{U3 - E(U3)}{[VAR(U3)]^{1/2}}. \quad (2.8)$$

2.2.7 Estatística H^T

Um tipo diferente de índice foi introduzido por Sijtsma (1986) e Sijtsma e Meijer (1992), denominado H_n^T . É calculado para o padrão de resposta do examinado n . A formulação matemática deste índice é dada por

$$H_n^T = \frac{\sum_{m \neq n} (\beta_{nm} - \beta_n \beta_m)}{\sum_{m \neq n} \min \{ \beta_n (1 - \beta_m), \beta_m (1 - \beta_n) \}}, \quad (2.9)$$

em que β_n é a proporção esperada de itens para os quais o examinado n forneceu respostas corretas, β_m é a proporção esperada de itens para os quais o examinado m forneceu respostas

corretas, enquanto

$$\beta_{nm} = \frac{1}{J} \sum_{j=1}^J X_{nj} X_{mj} \quad (2.10)$$

é a proporção esperada de itens para os quais tanto o examinado n quanto o examinado m forneceram respostas corretas. Com isso, $\sigma_{nm} := \beta_{nm} - \beta_n \beta_m$ é a covariância entre as pontuações dos examinados n e m . E, o máximo de H_n^T , igual a 1, ocorre quando cada uma das covariâncias entre os padrões de pontuação de item de examinados n e m ($\forall n, m; n \neq m$) atinge seu valor máximo. Já o valor nulo de H_n^T ocorre quando a média covariância (numerador) é 0; por fim, $H_n^T < 0$, se a covariância média for negativa. H_n^T não é padronizado contra o padrão Guttman. Sijtsma (1986) mostrou que $H_n^T = 1$ não é necessário para obter o padrão de pontuação de item perfeito.

2.2.8 Índice de Concordância A

Kane e Brennan (1980) discutiram o índice de concordância que mede a concordância entre o vetor de resposta do examinado n (X_n) e o vetor p de proporção das respostas corretas no item, onde p_j é a proporção de respostas corretas obtidas pelos N entrevistados no item j . O índice de concordância é dado por:

$$A_n = \sum_{j=1}^J X_{nj} p_j. \quad (2.11)$$

O valor do índice de concordância também varia 0 (zero) e 1 (um), onde 1 indica uma combinação perfeita e 0 indica nenhuma concordância.

Capítulo 3

Estatística R_n e Estudo de Simulação

3.1 Regressão isotônica

Em várias situações práticas, devido a alguma informação a priori, os parâmetros de interesse apresentam algum tipo de ordenação. Desse modo, deve-se buscar um modelo que preserve esta característica de ordem. O método para encontrar este modelo é denominado Regressão Isotônica, sendo que o termo isotônica (ou monotônica) refere-se ao fato de que um aumento na variável independente implique em um aumento na variável resposta. Se a variável resposta decresce com o aumento da variável preditora, utiliza-se o termo antitônica.

A regressão isotônica, que tornou-se mais conhecida, após trabalho de Barlow et al. (1972), refere-se a uma regressão não paramétrica que minimiza a soma dos quadrados de uma função não decrescente. Diferencia-se da regressão usualmente utilizada por se valer da informação de uma relação de ordem da variável resposta.

Seja X um conjunto nito x_1, \dots, x_n com relação de ordem estabelecida tal que $x_1 < x_2 < \dots < x_n$. Uma função real f em X é isotônica se $x, y \in X$ e $x < y$ implica que $f(x) \leq f(y)$, ou seja, é uma função não decrescente.

Se g é uma função em X e w uma função positiva em X , então uma função isotônica g^* em X é uma regressão isotônica de g , com peso w e com respeito a relação de ordem $x_1 < x_2 < \dots <$

x_n , se ela minimiza na classe das funções isotônicas f em X , a soma

$$\sum_{i=1}^n [g(x_i) - f(x_i)]^2 w(x_i).$$

Quando os parâmetros de interesse apresentam algum tipo de ordenação, espera-se que a regressão isotônica seja mais adequada, justamente por considerar essa relação de ordem. É o caso da relação entre a dificuldade do item e a probabilidade de acerto, pois quanto mais difícil é o item, menos acertos espera-se do respondente, mantido o parâmetro a constante, tendo, dessa maneira, uma relação de ordem decrescente, podendo tornar-se crescente simplesmente com uma inversão do eixo relativo à dificuldade do item.

Nesta subseção, utilizaremos temporariamente a notação G e G^* para designar quantidades distintas das estatísticas G de Gutmann e G^* de van der Flier.

Uma interpretação gráfica esclarecedora da regressão isotônica pode ser obtida ao plotar as seguintes somas cumulativas:

$$G_k = \sum_{i=1}^k g(x_i)w(x_i),$$

contra

$$W_k = \sum_{i=1}^k w(x_i), k = 1, 2, 3, \dots, n.$$

Esse gráfico é uma plotagem dos pontos $P_k = (W_k, G_k)$ no plano cartesiano. Esses pontos constituem o Diagrama da Soma Acumulada (DSA) de uma função dada g com peso w . A inclinação do segmento que liga o ponto P_{k-1} a P_k é dada por:

$$\frac{G_k - G_{k-1}}{W_k - W_{k-1}} = g(x_k).$$

Já a inclinação do segmento que liga P_{i-1} a P_k , para $i \leq k$, representa a média ponderada

$$Av\{x_i, x_{i+1}, \dots, x_k\} = \frac{\sum_{r=i}^k g(x_r)w(x_r)}{\sum_{r=i}^k w(x_r)}.$$

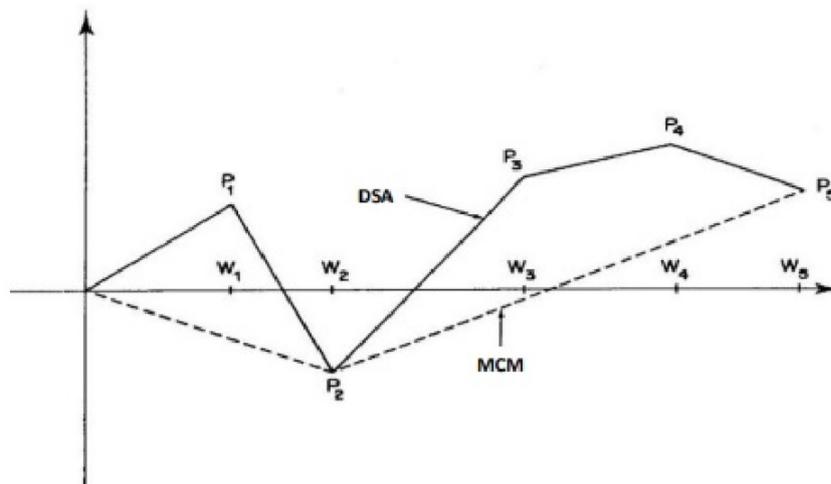
Será mostrado que a regressão isotônica de g é dada pela inclinação do Minorante Convexo Máximo (MCM) do DSA.

A função minorante convexa máxima (MCM) do DSA, no intervalo $[0, W_n]$, G^* , é o gráfico do supremo de todas as funções convexas que estão abaixo do DSA. Trata-se de uma linha que intercepta o DSA em alguns pontos, mas que nunca localiza-se acima dele.

A Figura (3.1) ilustra o DSA e a função MCM de um conjunto de dados hipotéticos. Por definição, como G^* é convexo no intervalo $[0, W_n]$, então também será diferenciável em cada um dos pontos W_i . Os pontos da MCM são representados por $P_k^* = (W_k, G_k^*)$ com

$$G_k^* = \sum_{i=1}^k g^*(x_i)w(x_i).$$

Desta forma, o valor da regressão isotônica g^* no ponto x_k é a inclinação da MCM no ponto $P_k^* = (W_k, G_k^*)$. Se P_k^* for um ponto não diferenciável do gráfico, $g^*(x_k)$ será a inclinação do seguimento prolongado à esquerda. Para algum i , se a MCM em W_i é colocada estritamente abaixo do DSA, então as inclinações da MCM à esquerda e à direita de W_i são as mesmas.



Fonte: Schlemper (2014)

Figura 3.1: Representação gráfica do Diagrama de Soma Acumulada (DSA) e a função Minorante Convexo Máximo (MCM).

Algumas propriedades do MCM serão expressas analiticamente. A primeira delas é que o DSA e o MCM coincidem em P_n^* , ou seja,

$$G_n^* = G_n. \quad (3.1)$$

A segunda propriedade é que, se para algum índice i , o MCM em P_{i-1}^* mantém-se estritamente abaixo do DSA em P_{i-1} , então a inclinação do MCM chegando em P_{i-1}^* e saindo do mesmo são iguais. Sendo assim,

$$G_{i-1}^* < G_{i-1} \Rightarrow g_i^* - g_{i-1}^* = 0, i = 1, 2, \dots, n. \quad (3.2)$$

Finalmente, se $P_r = P_r^*$, $P_s = P_s^*$, se $P_r P_s = P_r^* P_s^*$ é um vértice do MCM e $P_s P_t = P_s^* P_t^*$ é o vértice adjacente a direita, e se $r < q < s$, então a inclinação de $P_q P_s$ é menor que a inclinação de $P_q^* P_s^*$. Se $s < q < t$ então a inclinação de $P_s P_q$ é maior que a inclinação de $P_s^* P_q^*$. Essas propriedades auxiliam na prova de que g^* é uma regressão isotônica de g ponderado por w .

Teorema 3.1.1. *Se X tem ordenação simples, a inclinação g^* do MCM fornece a regressão isotônica de g . Ou seja, se f é isotônica em X então*

$$\sum_x [g(x) - f(x)]^2 w(x) \geq \sum_x [g(x) - g^*(x)]^2 w(x) + \sum_x [g^*(x) - f(x)]^2 w(x). \quad (3.3)$$

Além disso, a regressão isotônica é única (Barlow et al., 1972).

Demonstração. **(Teorema 3.1.1)** Por manipulações algébricas, a diferença entre as expressões dos membros esquerdo e direito da Eq. (3.3) reduz a

$$2 \sum_x [g(x) - g^*(x)][g^*(x) - f(x)]w(x).$$

Com isso, basta mostrar que

$$\sum_x [g(x) - g^*(x)][g^*(x) - f(x)]w(x) \geq 0,$$

para toda regressão isotônica f em X .

De fato, pois, pela fórmula da soma parcial de Abel, temos

$$\begin{aligned} \sum_{i=1}^n [g(x) - g^*(x)][g^*(x) - f(x)]w(x) &= \sum_{i=1}^n \{[f(x_i) - f(x_{i-1})] - [g^*(x_i) - g^*(x_{i-1})]\} \\ &\quad \times [G_{i-1} - G_{i-1}^*]w(x_i) + [g^*(x_n) - f(x_n)][G_n - G_n^*], \end{aligned} \tag{3.4}$$

em que $x_0 = f(x_0) = g^*(x_0) = G_0 = G_0^* = 0$.

O último termo na direita de (3.4) é nulo, por (3.1). Já pela propriedade apresentada em 3.2, segue que $[g^*(x_i) - g^*(x_{i-1})][G_{i-1} - G_{i-1}^*] = 0$, para todo $i \in \{1, 2, \dots, n\}$, uma vez que $G_{i-1}^* \leq G_{i-1}$, $i = 1, 2, \dots, n$. E, por fim, utilizando a hipótese de que f é isotônica, conclui-se a prova, pois, nesse caso, $f(x_{i-1}) \leq f(x_i)$, $i = 1, 2, \dots, n$.

Para mostrar a unicidade da regressão isotônica, suponha que g_1 seja uma outra função isotônica em X e que

$$\sum_x [g(x) - g_1(x)]^2 w(x) \leq \sum_x [g(x) - f(x)]^2 w(x),$$

para toda função isotônica f em X . Então, sendo g_1 e g^* isotônicas,

$$\sum_x [g(x) - g_1(x)]^2 w(x) = \sum_x [g(x) - g^*(x)]^2 w(x).$$

Porém, se g_1 é isotônica, então

$$\sum_x [g(x) - g_1(x)]^2 w(x) \geq \sum_x [g(x) - g^*(x)]^2 w(x) + \sum_x [g^*(x) - g_1(x)]^2 w(x).$$

E, isto implica que

$$\sum_x [g^*(x) - g_1(x)]^2 w(x) = 0,$$

consequentemente, $g^*(x) = g_1(x)$ para $x \in X$. □

Para executar o algoritmo PAV, para algum i tal que $g(x_{i-1}) > g(x_i)$, o gráfico do MCM entre os pontos P_{i-2}^* e P_i^* deverá ser um segmento de linha reta. Então, o gráfico do DSA pode ser alterado pela conexão de P_{i-2}^* e P_i^* por um segmento de linha reta, sem alteração no MCM. Este método de sucessivas aproximações do MCM pode ser descrito algebricamente como algoritmo *Pool-Adjacent-Violators* (PAV), proposto por Ayer et al. (1955).

Resumidamente, considerando X_i a variável explicativa e $Y_i = f(X_i)$ a variável explicada, o algoritmo PAV é dado pelos seguintes passos:

Passo 1: Ordene os pares (X_i, Y_i) de acordo com os valores de X_i , obtendo $(X_{(i)}, Y_{(i)})$, em que (i) é a i -ésima observação após a ordenação.

Passo 2: Começando por $Y_{(1)}$, mova para a direita e pare se o par $(Y_{(i)}, Y_{(i+1)})$ violar a restrição de monotonicidade, ou seja, se $Y_{(i)} > Y_{(i+1)}$. Se a restrição for violada, substitua os valores de $Y_{(i)}$ e $Y_{(i+1)}$ pela média de ambos, de tal forma que $Y_{(i)}^* = Y_{(i+1)}^* = (Y_{(i)} + Y_{(i+1)})/2$.

Passo 3: Confira se $Y_{(i-1)} \leq Y_{(i)}^*$; caso não seja, substitua os valores de $Y_{(i-1)}$, $Y_{(i)}$ e $Y_{(i+1)}$ pela média aritmética dos três, de tal forma que $Y_{(i-1)}^* = Y_{(i)}^* = Y_{(i+1)}^* = (Y_{(i-1)} + Y_{(i)} + Y_{(i+1)})/3$.

Seguindo esses passos, até obter a restrição de ordem necessária, o algoritmo fornecerá a regressão isotônica. O algoritmo PAV pode ser esquematizado, conforme segue:

Algoritmo 1: Pool-Adjacent-Violators (PAV)

Result: Regressão Isotônica

```

1 início: Ordene os pares  $(X_i, Y_i)$  de acordo com os valores de  $X_i$ , obtendo  $(X_{(i)}, Y_{(i)})$ ,
2 para Começo a partir de  $Y_{(1)}$ , i.e,  $(1 \leq i \leq n)$  faça
3   se  $Y_{(i)} > Y_{(i+1)}$  então
4     Mova para direita;
5     Substitua os valores de  $Y_{(i)}$  e  $Y_{(i+1)}$  pela média de ambos, de tal forma que
6        $Y_{(i)}^* = Y_{(i+1)}^* = (Y_{(i)} + Y_{(i+1)})/2$ .
7     se  $Y_{(i-1)} > Y_{(i)}^*$  então
8       Substitua os valores de  $Y_{(i-1)}$ ,  $Y_{(i)}$  e  $Y_{(i+1)}$  pela média aritmética dos três,
9       de tal forma que  $Y_{(i-1)}^* = Y_{(i)}^* = Y_{(i+1)}^* = (Y_{(i-1)} + Y_{(i)} + Y_{(i+1)})/3$ .
10    fim
11  fim

```

Härdle (1990) destaca que, se os dados seguem a restrição de ordem, então o algoritmo PAV irá simplesmente reproduzir os dados, e se a restrição for violada para algum Y_i , então uma função escada será produzida e degraus são obtidos no intervalo de violação. Essa função escada não é suavizada, embora seja não decrescente. Além disso, na presença de outliers ou valores discrepantes, o algoritmo PAV fornece uma função com uma longa linha reta na região desses valores. Uma solução proposta por Lee (2007) seria aplicar o algoritmo PAV para garantir a restrição e, depois, suavizar a regressão.

3.2 Uma nova estatística baseada na regressão isotônica R_n

Assim, seja $X_{jn} \in \{0, 1\}$ a resposta do n -ésimo respondente para o j -ésimo item, e $r_n = \sum_{j=1}^J X_{jn}$ o total de respostas corretas do n -ésimo indivíduo. Na situação ideal, esse

indivíduo deve fornecer as respostas corretas para os r_n itens mais fáceis (i.e., aqueles com os menores valores do parâmetro de dificuldade b_j , caso um modelo de TRI tenha sido ajustado, ou aqueles com as maiores proporções de respostas corretas, caso utilizemos uma estimação não paramétrica do grau de dificuldade do item). Seja g_{jn}^* , $j = 1, \dots, J$, a regressão isotônica (na verdade, antitônica) de X_{jn} , $j = 1, \dots, J$, seguindo a ordenação dos itens mais fáceis para os mais difíceis, e com pesos iguais para todos os itens. Podemos então, avaliar a coerência das respostas fornecidas pelo n -ésimo indivíduo comparando os valores g_j^* , $j = 1, \dots, J$, com aqueles da sequência ideal de respostas (r_n 1's seguidos de $J - r_n$ 0's). Uma estatística para avaliar a coerência das respostas é dada por

$$R_n = \frac{1}{J} \left(\sum_{j=1}^{r_n} (1 - g_{jn}^*) + \sum_{j=r_n+1}^J g_{jn}^* \right). \quad (3.5)$$

Quanto maior o valor de R_n , mais incoerente é a sequência de respostas do n -ésimo indivíduo.

3.3 Avaliando a Eficiência das estatísticas de ajuste individual

A eficiência geral das estatísticas de ajuste individual foi avaliada pela análise ROC (Receiver Operating Curve). A curva ROC produzida por uma estatística de ajuste individual é um gráfico bidimensional que traça pontos com as coordenadas $F(c)$ e $H(c)$, onde $F(c)$, no eixo x , é a proporção normal dos padrões de resposta incorreta identificados como aberrantes e $H(c)$, no eixo y , é a proporção de padrões de respostas aberrantes, e c é a pontuação de corte para qual a estatística de ajuste individual é comparada para tomada de decisão.

A proporção de $H(c)$ é a sensibilidade, ou "taxa de acerto", da estatística de ajuste individual condicionada ao valor c , a qual identifica corretamente um examinado que responde aberrante. Portanto, a taxa de erros $1 - H(c)$ é a probabilidade de que uma estatística de ajuste classifique incorretamente um examinado que responde aberrante como normal. A probabilidade $F(c)$ é a taxa de falso alarme da estatística de ajuste individual, condicionada ao valor c . Essa taxa

refere-se à probabilidade de uma estatística adequada ao indivíduo (no valor crítico c) classifica incorretamente um examinado normal como aberrante. A quantidade $1 - F(c)$ refere-se à especificidade, a probabilidade de uma estatística ajustada ao indivíduo classificar corretamente um examinado normal (Metz, 1986).

Para um conjunto de dados, a curva ROC de uma estatística ajustada ao indivíduo é representada por uma linha que conecta um conjunto dessas coordenadas em todos os possíveis valores c . Uma estatística adaptada à pessoa com alta sensibilidade e especificidade terá uma linha ROC que se curva próximo ao canto superior esquerdo do gráfico. Portanto, para uma estatística ajustada ao indivíduo, sua sensibilidade e especificidade são representadas em conjunto pela área $a \in [0, 1]$ sob a curva ROC. O valor a igual a 1 refere-se a uma estatística de ajuste com perfeitas sensibilidade e especificidade (Metz, 1986).

3.4 Simulações

Nesta seção, a definição dos casos é semelhante a que foi feito em Karabatsos (2003). As simulações incluíram 5 (cinco) tipos de examinados aberrantes (fraudadores, criativos, adivinhos, descuidados e aleatórios), 4 (quatro) porcentagens de examinados que respondem aberrantes (5%, 10%, 25% ou 50%) e 3 (três) períodos de teste (curto - 17 itens, médio - 33 itens ou longo - 65 itens) resultando em 60 conjuntos de dados com 500 examinados normais e aberrantes. Cada conjunto de dados foi replicado 100 vezes. A adequação do número escolhido de replicações foi baseado no artigo Tendeiro e Meijer (2014).

Assim, o conjunto de dados $X = (X_{nj} \mid n = 1, \dots, N; j = 1, \dots, J)$ foi pela geração de $X_{nj} = 1$ com probabilidade P_{nj1} e, $X_{nj} = 0$ com probabilidade $P_{nj0} = 1 - P_{nj1}$, independentemente para todos $n = 1, \dots, N$ e $j = 1, \dots, J$. Essas probabilidades $P = (P_{nj1} \mid n = 1, \dots, N; j = 1, \dots, J)$ foram obtidas a partir da definição dos modelos a serem utilizados (ML1P, ML2P ou ML3P), e através da introdução de valores particulares dos parâmetros de capacidade $\theta = (\theta_1, \dots, \theta_n, \dots, \theta_N)'$ e, os parâmetros de dificuldade do item

$b = (b_1, \dots, b_j, \dots, b_J)'$ para ML1P. Já o parâmetro de discriminação $a = (a_1, \dots, a_j, \dots, a_J)$ está presente no ML2P e ML3P e o parâmetro de acerto casual $c = (c_1, \dots, c_j, \dots, c_J)$ pertence somente aos casos ML3P.

Os examinados normais do conjunto de dados terão habilidades θ distribuídas uniformemente no intervalo $[-2, 2]$. Examinados fraudadores foram gerando respostas a itens de uma distribuição uniforme de baixa capacidade com um intervalo θ de $[-2, -0.5]$ e imputando respostas corretas para os 18% itens mais difíceis do teste (itens com $\delta_j > 1.5$). Examinados criativos foram gerando respostas de uma distribuição uniforme e de alta capacidade com um intervalo θ de $[0.5, 2]$ e imputando respostas incorretas aos itens 18% mais fáceis (itens com $\delta_j < -1.5$). Os examinados com respostas de adivinhação corretas foram gerados a partir de uma distribuição uniforme de baixa capacidade com um intervalo θ de $[-2, -0.5]$ e cada um dos 41% mais difíceis itens (itens com $\delta_j > 0.5$) foi atribuída uma probabilidade 0,25 de conter uma correta resposta. O valor 0,25 foi escolhido para simular itens de múltipla escolha, cada tendo quatro opções de resposta. Examinados descuidados foram a partir de uma distribuição uniforme de alta capacidade com um intervalo θ de $[0.5, 2]$ e cada um dos 41% de itens mais fáceis (com dificuldade $\delta_j < -0.5$) foi atribuída uma probabilidade 0,5 de conter uma resposta incorreta. Para gerar examinados que respondem aleatoriamente, a cada item, foi atribuída uma probabilidade 0,25 de conter uma resposta correta. Isso simula examinados que respondem cegamente a todos os itens de teste de múltipla escolha, tendo cada item quatro opções de resposta.

Para cada um dos conjuntos de dados, os parâmetros dos modelos ML1P, ML2P e ML3P foram estimados, e as 10 (dez) estatísticas de ajuste individual foram calculadas para cada examinado (simulado). A comparação das 10 (dez) estatísticas de ajuste individual ocorrerá pela área sob a curva ROC dos 60 conjuntos de dados gerados a partir da simulação de dados de 3 (três) tipos de teste (17, 33 e 65 itens), 4 (quatro) porcentagens de respondentes aberrantes (5%, 10%, 25% e 50%) e 5 (cinco) tipos de respondentes (fraudadores, descuidados, criativos, adivinhos e aleatórios).

Os métodos descritos neste trabalho foram implementados computacionalmente com auxílio software para computação estatística R, versão 4.0.3 (R Core Team, 2020).

3.5 Resultados

3.5.1 Simulação para modelos com 1 (um) parâmetro

A Tabela 3.1 mostra a média e o desvio padrão da área sob a ROC de dados com modelo de 1 (um) parâmetro para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n , em três comprimentos de teste (17, 33 e 65 itens) e quatro níveis de respondentes aberrantes (5%, 10%, 25% e 50%)

Os resultados das áreas sob a ROC na Tabela 3.1 mostram que as taxas de detecção das estatísticas aumentam com a duração do teste e diminuem com a porcentagem de respondente aberrante. Notou-se que a condição de 50% é a mais difícil de detectar os respondentes aberrantes. Para caso com tamanho de teste curto (17 itens) e 5% de respondentes aberrantes, a estatística H^T obteve o melhor desempenho e a estatística proposta R_n apresentou melhor taxa de detecção em relação às estatísticas G^* , CI , MCI , $U3$, $ZU3$ e A . Para o caso de teste médio (33 itens) e longo (65 itens) e 5% de respondentes aberrantes, as estatísticas G e R_n são mais eficazes.

Os casos com teste curto (17 itens) e 10% de respondentes aberrantes as H^T e G são as estatísticas com melhor atuação e a estatística proposta R_n apresentou a taxa de detecção semelhante à estatística r_{pbis} e melhor que as estatísticas G^* , CI , MCI , $U3$, $ZU3$, A . Para o caso de teste médio (33) e teste longo (65 itens) e 10% de respondentes aberrantes as estatísticas G e R_n são mais eficazes (Tabela 3.1).

Em relação aos casos com 25% de respondentes aberrantes e para teste curto com 17 itens as estatísticas G e H^T apresentou o melhor desempenho e a estatística R_n foi mais eficaz na identificação de indivíduos com padrão de resposta inadequada que as estatísticas G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, A . Para teste com 33 itens e com 65 itens as estatísticas G e R_n são os me-

lhores na detecção de respondente descuidos. Para casos com 50% de respondentes aberrantes as estatísticas G e R_n são mais eficazes para os três comprimentos de teste. (Tabela 3.1).

Assim, na Tabela 3.1, estatística H^T apresentou-se mais eficiente para casos com testes curtos com porcentagem máxima de 25% de respondentes aberrantes. As estatísticas G e R_n foram mais eficazes na detecção de examinandos que respondem com descuido, para testes médios e longos, para os 4 (quatro) níveis de respondentes aberrantes (5%, 10%, 25% e 50%) (Tabela 3.1).

Para os fraudadores, a Tabela 3.2 mostra a média e desvio padrão da área sob a ROC de dados com modelo de 1 (um) parâmetro para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n com três comprimentos de teste (17, 33 e 65 itens), 4 (quatro) porcentagens de respondentes aberrantes (5%, 10%, 25% e 50%) e considerando respostas corretas para 18% dos itens mais difíceis.

As médias das áreas sob a ROC na Tabela 3.2 mostraram-se semelhantes aos resultados da Tabela 3.1, onde o desempenho das estatísticas aumentam com a duração do teste e com a porcentagem de respondentes aberrantes. E, para o nível de 50% de respondentes, as estatísticas detectam, com maiores dificuldades, as respostas aberrantes.

Na Tabela 3.2, para 5% e 10% de respondentes aberrantes com tamanho de teste curto (17 itens), a estatística H^T apresentou o melhor desempenho. Para o caso de teste médio (33 itens) e teste longo (65 itens), as estatísticas G e G^* foram mais eficazes. A estatística proposta R_n apresentou-se melhor que o índice A para todos os tamanhos de teste. Nos casos com 25% de respondentes aberrantes e três comprimentos de teste (17, 33 e 65 itens), a estatística R_n foi mais eficaz na detecção de examinandos que respondem com fraude. Já para casos com 50% de respondentes fraudadores o índice A foi que apresentou melhor desempenho nos três tamanhos de teste.

Para os respondentes fraudadores, a estatística H^T apresentou-se mais eficiente para casos com testes curtos e com porcentagem máxima de 10% de respondentes aberrantes. As estatísticas G e G^* foram mais eficazes para testes médios e longos, considerando 5% e 10% de

respondentes aberrantes e, também, para casos com 25% de respondentes aberrantes a estatística R_n foi mais eficaz em todos os tamanhos de teste propostos e para casos com taxa 50% de respondentes aberrantes estatística A foi que apresentou melhor desempenho nos três tamanhos de teste.

Na Tabela 3.3 observa-se média e desvio padrão da área sob a curva ROC de dados para modelos de 1 (um) parâmetro para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com resposta criativa (5%, 10%, 25% e 50%) e considerando resposta incorreta para 18% dos itens fáceis.

Os resultados da Tabela 3.3 são semelhante aos vistos na Tabela 3.1 e 3.2, onde as ROC das estatísticas também mostram que a capacidade de detecção das estatísticas aumenta com a duração do teste e diminui com a porcentagem de respondentes aberrantes, onde os casos com 50% de respondente com resposta criativa é a mais difícil de detectar.

Considerando 5% de respondentes com resposta criativa e com tamanho de teste curto as estatísticas H^T e r_{pbis} apresentaram-se mais eficientes e a estatística proposta R_n apresentou a taxa de detecção semelhante à estatística G e melhor do que das estatísticas G^* , CI , MCI , $U3$, $ZU3$ e A . Já as estatísticas G e R_n apresentaram o melhor desempenho para os testes médios e longos. Os casos com 10 % de respondentes com resposta criativa e considerando os testes curtos as estatísticas H^T , G e R_n apresentaram média da área ROC superior às áreas ROC das demais estatísticas. Para os testes com 33 e 65 itens, as estatísticas G e R_n foram as mais eficazes. Sob a condição em que considera 25% de respondentes aberrantes, a estatística R_n teve a maior área ROC em todos os tamanhos de teste proposto e, para situações com taxa 50% de respondentes aberrantes, o índice A apresentou melhor desempenho nos três tamanhos de teste e, a estatística proposta R_n apresentou a melhor taxa de detecção em relação às estatísticas G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$ e H^T .

Na Tabela 3.4, observa-se média e desvio padrão da área sob a curva ROC de dados para modelo logístico de 1 (um) parâmetro para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI ,

MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com respostas de adivinhação (5%, 10%, 25% e 50%) e considerando uma probabilidade de 0,25 de conter a resposta correta para os 41% dos itens difíceis.

Os resultados da Tabela 3.4 mostram-se diferentes dos demais. Note que as estimativas das áreas sob a curva ROC aumentam com a duração do teste e não aumentam com o aumento da porcentagem de respondentes aberrantes. Observa-se ainda que para 5% de respondentes aberrantes, em teste com 17 itens, o índice A apresentou o melhor desempenho. Para testes médios e longos, as estatísticas G e H^T apresentaram mais eficazes. Para os testes curtos com 10% de respondentes aberrantes, o índice A foi melhor do que as demais e, para testes médios e longos, as estatísticas G e MCI tiveram uma atuação melhor na detecção de respondentes aberrantes com resposta de adivinhação. Para casos com 25% de respondentes aberrantes, o índice A foi a mais competente para testes curtos e, analisando para os testes de tamanhos médios, MCI foi a melhor. Já em testes com 65 itens, a estatística G foi a mais eficaz. Em casos com 50% de respondentes aberrantes e teste com 17 itens a estatística H^T apresentou mais eficiente e, para testes médios e longos, a estatística G mostrou-se melhor que as demais. A estatística proposta R_n apresentou o pior desempenho para três tamanhos de teste com porcentagem máxima de 25% de respondentes aberrantes, já para casos com 50% de respondentes aberrantes, a estatística proposta apresentou o pior desempenho em testes com 17 e 33 itens e para 65 itens mostrou-se melhor que o índice A .

Para a Tabela 3.5, analisou-se a média e o desvio padrão da área sob a curva ROC de dados para modelos logísticos de 1 (um) parâmetro para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com respostas aleatórias (5%, 10%, 25% e 50%) e considerando uma probabilidade de 0,25 para cada item em conter a resposta correta.

Na Tabela 3.5, as áreas sob as curvas ROC são bem parecidas para respondentes adivinhos e apresentam maiores diferenças nos demais. Note que o desempenho das estatísticas aumenta

com a duração do teste e não com a porcentagem de respondentes aberrantes.

