



Instituto de
MATEMÁTICA
E ESTATÍSTICA

UFRGS



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DEPARTAMENTO DE ESTATÍSTICA

**TRIDIF: UM APLICATIVO PARA TEORIA DE RESPOSTA AO ITEM E FUNCIONAMENTO
DIFERENCIAL DO ITEM**

MARIANA LIMA GARCIA

Porto Alegre
2021

MARIANA LIMA GARCIA

**TRIDIF: UM APLICATIVO PARA TEORIA DE RESPOSTA AO ITEM E
FUNCIONAMENTO DIFERENCIAL DO ITEM**

Trabalho de Conclusão submetido à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Profa. Dra. Stela Maris de Jesus Castro

Porto Alegre
2021

CIP - Catalogação na Publicação

Garcia, Mariana Lima
TRIDIF: UM APLICATIVO PARA TEORIA DE RESPOSTA AO
ITEM E FUNCIONAMENTO DIFERENCIAL DO ITEM / Mariana
Lima Garcia. -- 2021.
55 f.
Orientadora: Stela Maris de Jezus Castro.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto
de Matemática e Estatística, Curso de Estatística,
Porto Alegre, BR-RS, 2021.

1. Teoria da resposta ao item. 2. Funcionamento
diferencial do item. 3. Shiny. 4. TRI. 5. DIF. I.
Castro, Stela Maris de Jezus, orient. II. Título.

Instituto de Matemática e Estatística
Departamento de Estatística

**TRIDIF: Um aplicativo para Teoria de Resposta ao Item e
Funcionamento Diferencial do Item**
MARIANA LIMA GARCIA

Banca examinadora:

Rogério Boff Borges
UFRGS/HCPA

Stela Maris de Jezus Castro
UFRGS/IME

AGRADECIMENTOS

Agradeço à minha orientadora, professora Stela, por toda sua atenção e por ter compartilhado seus conhecimentos comigo durante a realização deste trabalho. Ao professor Marcio Valk pela compreensão e disponibilidade que sempre me despendeu. À professora Gabriela Cybis por ter me apresentado ao pacote *shiny* e aos gráficos do *plotly* durante o desenvolvimento do aplicativo que simula epidemias. A todos os professores da minha caminhada acadêmica que me serviram de exemplo e desenvolveram em mim o gosto pela busca de conhecimento que tenho hoje. E, em especial agradeço ao meu marido Alexandre Leite, que sempre me incentivou, além de me apoiar e auxiliar em tudo o que faço.

RESUMO

É crescente a utilização da teoria de resposta ao item (TRI) para análises de testes, principalmente em grandes avaliações educacionais e psicológicas. A necessidade de softwares estatísticos proprietários ou softwares gratuitos que requerem conhecimento de sua linguagem para a realização dessas análises, por vezes dificultam o trabalho de pesquisadores. Neste trabalho é proposto um aplicativo para a realização de todas as etapas de uma análise TRI. Desenvolvido em R Core Team (2020), a aplicação web criada a partir do pacote shiny (Chang et al., 2021) possibilita ao pesquisador uma interface amigável e interativa, com download facilitado dos resultados dos ajustes de modelos em tabelas e gráficos. O aplicativo e o código desenvolvido estão disponíveis em repositórios públicos.

Palavras-chave: Teoria da resposta ao item. TRI. Funcionamento diferencial do item. DIF. Shiny.

ABSTRACT

The use of item response theory (IRT) for tests analysis is increasing, mostly in field like educational and psychological assessments. The necessity of proprietary statistical softwares or free softwares that require knowledge of their language to carry out these analyses, sometimes hinders the researchers' work. In this work, an application is proposed to carry out all the steps of a TRI analysis. Developed in R Core Team (2020), has a web application created from the shiny package (Chang et al., 2021) provides the researcher an interface user-friendly and interactive, with easy download of the results of models adjustments in tables and graphs. The application and all the code developed are available at public repositories.

Keywords: Item Response Theory. IRT. Differential item functioning DIF. Shiny.

LISTA DE FIGURAS

Figura 1: Tela inicial do aplicativo	12
Figura 2: Topo de todas as telas, informações gerais	12
Figura 3: Menus do aplicativo.....	13
Figura 4: Exemplo de <i>box</i> retrátil do aplicativo.....	13
Figura 5: Exemplo de abas.....	14
Figura 6: Exemplo da aba “Ajuda”	14
Figura 7: Caixa informativa com o tipo de dado lido pelo aplicativo	15
Figura 8: <i>Box</i> “Matriz de dados” do menu “Entrada de dados”	16
Figura 9: Exclusão de Item.....	17
Figura 10: Menu Análises descritivas	19
Figura 11: Informação textual do Menu Análises descritivas	20
Figura 12: Tabela “Valores faltantes em cada item”	20
Figura 13: Tabela “Proporção de respostas” para itens politômicos	20
Figura 14: Tabela “Alfa de Cronbach”	21
Figura 15: Menu Unidimensionalidade	23
Figura 16: Exemplo de curva característica do item para itens dicotômicos	26
Figura 17: Exemplo de curva de informação do item para itens dicotômicos.....	27
Figura 18: Exemplo de curva de informação e erro padrão de estimação de teste...28	
Figura 19: Exemplo de curva característica do item para itens politômicos	31
Figura 20: Exemplo de curva de informação de item para itens politômicos.....	32
Figura 21: Menu Ajuste do modelo.....	33
Figura 22: Itens com DIF e escolha de itens para exclusão do modelo	34
Figura 23: <i>Box</i> Traços latentes.....	35
Figura 24: <i>Box</i> “Itens do modelo” para modelos dicotômicos.....	37
Figura 25: <i>Box</i> “Itens do modelo” para modelos politômicos.....	38
Figura 26: Visualização da detecção de DIF	44
Figura 27: Menu “Presença de DIF”	45
Figura 28: Opção de escolha “Métodos de Correção”	46
Figura 29: <i>Box</i> “Detecção dos Itens” com DIF.....	47
Figura 30: <i>Box</i> “Análise gráfica” DIF em itens politômicos	47
Figura 31: Menu “Exportação de dados”	49