Observando os casos com 5% de respondentes aberrantes, considerando os três comprimentos de teste, a estatística r_{pbis} foi a mais eficiente. Verificando os resultados para 10% de respondentes aberrantes e tamanhos de teste curto e médio, a estatística r_{pbis} também apresentou melhor desempenho. Mas, para teste longo, as estatísticas H^T e r_{pbis} foram mais eficazes. Já para casos com 25% de respondentes aberrantes, em testes de 17 e 33 itens, as estatísticas r_{pbis} e H^T foram melhores que as demais e, para teste com 65 itens, a estatística H^T foi mais eficaz. Os casos que consideram 50% de respondentes com resposta aleatória, a H^T e r_{pbis} mostraram-se mais competentes. A estatística proposta R_n apresentou o pior desempenho para três tamanhos de teste e para as 4 (quatro) porcentagens de respondentes aberrantes (Tabela 3.5).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.8774 (0.02468)	0.97976 (0.00885)	0.99665 (0.00159)
	G^*	0.8566 (0.02838)	0.96821 (0.01247)	0.99313 (0.00416)
	r_{pbis}	0.87234 (0.02875)	0.95583 (0.01775)	0.98772 (0.00839)
	CI	0.85442 (0.02757)	0.96745 (0.01353)	0.99296 (0.00427)
	MCI	0.85504 (0.02761)	0.96765 (0.01382)	0.99299 (0.00418)
	$U3$	0.8531 (0.02776)	0.96614 (0.01365)	0.99237 (0.00447)
	$ZU3$	0.8531 (0.02776)	0.96614 (0.01365)	0.99237 (0.00447)
	H^T	0.88872 (0.02582)	0.96868 (0.01289)	0.9935 (0.00369)
	A	0.52896 (0.03316)	0.53989 (0.03258)	0.53547 (0.03417)
	R_n	0.87028 (0.02667)	0.97558 (0.01214)	0.99426 (0.00447)
10%	G	0.88781 (0.01683)	0.97741 (0.00823)	0.99456 (0.00277)
	G^*	0.86673 (0.01829)	0.96451 (0.01072)	0.99014 (0.00476)
	r_{pbis}	0.88065 (0.0249)	0.94898 (0.0146)	0.98182 (0.007)
	CI	0.86273 (0.02039)	0.96195 (0.01136)	0.98914 (0.00466)
	MCI	0.86402 (0.02007)	0.96216 (0.01122)	0.9886 (0.00465)
	$U3$	0.86176 (0.02039)	0.96065 (0.01137)	0.98787 (0.0051)
	$ZU3$	0.86176 (0.02039)	0.96065 (0.01137)	0.98787 (0.0051)
	H^T	0.89842 (0.02042)	0.96385 (0.01065)	0.99016 (0.00443)
	A	0.52103 (0.02471)	0.52666 (0.02352)	0.53522 (0.02263)
	R_n	0.88011 (0.02066)	0.97082 (0.00915)	0.99283 (0.00267)
25%	G	0.88277 (0.02038)	0.97241 (0.00591)	0.99451 (0.00165)
	G^*	0.86078 (0.02117)	0.95724 (0.00751)	0.99014 (0.00306)
	r_{pbis}	0.85571 (0.01934)	0.92701 (0.00998)	0.97724 (0.00585)
	CI	0.84924 (0.02067)	0.94532 (0.00868)	0.98597 (0.00379)
	MCI	0.85046 (0.02038)	0.94164 (0.00975)	0.98153 (0.00489)
	$U3$	0.84842 (0.02032)	0.93984 (0.00959)	0.9807 (0.00488)
	$ZU3$	0.84842 (0.02032)	0.93984 (0.00959)	0.9807 (0.00488)
	H^T	0.88271 (0.01737)	0.94901 (0.00761)	0.98876 (0.00312)
	A	0.49015 (0.02099)	0.50643 (0.0339)	0.50874 (0.02566)
	R_n	0.87165 (0.02243)	0.96703 (0.00673)	0.99264 (0.00165)
50%	G	0.79408 (0.03236)	0.91076 (0.01383)	0.9674 (0.00331)
	G^*	0.7502 (0.03794)	0.86652 (0.023)	0.93591 (0.00849)
	r_{pbis}	0.74125 (0.03031)	0.83508 (0.02006)	0.90561 (0.01287)
	CI	0.76687 (0.02721)	0.86854 (0.01616)	0.93327 (0.00929)
	MCI	0.77629 (0.02472)	0.85894 (0.01786)	0.91347 (0.01195)
	$U3$	0.77169 (0.02521)	0.85573 (0.01807)	0.91124 (0.01189)
	$ZU3$	0.77169 (0.02521)	0.85573 (0.01807)	0.91124 (0.01189)
	H^T	0.78943 (0.02867)	0.88442 (0.01432)	0.95023 (0.00717)
	A	0.44001 (0.02255)	0.47553 (0.03046)	0.45789 (0.02815)
	R_n	0.79518 (0.03031)	0.91047 (0.01356)	0.96536 (0.00498)

Tabela 3.1: Média e Desvio Padrão da área sob a curva ROC para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes descuidados (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.94851 (0.01205)	0.99408 (0.0028)	0.99768 (0.00065)
	G^*	0.93603 (0.01299)	0.98779 (0.00504)	0.99743 (0.00048)
	r_{pbis}	0.95996 (0.00975)	0.97694 (0.0092)	0.99617 (0.00103)
	CI	0.92896 (0.01354)	0.98328 (0.00694)	0.99697 (7e-04)
	MCI	0.92772 (0.01348)	0.98288 (0.00723)	0.99677 (0.00096)
	$U3$	0.92882 (0.01357)	0.98281 (0.00703)	0.99673 (0.001)
	$ZU3$	0.92882 (0.01357)	0.98281 (0.00703)	0.99673 (0.001)
	H^T	0.96425 (0.0086)	0.98366 (0.00663)	0.99696 (0.00074)
	A	0.69492 (0.02829)	0.71402 (0.04195)	0.69721 (0.02455)
	R_n	0.86648 (0.02668)	0.90603 (0.02051)	0.97193 (0.00935)
10%	G	0.94542 (0.01353)	0.98917 (0.00409)	0.99747 (0.00046)
	G^*	0.92642 (0.01725)	0.97574 (0.00816)	0.99608 (0.00147)
	r_{pbis}	0.9392 (0.01654)	0.95834 (0.01186)	0.99233 (0.00291)
	CI	0.91618 (0.0163)	0.97137 (0.00888)	0.99536 (0.00174)
	MCI	0.91278 (0.01729)	0.96912 (0.00955)	0.99412 (0.00225)
	$U3$	0.91068 (0.01802)	0.96732 (0.00976)	0.99359 (0.00243)
	$ZU3$	0.91068 (0.01802)	0.96732 (0.00976)	0.99359 (0.00243)
	H^T	0.94883 (0.01426)	0.97177 (0.0089)	0.9955 (0.00163)
	A	0.67647 (0.02901)	0.69518 (0.02619)	0.68549 (0.02255)
	R_n	0.89114 (0.01344)	0.93854 (0.01698)	0.98616 (0.00506)
25%	G	0.82434 (0.04409)	0.88189 (0.02945)	0.93933 (0.01508)
	G^*	0.76188 (0.05005)	0.81487 (0.03469)	0.87909 (0.02384)
	r_{pbis}	0.75908 (0.03092)	0.77515 (0.02348)	0.85752 (0.0233)
	CI	0.77406 (0.02717)	0.82208 (0.02144)	0.89774 (0.02089)
	MCI	0.77927 (0.02795)	0.83177 (0.02064)	0.89764 (0.01939)
	$U3$	0.76609 (0.02985)	0.81549 (0.02111)	0.88518 (0.02111)
	$ZU3$	0.76609 (0.02985)	0.81549 (0.02111)	0.88518 (0.02111)
	H^T	0.80831 (0.02694)	0.84111 (0.0203)	0.91655 (0.0173)
	A	0.61553 (0.02901)	0.6512 (0.02292)	0.62327 (0.02377)
	R_n	0.85907 (0.03832)	0.91211 (0.01872)	0.9626 (0.01176)
50%	G	0.2945 (0.03365)	0.2522 (0.03621)	0.17776 (0.02501)
	G^*	0.22038 (0.03254)	0.16269 (0.02775)	0.06819 (0.01355)
	r_{pbis}	0.17579 (0.01822)	0.11978 (0.01689)	0.03449 (0.0083)
	CI	0.2012 (0.01955)	0.15466 (0.02118)	0.06878 (0.01138)
	MCI	0.19939 (0.019)	0.15335 (0.02207)	0.07867 (0.01221)
	$U3$	0.20596 (0.01944)	0.1601 (0.02257)	0.08688 (0.01309)
	$ZU3$	0.20596 (0.01944)	0.1601 (0.02257)	0.08688 (0.01309)
	H^T	0.23131 (0.01899)	0.18265 (0.02305)	0.07816 (0.0128)
	A	0.50429 (0.02934)	0.54246 (0.03025)	0.50202 (0.03017)
	R_n	0.30012 (0.03522)	0.25361 (0.0371)	0.18638 (0.02042)

Tabela 3.2: Média e Desvio Padrão da área sob a curva ROC para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes fraudadores (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.94625 (0.01343)	0.99395 (0.00302)	0.99787 (5e-05)
	G^*	0.93358 (0.01372)	0.98808 (0.00462)	0.99745 (0.00046)
	r_{pbis}	0.95592 (0.01321)	0.97811 (0.00837)	0.9966 (0.00147)
	CI	0.92508 (0.01756)	0.98382 (0.00632)	0.99718 (0.00072)
	MCI	0.92444 (0.0182)	0.98316 (0.00673)	0.99696 (9e-04)
	$U3$	0.92504 (0.01715)	0.98341 (0.00628)	0.99695 (0.00095)
	$ZU3$	0.92504 (0.01715)	0.98341 (0.00628)	0.99695 (0.00095)
	H^T	0.96082 (0.01188)	0.9842 (0.00612)	0.99721 (0.00074)
	A	0.57561 (0.03277)	0.52652 (0.02551)	0.54485 (0.03153)
	R_n	0.94322 (0.01302)	0.99013 (0.00444)	0.99768 (0.00037)
10%	G	0.9423 (0.01677)	0.99084 (0.00332)	0.99752 (0.00041)
	G^*	0.9208 (0.01771)	0.98037 (0.00703)	0.99623 (0.00098)
	r_{pbis}	0.93048 (0.02253)	0.96428 (0.00968)	0.99206 (0.00286)
	CI	0.90839 (0.02225)	0.97556 (0.00745)	0.99504 (0.00157)
	MCI	0.90587 (0.02375)	0.97395 (0.00766)	0.99341 (0.0024)
	$U3$	0.90445 (0.02424)	0.9724 (0.0078)	0.99283 (0.00254)
	$ZU3$	0.90445 (0.02424)	0.9724 (0.0078)	0.99283 (0.00254)
	H^T	0.94211 (0.01687)	0.97595 (0.00718)	0.99521 (0.00146)
	A	0.55945 (0.02359)	0.50222 (0.02223)	0.52663 (0.02432)
	R_n	0.94134 (0.0164)	0.98763 (0.00345)	0.99725 (0.00042)
25%	G	0.81841 (0.03224)	0.86368 (0.03135)	0.94736 (0.01534)
	G^*	0.75458 (0.04134)	0.78936 (0.03484)	0.89232 (0.02416)
	r_{pbis}	0.75294 (0.03099)	0.76212 (0.03011)	0.86393 (0.02175)
	CI	0.76857 (0.03257)	0.81141 (0.02959)	0.9022 (0.01966)
	MCI	0.77368 (0.03246)	0.82321 (0.02877)	0.9042 (0.01976)
	$U3$	0.76054 (0.03251)	0.80591 (0.03072)	0.89254 (0.02101)
	$ZU3$	0.76054 (0.03251)	0.80591 (0.03072)	0.89254 (0.02101)
	H^T	0.80421 (0.02734)	0.8303 (0.0258)	0.92025 (0.01611)
	A	0.48702 (0.02321)	0.43963 (0.01971)	0.45528 (0.02094)
	R_n	0.83246 (0.03102)	0.88865 (0.02858)	0.95526 (0.01464)
50%	G	0.29013 (0.0386)	0.25206 (0.03556)	0.17457 (0.02185)
	G^*	0.21846 (0.03766)	0.16052 (0.03178)	0.06657 (0.01674)
	r_{pbis}	0.17662 (0.02275)	0.11815 (0.01839)	0.03343 (0.01042)
	CI	0.20296 (0.02331)	0.15139 (0.0209)	0.06535 (0.01446)
	MCI	0.20156 (0.02325)	0.15064 (0.02072)	0.07446 (0.0148)
	$U3$	0.20755 (0.02338)	0.15764 (0.02135)	0.08153 (0.01522)
	$ZU3$	0.20755 (0.02338)	0.15764 (0.02135)	0.08153 (0.01522)
	H^T	0.23218 (0.02434)	0.17863 (0.02202)	0.07944 (0.01682)
	A	0.37894 (0.02505)	0.34666 (0.03104)	0.35213 (0.02158)
	R_n	0.2596 (0.03397)	0.21715 (0.02995)	0.14429 (0.02152)

Tabela 3.3: Média e Desvio Padrão da área sob a curva ROC para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes criativos (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.68623 (0.06164)	0.83017 (0.04914)	0.89886 (0.02811)
	G^*	0.67845 (0.061)	0.82276 (0.04834)	0.88598 (0.03525)
	r_{pbis}	0.69377 (0.05782)	0.80019 (0.05565)	0.84872 (0.04365)
	CI	0.68135 (0.06235)	0.82553 (0.05047)	0.88761 (0.03524)
	MCI	0.68322 (0.064)	0.82725 (0.0497)	0.89081 (0.0358)
	$U3$	0.67947 (0.06313)	0.82334 (0.05036)	0.88647 (0.03654)
	$ZU3$	0.67947 (0.06313)	0.82334 (0.05036)	0.88647 (0.03654)
	H^T	0.71839 (0.06035)	0.83133 (0.04965)	0.89222 (0.03486)
	A	0.75085 (0.03539)	0.76516 (0.03815)	0.76017 (0.02449)
	R_n	0.5749 (0.05889)	0.7043 (0.04744)	0.75399 (0.04638)
10%	G	0.69008 (0.04216)	0.80996 (0.01816)	0.90486 (0.02612)
	G^*	0.6827 (0.04512)	0.79884 (0.01998)	0.89464 (0.02781)
	r_{pbis}	0.6964 (0.04654)	0.76702 (0.02473)	0.8554 (0.03143)
	CI	0.68507 (0.04438)	0.80118 (0.02002)	0.89611 (0.02783)
	MCI	0.69118 (0.04382)	0.80809 (0.02023)	0.90163 (0.02727)
	$U3$	0.68793 (0.04486)	0.80422 (0.02057)	0.89752 (0.02759)
	$ZU3$	0.68793 (0.04486)	0.80422 (0.02057)	0.89752 (0.02759)
	H^T	0.71973 (0.0427)	0.80597 (0.01798)	0.90091 (0.02654)
	A	0.7477 (0.03086)	0.75383 (0.02111)	0.76722 (0.02058)
	R_n	0.58261 (0.04435)	0.68309 (0.03637)	0.76389 (0.03847)
25%	G	0.70788 (0.02688)	0.8237 (0.01427)	0.90006 (0.01542)
	G^*	0.70079 (0.02889)	0.81244 (0.01884)	0.88634 (0.01798)
	r_{pbis}	0.70243 (0.03008)	0.76782 (0.02282)	0.83477 (0.02206)
	CI	0.70451 (0.02779)	0.81448 (0.01882)	0.88753 (0.01865)
	MCI	0.71708 (0.02736)	0.83014 (0.01813)	0.89914 (0.01792)
	$U3$	0.71294 (0.02739)	0.82522 (0.01805)	0.89508 (0.01842)
	$ZU3$	0.71294 (0.02739)	0.82522 (0.01805)	0.89508 (0.01842)
	H^T	0.73892 (0.02528)	0.82043 (0.01913)	0.89764 (0.01652)
	A	0.74557 (0.02027)	0.75659 (0.02276)	0.76461 (0.01908)
	R_n	0.59728 (0.03113)	0.69166 (0.02267)	0.76223 (0.02475)
50%	G	0.70302 (0.02809)	0.81815 (0.01951)	0.89139 (0.01153)
	G^*	0.69236 (0.02741)	0.80795 (0.01857)	0.87707 (0.01359)
	r_{pbis}	0.66875 (0.02604)	0.7361 (0.02328)	0.793 (0.02224)
	CI	0.69774 (0.02476)	0.80982 (0.02013)	0.87614 (0.01491)
	MCI	0.71977 (0.02466)	0.83634 (0.01857)	0.89085 (0.01574)
	$U3$	0.71727 (0.02485)	0.8336 (0.01886)	0.89124 (0.01517)
	$ZU3$	0.71727 (0.02485)	0.8336 (0.01886)	0.89124 (0.01517)
	H^T	0.73054 (0.02283)	0.81596 (0.01911)	0.8887 (0.0142)
	A	0.72983 (0.02173)	0.75682 (0.0234)	0.74537 (0.02521)
	R_n	0.59812 (0.02701)	0.68972 (0.02162)	0.75706 (0.02008)

Tabela 3.4: Média e Desvio Padrão da área sob a curva ROC para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes adivinhos (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.79802 (0.04706)	0.90575 (0.03316)	0.96965 (0.01723)
	G^*	0.85568 (0.03295)	0.96329 (0.0212)	0.99374 (0.00434)
	r_{pbis}	0.90787 (0.02872)	0.97033 (0.0139)	0.99393 (0.00386)
	CI	0.86662 (0.03236)	0.96136 (0.02302)	0.99375 (0.00429)
	MCI	0.86625 (0.0323)	0.96083 (0.02338)	0.99383 (0.00417)
	$U3$	0.86539 (0.0298)	0.9612 (0.02205)	0.99372 (0.00429)
	$ZU3$	0.86539 (0.0298)	0.9612 (0.02205)	0.99372 (0.00429)
	H^T	0.8903 (0.04164)	0.96155 (0.02429)	0.99403 (0.00411)
	A	0.88877 (0.0175)	0.91751 (0.01689)	0.92709 (0.01825)
	R_n	0.68572 (0.06151)	0.77957 (0.05564)	0.87151 (0.04013)
10%	G	0.81104 (0.04246)	0.90808 (0.02499)	0.96557 (0.00971)
	G^*	0.87068 (0.02951)	0.96488 (0.01688)	0.99284 (0.00349)
	r_{pbis}	0.91533 (0.02261)	0.97092 (0.01187)	0.99323 (0.00333)
	CI	0.87431 (0.02717)	0.96464 (0.01698)	0.99258 (0.00363)
	MCI	0.8731 (0.02707)	0.965 (0.01708)	0.99235 (0.00359)
	$U3$	0.87071 (0.02668)	0.96277 (0.01758)	0.99156 (0.0039)
	$ZU3$	0.87071 (0.02668)	0.96277 (0.01758)	0.99156 (0.0039)
	H^T	0.89909 (0.02682)	0.96646 (0.01637)	0.99345 (0.00314)
	A	0.88739 (0.01323)	0.90924 (0.01387)	0.92942 (0.01099)
	R_n	0.69317 (0.0422)	0.77282 (0.04029)	0.85868 (0.02712)
25%	G	0.81328 (0.01805)	0.91415 (0.01737)	0.96643 (0.0069)
	G^*	0.87692 (0.01613)	0.96705 (0.00941)	0.99316 (0.00256)
	r_{pbis}	0.9207 (0.00964)	0.97084 (0.00791)	0.99312 (0.0025)
	CI	0.88177 (0.0181)	0.96613 (0.00968)	0.99308 (0.00264)
	MCI	0.87986 (0.0194)	0.966 (0.01034)	0.99284 (0.0029)
	$U3$	0.87054 (0.01977)	0.96026 (0.01158)	0.99078 (0.00391)
	$ZU3$	0.87054 (0.01977)	0.96026 (0.01158)	0.99078 (0.00391)
	H^T	0.91641 (0.00983)	0.9728 (0.00794)	0.9946 (0.00184)
	A	0.88975 (0.01312)	0.9051 (0.01203)	0.92138 (0.0109)
	R_n	0.68439 (0.02585)	0.7866 (0.0302)	0.86324 (0.01383)
50%	G	0.80031 (0.01723)	0.90928 (0.01244)	0.96411 (0.00732)
	G^*	0.86072 (0.01264)	0.96457 (0.00719)	0.99068 (0.00185)
	r_{pbis}	0.90974 (0.0127)	0.96926 (0.00738)	0.9912 (0.00175)
	CI	0.86528 (0.01422)	0.96431 (0.00778)	0.99069 (0.00202)
	MCI	0.86253 (0.01785)	0.96448 (0.00861)	0.99075 (0.00209)
	$U3$	0.84675 (0.01858)	0.95602 (0.01058)	0.98806 (0.00274)
	$ZU3$	0.84675 (0.01858)	0.95602 (0.01058)	0.98806 (0.00274)
	H^T	0.91285 (0.0107)	0.97559 (0.00583)	0.99322 (0.00132)
	A	0.87533 (0.01842)	0.89378 (0.01383)	0.9051 (0.01176)
	R_n	0.67915 (0.0245)	0.78287 (0.02188)	0.85792 (0.02357)

Tabela 3.5: Média e Desvio Padrão da área sob a curva ROC de dados em ML1P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aleatórios (5%, 10%, 25% e 50%).

Cada figura corresponde à comparação das estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n , com dados ML1P para respondentes aberrantes (descuidados, fraudadores, criativos, adivinhos e aleatórios). Em cada painel gráfico, temos para cada tipo de respondentes aberrantes a ROC para cada estatística de ajuste individual no eixo y e porcentagem de respondentes aberrantes ou tamanho de teste no eixo x . Observe que os resultados gráficos são um pouco mais geral do que os apresentados nas tabelas anteriores, e que podem demonstrar uma pequena diferença em relação aos resultados apresentados nas tabelas, pois, para gerar os gráficos, é necessário o agrupamento médio da curva ROC das estatísticas de ajuste individual.

Nas Figuras 3.2 e 3.3, nota-se que o índice A é bem menos eficaz para as quatro porcentagens de respondentes descuidados e os três comprimentos de teste. É possível afirmar também que para os casos com respondentes descuidados as estatísticas G , H^T e R_n apresentaram mais desempenho que as demais considerando os três tamanhos de teste e as quatro porcentagens de respondentes aberrantes. Também é possível confirmar, com base nos gráficos e nas tabelas dos respondentes descuidados, que o tamanho do teste e a porcentagem dos respondentes aberrantes influenciam no desempenho das estatísticas.

Para os respondentes fraudadores, as Figuras 3.4 e 3.5 mostram-se semelhantes aos dados da Tabela 3.2, onde a capacidade das estatísticas de ajuste individual aumentam com o aumento das porcentagens de respondentes fraudadores e dos comprimentos de teste. Para 50% de respondentes fraudadores, o índice A teve o melhor desempenho, seguido das estatísticas H^T e R_n , considerando o teste longo. Para os casos com 25% respondentes fraudadores, as estatísticas G e R_n foram mais eficazes. Observando os casos com 5% e 10% de respondentes fraudadores, a estatística proposta apresenta progresso no desempenho com o aumento do tamanho do teste.

Nas Figuras 3.6 e 3.7, para os casos com 5% e 10% de respondentes criativos nos três tamanhos de teste, as áreas sob a curva ROC das 10 (dez) estatísticas de ajuste individual são muito próximas, sem vencedores claros. Os testes com 25% respondentes criativos as estatísticas G e R_n tiveram o melhor desempenho e o índice A apresenta desempenho mais eficaz para casos

com 50% de respondentes criativos em teste longo seguido das estatísticas G e R_n . Nas figuras e tabela sobre os respondentes criativos, notou-se que as quatro porcentagens de respondentes aberrantes e os tamanhos de teste influenciam no desempenho das estatísticas.

Nos casos dos respondentes adivinhos das Figuras 3.8 e 3.9, o desempenho das estatísticas de ajuste individual tem pouco impacto com o aumento das porcentagens de respondentes adivinhos e melhora com o aumento do comprimento de teste. A estatística proposta apresenta um desempenho ruim para todos os casos, e a estatística H^T apresenta uma eficiência global melhor que as demais.

Os respondentes aleatórios nas Figuras 3.10 e 3.11 mostram que a eficiência das estatísticas de ajuste individual melhora com o aumento dos itens e que as quatro porcentagens consideradas não interferem no desempenho das estatísticas. A estatística proposta apresenta o pior desempenho para todas as situações e observa-se que, com aumento dos itens, as estatísticas apresentam áreas sob a curva ROC muito próximas. Nos casos em que há 5% e 10% de respondentes aleatórios, as estatísticas H^T e r_{pbis} tiveram os melhores desempenhos.

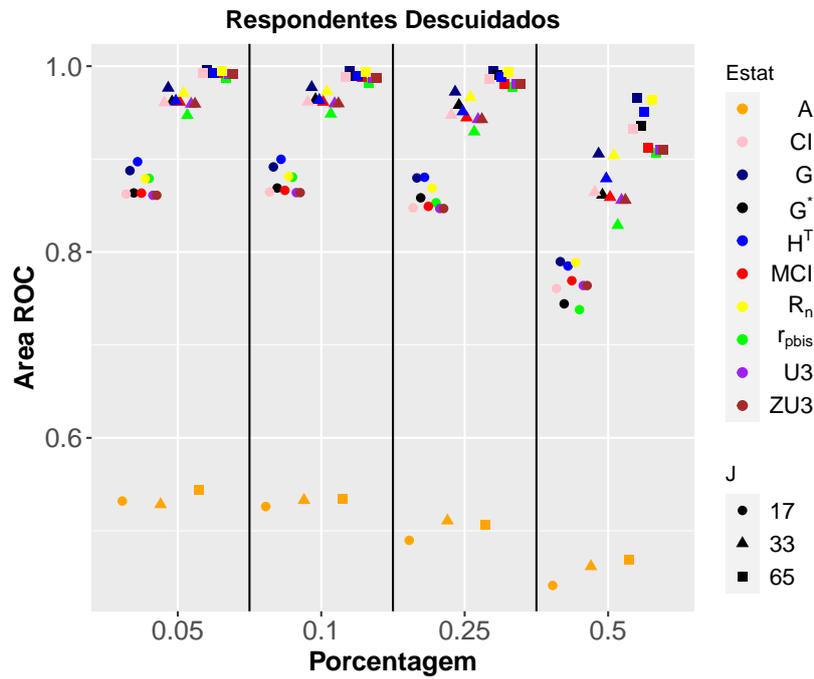


Figura 3.2: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes descuidados sob 4 (quatro) condições de porcentagem.

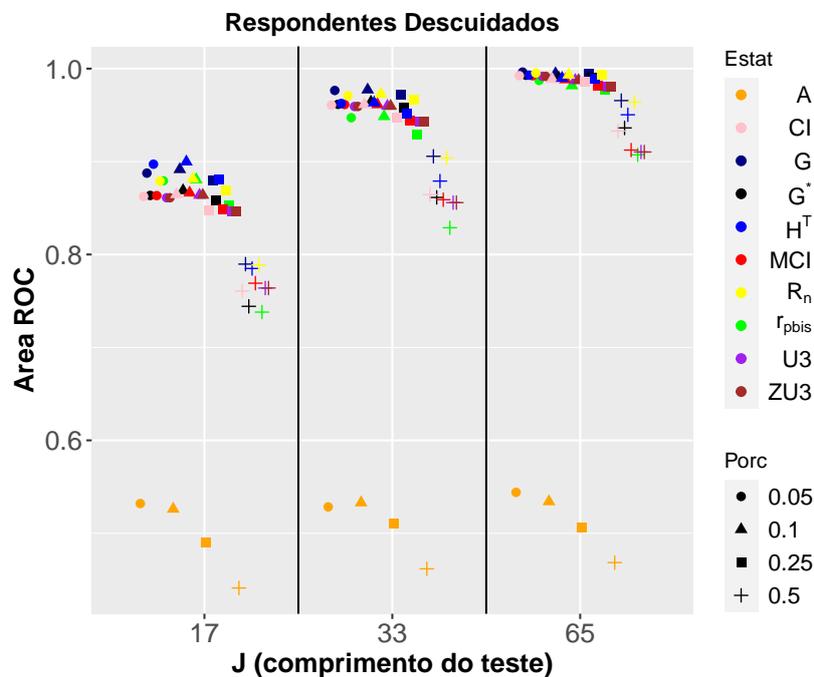


Figura 3.3: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes descuidados sob 3 (três) tamanhos do teste.

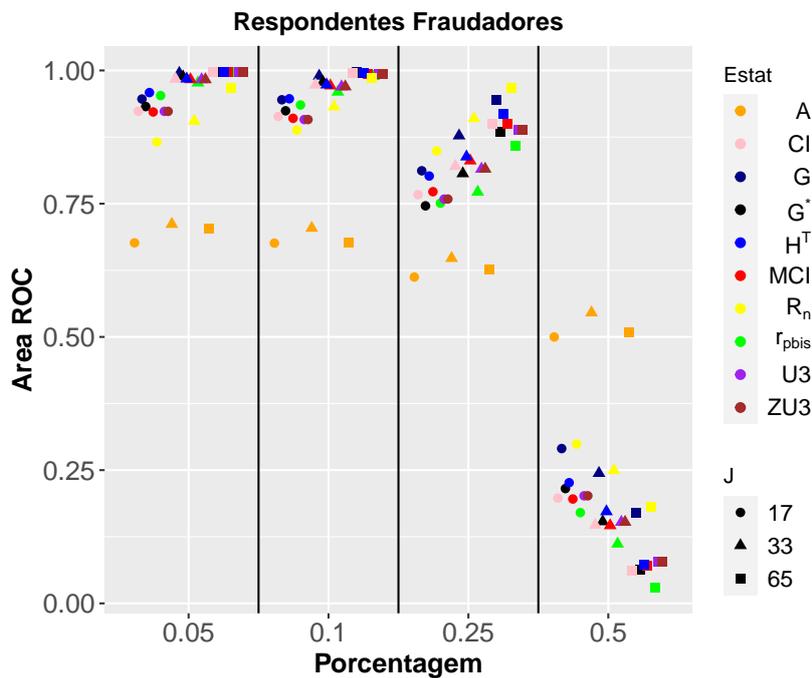


Figura 3.4: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes fraudulentos sob 4 (quatro) condições de porcentagem.

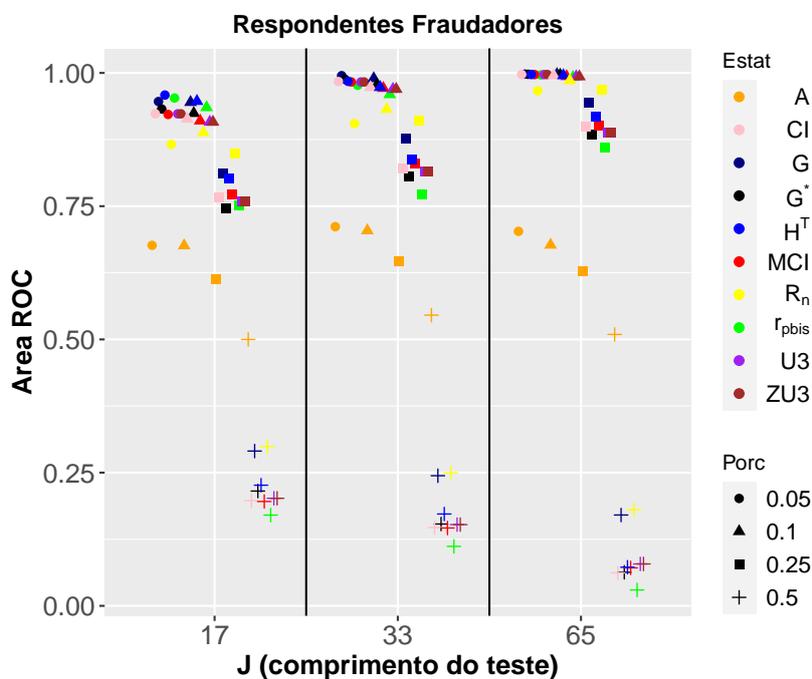


Figura 3.5: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes fraudulentos sob 3 (três) tamanhos do teste.

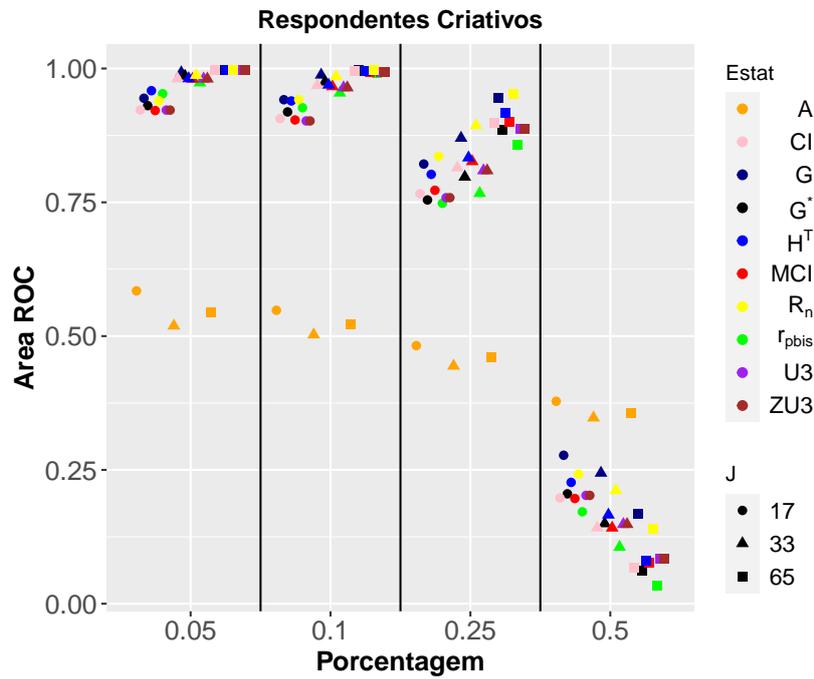


Figura 3.6: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes criativos sob 4 (quatro) condições de porcentagem.

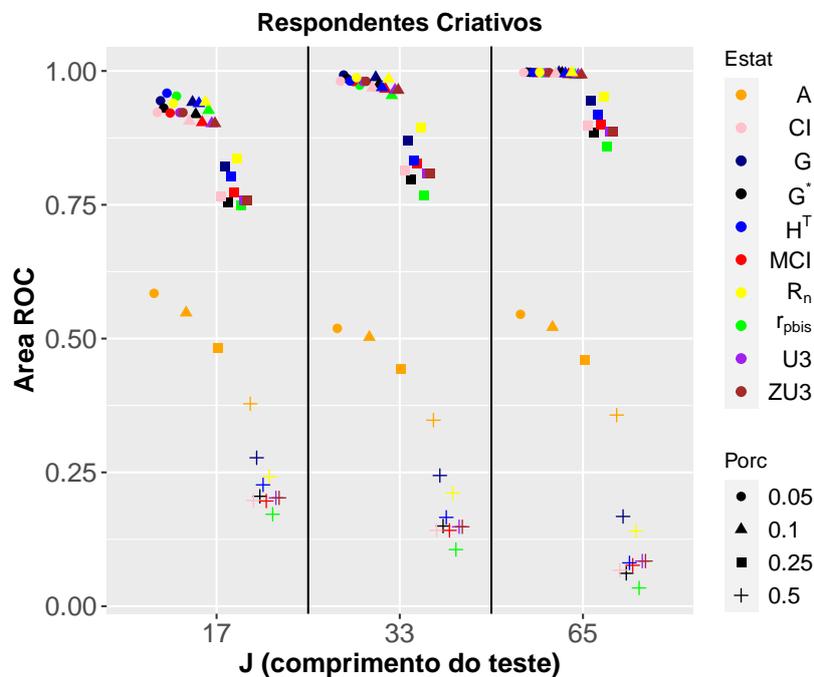


Figura 3.7: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes criativos sob 3 (três) tamanhos do teste.

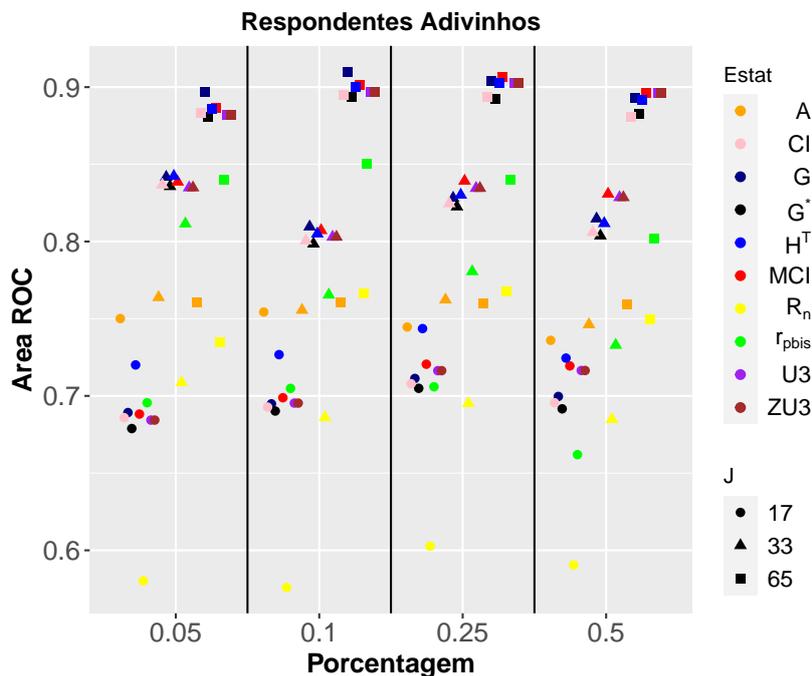


Figura 3.8: Comparação de 10 (dez) estatísticas de ajuste individual, comdados M1PL para respondentes adivinhos sob 4 (quatro) condições de porcentagem.

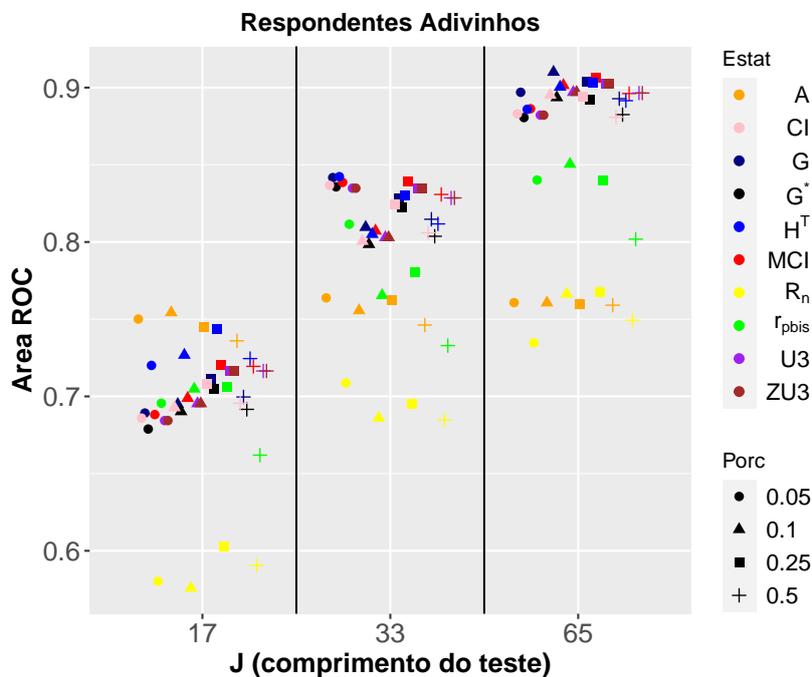


Figura 3.9: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes adivinhos sob 3 (três) tamanhos do teste.

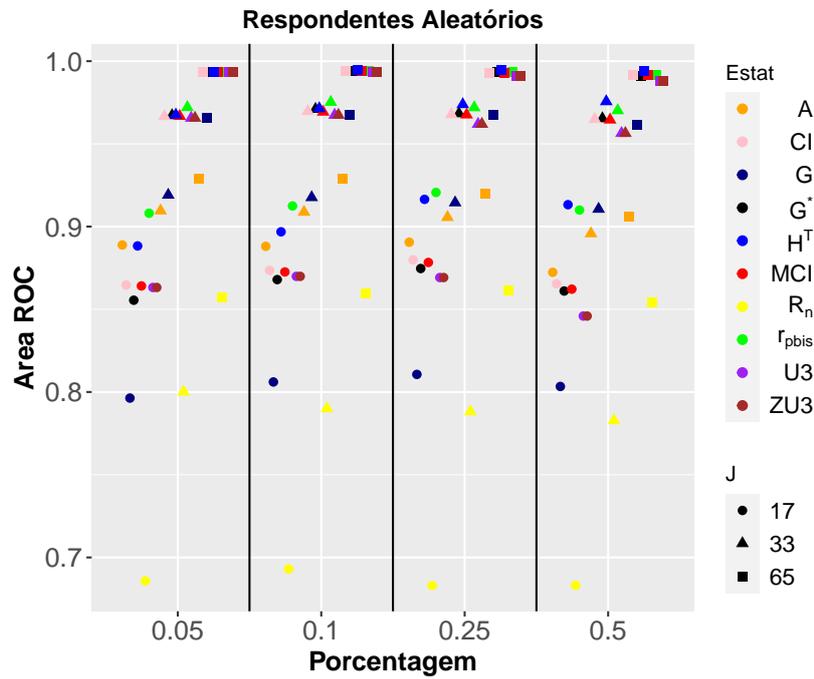


Figura 3.10: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes aleatórios sob 4 (quatro) condições de porcentagem.

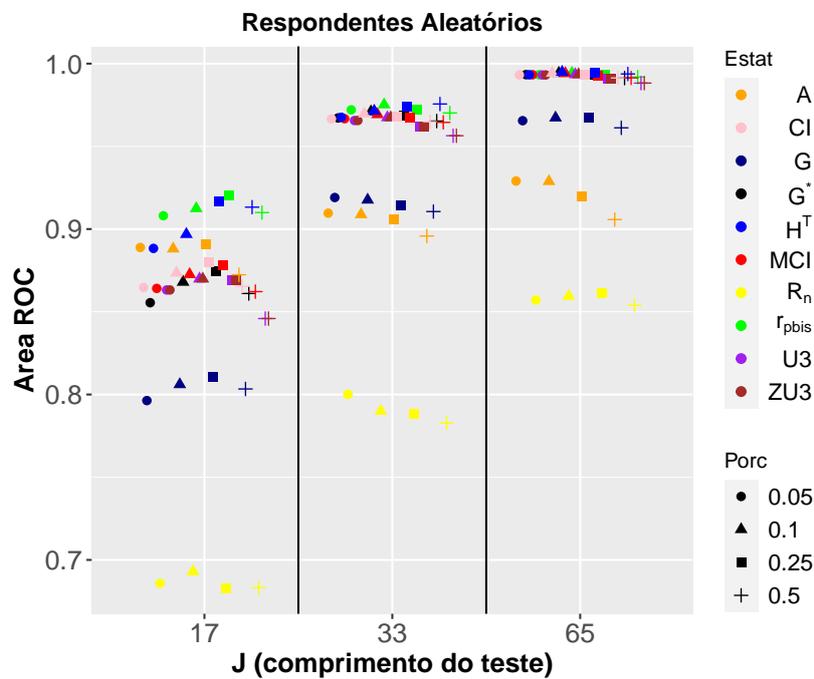


Figura 3.11: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML1P para respondentes aleatórios sob 3 (três) tamanhos do teste.

3.5.2 Simulação para modelos com 2 (dois) parâmetros

A Tabela 3.6 mostra a média e o desvio padrão da área sob a curva ROC de dados com modelo de 2 (dois) parâmetros das estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n , em três comprimentos de teste (17, 33 e 65 itens) e quatro níveis de (5%, 10%, 25% e 50%) para os respondentes com resposta descuidada com uma probabilidade de 0,5 de conter as respostas incorretas para os 41% dos itens fáceis.

Analisando Tabela 3.6, observa-se que as áreas sob a curva ROC aumentam com a duração do teste e não diminuem com as porcentagens de respondentes aberrantes. Os resultados são diferentes dos observados para o ML1P, onde os casos com 50% de respondentes aberrantes eram os mais difíceis de detectar. Nos casos com tamanhos de testes curtos (17 itens) e 5% de respondentes aberrantes, a estatística H^T teve o melhor desempenho e a estatística proposta R_n apresentou a taxa de detecção mais eficaz em relação às estatísticas G^* , CI , MCI , $U3$, $ZU3$ e A . Para os casos de testes médios e longos e com 5% de respondentes aberrantes, as estatísticas G e R_n são mais eficazes.

Nos casos com 10% e 25% de respondentes aberrantes em testes curtos, as estatísticas H^T e G apresentaram os melhores desempenhos e a estatística proposta R_n teve taxa de detecção semelhante às estatísticas H^T e r_{pbis} e melhor que as estatísticas G^* , CI , MCI , $U3$, $ZU3$ e A . Em casos com testes médios e longos, as estatísticas G e R_n obtiveram as maiores áreas sob a curva ROC. Observando a simulação com 50% de respondentes aberrantes em testes curtos, a estatística H^T foi mais eficaz e a estatística R_n apresentou-se melhor do que as estatísticas G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$ e A . Em testes com 33 e 65 itens, as estatísticas G e R_n são melhores na detecção de respondentes descuidados (Tabela 3.6).

Dessa maneira, a estatística H^T apresentou-se mais eficiente para casos em testes curtos e para as 4 (quatro) porcentagens de respondentes aberrantes. As estatísticas G e R_n foram mais eficazes na detecção de examinandos que respondem com descuido, para testes médios e longos, considerando também os 4 (quatro) níveis de respondentes aberrantes (5%, 10%, 25%

e 50%) (Tabela 3.6).

Considerando os respondentes fraudadores, a Tabela 3.7 mostra a média e o desvio padrão da área sob a curva ROC de dados com modelo de 2 (dois) parâmetros das estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n com três comprimentos de teste (17, 33 e 65 itens), 4 (quatro) porcentagens de respondentes aberrantes (5%, 10%, 25% e 50%) e considerando resposta correta para 18% dos itens mais difíceis.

A Tabela 3.7 mostra-se semelhantes aos resultados para ML1P, onde a eficiência das estatísticas aumenta com a duração do teste e diminui com a porcentagem de respondentes aberrantes. Além disso, os casos com 50% de respondentes aberrantes ainda é o mais difícil de detectar.

Os casos com 5% e 10% de respondentes aberrantes em teste curto (17 itens) apresentou-se muito próxima da análise feita para a Tabela 3.2, onde a estatística H^T apresentou o melhor desempenho. Em testes médios (33 itens) a estatística G foi mais eficaz. Para teste longo (65 itens) as estatísticas G e G^* tiveram o melhor desempenho. A estatística proposta R_n mostrou-se melhor que o índice A para todos os tamanhos de testes. Verificando os resultados para 25% de respondentes aberrantes em testes com 17 itens a estatística R_n foi melhor nos três comprimentos de teste, já em casos com 50% de respondentes aberrantes o resultado se tornou similar ao apresentado na simulação para casos com 1 (um) parâmetro, onde o índice A foi mais eficaz para três tamanhos de testes na detecção de examinandos que respondem com fraude.

Com isso, para os respondentes fraudadores a estatística H^T apresentou-se mais eficiente para casos com testes curtos e com porcentagem máxima de 10% de respondentes aberrantes. As estatísticas G e G^* foram mais eficazes para testes médios e longos com porcentagem máxima de 10% de respondente aberrantes, também, para casos com 25% de respondentes aberrantes a estatística R_n foi melhor em todos os tamanhos de testes propostos e, para casos com taxa 50% de respondentes aberrantes o índice A foi que apresentou melhor desempenho nos três tamanhos de testes (Tabela 3.7).

Na Tabela 3.8, observa-se a média e o desvio padrão da área sob a curva ROC de dados com ML2P das estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n

em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com resposta criativa (5%, 10%, 25% e 50%) e considerando resposta incorreta para 18% dos itens fáceis.

Os resultados das áreas sob a curva ROC também mostram-se próximos aos observados na Tabela 3.3 para ML1P, onde o desempenho das estatísticas aumenta com a duração do teste e diminui com a porcentagem de respondentes aberrantes. Nas áreas da Tabela 3.8, nota-se que, para 5% de respondentes com resposta criativa em teste curto, a estatística H^T apresenta o melhor desempenho e a estatística proposta R_n apresentou a taxa de detecção mais eficaz que as estatísticas G , G^* , CI , MCI , $U3$, $ZU3$ e A . Para o caso de teste médio e longo as estatísticas G e R_n são mais eficientes. Sob a condição em que considera 10% de respondentes com resposta criativa em teste curto, as estatísticas H^T e R_n tiveram os melhores desempenhos. Em testes com 33 e 65 itens, as estatísticas R_n e G tiveram as melhores atuações. Sob a condição em que considera 25% de respondentes aberrantes, a estatística R_n teve maior capacidade de detectar padrão de resposta aberrante em todos os tamanhos de teste propostos e para situações com 50% de respondentes aberrantes o índice A foi que apresentou melhor desempenho nos três tamanhos de testes.

Desse modo, para os respondentes do tipo criativo a estatística H^T apresentou mais eficiência para casos em testes curtos e porcentagem máxima de 10% de respondentes aberrantes. As estatísticas G e R_n foram mais eficazes para testes médios e longos com porcentagem máxima de 10% de respondentes, também, para casos com 25% de respondentes aberrantes, a estatística R_n foi mais eficaz em todos os tamanhos de testes propostos e para casos com taxa 50% de respondentes aberrantes o índice A foi que apresentou melhor desempenho nos três comprimentos de teste (Tabela 3.8).

Na Tabela 3.9, observa-se média e desvio padrão da área sob a curva ROC de dados com ML2P das estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com respostas de adivinhação (5%, 10%, 25% e 50%) e considerando uma probabilidade

de 0,25 de conter as resposta correta para os 41% dos itens difíceis.

As estimativas das áreas sob a curva ROC na Tabela 3.9 mostram-se bem próximas às dos dados para ML1P, onde as taxas de detecção das estatísticas aumentam com o aumento dos itens em teste e a porcentagem de respondentes aberrantes não influência no desempenho das mesmas.

Considerando a Tabela 3.9, observa-se que para 5% de respondentes aberrantes em testes com 17 itens a estatística H^T apresentou o melhor desempenho. Para testes médios as estatísticas G e H^T apresentaram mais eficazes e observando os testes longos a estatística G foi melhor que as demais. Nota-se para os casos com 10% de respondentes aberrantes em testes curtos o índice A foi melhor que as demais, em testes médios as estatísticas H^T e G tiveram uma atuação melhor na detecção de respondentes aberrantes com resposta de adivinhação e para teste longo a estatística G é a melhor.

Para casos com 25% de respondentes aberrantes a estatística H^T foi a mais competente para testes curtos, e analisando para os testes de tamanhos médios G e H^T foram melhor e teste com 65 itens a estatística G foi mais eficaz. Em situações com 50% de respondentes aberrantes em teste com 17 itens, o índice A apresentou-se mais eficiente e, para testes médios, a estatística MCI mostrou-se melhor que as demais e, considerando os exames com 65 itens, a estatística G foi a mais eficaz. A estatística proposta R_n apresentou o pior desempenho para testes 17 e 33 itens, já para testes com 65 itens, mostrou-se melhor que o índice A .

Na Tabela 3.10, analisou-se média e desvio padrão da área sob a curva ROC de dados com ML2P para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com respostas aleatórias (5%, 10%, 25% e 50%) e considerando uma probabilidade de 0,25 para cada item em conter a resposta correta.

Os resultados das áreas sob a curva ROC são semelhantes aos resultados apresentados para ML1P, onde a identificação de examinados aberrantes das estatísticas aumenta com a duração do teste e não altera com o aumento da porcentagem de respondentes aberrantes.

Observando os casos com 5% de respondentes aberrantes para os dois tipos de comprimentos de teste (17 e 33 itens) a estatística r_{pbis} foi a mais eficiente, e para teste longo (65 itens) a estatística H^T é melhor. Verificando os resultados para 10% de respondentes aberrantes para os tamanhos de testes curtos e médios as estatísticas r_{pbis} e H^T apresentaram o melhor desempenho. Mas, em teste longo as estatísticas H^T , CI e r_{pbis} foram mais eficazes. Nos casos com 25% de respondentes aberrantes em testes de 17 e 33 itens, as estatísticas r_{pbis} e H^T foram melhores que as demais e, para testes com 65 itens, a estatística H^T foi mais eficaz. Em casos de 50% de respondentes com resposta aleatória, a H^T mostrou-se mais competente para os três comprimentos de teste. E, a estatística proposta R_n apresentou o pior desempenho para três tamanhos de teste com porcentagem máxima de 10% de respondentes aberrantes, já para casos com 50% e 25% de respondentes aberrantes a estatística proposta apresentou o pior desempenho para testes com 17 e 33 itens e para 65 itens mostrou-se melhor que o índice A (Tabela 3.10).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.87107 (0.03049)	0.98698 (0.00966)	0.99748 (0.00074)
	G^*	0.85407 (0.03403)	0.97321 (0.01379)	0.99504 (0.00246)
	r_{pbis}	0.91327 (0.03432)	0.95492 (0.01836)	0.99008 (0.0059)
	CI	0.85665 (0.03307)	0.97141 (0.01411)	0.99508 (0.00238)
	MCI	0.85731 (0.03322)	0.97215 (0.01401)	0.99516 (0.0023)
	$U3$	0.85529 (0.03354)	0.96892 (0.01475)	0.99457 (0.00271)
	$ZU3$	0.85529 (0.03354)	0.96892 (0.01475)	0.99457 (0.00271)
	H^T	0.9378 (0.02632)	0.97358 (0.0129)	0.99541 (0.00215)
	A	0.548 (0.03209)	0.54931 (0.03231)	0.55567 (0.02915)
	R_n	0.86385 (0.03264)	0.98296 (0.01281)	0.9972 (0.00107)
10%	G	0.89128 (0.0192)	0.98771 (0.0054)	0.99696 (0.00104)
	G^*	0.86877 (0.01948)	0.97688 (0.00754)	0.99284 (0.00281)
	r_{pbis}	0.88612 (0.02158)	0.95498 (0.0162)	0.98218 (0.00684)
	CI	0.86524 (0.01884)	0.97352 (0.00929)	0.99208 (0.00332)
	MCI	0.86661 (0.01816)	0.97102 (0.01068)	0.99135 (0.00381)
	$U3$	0.86381 (0.01817)	0.9677 (0.0117)	0.99044 (0.00416)
	$ZU3$	0.86381 (0.01817)	0.9677 (0.0117)	0.99044 (0.00416)
	H^T	0.90674 (0.01967)	0.97662 (0.00823)	0.99272 (0.00304)
	A	0.49554 (0.02895)	0.52345 (0.03046)	0.54504 (0.02187)
	R_n	0.88224 (0.02206)	0.98369 (0.00663)	0.99653 (0.0011)
25%	G	0.88346 (0.01648)	0.97695 (0.00866)	0.99634 (0.00137)
	G^*	0.86192 (0.01798)	0.96002 (0.01153)	0.9921 (0.00248)
	r_{pbis}	0.87416 (0.01554)	0.92648 (0.01469)	0.97442 (0.00514)
	CI	0.85212 (0.01895)	0.94885 (0.01173)	0.98682 (0.00382)
	MCI	0.85493 (0.01875)	0.94387 (0.01269)	0.98031 (0.00428)
	$U3$	0.8511 (0.01835)	0.94121 (0.01323)	0.97827 (0.00433)
	$ZU3$	0.8511 (0.01835)	0.94121 (0.01323)	0.97827 (0.00433)
	H^T	0.90784 (0.01381)	0.95366 (0.01106)	0.99176 (0.00258)
	A	0.50383 (0.01854)	0.51838 (0.02673)	0.51627 (0.01901)
	R_n	0.87048 (0.01735)	0.97151 (0.0097)	0.99601 (0.00112)
50%	G	0.69966 (0.02734)	0.8525 (0.02279)	0.98321 (0.0047)
	G^*	0.63074 (0.02951)	0.77346 (0.03172)	0.95477 (0.00868)
	r_{pbis}	0.64844 (0.03003)	0.73906 (0.02229)	0.90095 (0.01223)
	CI	0.68235 (0.02968)	0.80245 (0.02441)	0.94415 (0.01018)
	MCI	0.71461 (0.02813)	0.82642 (0.02308)	0.92727 (0.01166)
	$U3$	0.70123 (0.03075)	0.81769 (0.02469)	0.92251 (0.01176)
	$ZU3$	0.70123 (0.03075)	0.81769 (0.02469)	0.92251 (0.01176)
	H^T	0.72269 (0.02707)	0.82712 (0.01907)	0.96765 (0.00591)
	A	0.41084 (0.02686)	0.44047 (0.02655)	0.47616 (0.03393)
	R_n	0.70679 (0.02675)	0.85097 (0.01972)	0.97855 (0.00614)

Tabela 3.6: Média e Desvio Padrão da área sob a curva ROC de dados com ML2P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes descuidados (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.90878 (0.01352)	0.99561 (0.002)	0.99787 (9e-05)
	G^*	0.89939 (0.01363)	0.98924 (0.00295)	0.99736 (0.00057)
	r_{pbis}	0.96215 (0.01227)	0.98177 (0.00533)	0.99392 (0.00305)
	CI	0.89752 (0.01642)	0.98799 (0.00355)	0.99656 (0.00126)
	MCI	0.896 (0.01697)	0.98747 (0.00373)	0.99575 (0.00163)
	$U3$	0.89508 (0.01647)	0.98773 (0.00397)	0.99489 (0.00217)
	$ZU3$	0.89508 (0.01647)	0.98773 (0.00397)	0.99489 (0.00217)
	H^T	0.97238 (0.00873)	0.98944 (0.00346)	0.99693 (0.00095)
	A	0.71347 (0.03565)	0.73305 (0.02859)	0.71907 (0.02353)
	R_n	0.87494 (0.02052)	0.94452 (0.0133)	0.98829 (0.00531)
10%	G	0.93704 (0.01288)	0.99246 (0.00283)	0.99741 (0.00082)
	G^*	0.91393 (0.01821)	0.97855 (0.0073)	0.99703 (0.00074)
	r_{pbis}	0.93582 (0.01663)	0.95973 (0.01204)	0.9919 (0.00307)
	CI	0.90925 (0.01775)	0.97293 (0.00824)	0.9961 (0.00152)
	MCI	0.90731 (0.01746)	0.96794 (0.01003)	0.99476 (0.00232)
	$U3$	0.90155 (0.01807)	0.96537 (0.01067)	0.99365 (0.00291)
	$ZU3$	0.90155 (0.01807)	0.96537 (0.01067)	0.99365 (0.00291)
	H^T	0.95448 (0.01165)	0.97542 (0.00729)	0.99654 (0.00116)
	A	0.70117 (0.0281)	0.72057 (0.01893)	0.70739 (0.01487)
	R_n	0.89872 (0.01751)	0.95187 (0.01622)	0.99165 (0.00275)
25%	G	0.84921 (0.0363)	0.9305 (0.0161)	0.95338 (0.00947)
	G^*	0.78671 (0.04338)	0.8613 (0.02536)	0.87322 (0.01931)
	r_{pbis}	0.73932 (0.03101)	0.80213 (0.02455)	0.8307 (0.01906)
	CI	0.7501 (0.03188)	0.8574 (0.01931)	0.89173 (0.01596)
	MCI	0.75659 (0.03171)	0.86458 (0.01809)	0.89983 (0.01504)
	$U3$	0.7362 (0.03305)	0.84625 (0.02078)	0.88449 (0.01681)
	$ZU3$	0.7362 (0.03305)	0.84625 (0.02078)	0.88449 (0.01681)
	H^T	0.79021 (0.02907)	0.87928 (0.01614)	0.91278 (0.01372)
	A	0.60148 (0.03099)	0.66659 (0.02354)	0.65079 (0.02096)
	R_n	0.86983 (0.02367)	0.9424 (0.01058)	0.97184 (0.00668)
50%	G	0.2705 (0.0275)	0.24306 (0.02797)	0.1586 (0.01717)
	G^*	0.20466 (0.0263)	0.15937 (0.02909)	0.06598 (0.01384)
	r_{pbis}	0.17045 (0.01945)	0.10193 (0.01978)	0.04059 (0.00881)
	CI	0.19523 (0.02053)	0.1472 (0.02088)	0.0714 (0.0128)
	MCI	0.19294 (0.02044)	0.14659 (0.02054)	0.07555 (0.01351)
	$U3$	0.19921 (0.02069)	0.15404 (0.02083)	0.08424 (0.01484)
	$ZU3$	0.19921 (0.02069)	0.15404 (0.02083)	0.08424 (0.01484)
	H^T	0.23284 (0.02219)	0.17498 (0.02245)	0.08675 (0.01226)
	A	0.51266 (0.0227)	0.55275 (0.03178)	0.51951 (0.02934)
	R_n	0.30809 (0.03027)	0.2823 (0.02706)	0.18388 (0.01744)