SUMÁRIO

1. INTRODUÇÃO	9
2. TERMOS E ESTRUTURA DO APLICATIVO	11
3. MENU ENTRADA DE DADOS	16
4. MENU ANÁLISES DESCRITIVAS	18
5. MENU UNIDIMENSIONALIDADE	22
6. MENU AJUSTE DO MODELO	24
6.1 MODELOS DICOTÔMICOS	24
6.2 MODELOS POLITÔMICOS ORDINAIS	28
6.3 AJUSTE DE MODELOS NO APLICATIVO TRIDIF	32
7. MENU PRESENÇA DE DIF	39
7.1 MÉTODOS DE DETECÇÃO DE DIF	40
7.2 DETECÇÃO DE DIF NO APLICATIVO TRIDIF	44
8. MENU EXPORTAÇÃO DE DADOS	48
9. CONCLUSÃO	50
REFERÊNCIAS	51

1. INTRODUÇÃO

A teoria de resposta ao item (TRI) vem sendo amplamente utilizada como método de estimação de traços latentes, pois permite que os itens que compõem o instrumento de medida (ou teste) sejam avaliados individualmente de modo que possam contribuir de modo diferenciado na estimativa dos mesmos.

De acordo com Castro (2008) a TRI compreende um grupo de modelos lineares generalizados e procedimentos estatísticos associados. Para um bom ajuste desses modelos, são necessárias algumas etapas: a primeira delas é a análise descritiva, onde se tem uma ideia inicial das respostas aos itens do teste; a segunda é uma etapa de verificação de suposições do modelo; a terceira, a avaliação de funcionamento diferencial de item (DIF) entre subgrupos da população alvo; e por fim é que então o modelo é ajustado e os traços latentes são estimados.

Todas essas etapas são realizadas em softwares estatísticos e, dentre eles, se destaca o R (R Core Team 2020). O R é um software gratuito, desenvolvido para lidar com ciência de dados e totalmente expansível através de seus pacotes. Apesar de suas amplas possibilidades, a utilização do R requer conhecimento de sua linguagem para a geração dos códigos que atendam à demanda do pesquisador.

Com o intuito de facilitar a utilização dos pacotes disponíveis no R para o ajuste de modelos unidimensionais da TRI é que surgiu o aplicativo TRIDIF. Nele é possível realizar todas as análises relacionadas aos modelos TRI em uma interface amigável e interativa, com download facilitado dos resultados em tabelas e gráficos.

O aplicativo TRIDIF¹ foi desenvolvido em código R Core Team (2020), sob o ambiente de desenvolvimento integrado do RStudio (2021). Para criar uma aplicação *web* interativa foi utilizado o pacote *shiny* (Chang et al., 2021). Outros pacotes foram utilizados para dar funcionalidades pontuais ao aplicativo, são eles: *shinydashboard* (Chang e Ribeiro 2021) para dar uma aparência de painel ao aplicativo, organizando gráficos e tabelas; *shinyAce* (Nijs et al., 2019) para a leitura facilitada de dados; *lrm* (Rizopoulos, 2006) para análises descritivas; *psych* (Revelle 2020) para a verificação das suposições do modelo; *difR* (Magis et al., 2010) para a análise DIF em testes de respostas dicotômicas; *difNLR* (Hladká & Martinkova 2021) para a análise DIF em

¹ Disponível em: <https://mariana-garcia.shinyapps.io/TRIDIF/>

testes de respostas politômicas; *mirt* (Chalmers, 2012) para os ajustes de modelos; *DT* (Xie et al., 2021) para a geração de todas as tabelas do aplicativo; *plotly* (Sievert, 2020) para a geração de todos os gráficos do aplicativo; e *xlsx* (Dragulescu & Arendt, 2020) para a exportação de tabelas de dados.

Este trabalho está organizado em nove capítulos, incluindo este primeiro. O capítulo 2 apresenta a definição de alguns termos, a estrutura geral do aplicativo e seus elementos de interface. Os capítulos 3 a 8 detalham os menus do aplicativo, “Entrada de dados”, “Análises descritivas”, “Ajuste do modelo”, “Presença de DIF” e “Exportação de dados”. Neste texto o menu “Presença de DIF” está depois do menu “Ajuste do modelo”, apenas para facilitar a utilização de algumas definições, porém, na prática do aplicativo, recomenda-se que o pesquisador primeiro realize a análise de presença de DIF para depois ajustar o modelo aos dados. E por fim, o capítulo 9 traz as conclusões deste trabalho.

2. TERMOS E ESTRUTURA DO APLICATIVO

Antes de começar o detalhamento do aplicativo, alguns termos precisam ser contextualizados. São eles:

- **Teste:** É o termo para designar a avaliação realizada pelo pesquisador através de um instrumento de medida composto por itens que medem os conteúdos relacionados ao traço latente. Pode ser uma prova (no contexto educacional) ou um questionário com auto resposta (no contexto psicométrico).
- **Item:** É o nome dado a cada questão (variável) do instrumento de medida (teste).
 - **Dicotômico:** Itens dicotômicos possuem duas categorias de respostas, sim e não ou certo e errado. Inclui-se nessa categoria itens que, apesar de serem de múltipla escolha, são dicotomizados para certo ou errado.
 - **Politômico:** Itens politômicos possuem mais de duas categorias de respostas. Itens politômicos que apresentam uma ordenação em suas categorias de respostas ainda são chamados de politômicos ordinais, como, por exemplo, itens cujas categorias são dadas conforme uma escala Likert.

Neste aplicativo, os itens podem ser do tipo dicotômicos ou politômicos ordinais.

- **Respondentes:** Todas as pessoas que responderam àquele teste (unidades amostrais).
- **Grupo:** Conjunto no qual são classificados os respondentes do teste, como por exemplo sexo, raça ou status socioeconômico. Neste aplicativo, o grupo deve ter apenas duas possibilidades de respostas, como por exemplo, sexo masculino ou feminino. Essas duas possibilidades são denominadas grupo focal e grupo referência. O grupo focal normalmente é minoritário em relação ao grupo de referência que é o grupo de comparação.
- **Traço latente:** Também conhecido como habilidade ou proficiência, é uma característica do indivíduo que não pode ser observada diretamente, mas

que podem ser estimada através de comportamentos ou de respostas a itens de um teste.