Tabela 3.7: Média e Desvio Padrão da área sob a curva ROC de dados com ML2P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes fraudadores (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.90661 (0.01374)	0.99464 (0.00379)	0.99779 (0.00047)
	G^*	0.89702 (0.01391)	0.98491 (0.00489)	0.99767 (0.00023)
	r_{pbis}	0.96165 (0.01081)	0.9702 (0.00886)	0.99713 (0.00063)
	CI	0.8955 (0.0157)	0.98254 (0.00577)	0.99765 (0.00025)
	MCI	0.89583 (0.01628)	0.98278 (0.0059)	0.99765 (0.00027)
	$U3$	0.89655 (0.01577)	0.98207 (0.00562)	0.99769 (0.00021)
	$ZU3$	0.89655 (0.01577)	0.98207 (0.00562)	0.99769 (0.00021)
	H^T	0.97268 (0.00772)	0.98315 (0.00578)	0.99769 (0.00023)
	A	0.5769 (0.02813)	0.54234 (0.02199)	0.56824 (0.01441)
	R_n	0.90744 (0.01388)	0.99254 (0.00322)	0.99789 (3e-05)
10%	G	0.93028 (0.0126)	0.99235 (0.00258)	0.99761 (5e-04)
	G^*	0.90406 (0.01932)	0.97914 (0.00639)	0.99639 (0.00098)
	r_{pbis}	0.92926 (0.02256)	0.96161 (0.00964)	0.98954 (0.00391)
	CI	0.90103 (0.02002)	0.97522 (0.00657)	0.99577 (0.00131)
	MCI	0.90128 (0.01977)	0.97651 (0.00624)	0.99541 (0.00134)
	$U3$	0.90012 (0.02096)	0.9753 (0.00659)	0.99463 (0.0016)
	$ZU3$	0.90012 (0.02096)	0.9753 (0.00659)	0.99463 (0.0016)
	H^T	0.94952 (0.01566)	0.97557 (0.00644)	0.99612 (0.00119)
	A	0.53833 (0.01977)	0.52397 (0.02084)	0.52805 (0.01598)
	R_n	0.93092 (0.01269)	0.98907 (0.00334)	0.99763 (0.00024)
25%	G	0.82165 (0.0301)	0.93237 (0.01227)	0.96319 (0.01051)
	G^*	0.74911 (0.03847)	0.86208 (0.02305)	0.89085 (0.0272)
	r_{pbis}	0.75424 (0.03135)	0.80908 (0.0223)	0.84758 (0.02411)
	CI	0.77522 (0.02642)	0.86159 (0.01893)	0.90358 (0.01964)
	MCI	0.77879 (0.02587)	0.86989 (0.01842)	0.91259 (0.0175)
	$U3$	0.76851 (0.02657)	0.84782 (0.02015)	0.89473 (0.02003)
	$ZU3$	0.76851 (0.02657)	0.84782 (0.02015)	0.89473 (0.02003)
	H^T	0.80399 (0.02577)	0.8831 (0.01698)	0.9233 (0.01701)
	A	0.48535 (0.02456)	0.46185 (0.02739)	0.45893 (0.02106)
	R_n	0.82716 (0.02882)	0.94831 (0.0095)	0.97159 (0.00921)
50%	G	0.28915 (0.05522)	0.26252 (0.02846)	0.15695 (0.01981)
	G^*	0.2196 (0.05391)	0.17087 (0.02623)	0.06185 (0.01137)
	r_{pbis}	0.19377 (0.02471)	0.12884 (0.01709)	0.04351 (0.00898)
	CI	0.2119 (0.0245)	0.16119 (0.02016)	0.07322 (0.01151)
	MCI	0.20881 (0.02432)	0.15902 (0.02029)	0.07651 (0.01141)
	$U3$	0.21863 (0.02426)	0.16667 (0.02078)	0.08528 (0.01242)
	$ZU3$	0.21863 (0.02426)	0.16667 (0.02078)	0.08528 (0.01242)
	H^T	0.25421 (0.02542)	0.19162 (0.0211)	0.09254 (0.01618)
	A	0.38151 (0.02468)	0.33458 (0.03195)	0.35827 (0.02734)
	R_n	0.28073 (0.04682)	0.26934 (0.02791)	0.14871 (0.01635)

Tabela 3.8: Média e Desvio Padrão da área sob a curva ROC de dados com ML2P para as 10 (dez) estatísticas de ajuste dindividual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes criativos (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.73131 (0.0465)	0.89488 (0.03901)	0.93937 (0.02001)
	G^*	0.71907 (0.04923)	0.88672 (0.02897)	0.90987 (0.02696)
	r_{pbis}	0.76206 (0.0548)	0.85898 (0.02885)	0.85192 (0.03748)
	CI	0.73183 (0.0432)	0.88766 (0.02791)	0.91156 (0.02675)
	MCI	0.7344 (0.04339)	0.88673 (0.02837)	0.91478 (0.02721)
	$U3$	0.72843 (0.04288)	0.88189 (0.02865)	0.90877 (0.02836)
	$ZU3$	0.72843 (0.04288)	0.88189 (0.02865)	0.90877 (0.02836)
	H^T	0.80386 (0.05321)	0.89342 (0.0273)	0.9112 (0.02624)
	A	0.76073 (0.0284)	0.77599 (0.02391)	0.76144 (0.0307)
	R_n	0.5947 (0.05654)	0.74868 (0.05321)	0.79395 (0.03632)
10%	G	0.68046 (0.04456)	0.8783 (0.02314)	0.95975 (0.01408)
	G^*	0.67298 (0.04503)	0.86861 (0.02365)	0.93593 (0.01541)
	r_{pbis}	0.6883 (0.04326)	0.83337 (0.03066)	0.88411 (0.01982)
	CI	0.67406 (0.04254)	0.8703 (0.02431)	0.93961 (0.01518)
	MCI	0.68266 (0.04259)	0.86944 (0.02542)	0.94266 (0.01495)
	$U3$	0.67967 (0.04025)	0.86174 (0.02664)	0.93773 (0.01564)
	$ZU3$	0.67967 (0.04025)	0.86174 (0.02664)	0.93773 (0.01564)
	H^T	0.71696 (0.04627)	0.88686 (0.02452)	0.94164 (0.0152)
	A	0.75934 (0.02617)	0.75763 (0.03316)	0.76739 (0.02829)
	R_n	0.53453 (0.04734)	0.72679 (0.02561)	0.81433 (0.03155)
25%	G	0.74619 (0.02237)	0.85607 (0.02074)	0.95282 (0.01061)
	G^*	0.73477 (0.02156)	0.8463 (0.01997)	0.93028 (0.01383)
	r_{pbis}	0.74336 (0.0222)	0.80509 (0.02232)	0.8666 (0.02004)
	CI	0.7359 (0.02089)	0.84598 (0.02065)	0.92673 (0.01386)
	MCI	0.7471 (0.0199)	0.85678 (0.02089)	0.9348 (0.01206)
	$U3$	0.74199 (0.01981)	0.85094 (0.02126)	0.92896 (0.01356)
	$ZU3$	0.74199 (0.01981)	0.85094 (0.02126)	0.92896 (0.01356)
	H^T	0.78751 (0.01939)	0.85512 (0.02006)	0.94049 (0.0115)
	A	0.75374 (0.02195)	0.76669 (0.01964)	0.76477 (0.02195)
	R_n	0.62978 (0.02873)	0.71266 (0.02705)	0.81064 (0.01899)
50%	G	0.65625 (0.02548)	0.81102 (0.02148)	0.95386 (0.00989)
	G^*	0.64701 (0.02575)	0.79754 (0.01722)	0.93075 (0.01205)
	r_{pbis}	0.60805 (0.02321)	0.70449 (0.01573)	0.83432 (0.02113)
	CI	0.66789 (0.02201)	0.80861 (0.01779)	0.92421 (0.0128)
	MCI	0.70333 (0.02202)	0.84285 (0.01668)	0.92928 (0.01115)
	$U3$	0.69695 (0.0226)	0.83865 (0.01643)	0.92935 (0.01113)
	$ZU3$	0.69695 (0.0226)	0.83865 (0.01643)	0.92935 (0.01113)
	H^T	0.70147 (0.02371)	0.81451 (0.01824)	0.93862 (0.01195)
	A	0.74422 (0.02653)	0.75022 (0.02016)	0.75768 (0.02065)
	R_n	0.54119 (0.03217)	0.66346 (0.02536)	0.83291 (0.022)

Tabela 3.9: Média e Desvio Padrão da área sob a curva ROC de dados com ML2P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes adivinhos (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.82094 (0.0381)	0.91359 (0.0326)	0.98322 (0.01198)
	G^*	0.84906 (0.03067)	0.96321 (0.01761)	0.99599 (0.002)
	r_{pbis}	0.93777 (0.02334)	0.96847 (0.0139)	0.99524 (0.00226)
	CI	0.85913 (0.02933)	0.95977 (0.01993)	0.99608 (0.00189)
	MCI	0.85942 (0.02925)	0.95101 (0.02207)	0.99591 (0.00205)
	$U3$	0.86032 (0.02822)	0.94642 (0.02269)	0.99539 (0.00231)
	$ZU3$	0.86032 (0.02822)	0.94642 (0.02269)	0.99539 (0.00231)
	H^T	0.92723 (0.03263)	0.9661 (0.01711)	0.99625 (0.00189)
	A	0.8772 (0.01785)	0.90923 (0.01775)	0.91995 (0.01628)
	R_n	0.74295 (0.05697)	0.77451 (0.05374)	0.90476 (0.03592)
10%	G	0.84089 (0.03402)	0.94602 (0.02538)	0.98376 (0.00712)
	G^*	0.88439 (0.02577)	0.97777 (0.01778)	0.99557 (0.0014)
	r_{pbis}	0.93767 (0.01927)	0.98252 (0.01257)	0.99484 (0.00189)
	CI	0.89759 (0.01897)	0.97795 (0.01653)	0.99557 (0.00153)
	MCI	0.89931 (0.01867)	0.97588 (0.01746)	0.9953 (0.00181)
	$U3$	0.89969 (0.01857)	0.97139 (0.01853)	0.99445 (0.00221)
	$ZU3$	0.89969 (0.01857)	0.97139 (0.01853)	0.99445 (0.00221)
	H^T	0.92589 (0.0233)	0.98258 (0.01371)	0.99604 (0.00122)
	A	0.90153 (0.01384)	0.90073 (0.01442)	0.92957 (0.00928)
	R_n	0.72864 (0.04431)	0.83445 (0.03858)	0.89863 (0.02372)
25%	G	0.83238 (0.02056)	0.93473 (0.01575)	0.98365 (0.0052)
	G^*	0.87808 (0.01722)	0.9719 (0.00992)	0.99495 (0.00196)
	r_{pbis}	0.93249 (0.00954)	0.97495 (0.00777)	0.99414 (0.002)
	CI	0.88936 (0.01736)	0.97206 (0.00964)	0.99504 (0.0018)
	MCI	0.89287 (0.01841)	0.97223 (0.00977)	0.99507 (0.00179)
	$U3$	0.88833 (0.01928)	0.96849 (0.01085)	0.99404 (0.00231)
	$ZU3$	0.88833 (0.01928)	0.96849 (0.01085)	0.99404 (0.00231)
	H^T	0.93317 (0.0105)	0.97643 (0.00868)	0.99569 (0.00135)
	A	0.88415 (0.01659)	0.89618 (0.0144)	0.9062 (0.0116)
	R_n	0.72532 (0.02235)	0.82852 (0.02992)	0.90905 (0.01514)
50%	G	0.82153 (0.01075)	0.94092 (0.00999)	0.98628 (0.00275)
	G^*	0.86551 (0.01394)	0.97502 (0.00577)	0.9939 (0.00126)
	r_{pbis}	0.92322 (0.00887)	0.97796 (0.00487)	0.99308 (0.00132)
	CI	0.87255 (0.01429)	0.97414 (0.00573)	0.99393 (0.00124)
	MCI	0.87066 (0.01554)	0.97119 (0.00682)	0.99392 (0.00129)
	$U3$	0.85577 (0.01528)	0.96165 (0.00821)	0.9921 (0.00186)
	$ZU3$	0.85577 (0.01528)	0.96165 (0.00821)	0.9921 (0.00186)
	H^T	0.92755 (0.0067)	0.98386 (0.00444)	0.9948 (0.00076)
	A	0.86919 (0.01575)	0.88635 (0.01589)	0.89402 (0.01513)
	R_n	0.70322 (0.01866)	0.8171 (0.02213)	0.91543 (0.01432)

Tabela 3.10: Média e Desvio Padrão da área sob a curva ROC de dados com ML2P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aleatórios (5%, 10%, 25% e 50%).

As figuras a seguir correspondem à comparação das estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n com dados ML2P para respondentes aberrantes com resposta descuidada, resposta com uso de fraudes, resposta criativa, resposta de adivinhação e resposta aleatória. Em cada gráfico, temos um tipo de respondente aberrante com os valores das áreas sob as curvas ROC para cada estatística de ajuste individual no eixo y e porcentagem de respondentes aberrantes ou tamanho de teste no eixo x . Observe que os resultados gráficos são um pouco mais gerais do que os apresentados nas tabelas anteriores e, que pode verificar uma pequena diferença dos resultados apresentados nas tabelas, pois para gerar os gráficos é necessário o agrupamento médio das áreas ROC das estatísticas de ajuste individual.

Para os respondentes descuidados, as Figuras 3.12 e 3.13 observou-se que o comprimento do teste tem uma pequena perturbação na eficiência das estatísticas, quando ocorre o aumento do comprimento do teste as estatísticas tendem a se igualar no desempenho. Em todos os casos, o índice A mostra-se menos eficaz. Para os casos com 33 e 65 itens e considerando as porcentagem de 25% e 50% de respondentes com resposta descuidada, as estatísticas G e R_n têm a mesma sensibilidade que as demais. Para os casos com 17 itens, a estatística H^T tem um destaque no desempenho, considerando as 4 (quatro) porcentagens de respondentes aberrantes. É possível afirmar que quando se considera teste longo (65 itens), as estatísticas de ajuste individual apresentam desempenhos semelhantes para três porcentagem de respondentes descuidados (5%, 10% e 25%) e que a condição de 50% de respondentes aberrantes é mais difícil de detectar respondentes descuidados.

As Figuras 3.14 e 3.15 mostram semelhantes aos resultados para ML2P, onde para os casos com 50% de respondentes fraudadores o índice A é mais eficaz seguidos das estatísticas R_n e G . É possível afirmar também que para os casos com 25% de respondentes descuidados, as estatísticas R_n e G tiveram mais eficácias do que as demais, considerando os três tamanhos de teste. Para as 4 (quatro) porcentagens de respondentes descuidados e comprimento de teste curto, a estatística H^T e r_{pbis} e para comprimento de testes médios e longos a estatística proposta é um pouco menos eficaz. Também é possível confirmar, com bases nos gráficos e tabelas dos

respondentes trapaceiros, que o tamanho do teste e a porcentagem dos respondentes aberrantes influenciam no desempenho das estatísticas.

Nas Figuras 3.16 e 3.17, a taxa de identificação das estatísticas de ajuste individual com o aumento das porcentagens de respondentes criativos, já em relação aos comprimentos de teste, não houve interferência significativa no desempenho da mesma. Para 50% de respondentes criativos, o índice A teve o melhor desempenho seguido das estatísticas R_n e G , considerando o tamanho de teste longo. Para os casos com 25% de respondentes criativos, as estatísticas R_n e G foram mais eficazes para os três comprimentos de teste. Observando os casos com 5% e 10% de respondentes fraudadores, a estatística proposta apresenta progresso no desempenho com o aumento do tamanho do teste.

Para os respondentes adivinhos nas Figuras 3.18 e 3.19, o desempenho das estatísticas de ajuste individual tem pouco impacto com o aumento das porcentagens, mas para o tamanho de teste o desempenho aumenta com o aumento dos itens em teste. A estatística proposta apresenta um desempenho ruim para todos os casos, e a estatística G apresenta uma eficiência melhor que as demais para teste longos. Para testes curtos, a estatística H^T teve um desempenho mais eficaz, considerando as porcentagens de 5% e 25% de respondente adivinhos e, além disso, nos casos com 10% e 50% de respondentes adivinhos, o índice A é o mais sensível.

Nos casos com respondentes aleatórios das Figuras 3.20 e 3.21, notam-se que as condições de porcentagem têm pouco impacto nas áreas sob a ROC das estatísticas, já quando aumenta os itens de teste, as áreas ROC das estatísticas aumentam. Considerando os casos com 33 e 65 itens, as estatísticas têm desempenho muito similar. Para teste com 17 itens, as estatísticas H^T e r_{pbis} tiveram melhor desempenho. A estatística R_n teve a pior taxa de identificação em todos os casos.

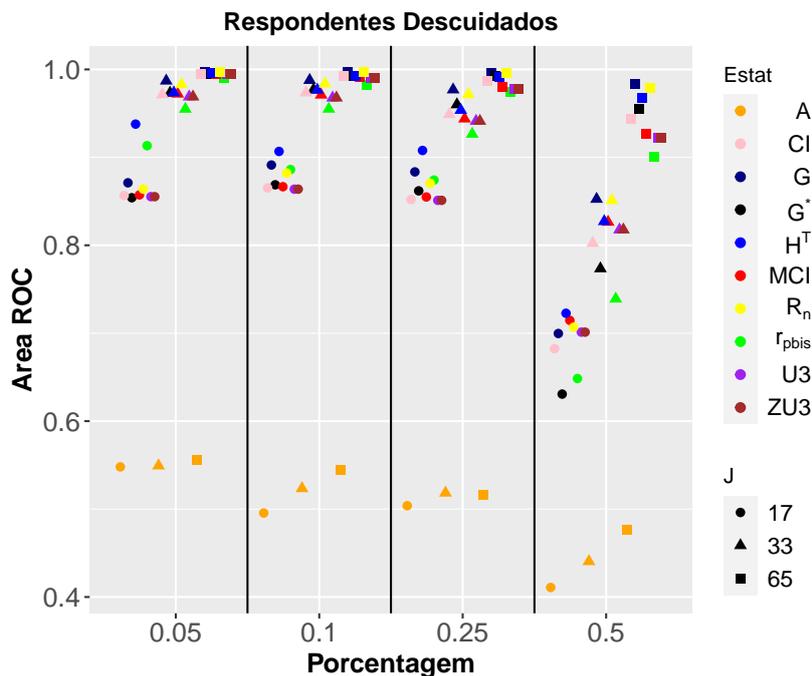


Figura 3.12: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes descuidados sob 4 (quatro) condições de porcentagem.

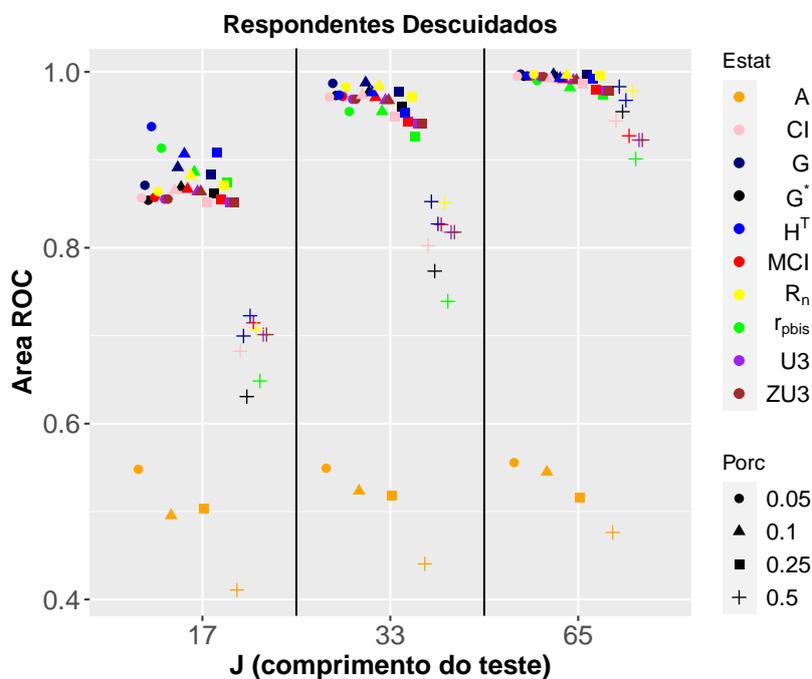


Figura 3.13: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes descuidados sob 3 (três) tamanhos do teste.

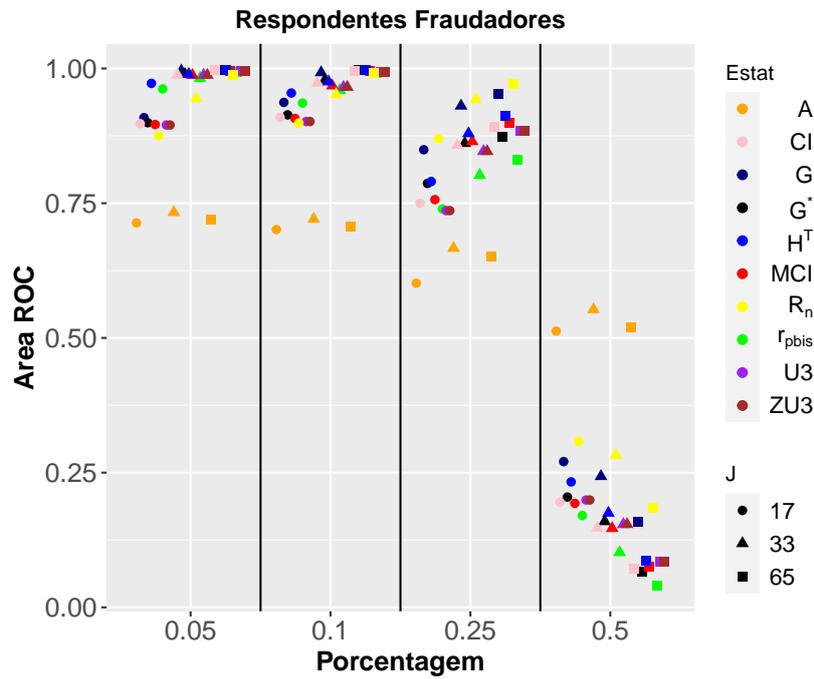


Figura 3.14: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes fraudulentos sob 4 (quatro) condições de porcentagem.

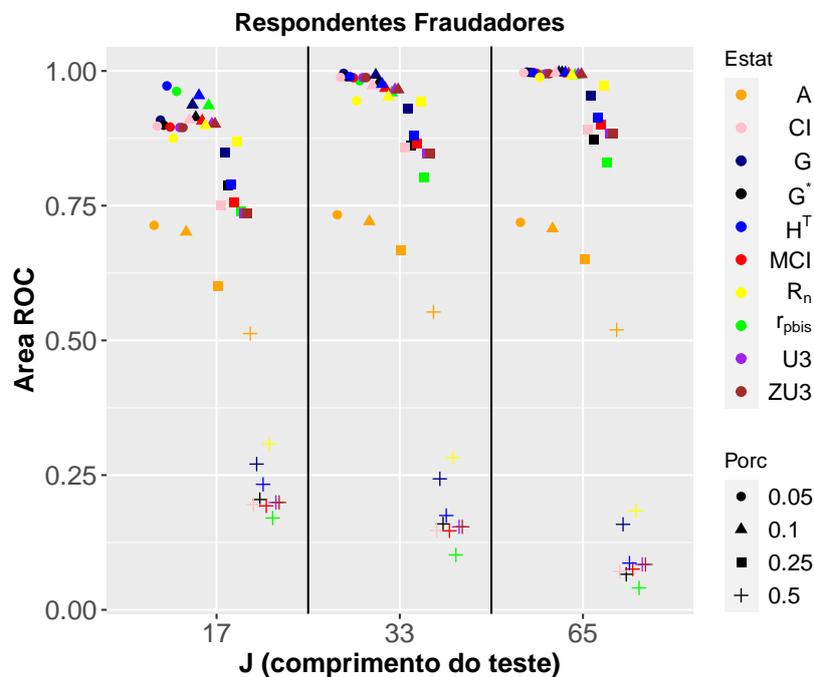


Figura 3.15: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes fraudulentos sob 3 (três) tamanhos do teste.

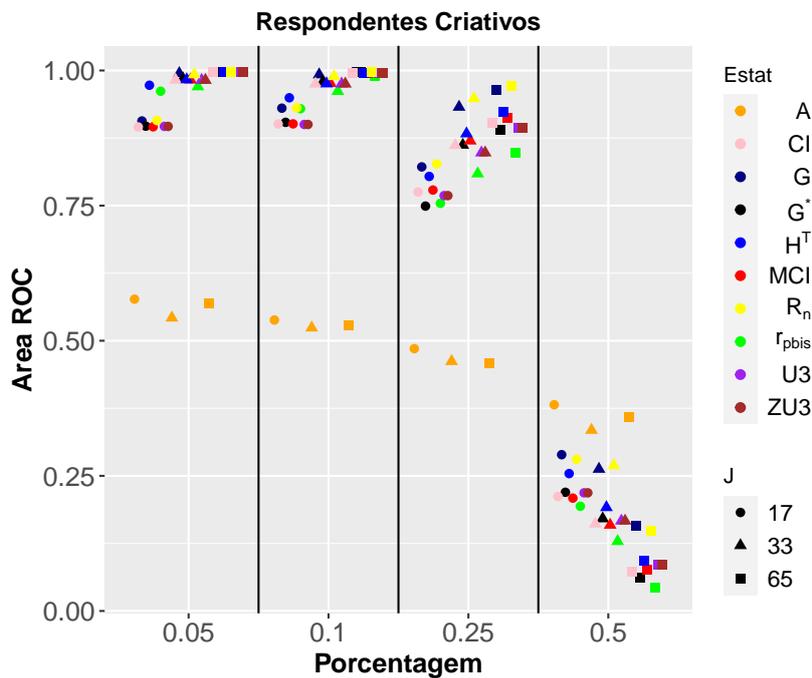


Figura 3.16: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes criativos sob 4 (quatro) condições de porcentagem.

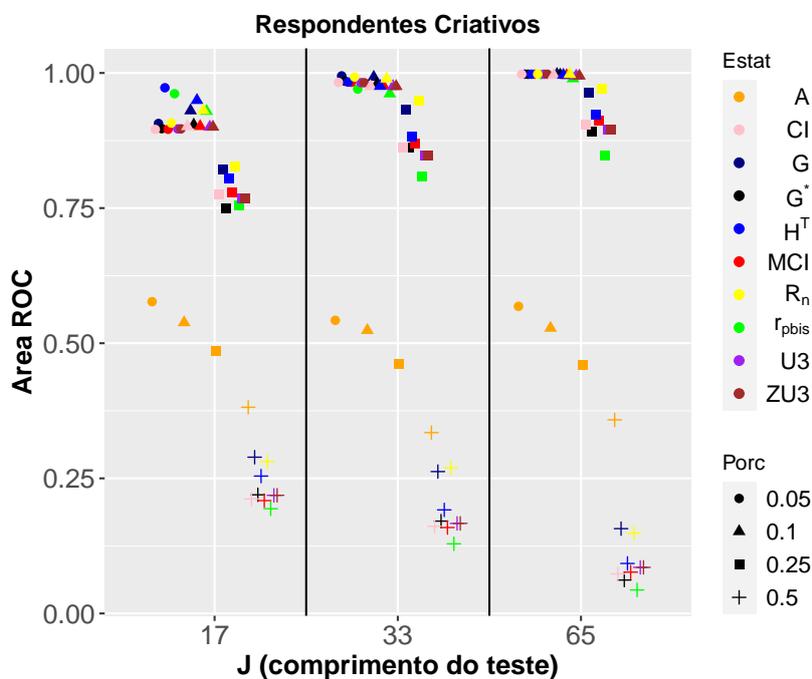


Figura 3.17: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes criativos sob 3 (três) tamanhos do teste.

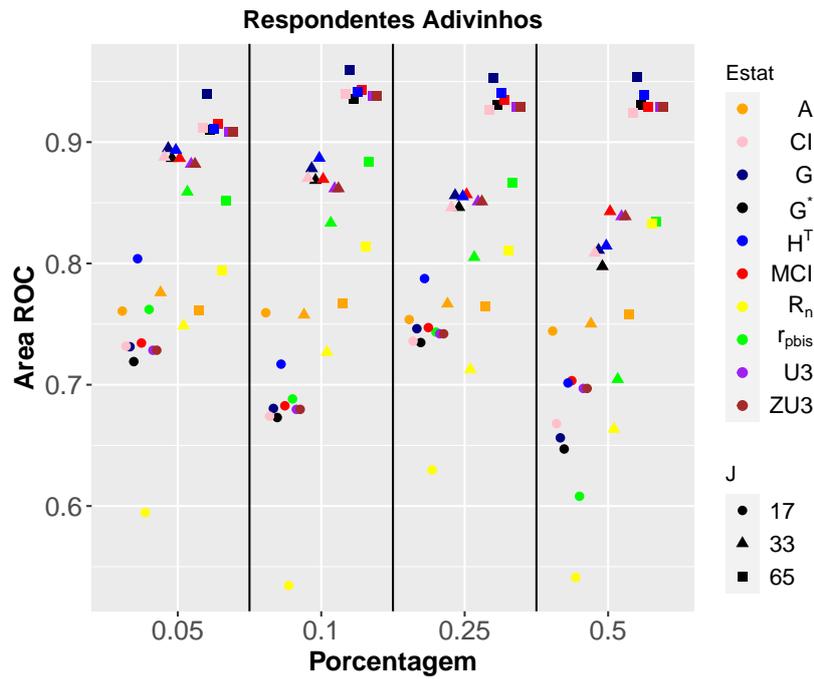


Figura 3.18: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes adivinhos sob 4 (quatro) condições de porcentagem.

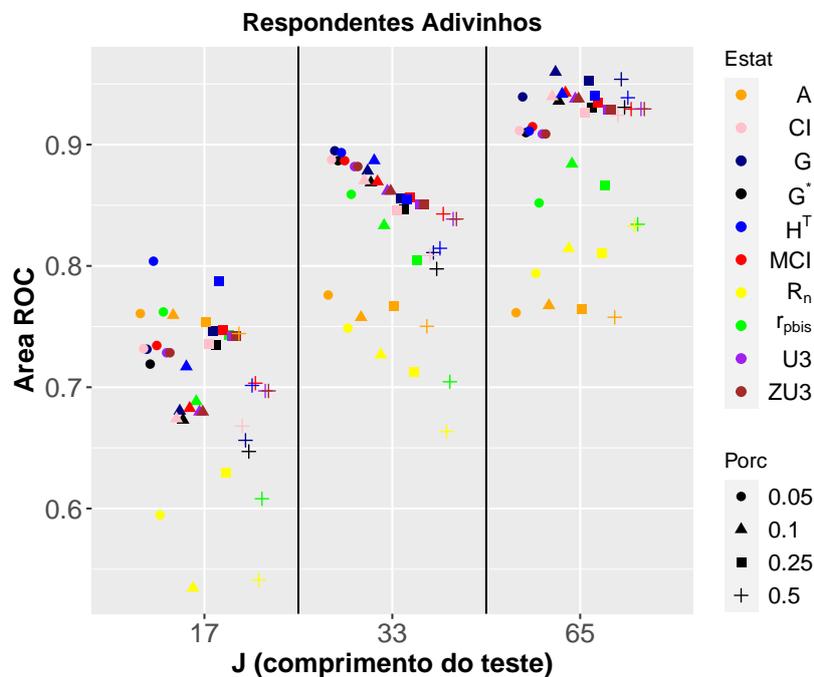


Figura 3.19: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes adivinhos sob 3 (três) tamanhos do teste.

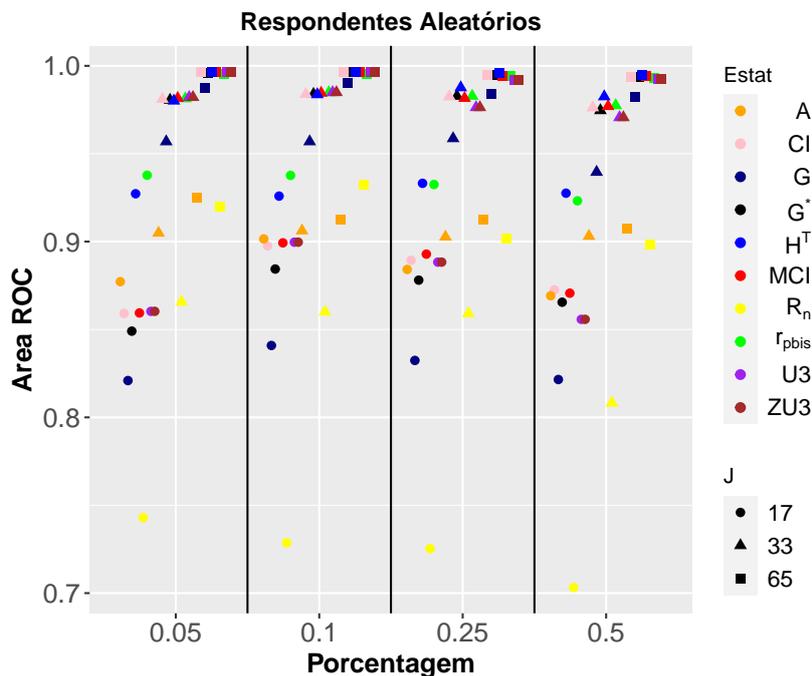


Figura 3.20: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes aleatórios sob 4 (quatro) condições de porcentagem.