- **Escore TRI:** É o traço latente do respondente estimado pelo modelo TRI.

Após essa contextualização, e antes da apresentação das funcionalidades do aplicativo, serão apresentados os elementos de estrutura do aplicativo. Na Figura 1 está a tela de entrada do aplicativo.

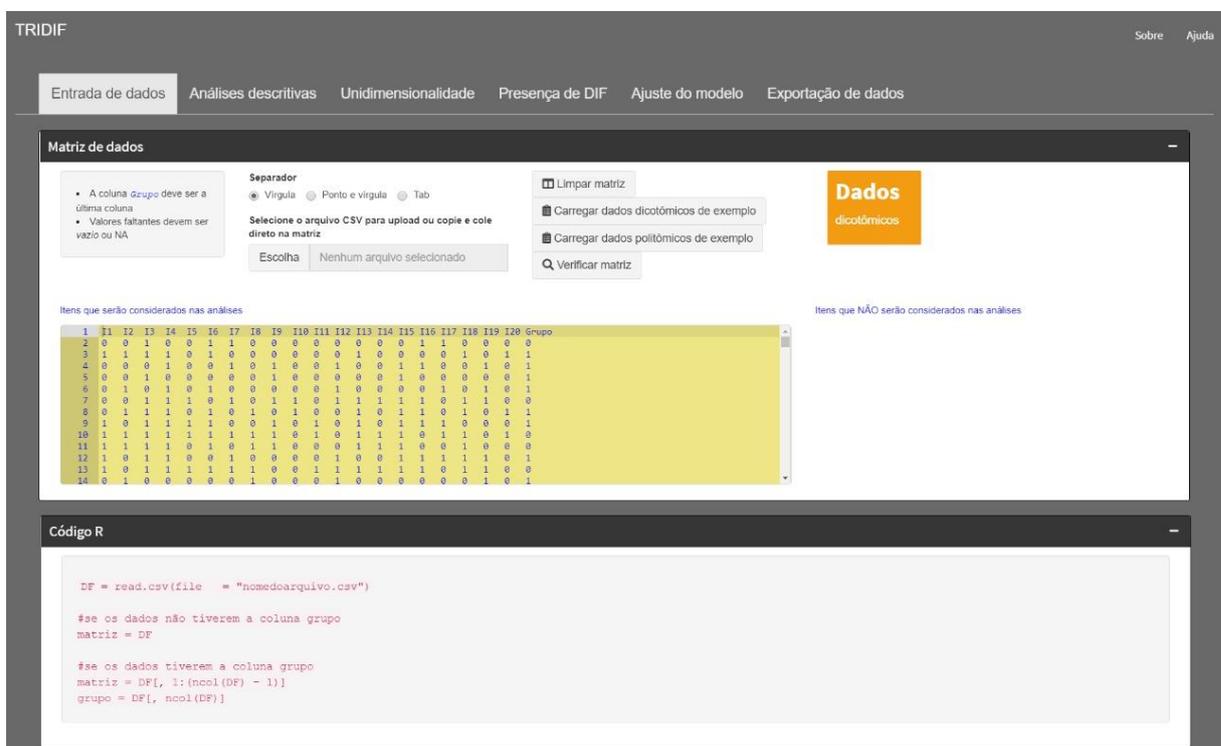


Figura 1: Tela inicial do aplicativo

A Figura 2 traz o topo de todas as telas do aplicativo, inclusive da tela inicial (Figura 1). Nela é possível identificar o nome do aplicativo “TRIDIF” e as opções “Sobre” e “Ajuda”. A opção “Sobre” reporta a um arquivo de texto com algumas informações técnicas do projeto, *link* para *download* e nomes dos responsáveis pelo desenvolvimento. A opção “Ajuda” detalha as funcionalidades do aplicativo e traz fundamentações teóricas para as análises.



Figura 2: Topo de todas as telas, informações gerais

O aplicativo possui elementos de interface que são comuns a todas as telas, menus, *box* retrátil e abas.

- **Menus:** As funcionalidades da aplicação são divididas em seis tópicos de menus que são “Entrada de dados”, “Análises descritivas”, “Unidimensionalidade”, “Presença de DIF”, “Ajuste do modelo” e “Exportação de dados”. Isso pode ser visto na Figura 3 . As funcionalidades de cada menu serão detalhadas nos capítulos seguintes.

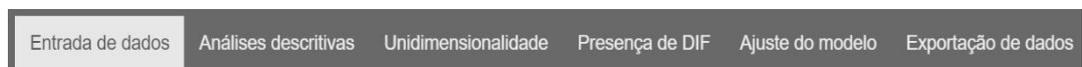


Figura 3: Menus do aplicativo

- **Box retrátil:** Dentro dos menus, as funcionalidades estão divididas em *boxes* retráteis. A Figura 1 mostra dois exemplos desse tipo de *box*, “Matriz de dados” e “Código R”. Eles são retráteis pois ao clicar no sinal de menos no canto superior direito do *box*, este é encolhido, para facilitar a visualização das demais funcionalidades. A Figura 4 demonstra essa situação.



Figura 4: Exemplo de *box* retrátil do aplicativo

- **Box retrátil “Código R”:** Esse *box* específico também pode ser visto na Figura 4 e está presente em todos os menus. Sua função é trazer um resumo de código que o pesquisador poderá utilizar no próprio console do R para reproduzir a análise realizada no menu do aplicativo.

- Abas: Dentro de *boxes*, as informações ainda podem ser organizadas em abas. No exemplo da Figura 5, cada aba possibilita ao pesquisador a exportação de um tipo de gráfico.