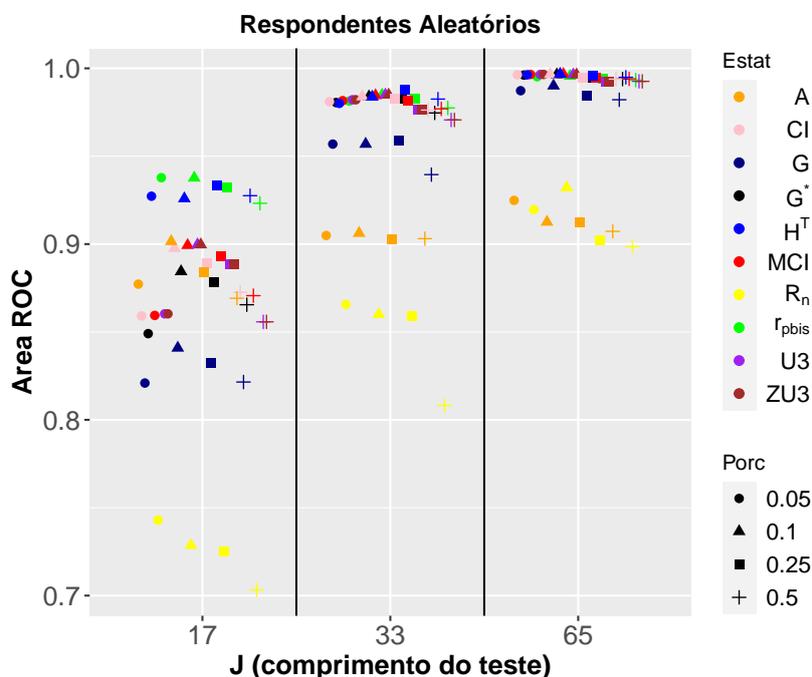


Figura 3.21: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML2P para respondentes aleatórios sob 3 (três) tamanhos do teste.

3.5.3 Simulação para modelos com 3 (três) parâmetros

A Tabela 3.11 mostra média e desvio padrão da área sob a curva ROC de dados com modelo de 3 (três) parâmetros para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e quatro níveis de respondentes descuidados (5%, 10%, 25% e 50%) com uma probabilidade de 0,5 de conter as resposta incorreta para os 41% dos itens fáceis.

As estimativas das áreas sob a curva ROC, na Tabela 3.11, aumentam com a duração do teste e diminuem com a porcentagem de respondente aberrante. E, observa-se que a condição de 50% é um pouco mais difícil de detectar os respondentes aberrantes. Para caso com 5% de respondentes aberrantes e tamanho de teste curto a estatística H^T apresentou mais sensível. Para o caso de teste médio e longo não houve diferença significativas entre as estatísticas. A estatística proposta R_n para testes curtos e longos é somente melhor que o índice A , mas para teste médio é melhor que as estatísticas r_{pbis} e A .

Os dados com 10% de respondentes aberrantes e 17 itens a estatística H^T teve a melhor atuação, a estatística proposta apresentou melhor desempenho que as estatística CI , MCI e A . Já para o caso com 33 itens G foi mais eficaz e a estatística R_n foi mais eficaz que as estatísticas r_{pbis} , CI , MCI , H^T e A . Para casos com 65 itens a estatística G foi a mais sensível, já a estatística proposta foi mais eficaz que as estatísticas CI e A . Em relação aos casos com 25% de respondentes aberrantes as estatísticas H^T , G e G^* foram melhores para testes curtos e a estatística proposta apresentou melhor que as estatística r_{pbis} , CI , MCI , H^T e A . Considerando os testes médios as estatísticas G e R_n tiveram melhor desempenho que as demais e para teste longo a estatística G teve a maior média das áreas sob a curva ROC, a estatística proposta foi mais eficaz que r_{pbis} , CI , MCI , H^T e A . Para casos com 50% de respondentes aberrantes, as estatísticas H^T e R_n são mais eficazes para comprimentos de testes curtos. Nos testes com 33 e 65 itens, a estatística proposta foi mais sensível que as demais na identificação de respondentes com resposta descuidada (Tabela 3.11).

Assim, na Tabela 3.11 estatística H^T apresentou mais eficiente para casos com testes curtos e as quatro porcentagens proposta neste trabalho para respondentes aberrantes. A estatística G foi mais eficaz na detecção de examinandos que respondem com descuido considerando testes médios e longos e 3 níveis de respondentes aberrantes (5%, 10% e 25%). A estatística proposta na maioria dos casos considerando os três tamanhos de teste e 5%, 10% e 25% respondentes aberrantes teve desempenho melhor que r_{pbis} , CI , MCI e A e para situações considerando 50% respondentes aberrantes a estatística R_n teve melhor desempenho para três comprimentos de teste (Tabela 3.11).

Para os fraudadores a Tabela 3.12 mostra a média e desvio padrão da área sob a curva ROC de dados com ML3P para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n com três comprimentos de teste (17, 33 e 65 itens), 4 (quatro) porcentagens de respondentes aberrantes (5%, 10%, 25% e 50%) e considerando resposta correta para 18% dos itens mais difíceis.

As curvas ROC na Tabela 3.12 mostram que o poder de detecção das estatísticas são semelhante aos resultados para ML2P para respondentes fraudadores, que aumentam com a duração do teste e diminuem com a porcentagem de respondentes aberrantes, por isso a condição de 50% é a mais difícil de detectar os respondentes aberrantes.

Na Tabela 3.12 para 5% de respondentes aberrantes com tamanho de teste de 17 itens a estatística H^T apresentou o melhor desempenho e a estatística proposta apresentou desempenho melhor que CI , MCI , $U3$ e A . Para o caso de teste médio (33) e longo (65 itens) a estatística G , H^T e G^* são mais eficazes. A estatística proposta R_n apresentou-se melhor que o índice A para testes curtos e melhor que as estatísticas MCI , $U3$ e A em teste longo. Para o conjunto de dados com 10% de respondentes aberrantes e testes curtos as estatísticas G , R_n e H^T foram mais eficientes e, para testes médios e longos, as estatísticas G , G^* , R_n e H^T tiveram o melhor desempenho.

Os casos da Tabela 3.12 com 25% de respondentes aberrantes e para os três comprimentos de teste (17, 33 e 65 itens) a estatística R_n foi mais eficaz na detecção de examinandos que

respondem com fraude. Já para casos com 50% de respondentes aberrantes do tipo fraudadores o índice A foi que apresentou melhor desempenho nos três tamanhos de teste, resultado este semelhante aos da simulação da Tabela 3.2 e 3.7.

Considerando os respondentes fraudadores a estatística H^T apresentou mais eficiente para casos com testes curtos e com porcentagem máxima de 5% de respondentes aberrantes e as estatísticas G , G^* e H^T foram mais eficazes para testes médios e longos. Para os casos com 10 % de respondentes aberrantes, as estatísticas G , R_n e G^* foram mais eficazes para os três tamanhos de teste e, também, para casos com 25% de respondentes aberrantes, a estatística R_n foi mais eficaz em todos os tamanhos de teste propostos e para casos com taxa 50% de respondentes aberrantes estatística A foi que apresentou melhor desempenho nos três tamanhos de teste.

Na Tabela 3.13 observa-se a média e desvio padrão da área sob a curva ROC de dados com ML3P para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com resposta criativa (5%, 10%, 25% e 50%) e considerando resposta incorreta para 18% dos itens fáceis.

Para a Tabela 3.13 as curvas ROC também mostram-se igual a simulação de dados de ML1P, onde os valores aumentam com a duração do teste e diminuem com a porcentagem de respondentes aberrantes.

Os resultados da Tabela 3.13, mostra que os casos com 5% de respondentes com resposta criativa em tamanho de testes curtos as estatísticas H^T e r_{pbis} apresentaram mais eficientes e a estatística proposta R_n apresentou a taxa de detecção semelhante à estatística G e melhor que as estatísticas G^* , CI , MCI , $U3$, $ZU3$ e A . Já as estatísticas G e G^* foram que apresentaram o melhor desempenho para os testes médios e longos e a estatística R_n apresentou a taxa de detecção semelhante r_{pbis} , CI , MCI , $U3$ e H^T e melhor que o índice A para testes médios e para teste longos semelhante r_{pbis} , CI e H^T e melhor que o índice A . Os casos com 10 % de respondentes com resposta criativa e considerando os testes curtos as estatísticas H^T e R_n

apresentaram média da área superior que as demais estatísticas. Para os testes com 33 e 65 itens as estatísticas G , G^* e R_n foram as mais eficazes. Sob a condição em que considera 25% de respondentes aberrantes, a estatística R_n apresenta área da curva ROC maior em todos os tamanhos de teste propostos e, para situações com taxa 50% de respondentes aberrantes, a estatística A foi a que apresentou melhor desempenho nos três tamanhos de teste e a estatística proposta R_n apresentou a taxa de detecção melhor que as estatísticas G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$ e H^T .

Na Tabela 3.14, observa-se a média e desvio padrão da área sob a curva ROC de dados com ML3P para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com respostas de adivinhação (5%, 10%, 25% e 50%) e considerando uma probabilidade de 0,25 de conter a resposta correta para os 41% dos itens difíceis.

As áreas sob a curva ROC, na Tabela 3.14, mostram-se diferentes das demais. Note que o poder de identificação de padrão aberrante das estatísticas aumenta quando aumenta a quantidade de itens do teste e não com a porcentagem de respondentes aberrantes.

Considerando a Tabela 3.14 observa-se que para as quatro porcentagens de respondentes aberrantes e três comprimentos de teste o índice A apresentou o melhor desempenho. A estatística proposta R_n para todas as situações apresentou-se melhor que a estatística r_{pbis} e para casos com testes médios e com 5% de aberrantes a estatística R_n teve desempenho semelhante H^T e, também, para casos com 10% aberrantes a estatística teve desempenho similar as estatísticas H^T e MCI . Para o conjunto de dados com testes médios e longos e 25% de aberrantes, a estatística proposta R_n teve desempenho semelhante à estatística G^* e, para casos com 50% de aberrantes em teste longo, também, teve desempenho semelhante à estatística G^* .

Para a Tabela 3.15 analisou-se média e desvio padrão da área sob a curva ROC de dados com ML3P para as estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aberrantes com respostas aleatórias (5%, 10%, 25% e 50%) e considerando uma probabilidade

de 0,25 para cada item em conter a resposta correta.

Na Tabela 3.15 as áreas sob a curva ROC mostram igual para respondentes com resposta de adivinhação e diferente das demais, e semelhante os dados para ML1P e ML2P. Note que as estimativas aumentam com a duração do teste e não com a porcentagem de respondentes aberrantes.

Observando a Tabela 3.15 e considerando as quatro porcentagens de respondentes aberrante e os três comprimentos de teste o índice A foi o mais eficiente. A estatística proposta R_n apresentou o pior desempenho para três tamanhos de teste e para as 4 (quatro) porcentagens de respondentes aberrantes.

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.8683 (0.0304)	0.9579 (0.01439)	0.99073 (0.00385)
	G^*	0.87033 (0.03357)	0.95579 (0.01427)	0.99092 (0.00403)
	r_{pbis}	0.89983 (0.03166)	0.94505 (0.01754)	0.98923 (0.00541)
	CI	0.86827 (0.03471)	0.95052 (0.01596)	0.99091 (0.00397)
	MCI	0.87061 (0.03457)	0.95362 (0.01607)	0.9915 (0.00381)
	$U3$	0.8731 (0.03274)	0.95549 (0.01538)	0.99014 (0.00428)
	$ZU3$	0.8731 (0.03274)	0.95549 (0.01538)	0.99014 (0.00428)
	H^T	0.90976 (0.03065)	0.95364 (0.01486)	0.99184 (0.00348)
	A	0.56482 (0.02581)	0.58448 (0.03375)	0.59044 (0.03009)
	R_n	0.85923 (0.03226)	0.95016 (0.01797)	0.98593 (0.00783)
10%	G	0.85736 (0.02163)	0.95825 (0.01103)	0.98582 (0.0051)
	G^*	0.85612 (0.01897)	0.95603 (0.01108)	0.98485 (0.00606)
	r_{pbis}	0.86433 (0.0194)	0.93736 (0.0119)	0.97825 (0.00879)
	CI	0.84498 (0.02068)	0.94675 (0.01013)	0.98245 (0.00703)
	MCI	0.84877 (0.02048)	0.95287 (0.01115)	0.98541 (0.00687)
	$U3$	0.85699 (0.02063)	0.95649 (0.01048)	0.98532 (0.00629)
	$ZU3$	0.85699 (0.02063)	0.95649 (0.01048)	0.98532 (0.00629)
	H^T	0.87367 (0.01714)	0.94902 (0.00951)	0.98368 (0.00667)
	A	0.53231 (0.03097)	0.55723 (0.02851)	0.57537 (0.02306)
	R_n	0.8505 (0.02413)	0.95456 (0.01344)	0.98317 (0.00677)
25%	G	0.84043 (0.01959)	0.934 (0.01336)	0.97976 (0.00567)
	G^*	0.84017 (0.02003)	0.928 (0.01494)	0.97702 (0.00729)
	r_{pbis}	0.83092 (0.01987)	0.8994 (0.01855)	0.96088 (0.01067)
	CI	0.81748 (0.01919)	0.90941 (0.01733)	0.9689 (0.00877)
	MCI	0.82482 (0.01965)	0.91417 (0.01776)	0.97401 (0.00712)
	$U3$	0.83245 (0.01864)	0.92061 (0.01752)	0.97788 (0.00648)
	$ZU3$	0.83245 (0.01864)	0.92061 (0.01752)	0.97788 (0.00648)
	H^T	0.8463 (0.01962)	0.91296 (0.01649)	0.97171 (0.00824)
	A	0.53626 (0.02286)	0.56726 (0.02683)	0.55345 (0.01853)
	R_n	0.83433 (0.01536)	0.93208 (0.01457)	0.97613 (0.00672)
50%	G	0.53678 (0.03829)	0.61276 (0.03)	0.80606 (0.02511)
	G^*	0.52612 (0.03799)	0.58316 (0.03908)	0.79216 (0.02183)
	r_{pbis}	0.54897 (0.02971)	0.59577 (0.03875)	0.77723 (0.02168)
	CI	0.53747 (0.02928)	0.60963 (0.0423)	0.79412 (0.02199)
	MCI	0.55557 (0.03324)	0.63054 (0.0515)	0.81406 (0.01977)
	$U3$	0.52875 (0.03549)	0.60379 (0.05297)	0.80025 (0.0215)
	$ZU3$	0.52875 (0.03549)	0.60379 (0.05297)	0.80025 (0.0215)
	H^T	0.5823 (0.02866)	0.63973 (0.03607)	0.81857 (0.02093)
	A	0.44268 (0.02853)	0.48335 (0.02463)	0.51343 (0.03142)
	R_n	0.58131 (0.03782)	0.6617 (0.02443)	0.84338 (0.02409)

Tabela 3.11: Média e Desvio Padrão da área sob a curva ROC de dados com ML3P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes descuidados (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.89091 (0.02282)	0.95692 (0.01323)	0.98651 (0.00584)
	G^*	0.88169 (0.02475)	0.95254 (0.01385)	0.98435 (0.00558)
	r_{pbis}	0.90263 (0.02239)	0.94267 (0.01714)	0.97623 (0.00872)
	CI	0.87121 (0.0239)	0.94654 (0.01607)	0.97798 (0.00765)
	MCI	0.86701 (0.02634)	0.9419 (0.01809)	0.97063 (0.009)
	$U3$	0.84851 (0.02778)	0.9266 (0.02161)	0.95237 (0.01251)
	$ZU3$	0.84851 (0.02778)	0.9266 (0.02161)	0.95237 (0.01251)
	H^T	0.91477 (0.01707)	0.95279 (0.01369)	0.98246 (0.00602)
	A	0.65392 (0.0333)	0.70276 (0.0385)	0.68615 (0.02324)
	R_n	0.87456 (0.02119)	0.92324 (0.01975)	0.97744 (0.00699)
10%	G	0.88998 (0.02202)	0.94003 (0.01979)	0.9839 (0.0054)
	G^*	0.8671 (0.02761)	0.92791 (0.02155)	0.98209 (0.00514)
	r_{pbis}	0.86521 (0.01995)	0.91048 (0.02397)	0.97214 (0.00777)
	CI	0.85855 (0.01964)	0.91908 (0.02283)	0.97447 (0.00704)
	MCI	0.8523 (0.02134)	0.90809 (0.02471)	0.96419 (0.00898)
	$U3$	0.8401 (0.02143)	0.89312 (0.02563)	0.94689 (0.01091)
	$ZU3$	0.8401 (0.02143)	0.89312 (0.02563)	0.94689 (0.01091)
	H^T	0.88489 (0.01657)	0.92696 (0.01993)	0.98031 (0.00575)
	A	0.6574 (0.02892)	0.69184 (0.02812)	0.66887 (0.01951)
	R_n	0.88754 (0.01645)	0.9273 (0.02436)	0.98231 (0.00662)
25%	G	0.76468 (0.03708)	0.84083 (0.02398)	0.86876 (0.0181)
	G^*	0.71559 (0.04466)	0.77873 (0.03207)	0.79458 (0.02555)
	r_{pbis}	0.67521 (0.04184)	0.73866 (0.03575)	0.77272 (0.02172)
	CI	0.68225 (0.0438)	0.78011 (0.03407)	0.81181 (0.01883)
	MCI	0.68092 (0.04521)	0.78159 (0.03276)	0.80188 (0.01974)
	$U3$	0.68493 (0.04398)	0.78396 (0.0305)	0.80535 (0.0193)
	$ZU3$	0.68493 (0.04398)	0.78396 (0.0305)	0.80535 (0.0193)
	H^T	0.70691 (0.04059)	0.79521 (0.03111)	0.83869 (0.01687)
	A	0.58407 (0.02819)	0.64848 (0.02522)	0.63294 (0.02081)
	R_n	0.79712 (0.03348)	0.87789 (0.01983)	0.90744 (0.02076)
50%	G	0.33826 (0.02992)	0.34842 (0.03202)	0.23787 (0.0318)
	G^*	0.25382 (0.02631)	0.22173 (0.02912)	0.10304 (0.01625)
	r_{pbis}	0.23972 (0.02265)	0.1869 (0.01996)	0.09276 (0.01179)
	CI	0.2553 (0.02236)	0.21512 (0.02469)	0.1068 (0.0151)
	MCI	0.25767 (0.02208)	0.21464 (0.02683)	0.11061 (0.01651)
	$U3$	0.27289 (0.02413)	0.23094 (0.0284)	0.12345 (0.01766)
	$ZU3$	0.27289 (0.02413)	0.23094 (0.0284)	0.12345 (0.01766)
	H^T	0.27323 (0.02459)	0.23178 (0.02598)	0.12262 (0.01739)
	A	0.53623 (0.01676)	0.5737 (0.03284)	0.54506 (0.02657)
	R_n	0.31858 (0.02436)	0.30329 (0.02532)	0.19248 (0.02741)

Tabela 3.12: Média e Desvio Padrão da área sob a curva ROC de dados com ML3P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes fraudadores (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.92343158 (0.0163)	0.97563158 (0.00734)	0.99715579 (0.00077)
	G^*	0.91218105 (0.0184)	0.97252842 (0.00893)	0.99638947 (0.00123)
	r_{pbis}	0.94161895 (0.01648)	0.96193053 (0.01215)	0.99572 (0.00147)
	CI	0.90505474 (0.01919)	0.96523368 (0.01065)	0.99589053 (0.00138)
	MCI	0.90646105 (0.019)	0.96406737 (0.01132)	0.99608842 (0.00137)
	$U3$	0.91383789 (0.01675)	0.96856421 (0.00875)	0.99651789 (0.00114)
	$ZU3$	0.91383789 (0.01675)	0.96856421 (0.00875)	0.99651789 (0.00114)
	H^T	0.94526737 (0.01446)	0.96716632 (0.0098)	0.99598947 (0.00137)
	A	0.58913474 (0.02599)	0.55994316 (0.02683)	0.58820421 (0.02031)
	R_n	0.92050947 (0.01763)	0.96215158 (0.00897)	0.99548 (0.00128)
10%	G	0.89897889 (0.01926)	0.95827444 (0.01297)	0.98933889 (0.00385)
	G^*	0.87989333 (0.02293)	0.95592556 (0.01136)	0.98431333 (0.00559)
	r_{pbis}	0.89219222 (0.02703)	0.93917889 (0.01278)	0.97510222 (0.0077)
	CI	0.8746 (0.02576)	0.94558222 (0.0114)	0.97929111 (0.00631)
	MCI	0.87706 (0.02513)	0.94728222 (0.01065)	0.98415778 (0.00566)
	$U3$	0.88189778 (0.02422)	0.95031222 (0.01032)	0.98396778 (0.00547)
	$ZU3$	0.88189778 (0.02422)	0.95031222 (0.01032)	0.98396778 (0.00547)
	H^T	0.90153778 (0.02469)	0.94715111 (0.0112)	0.98078444 (0.00591)
	A	0.56026778 (0.02512)	0.54552556 (0.02071)	0.54546556 (0.01886)
	R_n	0.90830667 (0.01786)	0.95122889 (0.01243)	0.98709222 (0.0038)
25%	G	0.648752 (0.04112)	0.7122928 (0.03329)	0.6893136 (0.02779)
	G^*	0.60673227 (0.04836)	0.700296 (0.0371)	0.683624 (0.03037)
	r_{pbis}	0.62174133 (0.03963)	0.6648784 (0.02872)	0.661456 (0.02636)
	CI	0.6250464 (0.03653)	0.6935552 (0.02861)	0.69515947 (0.02696)
	MCI	0.6299184 (0.03743)	0.7064752 (0.03094)	0.71575253 (0.03111)
	$U3$	0.60547733 (0.03816)	0.66461867 (0.03114)	0.6610656 (0.03094)
	$ZU3$	0.60547733 (0.03816)	0.66461867 (0.03114)	0.6610656 (0.03094)
	H^T	0.65340267 (0.03787)	0.71734933 (0.02708)	0.72341173 (0.02591)
	A	0.51386827 (0.02299)	0.4788224 (0.02823)	0.48148267 (0.02295)
	R_n	0.69543413 (0.04005)	0.78515627 (0.03046)	0.76077387 (0.02936)
50%	G	0.1416116 (0.01851)	0.1172664 (0.02236)	0.0712432 (0.01486)
	G^*	0.0825056 (0.01297)	0.045074 (0.01326)	0.0075948 (0.00352)
	r_{pbis}	0.0624716 (0.01526)	0.02283 (0.00499)	0.0013172 (0.00084)
	CI	0.086982 (0.01491)	0.0584496 (0.0141)	0.0118728 (0.00393)
	MCI	0.0885492 (0.01494)	0.0606692 (0.01488)	0.0130848 (0.00426)
	$U3$	0.0987536 (0.01538)	0.0708624 (0.01628)	0.0191296 (0.00525)
	$ZU3$	0.0987536 (0.01538)	0.0708624 (0.01628)	0.0191296 (0.00525)
	H^T	0.1028352 (0.02244)	0.0565456 (0.01194)	0.0116088 (0.00451)
	A	0.4056316 (0.02827)	0.3508176 (0.0318)	0.3773288 (0.02796)
	R_n	0.10562 (0.01622)	0.0963844 (0.01997)	0.0549372 (0.01195)

Tabela 3.13: Média e Desvio Padrão da área sob a curva ROC de dados com ML3P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes criativos (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.63272 (0.05198)	0.72883 (0.04847)	0.72539 (0.04026)
	G^*	0.59435 (0.04974)	0.68851 (0.04955)	0.66697 (0.04937)
	r_{pbis}	0.57784 (0.04645)	0.64258 (0.05321)	0.58763 (0.05914)
	CI	0.60416 (0.0469)	0.68808 (0.05113)	0.67518 (0.04782)
	MCI	0.60582 (0.04775)	0.68444 (0.05176)	0.68704 (0.04707)
	$U3$	0.62542 (0.04394)	0.70583 (0.04749)	0.71604 (0.04458)
	$ZU3$	0.62542 (0.04394)	0.70583 (0.04749)	0.71604 (0.04458)
	H^T	0.60505 (0.05216)	0.66368 (0.05502)	0.62347 (0.05274)
	A	0.76897 (0.02408)	0.80988 (0.02685)	0.80831 (0.027)
	R_n	0.58595 (0.05027)	0.66215 (0.05301)	0.65938 (0.05183)
10%	G	0.585 (0.03658)	0.70982 (0.03456)	0.74992 (0.02757)
	G^*	0.54088 (0.03828)	0.66224 (0.04076)	0.69352 (0.03283)
	r_{pbis}	0.52167 (0.0433)	0.61175 (0.04835)	0.6206 (0.04058)
	CI	0.54816 (0.03713)	0.66423 (0.04052)	0.70395 (0.03193)
	MCI	0.5566 (0.03617)	0.65985 (0.04117)	0.70948 (0.03025)
	$U3$	0.57546 (0.03374)	0.67904 (0.04043)	0.72781 (0.028)
	$ZU3$	0.57546 (0.03374)	0.67904 (0.04043)	0.72781 (0.028)
	H^T	0.54091 (0.04663)	0.65402 (0.04077)	0.67671 (0.03633)
	A	0.78004 (0.02083)	0.78885 (0.03076)	0.80161 (0.0266)
	R_n	0.5259 (0.04123)	0.65685 (0.0362)	0.68778 (0.02682)
25%	G	0.65124 (0.02678)	0.6907 (0.0223)	0.75271 (0.02137)
	G^*	0.61126 (0.02754)	0.63629 (0.02327)	0.69588 (0.02558)
	r_{pbis}	0.58636 (0.03145)	0.58441 (0.02429)	0.61875 (0.03099)
	CI	0.61778 (0.02666)	0.64399 (0.02269)	0.70279 (0.02745)
	MCI	0.62921 (0.02666)	0.65273 (0.02356)	0.71739 (0.02658)
	$U3$	0.64104 (0.02586)	0.66723 (0.02285)	0.73154 (0.0254)
	$ZU3$	0.64104 (0.02586)	0.66723 (0.02285)	0.73154 (0.0254)
	H^T	0.62685 (0.02929)	0.62999 (0.02551)	0.70114 (0.02447)
	A	0.76948 (0.02391)	0.80266 (0.02078)	0.80704 (0.01846)
	R_n	0.60555 (0.03016)	0.63481 (0.02549)	0.69539 (0.02423)
50%	G	0.59838 (0.02229)	0.68431 (0.01946)	0.77034 (0.02362)
	G^*	0.55714 (0.02387)	0.6283 (0.02132)	0.71646 (0.02434)
	r_{pbis}	0.5074 (0.0238)	0.53621 (0.0245)	0.61467 (0.02869)
	CI	0.58234 (0.02148)	0.65254 (0.01867)	0.73277 (0.02488)
	MCI	0.61492 (0.02297)	0.69194 (0.01746)	0.75642 (0.02415)
	$U3$	0.61716 (0.02335)	0.69361 (0.01695)	0.75792 (0.0241)
	$ZU3$	0.61716 (0.02335)	0.69361 (0.01695)	0.75792 (0.0241)
	H^T	0.59149 (0.02184)	0.64514 (0.02032)	0.74448 (0.02386)
	A	0.77007 (0.02551)	0.79507 (0.02091)	0.80301 (0.02059)
	R_n	0.54646 (0.0204)	0.61785 (0.01932)	0.71405 (0.02892)

Tabela 3.14: Média e Desvio Padrão da área sob a curva ROC de dados com ML3P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes adivinhos (5%, 10%, 25% e 50%).

% de respondentes aberrantes	Estatística	Número de itens		
		17	33	65
5%	G	0.76149 (0.04286)	0.85908 (0.0322)	0.89788 (0.02147)
	G^*	0.82943 (0.03555)	0.94018 (0.01967)	0.97726 (0.0091)
	r_{pbis}	0.88118 (0.0317)	0.94638 (0.01739)	0.98013 (0.00796)
	CI	0.83837 (0.03703)	0.94155 (0.01932)	0.97887 (0.00853)
	MCI	0.83871 (0.03851)	0.94315 (0.01896)	0.98032 (0.00789)
	$U3$	0.8594 (0.03409)	0.95835 (0.01405)	0.9855 (0.00564)
	$ZU3$	0.8594 (0.03409)	0.95835 (0.01405)	0.9855 (0.00564)
	H^T	0.84071 (0.0468)	0.91981 (0.03104)	0.96764 (0.01581)
	A	0.93755 (0.01043)	0.9744 (0.00684)	0.99003 (0.00465)
	R_n	0.66587 (0.06356)	0.7252 (0.04072)	0.75219 (0.044)
10%	G	0.75548 (0.0404)	0.86176 (0.02542)	0.90098 (0.02048)
	G^*	0.83973 (0.02893)	0.94572 (0.01453)	0.97915 (0.00692)
	r_{pbis}	0.88192 (0.02566)	0.95303 (0.01275)	0.98195 (0.00625)
	CI	0.85322 (0.02477)	0.94855 (0.01359)	0.97956 (0.00678)
	MCI	0.85778 (0.02441)	0.95136 (0.01249)	0.98032 (0.00657)
	$U3$	0.88007 (0.02199)	0.96216 (0.00951)	0.98461 (0.00525)
	$ZU3$	0.88007 (0.02199)	0.96216 (0.00951)	0.98461 (0.00525)
	H^T	0.84171 (0.03483)	0.93174 (0.02143)	0.97404 (0.00968)
	A	0.95609 (0.00828)	0.97517 (0.00386)	0.98929 (0.00311)
	R_n	0.6266 (0.04447)	0.72465 (0.03025)	0.76972 (0.03695)
25%	G	0.75445 (0.02381)	0.8574 (0.01653)	0.88909 (0.01432)
	G^*	0.83866 (0.0218)	0.94295 (0.01073)	0.97566 (0.0052)
	r_{pbis}	0.87371 (0.01693)	0.94997 (0.00954)	0.97868 (0.00465)
	CI	0.84961 (0.02159)	0.94229 (0.01083)	0.97511 (0.00521)
	MCI	0.85695 (0.02453)	0.94085 (0.01167)	0.97244 (0.00599)
	$U3$	0.86152 (0.02417)	0.94362 (0.01135)	0.97374 (0.00575)
	$ZU3$	0.86152 (0.02417)	0.94362 (0.01135)	0.97374 (0.00575)
	H^T	0.85785 (0.01826)	0.94342 (0.01055)	0.9754 (0.00511)
	A	0.94787 (0.01144)	0.97555 (0.0042)	0.98861 (0.00287)
	R_n	0.6302 (0.02708)	0.7254 (0.02306)	0.73576 (0.02572)
50%	G	0.74151 (0.01866)	0.839 (0.01755)	0.89214 (0.01467)
	G^*	0.82344 (0.01437)	0.93369 (0.01191)	0.97673 (0.00534)
	r_{pbis}	0.86295 (0.0107)	0.94128 (0.01215)	0.98 (0.00485)
	CI	0.8294 (0.01673)	0.93527 (0.01264)	0.97745 (0.00514)
	MCI	0.82589 (0.02919)	0.93437 (0.01621)	0.97767 (0.00562)
	$U3$	0.81215 (0.03125)	0.92565 (0.01873)	0.9741 (0.00665)
	$ZU3$	0.81215 (0.03125)	0.92565 (0.01873)	0.9741 (0.00665)
	H^T	0.86626 (0.01053)	0.94678 (0.01046)	0.98152 (0.00475)
	A	0.94362 (0.0114)	0.9763 (0.00711)	0.98784 (0.0028)
	R_n	0.61156 (0.02459)	0.67094 (0.02832)	0.73876 (0.02742)

Tabela 3.15: Média e Desvio Padrão da área sob a curva ROC de dados com ML3P para as 10 (dez) estatísticas de ajuste individual em três comprimentos de teste (17, 33 e 65 itens) e 4 (quatro) porcentagens de respondentes aleatórios (5%, 10%, 25% e 50%).