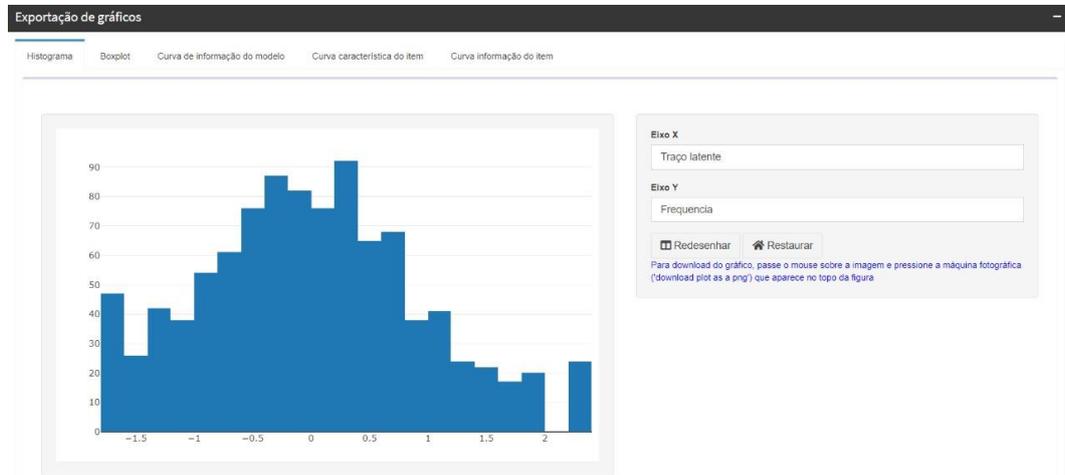


Figura 5: Exemplo de abas

- Aba “Ajuda”: Esta aba está presente em todas as tabelas e gráficos gerados pelo aplicativo. Sua função é ser uma ajuda rápida e sucinta para colunas de tabelas ou eixos de gráficos. A Figura 6 mostra um exemplo de aba “Ajuda” para o histograma dos traços latentes.

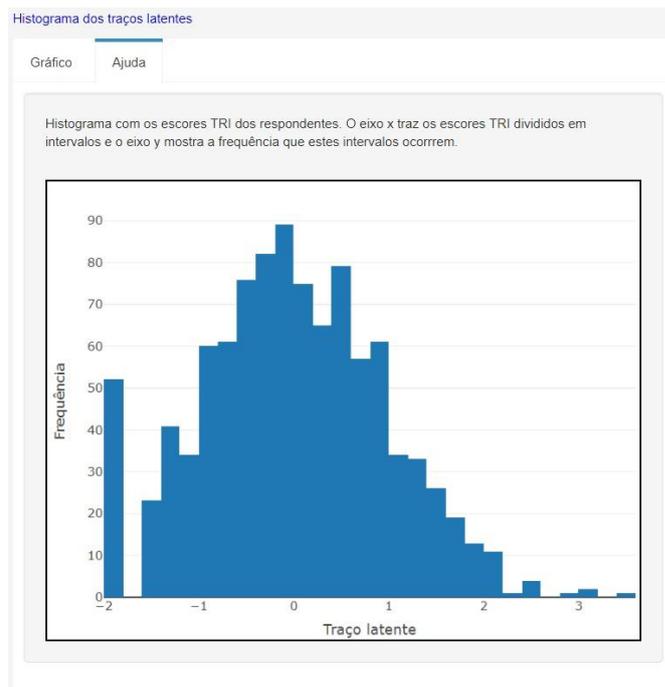


Figura 6: Exemplo da aba “Ajuda”

- Caixa informativa: Em todos os menus é possível identificar se os dados lidos pelo aplicativo são dicotômicos, politômicos ou indefinidos (Figura 1). Essa identificação é feita por uma caixa colorida informativa. As três possibilidades para esta informação são apresentadas na Figura 7.



Figura 7: Caixa informativa com o tipo de dado lido pelo aplicativo

Com os elementos de interface apresentados, no próximo capítulo será detalhado como deve se dar a entrada de dados no aplicativo.

3. MENU ENTRADA DE DADOS

Para a realização de qualquer análise, é preciso ter os dados coletados pelo pesquisador. Esses dados são resultados da aplicação do teste, e comumente são dispostos em uma tabela.

O menu “Entrada de dados” é o ponto de partida do pesquisador e já pôde ser visto na tela inicial do aplicativo (Figura 1). Possui apenas um *box*, o “Matriz de Dados” que é apresentado na Figura 8 (além do “Código R”, presente em todos os menus).

The screenshot shows the 'Matriz de dados' interface. It includes a 'Separador' section with radio buttons for 'Virgula', 'Ponto e virgula', and 'Tab'. Below it is a text box for 'Seleção o arquivo CSV para upload ou copie e cole direto na matriz' and a file selection button. To the right are buttons for 'Limpar matriz', 'Carregar dados dicotômicos de exemplo', 'Carregar dados politômicos de exemplo', and 'Verificar matriz'. A central table displays 14 rows of data with columns labeled I1 to I20 and a final 'Grupo' column. A legend on the right indicates 'Dados dicotômicos'.

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20	Grupo
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0
3	1	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1
4	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	1	0	0	1	0	1
5	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
6	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1
7	0	0	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0	1	1	0	0
8	0	1	1	1	0	1	0	1	0	1	0	0	1	1	1	0	1	0	1	1	1
9	1	0	1	1	1	1	0	0	1	0	1	0	1	0	1	1	1	0	0	0	1
10	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	0	1	0
11	1	1	1	1	0	1	0	1	1	0	0	0	1	1	1	0	0	1	0	0	0
12	1	0	1	1	0	0	1	0	0	0	1	0	0	1	1	0	1	1	1	0	1
13	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	0	1	0
14	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1

Figura 8: Box “Matriz de dados” do menu “Entrada de dados”

Neste *box*, o pesquisador deve entrar com sua tabela de dados, onde cada linha deverá representar um respondente e cada coluna um item. Caso haja uma separação em grupos dos respondentes, a última coluna dessa tabela poderá representar esse grupo e assim será possível realizar a análise de funcionamento diferencial de item (DIF) antes do ajuste do modelo.

A coluna grupo é opcional, mas caso haja, deve se chamar “grupo” e ser a última coluna da tabela de dados. Seus valores devem ser 0, para o grupo tido como referência e 1 para o grupo focal, assim diferenciando os respondentes nos grupos.

No caso de testes com respostas dicotômicas os valores das colunas dos itens devem ser 0 ou 1. No caso de teste com respostas politômicas os valores podem variar, mas devem obedecer a ordinalidade, sendo sequenciais como em uma escala Likert, como por exemplo 0, 1, 2, 3 ou 1, 2, 3, etc.

A entrada dos dados neste menu pode ser feita de duas formas, copiando os valores de uma planilha e colando no espaço para a tabela, ou ainda carregando um arquivo do tipo csv e especificando o separador de colunas utilizado (vírgula, ponto e

vírgula ou tabulação). Após esse carregamento, é preciso selecionar o botão “Verificar matriz”.

Para realizar ajustes de modelos TRI, os itens precisam ter um mínimo de variabilidade, ou seja, precisam ter ao menos duas categorias de respostas em cada item. O aplicativo faz essa verificação, e caso encontre algum item nesta condição, retira esse item da matriz de dados e avisa ao pesquisador, como mostra a Figura 9.

The screenshot shows the 'Matriz de dados' application interface. It features a data matrix with columns labeled I1, I2, I4, and Grupo. The matrix contains 14 rows of data. Item 13 is highlighted in yellow, indicating it is excluded from analysis. The interface also includes a separator selection (Virgula, Ponto e virgula, Tab), a file upload field (abortoparte.csv), and buttons for 'Limpar matriz', 'Carregar dados dicotômicos de exemplo', 'Carregar dados politômicos de exemplo', and 'Verificar matriz'. A message on the right states: 'Item 13: não tem valor '0''.

Item	I1	I2	I4	Grupo
1	1	1	1	1
2	1	1	1	1
3	1	1	1	0
4	1	1	1	0
5	1	1	1	0
6	0	1	1	1
7	1	0	1	0
8	1	1	1	1
9	1	1	0	1
10	1	1	1	0
11	1	1	1	1
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1

Figura 9: Exclusão de Item

No ajuste do modelo politômico de Escala Gradual (Andrich, 1978), cada item precisa apresentar todas as categorias de resposta. O aplicativo faz essa verificação, e caso encontre algum item fora desta condição, mostra um alerta de que o modelo de Escala Gradual não poderá ser ajustado, mas não retira o item da matriz de dados.

Para ilustrar o funcionamento do aplicativo, ainda é possível utilizar dados de exemplo através das opções “Carregar dados dicotômicos de exemplo” ou “Carregar dados politômicos de exemplo”.

A maior parte das funcionalidades deste menu utiliza funções base do R, exceto pela facilidade de copiar e colar dados de uma planilha que só é possível com a utilização de funções do pacote *shinyAce* (Nijs et al., 2019).

4. MENU ANÁLISES DESCRITIVAS

Antes de qualquer análise estatística mais complexa, é fundamental organizar, sumarizar e descrever os dados do banco, processo denominado análise descritiva. Assim, antes de qualquer ajuste de modelo TRI, faz-se essa organização verificando o número de itens do teste, o número de respondentes, contabilizando as não respostas de cada item (*missing* ou valores faltantes), a proporção de respostas em cada categoria de cada item e calculando o coeficiente alfa de Cronbach.

As funções utilizadas no aplicativo TRIDIF tanto para cálculo do modelo quanto para análise de funcionamento diferencial do item (DIF) permitem a ausência de valores nos dados. Exceto quando um respondente possui ausência em todas suas respostas, ou quando o valor ausente é no grupo e se deseja fazer análise DIF. Na estimação dos parâmetros de item, a resposta do indivíduo com valor faltante naquele item é descartada. Já na estimação do traço latente de um respondente com valor faltante, são consideradas apenas as respostas preenchidas, e, portanto, são calculadas apenas as probabilidades para estes itens². Assim, para obter boas estimativas é importante que a quantidade de valores faltantes no banco não seja alta.

O coeficiente alfa de Cronbach é uma medida de confiabilidade desenvolvida por Cronbach (1951) e é utilizado no contexto de modelos TRI para estimar a consistência interna de um teste, ou seja, o quanto os itens que o compõem medem o mesmo traço latente. Este coeficiente é obtido através da expressão:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k V_i^2}{V_t^2} \right) \quad (1)$$

onde k é o número de itens; V_i^2 é a variância dos escores dos respondentes ao i -ésimo item e V_t^2 é a variância dos escores totais de cada respondente.

O valor calculado de alfa varia de 0 a 1. Quanto mais próximo de 1 é este valor, maior é a consistência interna do teste.

É comum estimar o valor do coeficiente alfa considerando todos os itens do teste, e, reestimá-lo considerando a exclusão de cada item. Assim, devem ser

2

<https://cran.r-project.org/web/packages/difR/difR.pdf> acessado em 13/11/2021

<https://cran.r-project.org/web/packages/difNLR/difNLR.pdf> acessado em 13/11/2021

https://groups.google.com/g/mirt-package/c/slssi-7_koA?pli=1 acessado em 13/11/2021

identificados itens cujo valor do coeficiente, considerando sua exclusão, seja maior que o valor de coeficiente considerando todos os itens. Estes itens podem estar atrapalhando a consistência interna do teste (podendo até não estar medindo o mesmo traço latente dos demais itens), por isso sua exclusão aumenta o valor do coeficiente alfa de Cronbach. No contexto da TRI, estes itens podem atrapalhar o ajuste dos modelos ou contribuir muito pouco na estimativa do traço latente.

O menu “Análises descritivas” do aplicativo TRIDIF pode ser visto na Figura 10. Suas informações são apresentadas em um texto e três tabelas, detalhados a seguir:

Dados dicotômicos

O banco de dados contém 20 questões e 1000 respondentes. Apresentando 0 valores faltantes.