A comparação das estatísticas de ajuste individual G , G^* , r_{pbis} , CI , MCI , $U3$, $ZU3$, H^T , A e R_n nas Figuras a seguir são ML3P para respondentes aberrantes com resposta descuidada, resposta com uso de fraude, resposta criativa, resposta de adivinhação e resposta aleatória. Em cada gráfico, temos um tipo de respondente aberrante com áreas sob a curva ROC, para cada estatística de ajuste individual, no eixo y e porcentagem de respondentes aberrantes ou tamanho de teste no eixo x . Observe que os resultados gráficos são um pouco mais gerais do que os apresentados nas tabelas anteriores e, que pode verificar uma pequena diferença em relação aos resultados apresentados nas tabelas, pois para gerar os gráficos, é necessário o agrupamento médio da curva ROC das estatísticas de ajuste individual.

Nas Figuras 3.22 e 3.23 notou-se que o comprimento do teste tem impacto sobre a eficiência das estatísticas, quando ocorre o aumento do comprimento do teste as estatísticas vão se igualando no desempenho. Em todos os casos o índice A se mostra menos eficaz. Na maioria das situações, as ROC apresentam valores semelhantes, ficando difícil de identificar uma mais eficaz. Para os casos de testes com 65 itens e considerando as porcentagem de 50% de respondentes com resposta descuidada, as estatísticas H^T e R_n tiveram maior identificação de indivíduos inadequados que as demais. Para os casos com 17 itens, a estatística H^T tem um destaque no desempenho considerando as 4 (quatro) porcentagens de respondentes aberrantes. É possível afirmar que quando se considera a condição de 50% de respondentes aberrantes é mais difícil de detectar respondentes descuidados.

Para os respondentes fraudadores as Figuras 3.24 e 3.25, mostram-se semelhantes aos resultados das simulações de ML1P e ML2P, onde para os casos com 50% de respondentes fraudadores o índice A é mais eficaz seguidos das estatísticas G e R_n . Além disso, para os casos com 25% respondentes descuidados, as estatísticas R_n e G também apresentaram mais eficiência para identificar respondentes aberrantes considerando os três tamanhos de teste. Para a condição de 5% de respondentes descuidados e comprimento de teste curto, a estatística H^T e r_{pbis} foram que apresentaram melhor desempenho e para comprimento de testes médios e longos as estatísticas G e R_n é ligeiramente mais eficaz. Também é possível confirmar, com base

nos gráficos e tabelas dos respondentes trapaceiros, que o tamanho do teste e a porcentagem dos respondentes aberrantes influenciam no desempenho das estatísticas.

As figuras dos respondentes criativos 3.26 e 3.27 mostram que a porcentagem de 50% de respondentes aberrantes é mais difícil de detectar e que a detecção das estatísticas de ajuste individual com o aumento das porcentagens de respondentes criativos, já em relação aos comprimentos de teste, não houve interferência significativa no desempenho das mesmas. Para 50% de respondentes criativos, o índice A teve o melhor desempenho seguido das estatísticas G e R_n considerando o tamanho de teste longo. Para os casos com 25% respondentes criativos a estatística R_n foi mais sensível para os três comprimentos de teste. Além disso, os casos com 5% e 10% de respondentes fraudadores as estatísticas não apresentam diferença significativa e a estatística R_n apresenta uma evolução no desempenho com o aumento do tamanho do teste.

Para os respondentes adivinhos nas Figuras 3.28 e 3.29 o índice A teve a melhor eficiência global em todos os caso. A estatística proposta apresenta uma evolução no desempenho com o aumento de itens no teste. Além disso, o comprimento de teste apresenta um efeito sobre o desempenho das estatísticas, quanto maior a quantidade de itens no teste maior é a identificação de examinados aberrantes .

Os casos com respondentes aleatórios nas Figuras 3.30 e 3.31 notam-se que as condições de porcentagem têm impacto nas ROC das estatísticas, e que o aumento dos itens de teste aumentam as áreas sob a curva ROC das estatísticas. O índice A teve o melhor desempenho em todas as situações. Já a estatística proposta teve o pior desempenho em todas as ocasiões.

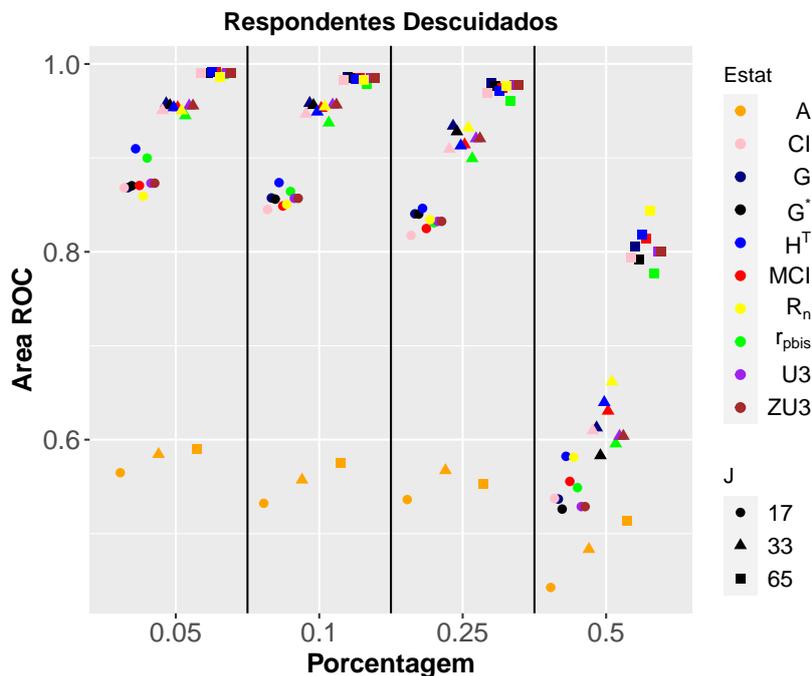


Figura 3.22: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes descuidados sob 4 (quatro) condições de porcentagem.

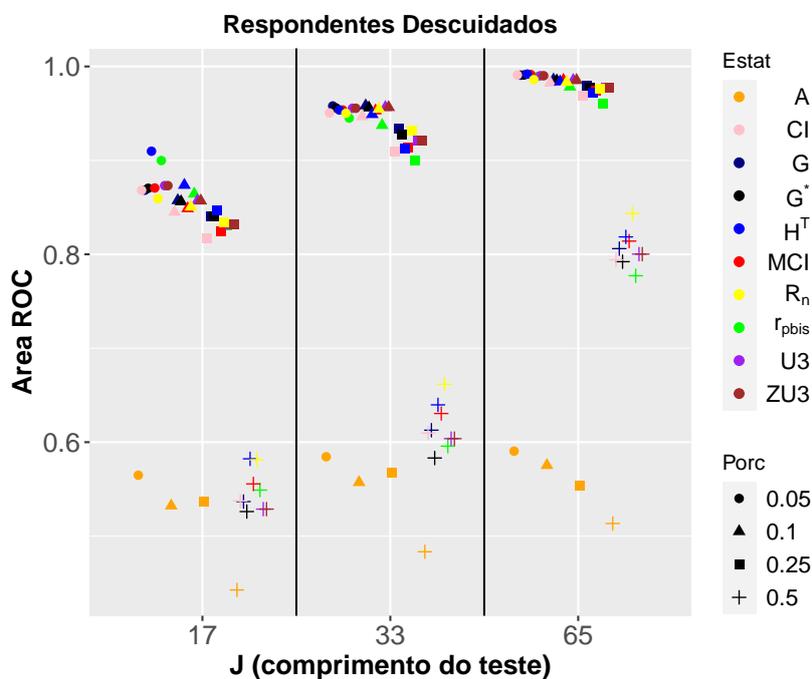


Figura 3.23: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes descuidados sob 3 (três) tamanhos do teste.

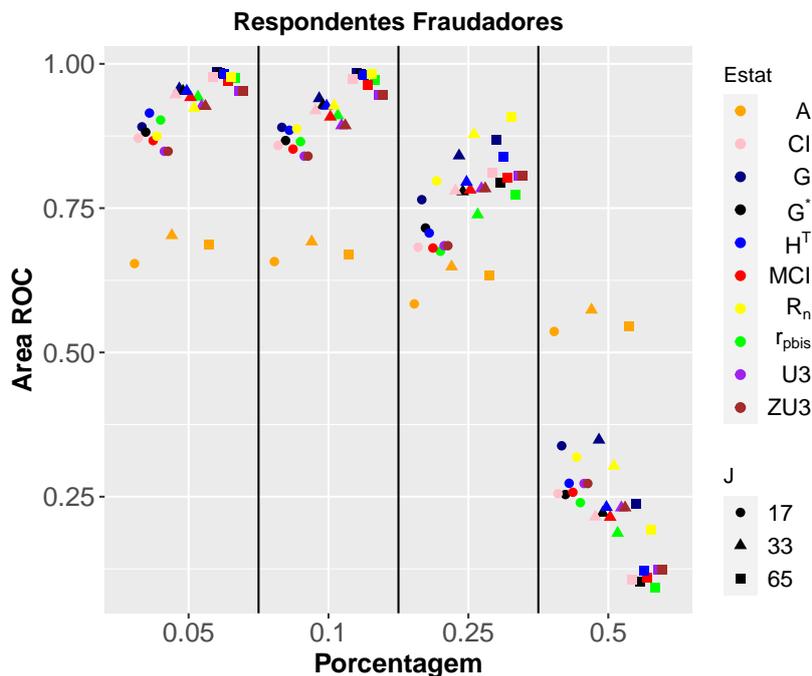


Figura 3.24: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes fraudulentos sob 4 (quatro) condições de porcentagem.

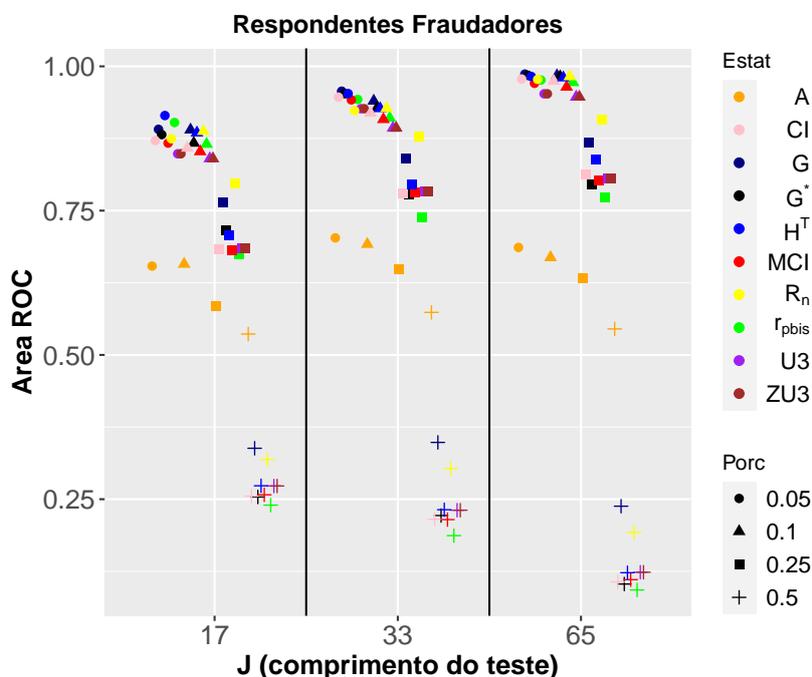


Figura 3.25: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes fraudulentos sob 3 (três) tamanhos do teste.

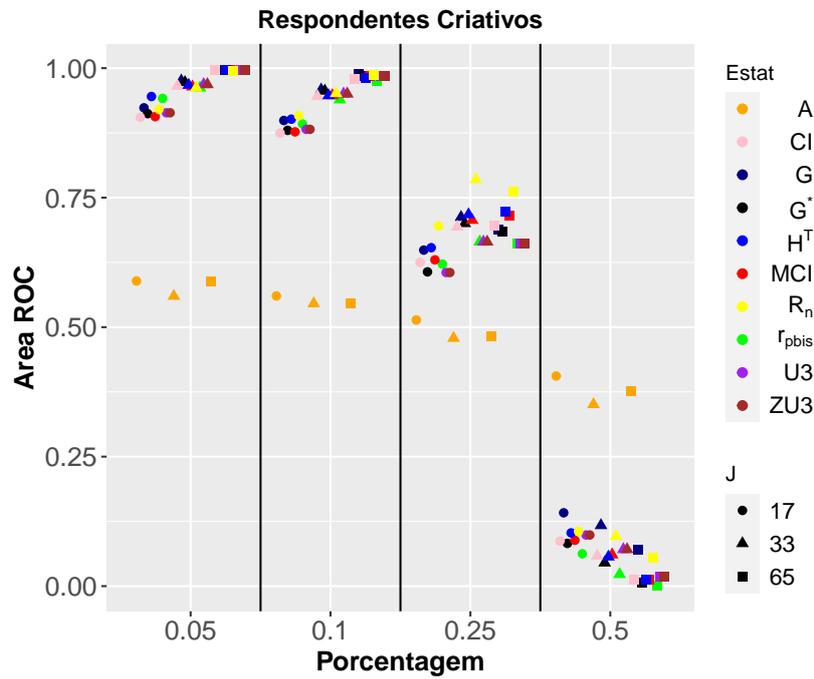


Figura 3.26: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes criativos sob 4 (quatro) condições de porcentagem.

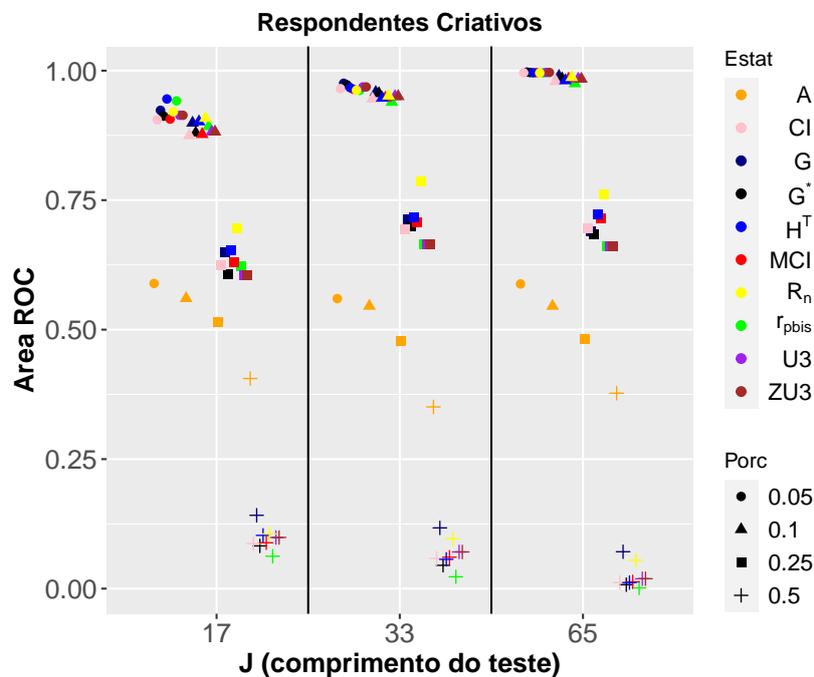


Figura 3.27: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes criativos sob 3 (três) tamanhos do teste.

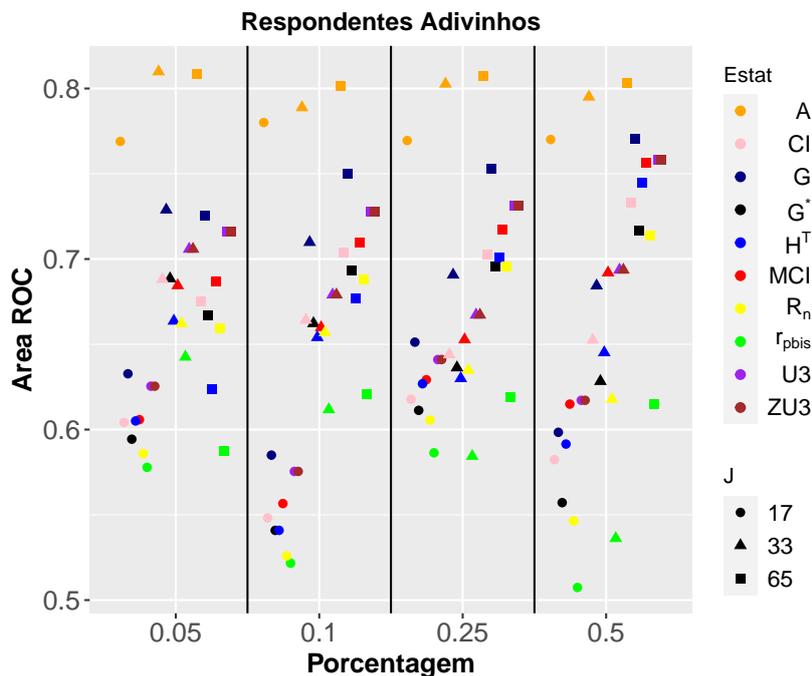


Figura 3.28: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes adivinhos sob 4 (quatro) condições de porcentagem.

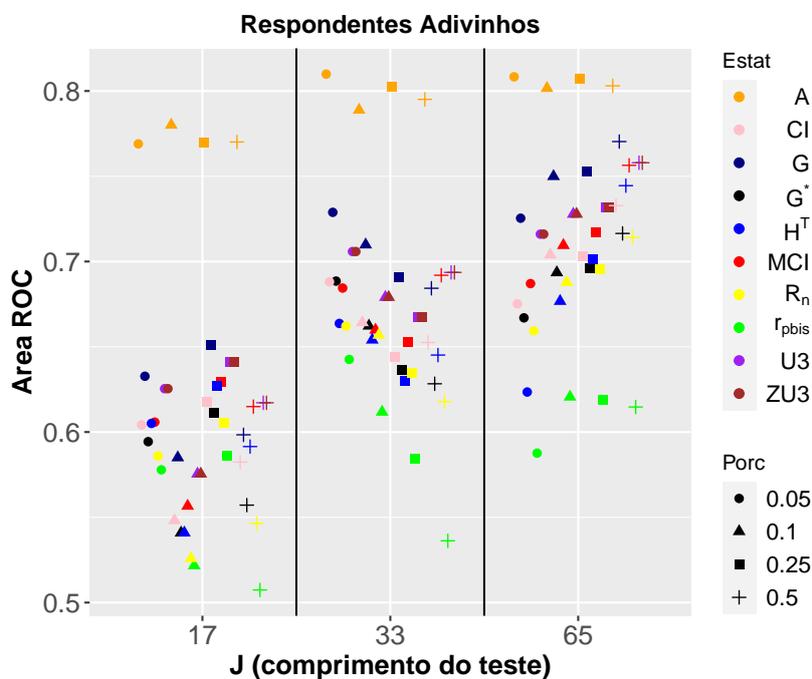


Figura 3.29: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes adivinhos sob 3 (três) tamanhos do teste.

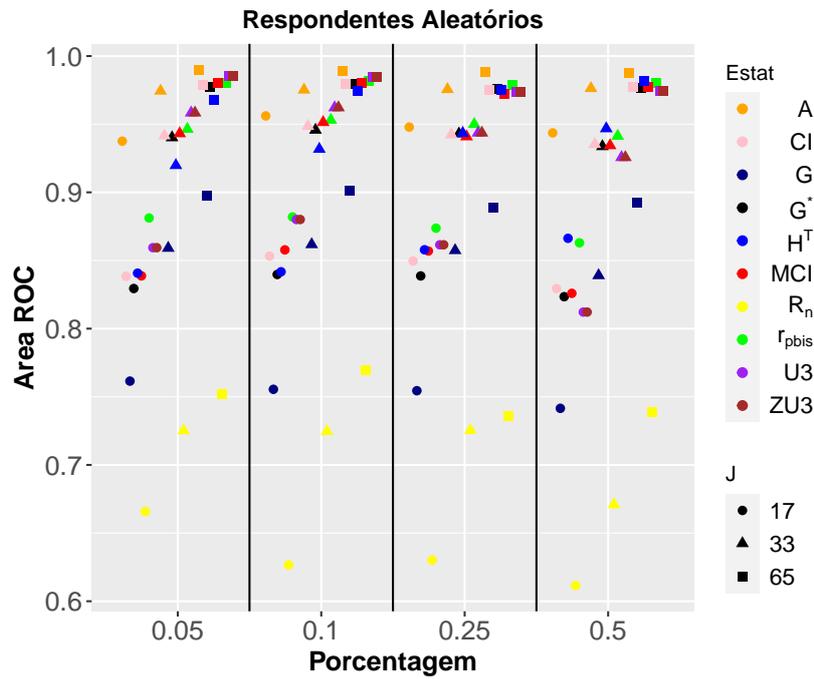


Figura 3.30: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes aleatórios sob 4 (quatro) condições de porcentagem.

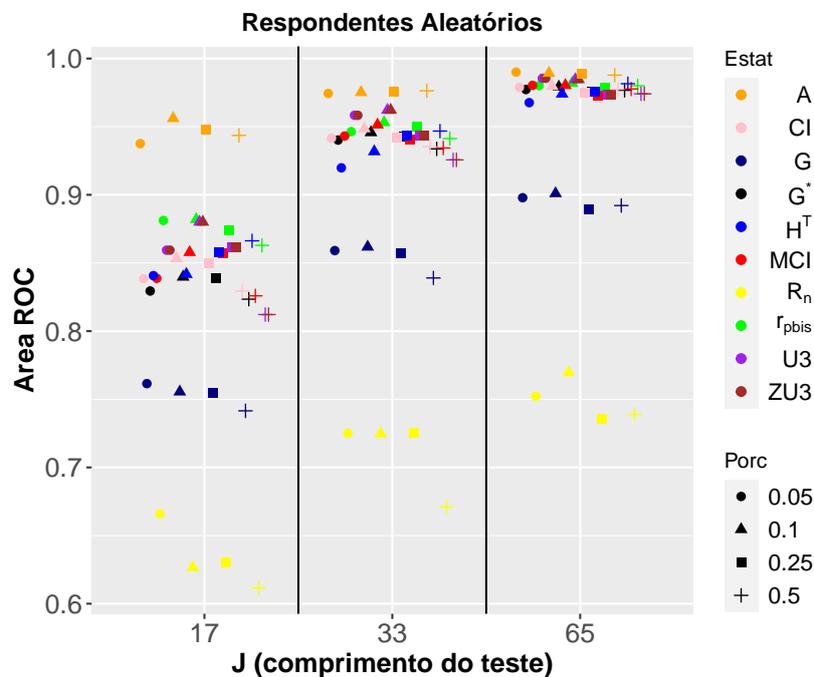


Figura 3.31: Comparação de 10 (dez) estatísticas de ajuste individual, comdados ML3P para respondentes aleatórios sob 3 (três) tamanhos do teste.

3.6 Discussão

As simulações do estudo de Karabatsos (2003) foram baseadas nas áreas sob a curva ROC. Karabatsos (2003) comparou 36 estatísticas de ajuste individual (paramétricas e não paramétricas), considerando o modelo Rasch na simulação dos conjuntos de dados. Ele descobriu que quando fixa-se a porcentagem de respondentes aberrantes e varia-se o tamanho de teste e o tipo de respondente a média das áreas aumenta conforme ocorre o aumento da porcentagem de respondentes aberrantes. Neste trabalho, o impacto das porcentagens de respondentes aberrantes ocorre para três modelos (ML1P, ML2P e ML3P), principalmente nas Tabelas 3.2, 3.3, 3.7, 3.8, 3.12 e 3.13 referentes aos casos com respondentes fraudadores e criativos.

Outro resultado semelhante neste trabalho ao estudo de Karabatsos (2003) ocorre quando o aumento dos itens de teste aumenta as áreas ROC das estatísticas de ajuste individual. Como as tabelas foram baseadas no estudo de Tendeiro e Meijer (2014), podemos observar, de forma mais clara, os impactos das porcentagens de respondentes aberrantes e do tamanho de teste. Além disso, uma diferença observada neste estudo com o estudo Karabatsos (2003) é que os indivíduos fraudadores e criativos com porcentagem a partir de 50% dos respondentes aberrantes apresentam a diminuição da taxa de desempenho das estatísticas de ajuste individual com o aumento dos itens de teste. Além destes, os estudos de Meijer (1994) e Dimitrov e Smith (2006) também mostram que o poder em detectar padrões de resposta aberrante das estatísticas de ajuste individual para respondentes adivinhos e fraudadores aumenta com os aumentos dos itens em teste e do nível de respondentes aberrantes.

Segundo Karabatsos (2003), a estatística de ajuste individual H^T mostrou-se mais sensível na detecção de padrões de resposta incoerentes de um examinado, superando 35 estatísticas (não paramétricas e paramétricas) de ajuste individual. Para o estudo de Dimitrov e Smith (2006), a estatística H^T superou as estatísticas paramétricas t , t^* , $Z3$ e $Z3^*$ para os casos com 20 e 30 itens, mas para teste com 10 itens as quatro estatísticas paramétricas têm melhor desempenho do que H^T . O que confirma os resultados deste estudo, pois maioria dos casos a estatística H^T

apresentou-se mais sensível para teste curto com 17 itens.

O estudo de Tendeiro e Meijer (2014) utilizou simulação com ML3P para 1.000 examinados, onde considerou vetores de resposta para pessoas com alta, baixa e mista capacidade. E, a estatística H^T apresentou maior potência do que os outros índices baseados em Guttman para identificar vetor de resposta inadequado. Este resultado é muito semelhante ao apresentado neste estudo, pois na maioria dos casos, a estatística H^T apresentou a maior eficiência global, estando sempre presente entre os 3 (três) índices mais eficazes para qualquer situação.

Para a geração dos gráficos nesta dissertação, utilizou-se o estudo de Sinharay (2017) que demonstrou a importância de uma análise mais tradicional e justa. Segundo Sinharay (2017), o agrupamento das áreas sob a curva ROC que combina a informação de vários conjuntos de dados para diferentes situações devem ser geradas sobre uma média ponderada. Desta forma, com base no resultado e análise das tabelas e gráficos deste trabalho, pode-se afirmar que a estatística G também é considerada uma estatística bastante relevante.

No estudo Meijer (1994), comparou-se o uso das estatísticas de ajuste individual G , G^* e $U3$ em uma simulação para modelo logístico de 2 (dois) parâmetros com 450 examinados, dois comprimentos de teste e dois tipos de respondente aberrante (fraudadores e adivinhos), onde as estatísticas de ajuste individual apresentaram desempenho semelhantes, em condições idênticas, para os respondentes fraudadores. O que reforça o desempenho das estatísticas G e G^* superior para algumas situações deste estudo.

Harnisch e Linn (1981) utilizaram dados referentes a exames de matemática e leitura aplicados a 6.300 alunos de três séries distintas. A comparação considerou as correlações entre as estatísticas de ajuste individual NCL , C , C^* , r_{pbis} , r_{bis} , A , D e E e a pontuação total dos examinados. Os resultados apontaram que o índice A apresentou correlação de 0,99 para ambos os testes, isto é, bem semelhante às tabelas e aos gráficos para os respondentes adivinhos e aleatórios deste estudo, onde o índice A apresentou as melhores estimativas em todas as situações simuladas.

Capítulo 4

Aplicações a Dados Reais

Utilizou-se um conjunto de dados que corresponde a uma amostra de 1001 respondentes de uma prova de Língua Portuguesa com 30 itens aplicada a alunos do terceiro ano do turno matutino de escolas públicas de São Paulo em 2007.

Para decidir se as respostas de cada respondente são aberrantes ou não, foi realizado um teste de hipótese cuja a hipótese nula H_0 diz que as respostas não são aberrantes. Para calcular o p -valor do teste, foram geradas respostas não aberrantes para os 30 itens de cada indivíduo. Ajustou-se o modelo de 3 parâmetros para os 30 itens, considerando o parâmetro de acerto ao acaso igual a 0.25 para todos os itens e os parâmetros de dificuldade e de discriminação para cada item. Além disso, estimou-se as proficiências de cada respondente e, com os parâmetros estimados, obteve-se as respostas de cada indivíduo para cada questão, utilizando a probabilidade de resposta positiva obtida ao substituir o valor estimado da proficiência do indivíduo e as estimativas dos parâmetros na fórmula do modelo logístico de 3 parâmetros.

Para obter o p -valor, são realizadas 1.000 repetições da amostra. A cada amostra, determinam-se os valores das 10 estatísticas. Com isso, temos 1.000 valores de cada estatística para cada respondente, sob a hipótese das respostas não serem aberrantes. Para uma determinada estatística e um determinado respondente, o p -valor é dado pela proporção dos 1.000 valores que estão acima do valor observado para a estatística que foi calculada com base nas respostas reais

do indivíduo. A estatística indica respostas aberrantes quando assume um valor muito pequeno.

Assim, comparou-se o desempenho das 10 (dez) estatísticas de ajuste individual na detecção de examinados com respostas inadequadas. A Tabela 4.1 mostra que, quando as estatísticas de ajuste individual apresentam p -valor menor que 1%, as respostas dos examinados são classificadas como aberrantes. Nesta aplicação, temos que os indivíduos com IDs (500, 532, 832, 987 e 992) foram considerados pela maioria das estatísticas como respondentes aberrantes, exceto pelo índice A que apresentou p -valor que não rejeita a hipótese nula, mas como o índice A foi considerado a estatística de ajuste individual com baixo desempenho para maioria das situações apresentadas nas simulações deste estudo, na identificação de indivíduos com padrão de resposta inadequada, considerou os IDs como aberrantes.

A Figura 4.1 mostra o padrão de resposta do examinado versus a dificuldade do item para os IDs considerados como aberrantes pelas estatísticas de ajuste individual. Nota-se, que para todos os indivíduos considerados aberrantes, os padrões de respostas apresentam-se comportamento diferente do esperado, pois quando se ordena os itens pelo número de acertos, espera-se que o padrão de resposta do examinado acerte os primeiros itens (considerados fáceis), no meio, oscile (com acertos e erros) e, no fim, erre os itens difíceis. Por isso, os padrões de respostas são considerados inadequados. Na Figura 4.2, observa-se o comportamento esperado dos examinados considerados com o comportamento adequado pelas estatísticas de ajuste individual, onde o padrão de respostas dos examinados apresentam acertos e erros de acordo com a dificuldade do item.

id	$p.G$	$p.G^*$	$p.r_{pbis}$	$p.CI$	$p.MCI$	$p.U3$	$p.ZU3$	$p.H^T$	$p.R_n$	$p.A$
500	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.004	0.117
535	0.004	0.006	0.003	0.003	0.004	0.004	0.006	0.003	0.000	0.145
832	0.001	0.000	0.001	0.001	0.000	0.000	0.001	0.000	0.008	0.085
987	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.002	0.381
992	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.275
993	0.827	0.797	0.802	0.806	0.800	0.816	0.809	0.850	0.833	0.322
995	0.935	0.941	0.940	0.936	0.937	0.930	0.912	0.912	0.745	0.605
997	0.413	0.426	0.414	0.417	0.413	0.441	0.458	0.416	0.261	0.402
998	0.605	0.617	0.593	0.609	0.611	0.567	0.569	0.582	0.970	0.736
999	0.395	0.392	0.343	0.334	0.326	0.258	0.277	0.343	0.270	0.243
708	0.995	0.995	0.997	0.996	0.996	0.996	0.991	0.993	1.000	0.604
369	0.994	0.990	0.955	0.994	0.994	0.995	0.989	0.994	0.994	0.951
924	0.995	0.990	0.985	0.988	0.988	0.986	0.984	0.994	0.992	0.325
743	0.988	0.991	0.995	0.994	0.994	0.994	0.985	0.988	0.955	0.740
21	0.989	0.989	0.991	0.991	0.991	0.991	0.985	0.990	0.932	0.545

Tabela 4.1: Referente aos p -valores associados a cada uma das estatísticas, para os 15 (quinze) respondentes do turno matutino submetidos ao exame de 30 itens do SARESP.

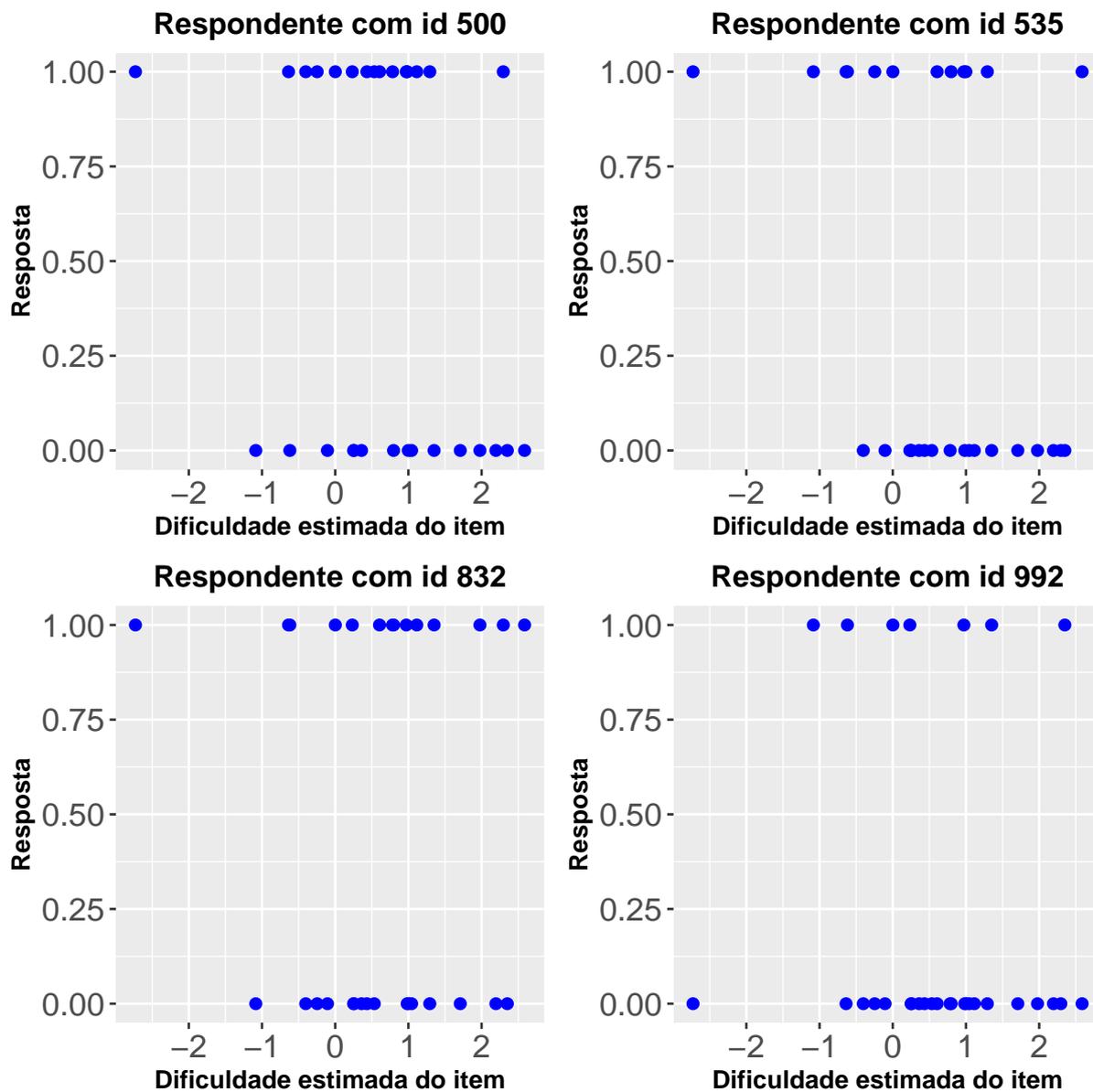


Figura 4.1: Padrão de respostas dos examinados com identificações 500, 535, 832 e 992 considerados como aberrantes, de acordo com o Teste de Hipótese, para todas as estatísticas exceto o índice A.

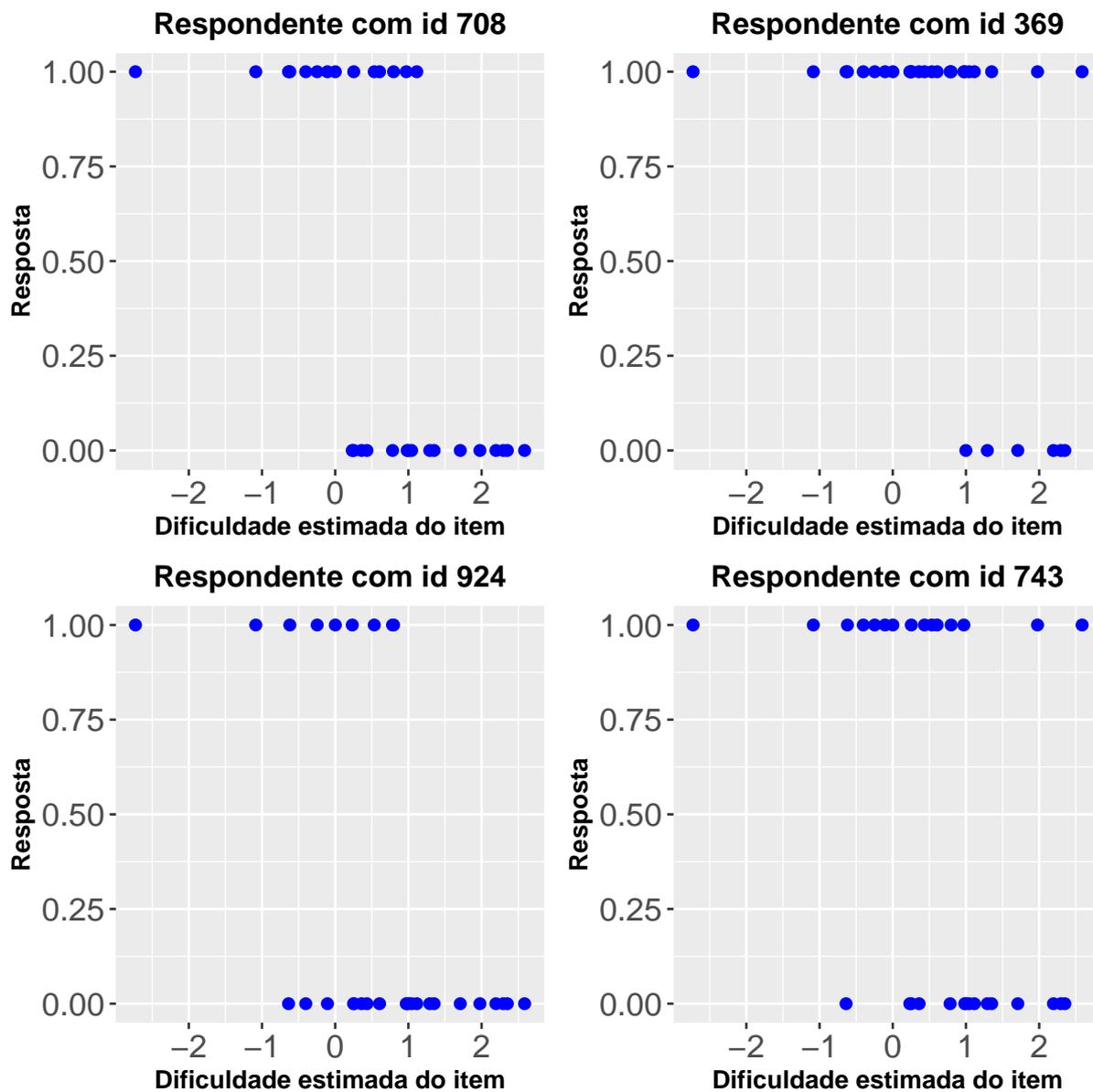


Figura 4.2: Padrão de respostas dos examinados com identificações 708, 369, 924 e 743 considerados como não aberrantes, de acordo com o Teste de Hipótese, para todas as estatísticas exceto o índice A.

A Tabela 4.2 mostra a quantidade de examinados que foram classificados como aberrantes e não aberrantes para cada estatística de ajuste individual, com base no p -valor e no nível descritivo $\alpha = 1\%$ do Teste de Hipótese. Observe que as estatísticas G , G^* , r_{pbis} , CI , MCI e H^T apresentam aproximadamente o mesmo número de examinados aberrantes. A estatística R_n teve desempenho superior que as estatísticas $U3$ e $ZU3$ na identificação de respondentes aberrantes. O índice A mostrou-se incapaz na detecção de padrões de respostas inadequados dos examinados, pois como visto na Figura 4.1 o padrão das respostas dos examinados com IDs 500, 535, 832 e 992 apresentam comportamento aberrante.

	G	G^*	r_{pbis}	CI	MCI	$U3$	$ZU3$	H^T	A	R_n
Respostas aberrantes	15	14	14	14	13	10	9	14	0	11
Respostas não aberrantes	986	987	987	987	988	991	992	987	1001	990
Porcentagem	1,5	1,4	1,4	1,4	1,3	1,0	0,9	1,4	0	1,1

Tabela 4.2: Quantidades de indivíduos aberrantes e não aberrantes e a porcentagem das estatísticas que foi obtida com base no p -valor e no nível descritivo $\alpha = 1\%$ do Teste de Hipótese.

Capítulo 5

Conclusões

5.1 Considerações Finais

Em várias diretrizes de testes educacionais e psicológicos, é recomendado verificar a qualidade dos dados do indivíduo. Conforme discutido, existem diversas estatísticas que podem ser utilizadas para identificar padrões de resposta incoerente em relação ao padrão esperado. Como uma medida de ajuste individual para detectar padrões de resposta aberrante em modelos de TRI, a estatística R_n , baseada em regressão isotônica, foi proposta e comparada, por meio de estudo de simulação computacional, com as 9 (nove) estatísticas não-paramétricas de ajuste individual já existentes na literatura. Além disso, este estudo propôs de forma precisa e minuciosa mostrar as vantagens e desvantagens das estatísticas de ajuste individual, ou seja, escolher as mais poderosas sob 60 (sessenta) condições de teste. Desse modo, a pesquisa indicou que a sensibilidade de detecção das estatísticas de ajuste individual depende das seguintes variáveis: (1) tipo de respondente, (2) comprimento do teste e (3) porcentagem de respondentes aberrantes. Além disso, a estatística proposta R_n demonstrou desempenho superior para simulações considerado respondentes com resposta descuidada, com uso de fraude e com resposta criativa, e 25% e 50% de respondentes aberrantes para teste com 33 itens ou mais. Em conclusão, espera-se que os resultados fornecidos neste artigo possam ter implicações metodológicas e práticas

úteis para encontrar estatísticas de ajuste individual que seja mais poderoso para todos os tipos de exames e examinados.

5.2 Trabalhos Futuros

- Comparar a estatística proposta R_n com as estatísticas l_z , M e ECI do estudo de Meijer e Sijtsma (2001);
- Examinar os efeitos do tamanho da amostra sobre as estatísticas de ajuste individual;
- Comparar a estatística proposta R_n com outras estatísticas de ajuste individual, para itens politômicos, com base nos estudos de Drasgow e Kanfer (1985) e de Emons (2008).

Referências Bibliográficas

- Andrade, D. F. de, Tavares, H. R. e Valle, R. da C. (2000). “Teoria da Resposta ao Item: conceitos e aplicações”. *ABE, Sao Paulo*.
- Ayer, Miriam et al. (1955). “An empirical distribution function for sampling with incomplete information”. *The annals of mathematical statistics*, pp. 641–647.
- Barlow, R. E. et al. (1972). *Statistical inference under order restrictions: The theory and application of isotonic regression*. London: Wiley.
- Dimitrov, Dimiter M e Smith, Richard M (2006). “Adjusted Rasch person-fit statistics”. *Journal of Applied measurement* 7.2, p. 170.
- Donlon, Thomas F e Fischer, Frederic E (1968). “An index of an individual’s agreement with group-determined item difficulties”. *Educational and Psychological Measurement* 28.1, pp. 105–113.
- Drasgow, Fritz e Kanfer, Ruth (1985). “Equivalence of psychological measurement in heterogeneous populations.” *Journal of applied psychology* 70.4, p. 662.
- Emons, Wilco HM (2008). “Nonparametric person-fit analysis of polytomous item scores”. *Applied Psychological Measurement* 32.3, pp. 224–247.
- Felt, John M et al. (2017). “Using person fit statistics to detect outliers in survey research”. *Frontiers in psychology* 8, p. 863.
- Guttman, Louis (1944). “A basis for scaling qualitative data”. *American sociological review* 9.2, pp. 139–150.

- Guttman, Louis (1950). “The basis for scalogram analysis: Measurement and prediction”. *Studies in social psychology in World War II*, pp. 60–90.
- Härdle, W. (1990). *Applied nonparametric regression*. 19. Cambridge university press.
- Harnisch, Delwyn L e Linn, Robert L (1981). “Analysis of item response patterns: Questionable test data and dissimilar curriculum practices”. *Journal of Educational Measurement*, pp. 133–146.
- Kane, Michael T e Brennan, Robert L (1980). “Agreement coefficients as indices of dependability for domain-referenced tests”. *Applied Psychological Measurement* 4.1, pp. 105–126.
- Karabatsos, George (2003). “Comparing the aberrant response detection performance of thirty-six person-fit statistics”. *Applied Measurement in Education* 16.4, pp. 277–298.
- Klauer, Karl Christoph e Rettig, Klaus (1990). “An approximately standardized person test for assessing consistency with a latent trait model”. *British Journal of Mathematical and Statistical Psychology* 43.2, pp. 193–206.
- Lee, Young-Sun (2007). “A comparison of methods for nonparametric estimation of item characteristic curves for binary items”. *Applied Psychological Measurement* 31.2, pp. 121–134.
- Li, Mao-neng Fred e Olejnik, Stephen (1997). “The power of Rasch person-fit statistics in detecting unusual response patterns”. *Applied Psychological Measurement* 21.3, pp. 215–231.
- Meijer, Rob R (1994). “The number of Guttman errors as a simple and powerful person-fit statistic”. *Applied Psychological Measurement* 18.4, pp. 311–314.
- (1996). “Person-fit research: An introduction”. *Applied Measurement in Education* 9.1, pp. 3–8.
- (2002). “Outlier detection in high-stakes certification testing”. *Journal of Educational Measurement* 39.3, pp. 219–233.
- Meijer, Rob R e Sijtsma, Klaas (2001). “Methodology review: Evaluating person fit”. *Applied psychological measurement* 25.2, pp. 107–135.

- Meijer, Rob R et al. (2008). “Detection and validation of unscalable item score patterns using item response theory: an illustration with Harter’s Self-Perception Profile for Children”. *Journal of Personality Assessment* 90.3, pp. 227–238.
- Metz, Charles E (1986). “Statistical analysis of ROC data in evaluating diagnostic performance”. *Multiple regression analysis: Applications in the health sciences* 13, pp. 365–384.
- Muniz, J. F. (1990). *Teoría de respuesta a los ítems: un nuevo enfoque en la evolución [ie evaluación] psicológica y educativa*. Pirámide.
- van der Flier, H. (1977). “Environmental factors and deviant response patterns”. *Basic problems in cross cultural psychology*, Amsterdam: Swets & Zeitlinger.
- van der Flier, H. (1980). *"Vergelijkbaarheid van individuele testprestaties"*. Swets & Zeitlinger.
- van der Flier, H. (1982). “Deviant response patterns and comparability of test scores”. *Journal of Cross-Cultural Psychology* 13.3, pp. 267–298.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Rasch, Georg (1960). “Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.”
- Reise, Steven P (1990). “A comparison of item-and person-fit methods of assessing model-data fit in IRT”. *Applied Psychological Measurement* 14.2, pp. 127–137.
- Sato, T. (1975). “The construction and interpretation of SP tables”. *Tokyo: Meiji Tosho*.
- Sijtsma, K. (1986). “A coefficient of deviance of response patterns”. *Kwantitatieve Methoden* 7.22, pp. 131–145.
- Sijtsma, Klaas e Meijer, Rob R (1992). “A method for investigating the intersection of item response functions in Mokken’s nonparametric IRT model”. *Applied Psychological Measurement* 16.2, pp. 149–157.
- Sinharay, Sandip (2017). “As estatísticas não paramétricas de ajuste de pessoa são mais poderosas do que suas contrapartes paramétricas? Revisitando as simulações em Karabatsos (2003)”. *Applied Measurement in Education* 30.4, pp. 314–328.

- Smith, R. M. (1991). *IPARM: Item and person analysis with the Rasch model*. JAM Press.
- Spearman, Charles (1910). “Correlation calculated from faulty data”. *British Journal of Psychology, 1904-1920* 3.3, pp. 271–295.
- Tendeiro, Jorge N e Meijer, Rob R (2014). “Detection of invalid test scores: The usefulness of simple nonparametric statistics”. *Journal of Educational Measurement* 51.3, pp. 239–259.
- Thurstone, Louis L (1927). “The method of paired comparisons for social values.” *The Journal of Abnormal and Social Psychology* 21.4, p. 384.
- Trabin, Tom E e Weiss, David J (1983). “The person response curve: Fit of individuals to item response theory models”. Em: *New horizons in testing*. Elsevier, pp. 83–108.
- Wollack, JA e Maynes, D Dennis (2017). *Detection of test collusion using cluster analysis. Handbook of quantitative methods for detecting cheating on tests (pp. 124–150)*.

Apêndice A

Código da Programação em R

A.1 Programação para Dados Simulados

```
1 #=====#
2 #----- Funcao "reg.isot" -----#
3 #  Calcula a regressao isotonica da funcao "g" com pesos "w". #
4 #=====#
5 reg.isot <- function(g,w) {
6   n <- length(g)
7   gw <- g*w
8   gwac <- cumsum(gw)
9   wac <- cumsum(w)
10  matslopes <- cbind(c(1:n), gwac/wac)
11  indice <- indaux <- 1
12  ### "isot" e' o vetor com os valores da regressao isotonica
13  isot <- rep(0,n)
14  while(indice < n) {
15    indice <- max(matslopes[matslopes[,2]==min(matslopes[,2]),1])
16    isot[indaux:indice] <- min(matslopes[,2])
```

```
17   indaux <- indice + 1
18   matslopes <- cbind(c(indaux:n), (gwac[indaux:n]-gwac[indice])/
19                       (wac[indaux:n]-wac[indice]))
20   # print(matslopes)
21 }
22 return(isot)
23 }
24 #=====
25 #===== FUNCAO aberrantes =====
26 #   Gera as Estatisticas e Area Roc para tabelas e graficos   #
27 #=====
28 # tipo: 1=careless, 2=cheaters, 3=creative, 4=lucky, 5=random
29 # model=1: 1PL ; model=2: 2PL ; model=3: 3PL
30 # tabel=1: tabela latex ; tabel=2: tabela para grafico
31 aberrantes<-function(tipo,model,tabel){
32   set.seed(23456)
33   ## numero de respondentes
34   N <- 500
35   areas.G<-areas.G.star<-areas.r.pbis<-areas.CI<-areas.MCI<-NULL
36   areas.U3<-areas.ZU3<-areas.H.line<-areas.A<-areas.Rn<-NULL
37   Areas.G<-Areas.G.star<-Areas.r.pbis<-Areas.CI<-Areas.MCI<-NULL
38   Areas.U3<-Areas.ZU3<-Areas.H.line<-Areas.A<-Areas.Rn<-NULL
39   for (t in 1:3){
40     J<- c(17,33,65)
41     J<-J[t]
42     for (s in 1:4){
43       pp <- c(0.05,0.10,0.25,0.50)
44       pp<-pp[s]
```

```
45     # parametro de dificuldade
46     delta <- seq(-2,2,length.out=J)
47     # parametro de discriminacao
48     a <- runif(J,0.5,2)
49     # parametro de acerto ao acaso
50     cc <- rep(0.2,J)
51     n.rep <- 20
52     G.rep <- G.star.rep <- r.pbis.rep <- CI.rep <- MCI.rep <-
53     U3.rep <- ZU3.rep <- H.line.rep <- A.rep <- D.rep <-
54     E.rep <- med.disc.vec.rep <- rep(0,n.rep)
55     for (k in 1:n.rep) {
56         if(tipo==1 || tipo==3){
57             NC <- N*pp
58             ## proficiencias
59             theta <- c(runif(NC,0.5,2),c(runif((N-NC),-2,2)))
60         }else if (tipo==2 || tipo==4){
61             NC <- N*pp
62             ## proficiencias
63             theta <- c(runif(NC,-2,-0.5),c(runif((N-NC),-2,2)))
64         }else if (tipo==5){
65             NC <- N*pp
66             ## proficiencias
67             theta <- runif(N,-2,2)
68         }
69         # respostas
70         if (model==1){
71             X <- matrix(0,N,J)
72             for (j in 1:J) {
```

```
73     prob <- 1/(1+exp(-(theta-delta[j])))
74     X[,j] <- rbinom(N,1,prob)
75   }
76 }else if (model==2){
77   X <- matrix(0,N,J)
78   for (j in 1:J) {
79     prob <- 1/(1+exp(-a[j]*(theta-delta[j])))
80     X[,j] <- rbinom(N,1,prob)
81   }
82 }else{
83   X <- matrix(0,N,J)
84   for (j in 1:J) {
85     prob <- cc[j]+(1-cc[j])/(1+exp(-a[j]*(theta-delta[j])))
86     X[,j] <- rbinom(N,1,prob)
87   }
88 }
89 if (tipo==1){
90   # indices dos 41% itens mais faceis
91   JD <- floor(J*0.41)
92   X[c(1:NC),c(1:JD)] <- rbinom(NC*JD,1,0.5)
93 }else if (tipo==2){
94   # indices dos 18% itens mais dificeis
95   JD <- ceiling(J*(1-0.18))+1
96   X[c(1:NC),c(JD:J)] <- 1
97 }else if (tipo==3){
98   # indices dos 18% itens mais faceis
99   JD <- floor(J*0.18)
100  X[c(1:NC),c(1:JD)] <- 0
```

```

101     }else if (tipo==4){
102         # indices dos 41% itens mais dificeis
103         JD <- ceiling(J*(1-0.41))+1
104         X[c(1:NC),c(JD:J)] <- rbinom(NC*(J-JD+1),1,0.25)
105     }else if (tipo==5){
106         # indices dos 18% itens mais dificeis
107         JD <- ceiling(J*(1-0.18))+1
108         X[c(1:NC),] <- rbinom(NC*J,1,0.25)
109     }
110     # estimativas das proficiencias e dos parametros de
111         dificuldade
112     theta.est <- log((apply(X,1,sum)+0.1)/(J-apply(X,1,sum)))
113     delta.est <- log( ( N-apply(X,2,sum) )/apply(X,2,sum) )
114 #=====
115 #=====  ESTATISTICAS NAO PARAMETRICAS  =====#
116 #=====
117 #          G (Guttman, 1944, 1950)          #
118 #=====
119     # Numero de erros de respostas (Guttman, 1944, 1950)
120     G <- rep(0,N)
121     soma <- apply(X,2,sum)
122     for (j1 in 1:(J-1)) {
123         for (j2 in (j1+1):J) {
124             G<-G+X[,j1]*(1-X[,j2])*sign(sign(soma[j2]-soma[j1]-1)+1)+
125                 (1-X[,j1])*X[,j2]*sign(sign(soma[j1]-soma[j2]-1)+1)
126         }
127     }

```

```

128 ## Calculo da area sob a curva ROC
129 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
130 # H(c)=sensibilidade = P( G > c | aberrante) ou P( G < c | aberrante)
131 Estat <- G
132 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
133 FF <- HH <- rep(0,length(c.valor))
134 Estat.aberr <- Estat[1:NC]
135 Estat.normal <- Estat[(NC+1):N]
136 for (ind in 1:length(c.valor)) {
137     FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]]) / (N-NC)
138     HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]]) / NC
139 }
140 area <- -sum(diff(FF) * (HH[-length(HH)]+HH[-1])) / 2)
141 G.rep[k] <- area
142 #=====
143 #           G normado (van der Flier, 1977)           #
144 #=====
145 escore <- apply(X,1,sum)
146 G.star <- G/(escore*(J-escore))
147 G.star[G.star=="NaN"] <- 0
148 Estat <- G.star
149 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
150 FF <- HH <- rep(0,length(c.valor))
151 Estat.aberr <- Estat[1:NC]
152 Estat.normal <- Estat[(NC+1):N]
153 for (ind in 1:length(c.valor)) {
154     FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]]) / (N-NC)
155     HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]]) / NC

```

```

156 }
157 area <- -sum(diff(FF) * (HH[-length(HH)] + HH[-1])) / 2)
158 G.star.rep[k] <- area
159 #=====
160 # Estatística ponto bisserial pessoal (Donlon e Fischer, 1968) #
161 # (pode ter desvio-padrão igual a zero para X[n,]) #
162 #=====
163 prop <- soma/N
164 r.pbis <- rep(0,N)
165 for (n in 1:N) r.pbis[n] <- cor(X[n,],prop)
166 r.pbis[is.na(r.pbis)] <- 1
167 ## Calculo da area sob a curva ROC
168 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
169 # H(c)=sensibilidade = P(G > c | aberrante) ou P(G < c | aberrante)
170 Estat <- r.pbis
171 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
172 FF <- HH <- rep(0,length(c.valor))
173 Estat.aberr <- Estat[1:NC]
174 Estat.normal <- Estat[(NC+1):N]
175 for (ind in 1:length(c.valor)) {
176   FF[ind] <- length(Estat.normal[Estat.normal < c.valor[ind]]) / (N-NC)
177   HH[ind] <- length(Estat.aberr[Estat.aberr < c.valor[ind]]) / NC
178 }
179 area <- sum(diff(FF) * (HH[-length(HH)] + HH[-1])) / 2)
180 r.pbis.rep[k] <- area
181 ## Posicao dos itens em ordem decrescente das somas de acertos
182 ord <- order(soma,decreasing=T)
183 #=====

```

```

184 #           Índice de cautela (Caution Index) (Sato, 1975)           #
185 # (pode ter desvio-padrão igual a zero para X[n,] e X.star[n,])#
186 #=====#
187 X.star <- matrix(0,N,J)
188 CI <- rep(0,N)
189   for (n in 1:N) {
190     X.star[n,ord[1:escore[n]]] <- 1
191     CI[n] <- 1-cov(X[n,],prop)/cov(X.star[n,],prop)
192   }
193 CI[CI=="NaN"] <- 0
194 ## Calculo da area sob a curva ROC
195 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
196 # H(c)=sensibilidade = P( G > c | aberrante) ou P( G < c | aberrante)
197 Estat <- CI
198 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
199 FF <- HH <- rep(0,length(c.valor))
200 Estat.aberr <- Estat[1:NC]
201 Estat.normal <- Estat[(NC+1):N]
202 for (ind in 1:length(c.valor)) {
203   FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]])/(N-NC)
204   HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]])/NC
205 }
206 area <- -sum(diff(FF) * (HH[-length(HH)]+HH[-1])/2)
207 CI.rep[k] <- area
208 #=====#
209 # Índice de atencao modificado (Modified Caution Index)           #
210 #           (Harnisch e Linn, 1981)           #
211 #           desvio-padrão pode ser zero para X.star[n,] e X.line[n,]) #

```

```

212 #=====
213 X.line <- matrix(0,N,J)
214 MCI <- rep(0,N)
215   for (n in 1:N) {
216     X.line[n,ord[(J-escore[n]+1):J]] <- 1
217     MCI[n] <- ( cov(X.star[n,],prop) - cov(X[n,],prop) ) /
218               (cov(X.star[n,],prop) - cov(X.line[n,],prop) )
219   }
220 MCI[MCI=="NaN"] <- 0
221 ## Calculo da area sob a curva ROC
222 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
223 # H(c)=sensibilidade = P( G > c | aberrante) ou P( G < c | aberrante)
224 Estat <- MCI
225 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
226 FF <- HH <- rep(0,length(c.valor))
227 Estat.aberr <- Estat[1:NC]
228 Estat.normal <- Estat[(NC+1):N]
229 for (ind in 1:length(c.valor)) {
230   FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]]) / (N-NC)
231   HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]]) / NC
232 }
233 area <- -sum(diff(FF) * (HH[-length(HH)] + HH[-1])) / 2)
234 MCI.rep[k] <- area
235 #=====
236 #           U3 (van der Flier, 1980)           #
237 #=====
238 U3 <- rep(0,N)
239 lnX <- lnX.star <- lnX.line <- rep(0,N)

```

```

240 for (n in 1:N) {
241   lnX[n] <- sum(X[n,]*log(prop/(1-prop)))
242   lnX.star[n] <- sum(X.star[n,]*log(prop/(1-prop)))
243   lnX.line[n] <- sum(X.line[n,]*log(prop/(1-prop)))
244   U3[n] <- ( lnX.star[n] - lnX[n] )/( lnX.star[n] - lnX.line[
      n] )
245 }
246 U3[U3=="NaN"] <- 0
247 ## Calculo da area sob a curva ROC
248 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
249 # H(c)=sensibilidade = P( G > c | aberrante) ou P( G < c | aberrante)
250 Estat <- U3
251 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
252 FF <- HH <- rep(0,length(c.valor))
253 Estat.aberr <- Estat[1:NC]
254 Estat.normal <- Estat[(NC+1):N]
255   for (ind in 1:length(c.valor)) {
256     FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]])/(N-NC)
257     HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]])/NC
258   }
259 area <- -sum(diff(FF) * (HH[-length(HH)]+HH[-1])/2)
260 U3.rep[k] <- area
261 #=====#
262 #           ZU3 (van der Flier, 1982)           #
263 #=====#
264   ZU3 <- (U3-mean(U3))/sd(U3)
265 ## Calculo da area sob a curva ROC
266 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)

```

```

267 # H(c)=sensibilidade = P( G > c | aberrante) ou P( G < c | aberrante)
268 Estat <- ZU3
269 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
270 FF <- HH <- rep(0,length(c.valor))
271 Estat.aberr <- Estat[1:NC]
272 Estat.normal <- Estat[(NC+1):N]
273 for (ind in 1:length(c.valor)) {
274   FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]])/(N-NC)
275   HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]])/NC
276 }
277 area <- -sum(diff(FF) * (HH[-length(HH)]+HH[-1])/2)
278 ZU3.rep[k] <- area
279 #=====
280 #           H.line (Sijtsma, 1986)           #
281 #=====
282 H.line <- rep(0,N)
283 beta.n <- escore/J
284 beta.nm <- matrix(0,N,N-1)
285 for (n in 1:N) {
286   beta.nm[n,] <- apply(t(X[n,]*t(X[-n,])),1,sum)/J
287   H.line[n] <- sum(beta.nm[n,]-beta.n[n]*beta.n[-n])/
288     sum(apply(cbind(beta.n[n]*(1-beta.n[-n]),(1-beta.n[n])*
289       beta.n[-n]),1,min))
289 }
290 H.line[H.line=="NaN"] <- 1
291 ## Calculo da area sob a curva ROC
292 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
293 # H(c)=sensibilidade = P( G > c | aberrante) ou P( G < c | aberrante)