Valores faltantes em cada item:

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20
Freq	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Proporção de respostas

	0	1
I1	0.713	0.287
I2	0.696	0.304
I3	0.405	0.595
I4	0.221	0.779
I5	0.724	0.276
I6	0.528	0.472
I7	0.557	0.443
I8	0.789	0.211
I9	0.616	0.384
I10	0.736	0.264

Alfa de Cronbach

	value
All Items	0.9
Excluding I1	0.894
Excluding I2	0.895
Excluding I3	0.896
Excluding I4	0.897
Excluding I5	0.893
Excluding I6	0.898
Excluding I7	0.894
Excluding I8	0.895
Excluding I9	0.898

Código R

```
library(ltm)
descript(matriz)
```

Figura 10: Menu Análises descritivas

- Informação textual: Logo na parte superior do *box*, informações de número de itens, número de respondentes e número total de valores faltantes da tabela de dados são apresentados em forma de texto, como mostra a Figura 11.

O banco de dados contém 20 questões e 1000 respondentes. Apresentando 15 valores faltantes.

Figura 11: Informação textual do Menu Análises descritivas

- Tabela “Valores faltantes em cada item”: Esta tabela traz a frequência de valores faltantes, *missing* de cada item e o percentual que esta frequência representa. Como exemplo, na Figura 12 temos que o item I1 possui valor faltante em 15 dos 1000 respondentes, o que representa 1.5%.

Valores faltantes em cada item:

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
Freq	15	0	0	0	0	0	0	0	0	0
(%)	1.5	0	0	0	0	0	0	0	0	0

Figura 12: Tabela “Valores faltantes em cada item”

- Tabela “Proporção de respostas”: Esta tabela mostra a proporção de respostas em cada categoria no item. A **Erro! Fonte de referência não encontrada.** mostra um exemplo desta tabela. No item I1, 770 dos 1000 respondentes estão na categoria 0 ($770/1000 = 0.77$), 125 estão na categoria 1 ($125/1000 = 0.125$), 71 na categoria 2 ($71/1000 = 0.071$) e 34 na categoria 3 ($34/1000 = 0.034$).

Proporção de respostas

10 resultados por página Pesquisar

	0	1	2	3
I1	0.77	0.125	0.071	0.034
I2	0.629	0.186	0.123	0.062
I3	0.687	0.157	0.112	0.044

Figura 13: Tabela “Proporção de respostas” para itens politômicos

- Tabela “Alfa de Cronbach”: A primeira linha desta tabela traz o alfa de Cronbach considerando todos os itens do teste, e, em cada uma das demais linhas o valor corresponde ao coeficiente recalculado considerando a

exclusão daquele item. Um exemplo desta tabela é apresentado na Figura 14.

Alfa de Cronbach

10 resultados por página

Pesquisar

	value
All Items	0.9
Excluding I1	0.894
Excluding I2	0.895
Excluding I3	0.896

Figura 14: Tabela “Alfa de Cronbach”

Os resultados descritos neste menu utilizam a função *descript* do pacote *Itm* (Rizopoulos, 2006) do R. Este pacote possibilita além de análises descritivas, a geração de modelos TRI, entre outras funcionalidades, e só não foi utilizado nos ajustes de modelos deste aplicativo porque não apresenta todas as demandas propostas neste trabalho.

5. MENU UNIDIMENSIONALIDADE

Os modelos TRI propostos na literatura dependem da quantidade de traços latentes que está sendo medida, ou seja, o número de habilidades que o teste está mensurando. Um único traço latente caracteriza um modelo unidimensional, enquanto que mais de um caracteriza um modelo multidimensional. Os modelos disponibilizados no aplicativo TRIDIF supõem unidimensionalidade.

Outra suposição que todos os modelos assumem é a independência local, ou seja, em um certo nível de traço latente, os itens não serem correlacionados uns com os outros. De acordo com Hays et al. (2000), as suposições de unidimensionalidade e de independência local estão relacionadas, se uma está satisfeita, a outra também está. Assim, na prática, pode-se fazer apenas a verificação de uma delas.

A unidimensionalidade de um teste é difícil de ser verificada, já que o ser humano utiliza sempre mais de uma habilidade na execução de qualquer tarefa. Mas, segundo Mchorney & Cohen (2000) a suposição de unidimensionalidade pode ser relaxada, bastando que ela seja suficiente, ou seja, que haja um traço latente predominante para que os modelos TRI unidimensionais possam ser utilizados. Segundo os autores, a suposição de unidimensionalidade suficiente é atendida quando a proporção de explicação da primeira dimensão é de, no mínimo, 20%.

O aplicativo TRIDIF propõem ajustes de modelos unidimensionais, por isso é tão importante certificar-se dessa suposição antes de prosseguir com qualquer ajuste. Para a verificação da unidimensionalidade é utilizada a suposição de unidimensionalidade suficiente proposta por Mchorney & Cohen (2000).

Para a verificação desta suposição, no menu “Unidimensionalidade”, primeiro é calculada a matriz de correlação dos itens do teste (correlação tetracórica para itens dicotômicos e policórica para itens politômicos). Após são determinados os autovalores da matriz de correlação e em seguida é calculada a proporção de explicação da primeira dimensão através do autovalor da mesma. Se a proporção de explicação for maior que 20%, então a suposição de unidimensionalidade suficiente está atendida. Este resultado é apresentado em forma de texto no *box* “Verificando a Unidimensionalidade”, como pode ser visto na Figura 15.

Adicionalmente é comum a utilização de um gráfico de auto valores (*scree plot*). Em análise fatorial esse gráfico é utilizado para auxiliar na determinação do número de fatores a serem retidos e faz isso visualmente através do “cotovelo” que a curva

apresenta. Esse gráfico pode ser visto na Figura 15. No eixo x são apresentados os fatores da análise e no eixo y a proporção da variabilidade explicada por cada fator.

Para o cálculo das matrizes de correlação é utilizado o pacote *psych* (Revelle 2020) e para a geração do gráfico de *scree* o pacote *plotly* (Sievert, 2020).



Figura 15: Menu Unidimensionalidade

6. MENU AJUSTE DO MODELO

Segundo Andrade et al. (2000), modelos TRI são um conjunto de modelos matemáticos que representam a probabilidade de um indivíduo dar determinada resposta em função dos parâmetros dos itens e do traço latente do respondente.

Diferente da teoria clássica do teste (TCT), onde os traços latentes são estimados através do escore bruto dos respondentes, a TRI estima os traços latentes dos respondentes conectando-os às respostas aos itens do teste e às características dos itens, fazendo com que cada item contribua de forma diferente para a medida do traço latente.

Os modelos TRI unidimensionais, para uma única população envolvida, podem ser divididos de acordo com o tipo de resposta ao item em modelos dicotômicos e modelos politômicos. Modelos dicotômicos são para itens que contenham apenas duas categorias de resposta. E, modelos politômicos, são para itens que contenham mais de duas categorias de respostas. Modelos politômicos ainda se dividem em ordinais, para itens que apresentam uma ordenação em suas categorias de respostas, e nominais, para itens que não apresentam essa ordenação entre as categorias. O aplicativo TRIDIF propõem ajustes de modelos dicotômicos e politômicos ordinais.

Assim, este capítulo será dividido em três seções, a primeira com definições e representações de modelos dicotômicos, a segunda com essas mesmas definições e representações para modelos politômicos ordinais e a terceira e última com a demonstração de como ajustar os modelos no aplicativo TRIDIF.

6.1 MODELOS DICOTÔMICOS

Os modelos dicotômicos que podem ser ajustados no aplicativo TRIDIF são Rasch (1960), e Logísticos de 1, 2 e 3 parâmetros (Lord (1952) e Birnbaum (1968)). Estes modelos se diferenciam pelo número de parâmetros que utilizam para descrever o item. O modelo logístico de 1 parâmetro (1PL) e o modelo de Rasch estimam somente o parâmetro de dificuldade (ou posição) do item. O modelo logístico de dois parâmetros (2PL) estima os parâmetros de dificuldade e de discriminação (ou inclinação). O modelo logístico de três parâmetros (3PL) estima o parâmetro de dificuldade, de discriminação e a probabilidade de resposta na categoria 1 do item dada por indivíduos de baixo traço latente (acerto ao acaso). A definição destes

modelos é realizada a seguir e está baseada em Andrade *et. al* (2000). O primeiro modelo apresentado é o de 3PL, pois é o mais completo e os demais derivam dele.

- Modelo Logístico de três parâmetros (3PL)

$$P(X_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (2)$$

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

onde X_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j acerta o item i ou o responde de forma afirmativa e 0 caso contrário.

θ_j representa o traço latente do j -ésimo indivíduo.

$P(X_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j , com traço latente θ_j responder a categoria 1 do item i .

b_i é o parâmetro de dificuldade (ou de posição) do item i .

a_i é o parâmetro de discriminação (ou inclinação) do item i

c_i é o parâmetro que representa a probabilidade de indivíduos com baixo traço latente responderem a categoria 1 do item i .

D é o fator de escala, constante e igual a 1. Utiliza-se 1,7 quando se deseja que a função logística forneça resultados semelhantes ao de uma ogiva normal.

- Modelo Logístico de dois parâmetros (2PL)

$$P(X_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (3)$$

- Modelo Logístico de um parâmetro ou RASCH (1PL)

$$P(X_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}} \quad (4)$$

No aplicativo TRIDIF, a diferença entre o modelo 1PL e o Rasch é que o primeiro utiliza o fator de escala constante igual a 1,7, já o segundo utiliza o fator de escala constante de 1.

Para visualizar o ajuste do modelo TRI aos itens utiliza-se a curva característica do item que descreve o quanto mudanças no nível do traço latente se relacionam com mudanças na probabilidade de uma resposta específica (Castro, 2008). A Figura 16 mostra um exemplo de curva característica do item (CCI) para um modelo com respostas dicotômicas.

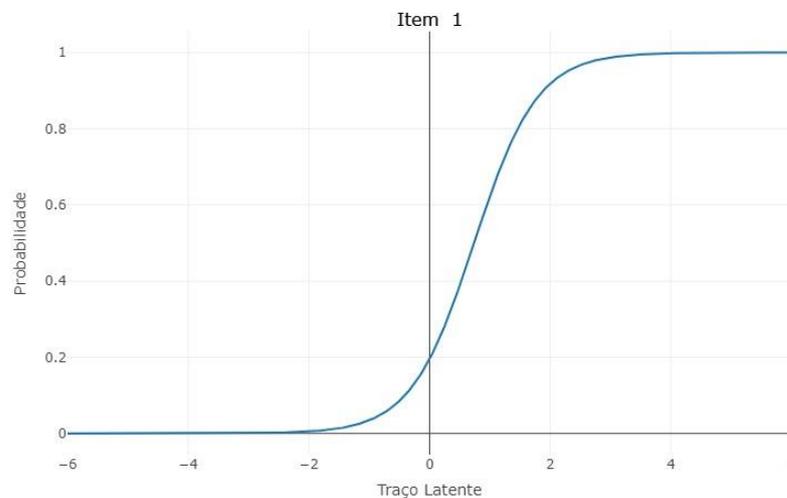


Figura 16: Exemplo de curva característica do item para itens dicotômicos

Outra representação gráfica importante em modelos TRI é a curva de informação do item que permite verificar quanto um item contém de informação para cada medida de traço latente. A informação é obtida pela seguinte expressão:

$$I_i(\theta) = \left[\frac{\frac{d}{d\theta} P_i(\theta)}{P_i(\theta) Q_i(\theta)} \right]^2 \quad (5)$$

onde,

$I_i(\theta)$ é a informação fornecida pelo item i no traço latente θ

$P_i(\theta) = P(X_{ij} = 1|\theta_j)$ e $Q_i(\theta) = 1 - P_i(\theta)$

A Figura 17 mostra a curva de informação do item (CCI) para o mesmo item representado pela curva característica da Figura 16.