```

```

294 Estat <- H.line
295 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
296 FF <- HH <- rep(0,length(c.valor))
297 Estat.aberr <- Estat[1:NC]
298 Estat.normal <- Estat[(NC+1):N]
299 for (ind in 1:length(c.valor)) {
300     FF[ind] <- length(Estat.normal[Estat.normal<c.valor[ind]]) / (N-NC)
301     HH[ind] <- length(Estat.aberr[Estat.aberr<c.valor[ind]]) / NC
302 }
303 area <- sum(diff(FF) * (HH[-length(HH)] + HH[-1]) / 2)
304 H.line.rep[k] <- area
305 #=====#
306 #     Indice de concordancia A (Kane e Brennan, 1980)     #
307 #     Indice de discordancia D (Kane e Brennan, 1980)     #
308 #     Indice de "dependability" E (Kane e Brennan, 1980)  #
309 #=====#
310 A <- D <- E <- rep(0,N)
311 for (n in 1:N) A[n] <- sum(X[n,]*prop)
312 D <- sum(prop) - A
313 E <- A/sum(prop)
314 ## Calculo da area sob a curva ROC
315 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
316 # H(c)=sensibilidade = P( G > c | aberrante) ou P( G < c | aberrante)
317 Estat <- A
318 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
319 FF <- HH <- rep(0,length(c.valor))
320 Estat.aberr <- Estat[1:NC]
321 Estat.normal <- Estat[(NC+1):N]

```

```
322 for (ind in 1:length(c.valor)) {
323   FF[ind] <- length(Estat.normal[Estat.normal<c.valor[ind]]) / (N-NC)
324   HH[ind] <- length(Estat.aberr[Estat.aberr<c.valor[ind]]) / NC
325 }
326 area <- sum(diff(FF) * (HH[-length(HH)] + HH[-1])) / 2)
327 A.rep[k] <- area
328 ##
329 Estat <- D
330 c.valor <- seq(min(Estat), max(Estat), length.out=1000)
331 FF <- HH <- rep(0, length(c.valor))
332 Estat.aberr <- Estat[1:NC]
333 Estat.normal <- Estat[(NC+1):N]
334 for (ind in 1:length(c.valor)) {
335   FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]]) / (N-NC)
336   HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]]) / NC
337 }
338 area <- -sum(diff(FF) * (HH[-length(HH)] + HH[-1])) / 2)
339 D.rep[k] <- area
340 ##
341 Estat <- E
342 c.valor <- seq(min(Estat), max(Estat), length.out=1000)
343 FF <- HH <- rep(0, length(c.valor))
344 Estat.aberr <- Estat[1:NC]
345 Estat.normal <- Estat[(NC+1):N]
346 for (ind in 1:length(c.valor)) {
347   FF[ind] <- length(Estat.normal[Estat.normal<c.valor[ind]]) / (N-NC)
348   HH[ind] <- length(Estat.aberr[Estat.aberr<c.valor[ind]]) / NC
349 }
```

```

350 area <- sum(diff(FF) * (HH[-length(HH)] + HH[-1]) / 2)
351 E.rep[k] <- area
352 #=====
353 #           Regressao isotonica           #
354 #=====
355 X.aux <- X[,order(delta.est,decreasing=T)]
356 ## Estatistica que mede a discrepancia entre as respostas perfeitas
357 ## (somente 0's, seguidos somente de 1's) e a regressao isotonica
358 ## das respostas ordenadas pelos valores decrescentes do parametro
359 ## de dificuldade, para todos os individuos
360 med.disc.vec <- rep(0,N)
361 for (n in 1:N) {
362   ri <- reg.isot(X.aux[n,],c(1:J))
363   med.disc.vec[n] <- sum(abs(ri-c(rep(0,(J-sum(X[n,]))),
364                               rep(1,sum(X[n,])))))
365 }
366 med.disc.vec <- med.disc.vec/J
367 ## Calculo da area sob a curva ROC
368 # F(c) = 1 - especificidade = P(G > c | normal) ou P(G < c | normal)
369 # H(c)=sensibilidade = P(G > c | aberrante) ou P(G < c | aberrante)
370 Estat <- med.disc.vec
371 c.valor <- seq(min(Estat),max(Estat),length.out=1000)
372 FF <- HH <- rep(0,length(c.valor))
373 Estat.aberr <- Estat[1:NC]
374 Estat.normal <- Estat[(NC+1):N]
375 for (ind in 1:length(c.valor)) {
376   FF[ind] <- length(Estat.normal[Estat.normal>c.valor[ind]]) / (N-NC)
377   HH[ind] <- length(Estat.aberr[Estat.aberr>c.valor[ind]]) / NC

```

```
378     }
379     area <- -sum(diff(FF) * (HH[-length(HH)] + HH[-1]) / 2)
380     med.disc.vec.rep[k] <- area
381   }
382   ## Para tabela (latex)
383   areas.G<-c(areas.G, paste(round(mean(G.rep), 5), "(", round(sd(G.rep), 5)
384     , ")"))
385   areas.G.star<-c(areas.G.star, paste(round(mean(G.star.rep), 5), "(",
386     round(sd(G.star.rep), 5), ")"))
387   areas.r.pbis<-c(areas.r.pbis, paste(round(mean(r.pbis.rep), 5), "(",
388     round(sd(r.pbis.rep), 5), ")"))
389   areas.CI<-c(areas.CI, paste(round(mean(CI.rep), 5), "(", round(sd(CI.rep
390     ), 5), ")"))
391   areas.MCI<-c(areas.MCI, paste(round(mean(MCI.rep), 5), "(", round(sd(MCI.
392     rep), 5), ")"))
393   areas.U3<-c(areas.U3, paste(round(mean(U3.rep), 5), "(", round(sd(U3.rep
394     ), 5), ")"))
395   areas.ZU3<-c(areas.ZU3, paste(round(mean(ZU3.rep), 5), "(", round(sd(ZU3.
396     rep), 5), ")"))
397   areas.H.line<-c(areas.H.line, paste(round(mean(H.line.rep), 5), "(",
398     round(sd(H.line.rep), 5), ")"))
399   areas.A<-c(areas.A, paste(round(mean(A.rep), 5), "(", round(sd(A.rep), 5)
400     , ")"))
401   areas.Rn<-c(areas.Rn, paste(round(mean(med.disc.vec.rep), 5), "(", round(
402     sd(med.disc.vec.rep), 5), ")"))
403   ## Para contrucao de gráfico
404   Areas.G<-c(Areas.G, mean(G.rep))
405   Areas.G.star<-c(Areas.G.star, mean(G.star.rep))
```

```
396 Areas.r.pbis<-c(Areas.r.pbis, mean(r.pbis.rep))
397 Areas.CI<-c(Areas.CI, mean(CI.rep))
398 Areas.MCI<-c(Areas.MCI, mean(MCI.rep))
399 Areas.U3<-c(Areas.U3, mean(U3.rep))
400 Areas.ZU3<-c(Areas.ZU3, mean(ZU3.rep))
401 Areas.H.line<-c(Areas.H.line, mean(H.line.rep))
402 Areas.A<-c(Areas.A, mean(A.rep))
403 Areas.Rn<-c(Areas.Rn, mean(med.disc.vec.rep))
404     }
405   }
406   if (tabel==1){
407     mat.dados<-rbind(areas.G,areas.G.star,areas.r.pbis,areas.CI,areas.
408       MCI,areas.U3,areas.ZU3,areas.H.line,areas.A,areas.Rn)
409     nomes<-rep(c("G","G*","rpbis","CI","MCI","U3","ZU3","Hprime","A","Rn"
410       ),4)
411     coluna1.table<-c(mat.dados[,1], mat.dados[,2], mat.dados[,3],
412       mat.dados[,4])
413     coluna2.table<-c(mat.dados[,5], mat.dados[,6], mat.dados[,7],
414       mat.dados[,8])
415     coluna3.table<-c(mat.dados[,9], mat.dados[,10], mat.dados[,11],
416       mat.dados[,12])
417     tabela<-cbind(nomes,coluna1.table,coluna2.table,coluna3.table)
418     rownames(tabela)<-NULL
419     library(xtable)
420     xtable(tabela)
421   }else{
422     Mat.dados<-rbind(Areas.G,Areas.G.star,Areas.r.pbis,Areas.CI,Areas.
423       MCI,Areas.U3,Areas.ZU3,Areas.H.line,Areas.A,Areas.Rn)
```

```
421   Mat.dados
422   }
423 }
424 #bd.careless1<-aberrantes(1,1,2) # careless; 1PL e tabela p/ grafico
425 #bd.careless2<-aberrantes(1,2,2) # careless; 2PL e tabela p/ grafico
426 bd.careless3<-aberrantes(1,3,2) # careless; 3PL e tabela p/ grafico
427
428 J<-as.factor(rep(c(17,33,65),each=40))
429 Porc<-rep(as.factor(rep(c(0.05,0.10,0.25,0.50), each=10)),3)
430 Estat<-rep(c("G", "G*", "rpbis", "CI", "MCI", "U3", "ZU3", "Hprime", "A", "Rn"
431             ),12)
432 ddd.careless1<-data.frame(J, Porc, Area.Roc=c(bd.careless1[,1:12]),
433                           Estat)
434 ddd.careless2<-data.frame(J, Porc, Area.Roc=c(bd.careless2[,1:12]),
435                           Estat)
436 ddd.careless3<-data.frame(J, Porc, Area.Roc=c(bd.careless3[,1:12]),
437                           Estat)
438 x11()
439 library(ggplot2)
440 # MODELO GGLOT PARA PORCENTAGEM
441 ggplot(data= ddd.careless3, aes(x=J, y=Area.Roc, color= Estat)) +
442   geom_point(size=2, aes(shape=Porc), position=position_dodge(width
443     =0.7))+
444   geom_vline(xintercept = c(1.5,2.5))+
445   labs(title = "Respondentes Descuidados", y="Area ROC",
446         x="J (comprimento do teste)")+
447   theme(plot.title = element_text(hjust = 0.5, face="bold"), axis.text=
448     element_text(size=14), axis.title=element_text(size=14, face="bold
```

```
"), legend.text = element_text(size=12))+  
446 scale_color_manual(labels = expression(A, CI, G, G^"*", H^T, MCI, R[n],  
447 r[pbis], U3, ZU3), values = c("orange", "pink", "navy blue", "  
black", "blue", "red", "yellow", "green", "purple", "brown"))  
448 path<-getwd()  
449 ggsave(filename= "descuidados3PLJ.pdf", path = path, width = 6,  
450 height = 5, device='pdf', dpi=700)  
451 # MODELO GGLOT PARA COMPRIMENTO DE TESTE  
452 ggplot(data= ddd.careless3, aes(x=Porc, y=Area.Roc, color= Estat)) +  
453 geom_point(size=2, aes(shape=J), position=position_dodge(width=0.8)) +  
454 geom_vline(xintercept = c(1.5, 2.5, 3.5)) +  
455 labs(title = "Respondentes Descuidados", y="Area ROC",  
456 x="Porcentagem") +  
457 theme(plot.title = element_text(hjust = 0.5, face="bold"), axis.text=  
element_text(size=14), axis.title=element_text(size=14, face="bold"  
) , legend.text = element_text(size=12)) +  
458 scale_color_manual(labels=expression(A, CI, G, G^"*", H^T, MCI, R[n], r[  
pbis], U3, ZU3), values = c("orange", "pink", "navy blue", "black", "  
blue", "red", "yellow", "green", "purple", "brown"))  
459 path<-getwd()  
460 ggsave(filename= "descuidados3PLPORC.pdf", path = path, width = 6,  
461 height = 5, device='pdf', dpi=700)
```

A.2 Programação para Dados Reais

```
1 setwd("C:\\Users\\thays\\Documents") ; getwd()
2 saresp <- read.table(file="saresp.txt")
3 gab2 <- read.table(file="saresp-gabarito.txt")
4 dados <- saresp
5 manha <- dados[dados[,4]=="m07",]
6 tarde <- dados[dados[,4]=="t07",]
7 noite <- dados[dados[,4]=="n07",]
8 gab3 <- matrix(9,nrow(gab2),ncol(gab2))
9 for (i in 1:nrow(gab3)) {
10   for (j in 1:ncol(gab3)) {
11     if (gab2[i,j]=="A") gab3[i,j] <- 1
12     if (gab2[i,j]=="B") gab3[i,j] <- 2
13     if (gab2[i,j]=="C") gab3[i,j] <- 3
14     if (gab2[i,j]=="D") gab3[i,j] <- 4
15   }
16 }
17 resp.manha <- manha[,5:34] # resposta aos 30 itens dos 1001 resp.
18 resp.manha <- as.matrix(resp.manha)
19 resp.m <- matrix(9,nrow(resp.manha),ncol(resp.manha))
20 resp.m[resp.manha=="A"] <- 1
21 resp.m[resp.manha=="B"] <- 2
22 resp.m[resp.manha=="C"] <- 3
23 resp.m[resp.manha=="D"] <- 4
24 for (i in 1:nrow(resp.m)) {
25   for (j in 1:ncol(resp.m)) {
26     if ((resp.m[i,j]!=gab3[1,j])&&(resp.m[i,j]!=9)) resp.m[i,j] <- 0
27     if (resp.m[i,j]==gab3[1,j]) resp.m[i,j] <- 1
```

```
28 }
29 }
30 resp.m[resp.m==9] <- 0
31 resp.tarde <- tarde[,5:34]
32 resp.tarde <- as.matrix(resp.tarde)
33 resp.t <- matrix(9,nrow(resp.tarde),ncol(resp.tarde))
34 resp.t[resp.tarde=="A"] <- 1
35 resp.t[resp.tarde=="B"] <- 2
36 resp.t[resp.tarde=="C"] <- 3
37 resp.t[resp.tarde=="D"] <- 4
38 for (i in 1:nrow(resp.t)) {
39   for (j in 1:ncol(resp.t)) {
40     if ((resp.t[i,j]!=gab3[2,j])&&(resp.t[i,j]!=9)) resp.t[i,j] <- 0
41     if (resp.t[i,j]==gab3[2,j]) resp.t[i,j] <- 1
42   }
43 }
44 resp.t[resp.t==9] <- 0
45 resp.noite <- noite[,5:34]
46 resp.noite <- as.matrix(resp.noite)
47 resp.n <- matrix(9,nrow(resp.noite),ncol(resp.noite))
48 resp.n[resp.noite=="A"] <- 1
49 resp.n[resp.noite=="B"] <- 2
50 resp.n[resp.noite=="C"] <- 3
51 resp.n[resp.noite=="D"] <- 4
52 for (i in 1:nrow(resp.n)) {
53   for (j in 1:ncol(resp.n)) {
54     if ((resp.n[i,j]!=gab3[3,j])&&(resp.n[i,j]!=9)) resp.n[i,j] <- 0
55     if (resp.n[i,j]==gab3[3,j]) resp.n[i,j] <- 1
```

```

56   }
57 }
58 resp.n[resp.n==9] <- 0
59 #=====
60 #           Funcao "reg.isot"           #
61 # Calcula a regressao isotonica da funcao "g" com pesos "w" #
62 #=====
63 reg.isot <- function(g,w) {
64   n <- length(g)
65   gw <- g*w
66   gwac <- cumsum(gw)
67   wac <- cumsum(w)
68   matslopes <- cbind(c(1:n),gwac/wac)
69   indice <- indaux <- 1
70   ### "isot" e' o vetor com os valores da regressao isotonica
71   isot <- rep(0,n)
72   while(indice < n) {
73     indice <- max(matslopes[matslopes[,2]==min(matslopes[,2]),1])
74     isot[indaux:indice] <- min(matslopes[,2])
75     indaux <- indice + 1
76     matslopes <- cbind(c(indaux:n), (gwac[indaux:n]-gwac[indice]) / (wac
77       [indaux:n]-wac[indice]))
77   }
78   return(isot)
79 }
80 ## numero de itens
81 J <- ncol(resp.m)
82 ## numero de respondentes

```

```

83 N <- nrow(resp.m)
84 # Respostas
85 (X <- resp.m)
86 # estimativas das proficiencias e dos parametros de dificuldade
87 theta.est <- log((apply(X,1,sum)+0.1)/(J-apply(X,1,sum)))
88 delta.est <- log((N-apply(X,2,sum))/apply(X,2,sum))
89 #=====#
90 #=====  ESTATISTICAS NAO PARAMETRICAS  =====#
91
92 #=====#
93 #           G (Guttman, 1944, 1950)           #
94 #=====#
95 ## Numero de erros de respostas (Guttman, 1944, 1950)
96 G <- rep(0,N) ; soma <- apply(X,2,sum)
97 for (j1 in 1:(J-1)) {
98   for (j2 in (j1+1):J) {
99     G <- G + X[,j1]*(1 - X[,j2])*sign(sign(soma[j2]-soma[j1]-1)+1)+
100       (1 - X[,j1])*X[,j2]*sign(sign(soma[j1]-soma[j2]-1)+1)
101   }
102 }
103 #=====#
104 #           G normado (van der Flier, 1977)           #
105 #=====#
106 escore <- apply(X,1,sum)
107 G.star <- G/(escore*(J-escore))
108 G.star[G.star=="NaN"] <- 0
109 #=====#
110 ## Estatistica ponto bisserial pessoal (Donlon e Fischer, 1968)#

```

```

111 #           (pode ter desvio-padrão igual a zero para X[n,])           #
112 #=====#
113 prop <- soma/N
114 r.pbis <- rep(0,N)
115 for (n in 1:N) r.pbis[n] <- cor(X[n,],prop)
116 r.pbis[is.na(r.pbis)] <- 1
117 ## Posicao dos itens em ordem decrescente das somas de acertos
118 ord <- order(soma,decreasing=T)
119 #=====#
120 #           Índice de cautela (Caution Index) (Sato, 1975)           #
121 # (pode ter desvio-padrão igual a zero para X[n,] e X.star[n,])#
122 #=====#
123 X.star <- matrix(0,N,J)
124 CI <- rep(0,N)
125 for (n in 1:N) {
126   X.star[n,ord[1:escore[n]]] <- 1
127   CI[n] <- 1-cov(X[n,],prop)/cov(X.star[n,],prop)
128 }
129 CI[CI=="NaN"] <- 0
130 #=====#
131 #           Índice de atencao modificado (Modified Caution Index)     #
132 #           (Harnisch e Linn, 1981)                                     #
133 # (pode ter desvio-padrão igual a zero para X.star[n,] e X.line[n,])#
134 #=====#
135 X.line <- matrix(0,N,J)
136 MCI <- rep(0,N)
137 for (n in 1:N) {
138   X.line[n,ord[(J-escore[n]+1):J]] <- 1

```

```

139   MCI[n] <- ( cov(X.star[n,],prop) - cov(X[n,],prop) ) / ( cov(X.star[n
      ],prop) - cov(X.line[n,],prop) )
140 }
141 MCI[MCI=="NaN"] <- 0
142 #=====
143 #           U3 (van der Flier, 1980)           #
144 #=====
145 U3 <- rep(0,N)
146 lnX <- lnX.star <- lnX.line <- rep(0,N)
147 for (n in 1:N) {
148   lnX[n] <- sum(X[n,]*log(prop/(1-prop)))
149   lnX.star[n] <- sum(X.star[n,]*log(prop/(1-prop)))
150   lnX.line[n] <- sum(X.line[n,]*log(prop/(1-prop)))
151   U3[n] <- ( lnX.star[n] - lnX[n] ) / ( lnX.star[n] - lnX.line[n] )
152 }
153 U3[U3=="NaN"] <- 0
154 #=====
155 #           ZU3 (van der Flier, 1982)           #
156 #=====
157 ZU3 <- (U3-mean(U3))/sd(U3)
158 #=====
159 #           H.line (Sijtsma, 1986)           #
160 #=====
161 H.line <- rep(0,N)
162 beta.n <- escore/J
163 beta.nm <- matrix(0,N,N-1)
164 for (n in 1:N) {
165   beta.nm[n,] <- apply(t(X[n,]*t(X[-n,])),1,sum)/J

```

```

166   H.line[n] <- sum(beta.nm[n,]-beta.n[n]*beta.n[-n])/sum(apply(cbind(
      beta.n[n]*(1-beta.n[-n]),(1-beta.n[n])*beta.n[-n]),1,min))
167 }
168 H.line[H.line=="NaN"] <- 1
169 #=====
170 #       Indice de concordancia A (Kane e Brennan, 1980)           #
171 #       Indice de discordancia D (Kane e Brennan, 1980)         #
172 #       Indice de "dependability" E (Kane e Brennan, 1980)      #
173 #=====
174 A <- D <- E <- rep(0,N)
175 for (n in 1:N) A[n] <- sum(X[n,]*prop)
176 D <- sum(prop) - A
177 E <- A/sum(prop)
178 #=====
179 #                   Regressao isotonica                           #
180 #=====
181 X.aux <- X[,order(delta.est,decreasing=T)]
182 ## Estatistica que mede a discrepancia entre as respostas perfeitas
183 ## (somente 0's, seguidos somente de 1's) e a regressao isotonica
184 ## das respostas ordenadas pelos valores decrescentes do parametro
185 ## de dificuldade, para todos os individuos
186 med.disc.vec <- rep(0,N)
187 for (n in 1:N) {
188   ri <- reg.isot(X.aux[n,],c(1:J))
189   med.disc.vec[n] <- sum(abs(ri-c(rep(0,(J-sum(X[n,]))),
190                               rep(1,sum(X[n,])))))
191 }
192 med.disc.vec <- med.disc.vec/J

```

```
193 #=====#
194 #           Cálculo dos p-valores individuais           #
195 #=====#
196 #install.packages("irtoys")
197 library(irtoys)
198 ##### Obtencao das estimativas dos parametros dos itens para turno
        da manha via pacote "irtoys"
199 #resp.m.tpm <- tpm(resp.m,constraint=cbind(1:ncol(resp.m),1,0.25))
200 (resp.m.tpm <- tpm(resp.m))
201 (par.m.est <- coef(resp.m.tpm)) # cc, bb, aa
202 ### estimacao da proficiencia (funcao "eap" serve apenas para o
        modelo 3PL)
203 ### matriz com a estimativa da proficiencia e o erro-padrão (1/sqrt(
        informacao(theta.est)))
204 (theta.m.est <- eap(resp.m, cbind(par.m.est[,3],par.m.est[,2],par.m.
        est[,1]), qu=normal.qu()))
205 prof.m.est <- theta.m.est[,1]
206 G.real <- G
207 G.star.real <- G.star
208 r.pbis.real <- r.pbis
209 CI.real <- CI
210 MCI.real <- MCI
211 U3.real <- U3
212 ZU3.real <- ZU3
213 H.line.real <- H.line
214 A.real <- A
215 D.real <- D
216 E.real <- E
```

```

217 med.disc.vec.real <- med.disc.vec
218 n.rep <- 1000 # repeticoes para gerar o p-valor
219 p.G <- p.G.star <- p.r.pbis <- p.CI <- p.MCI <- p.U3 <- p.ZU3 <- p.H.
      line <- p.A <- p.D <- p.E <- p.med.disc.vec <- rep(0,J)
220 a <- par.m.est[,3]
221 delta <- par.m.est[,2]
222 cc <- par.m.est[,1]
223 for (i in 1:n.rep) {
224   print(i)
225   # respostas
226   X <- matrix(0,N,J)
227   for (j in 1:J) {
228     prob <- cc[j]+(1-cc[j])/(1+exp(-a[j]*(prof.m.est-delta[j])))
229     X[,j] <- rbinom(N,1,prob)
230   }
231   ## Numero de erros de respostas (Guttman, 1944, 1950)
232   G <- rep(0,N)
233   soma <- apply(X,2,sum)
234   for (j1 in 1:(J-1)) {
235     for (j2 in (j1+1):J) {
236       G <- G + X[,j1]*(1 - X[,j2])*sign(sign(soma[j2]-soma[j1]-1)+1)+
237         (1 - X[,j1])*X[,j2]*sign(sign(soma[j1]-soma[j2]-1)+1)
238     }
239   }
240   p.G <- p.G + sign(sign(G-G.real)+1)
241   #=====#
242   #           G normado (van der Flier, 1977)           #
243   #=====#

```

```

244  escore <- apply(X,1,sum) ; G.star <- G/(escore*(J-escore))
245  G.star[G.star=="NaN"] <- 0
246  p.G.star <- p.G.star + sign(sign(G.star-G.star.real)+1)
247  #=====#
248  ## Estatistica ponto bisserial pessoal (Donlon e Fischer, 1968)#
249  ## (pode ter desvio-padrao igual a zero para X[n,]) #
250  #=====#
251  prop <- soma/N ; r.pbis <- rep(0,N)
252  for (n in 1:N) r.pbis[n] <- cor(X[n,],prop)
253  r.pbis[is.na(r.pbis)] <- 1
254  p.r.pbis <- p.r.pbis + sign(sign(r.pbis.real-r.pbis)+1)
255  ## Posicao dos itens em ordem decrescente das somas de acertos
256  ord <- order(soma,decreasing=T)
257  ## Indice de cautela (Caution Index) (Sato, 1975)
258  ## (pode ter desvio-padrao igual a zero para X[n,] e X.star[n,])
259  X.star <- matrix(0,N,J)
260  CI <- rep(0,N)
261  for (n in 1:N) {
262    X.star[n,ord[1:escore[n]]] <- 1
263    CI[n] <- 1-cov(X[n,],prop)/cov(X.star[n,],prop)
264  }
265  CI[CI=="NaN"] <- 0
266  p.CI <- p.CI + sign(sign(CI-CI.real)+1)
267  #=====#
268  #   Indice de atencao modificado (Modified Caution Index) #
269  #           (Harnisch e Linn, 1981) #
270  #   desvio-padrao pode ser zero para X.star[n,] e X.line[n,]) #
271  #=====#

```

```

272 X.line <- matrix(0,N,J)
273 MCI <- rep(0,N)
274 for (n in 1:N) {
275   X.line[n,ord[(J-escore[n]+1):J]] <- 1
276   MCI[n] <- ( cov(X.star[n,],prop) - cov(X[n,],prop) ) / ( cov(X.star
      [n,],prop) - cov(X.line[n,],prop) )
277 }
278 MCI[MCI=="NaN"] <- 0
279 p.MCI <- p.MCI + sign(sign(MCI-MCI.real)+1)
280 #=====
281 #           U3 (van der Flier, 1980)           #
282 #=====
283 U3 <- rep(0,N)
284 lnX <- lnX.star <- lnX.line <- rep(0,N)
285 for (n in 1:N) {
286   lnX[n] <- sum(X[n,]*log(prop/(1-prop)))
287   lnX.star[n] <- sum(X.star[n,]*log(prop/(1-prop)))
288   lnX.line[n] <- sum(X.line[n,]*log(prop/(1-prop)))
289   U3[n] <- ( lnX.star[n] - lnX[n] ) / ( lnX.star[n] - lnX.line[n] )
290 }
291 U3[U3=="NaN"] <- 0
292 p.U3 <- p.U3 + sign(sign(U3-U3.real)+1)
293 #=====
294 #           ZU3 (van der Flier, 1982)           #
295 #=====
296 ZU3 <- (U3-mean(U3))/sd(U3)
297 p.ZU3 <- p.ZU3 + sign(sign(ZU3-ZU3.real)+1)
298 #=====

```

```

299 #           H.line (Sijtsma, 1986)           #
300 #=====
301 H.line <- rep(0,N)
302 beta.n <- escore/J
303 beta.nm <- matrix(0,N,N-1)
304 for (n in 1:N) {
305     beta.nm[n,] <- apply(t(X[n,]*t(X[-n,])),1,sum)/J
306     H.line[n] <- sum(beta.nm[n,]-beta.n[n]*beta.n[-n])/
307         sum(apply(cbind(beta.n[n]*(1-beta.n[-n]),
308             (1-beta.n[n])*beta.n[-n]),1,min))
309 }
310 H.line[H.line=="NaN"] <- 1
311 p.H.line <- p.H.line + sign(sign(H.line.real-H.line)+1)
312 #=====
313 ## Indice de concordancia A (Kane e Brennan, 1980)
314 ## Indice de discordancia D (Kane e Brennan, 1980)
315 ## Indice de "dependability" E (Kane e Brennan, 1980)
316 #=====
317 A <- D <- E <- rep(0,N)
318 for (n in 1:N) A[n] <- sum(X[n,]*prop)
319 D <- sum(prop) - A
320 E <- A/sum(prop)
321 p.A <- p.A + sign(sign(A-A.real)+1)
322 p.D <- p.D + sign(sign(D-D.real)+1)
323 p.E <- p.E + sign(sign(E.real-E)+1)
324 #=====
325 #           Regressao isotonica           #
326 #=====

```

```
327 X.aux <- X[,order(delta.est,decreasing=T)]
328 ## Estatística que mede a discrepancia entre as respostas perfeitas
329 ## (somente 0's, seguidos somente de 1's) e a regressao isotonica
330 ## das respostas ordenadas pelos valores decrescentes do parametro
331 ## de dificuldade, para todos os individuos
332 med.disc.vec <- rep(0,N)
333 for (n in 1:N) {
334   ri <- reg.isot(X.aux[n,],c(1:J))
335   med.disc.vec[n] <- sum(abs(ri-c(rep(0,(J-sum(X[n,]))),
336                               rep(1,sum(X[n,])))))
337 }
338 med.disc.vec <- med.disc.vec/J
339 p.med.disc.vec <- p.med.disc.vec +
340   sign(sign(med.disc.vec-med.disc.vec.real)+1)
341 }
342 p.G <- p.G/n.rep
343 p.G.star <- p.G.star/n.rep
344 p.r.pbis <- p.r.pbis/n.rep
345 p.CI <- p.CI/n.rep
346 p.MCI <- p.MCI/n.rep
347 p.U3 <- p.U3/n.rep
348 p.ZU3 <- p.ZU3/n.rep
349 p.H.line <- p.H.line/n.rep
350 p.A <- p.A/n.rep
351 p.D <- p.D/n.rep
352 p.E <- p.E/n.rep
353 p.med.disc.vec <- p.med.disc.vec/n.rep
354 p.valor.mat <- cbind(c(1:N),p.G,p.G.star,p.r.pbis,p.CI,p.MCI,p.U3,
```

```

355         p.ZU3,p.H.line,p.A,p.med.disc.vec)
356 write(t(p.valor.mat), file="p-valor-matriz.txt", ncol=11)
357 p.valor.Est<-matrix(c(p.G,p.G.star,p.r.pbis,p.CI,p.MCI,p.U3,
358         p.ZU3,p.H.line,p.A,p.med.disc.vec), ncol=10)
359 x11()
360 plot(p.G)
361 for (j in 2:10){
362     lines(p.valor.Est[,j])
363 }
364 p.valor.mat <- read.table(file="p-valor-matriz.txt")
365 p.valor.mat<-p.valor.mat[,-c(1,11,12)]
366 #=====#
367 #     Sensibilidade baseada nos p-valores     #
368 #=====#
369 (qtd.aberr<-apply(p.valor.mat <= 0.01,2,sum))
370 (qtd.n.aberr<-apply(p.valor.mat > 0.01,2,sum))
371 (sensibilidade<-qtd.aberr/1001)
372 Sens<-rbind(qtd.aberr,qtd.n.aberr,sensibilidade)
373 colnames(Sens)<-c("G","Gstar","rpbis","CI", "MCI","U3","ZU3","Hline",
374         "A","Rn" )
375 library(xtable)
376 mat.dec<-matrix( c(rep(0,2*(dim(Sens)[2]+1)),rep(4,(dim(Sens)[2]+1)))
377         ,nrow=3,byrow=TRUE )
378 xtable(Sens, digits = mat.dec)

```