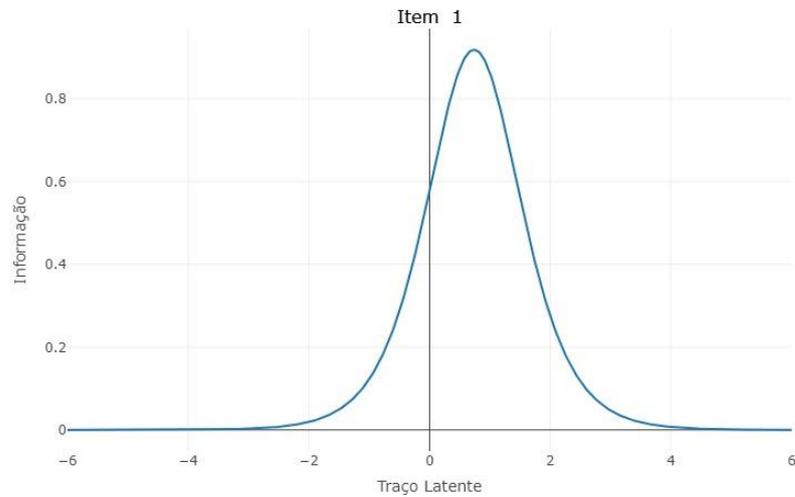


Figura 17: Exemplo de curva de informação do item para itens dicotômicos

Com a informação fornecida por cada item, soma-se essas informações e obtém-se a informação fornecida pelo teste.

$$I(\theta) = \sum_{i=1}^I I_i(\theta) \quad (6)$$

Outra forma de representar a informação do teste é através do erro-padrão de estimação, que é dado por:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (7)$$

A Figura 18 mostra um exemplo das curvas de informação e de erro padrão de estimação do teste.

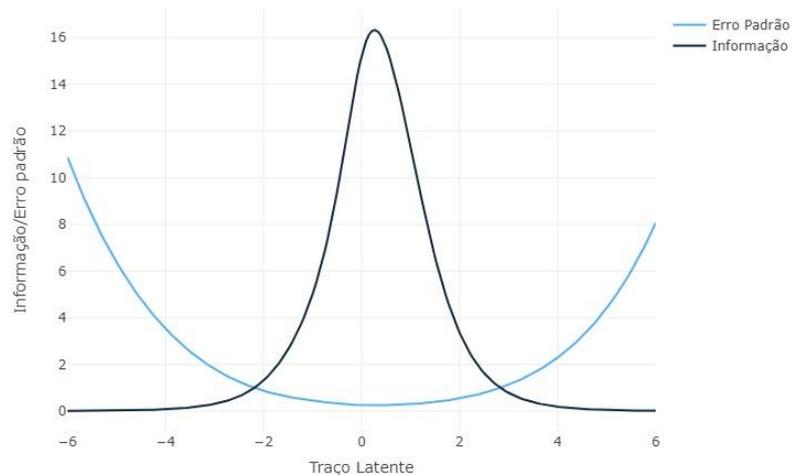


Figura 18: Exemplo de curva de informação e erro padrão de estimação de teste

6.2 MODELOS POLITÔMICOS ORDINAIS

Os modelos politômicos ordinais disponíveis para ajuste no aplicativo TRIDIF são Resposta Gradual (Samejima, 1969), Crédito Parcial (Masters, 1982), Crédito Parcial Generalizado (Muraki, 1992) e Escala Gradual (Andrich, 1978). Todos eles tentam obter mais informações do que apenas resposta correta ou incorreta (ou afirmativo ou falso), pois as respostas aos itens possuem categorias intermediárias ordenadas entre o certo e errado ou afirmativo e falso. A definição destes modelos é realizada a seguir e está baseada em Andrade *et. al* (2000) e Embretson e Reise (2000).

- Modelo de Resposta Gradual (Samejima, 1969)

É uma generalização do modelo logístico de 2 parâmetros e para seu ajuste os itens não precisam ter o mesmo número de categorias de respostas. A probabilidade de um indivíduo j escolher a categoria k ou outra mais alta no item i é dada por:

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

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, e $k = 0, 1, \dots, m_i$, onde:

$b_{i,k}$ é o parâmetro de dificuldade da k -ésima categoria do item i .

Os demais parâmetros são os mesmos do modelo 2PL já definido anteriormente. O parâmetro de inclinação em modelos politômicos é o

mesmo para todas as categorias de um item e não representa diretamente a discriminação do item.

A probabilidade de um indivíduo j responder exatamente a categoria k no item i é dada por:

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j)$$

Por definição $P_{i,0}^+(\theta_j) = 1$ e $P_{i,m_i+1}^+(\theta_j) = 0$. Assim,

$$P_{i,k}(\theta_j) = \frac{1}{1+e^{-Da_i(\theta_j-b_{i,k})}} - \frac{1}{1+e^{-Da_i(\theta_j-b_{i,k+1})}} \quad (8)$$

Em um item com $(m_i + 1)$ categorias, m_i valores de dificuldade deverão ser estimados, além do parâmetro de inclinação do item.

- Modelo de Crédito Parcial (Masters, 1982)

É uma extensão do modelo de Rasch para itens dicotômicos. Todos os parâmetros no modelo são de localização, sendo que a inclinação é assumida igual para todos os itens. Os itens não precisam ter o mesmo número de categorias de respostas. Supondo que o item i tem $(m_i + 1)$ categorias de resposta ordenáveis ($k = 0, 1, \dots, m_i$), o modelo de crédito parcial é dado por:

$$P_{i,k}(\theta_j) = \frac{\exp[\sum_{u=0}^k(\theta_j-b_{i,u})]}{\sum_{u=0}^{m_i} \exp[\sum_{v=0}^k(\theta_j-b_{i,v})]} \quad (9)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, e $k = 0, 1, \dots, m_i$, e $b_{i,0} \equiv 0$ onde:

$P_{i,k}(\theta_j)$ é a probabilidade de um indivíduo com traço latente θ_j escolher a categoria de resposta k , dentre as $(m_i + 1)$ categorias do item i .

$b_{i,k}$ é o parâmetro do item que regula a probabilidade de escolher a categoria de resposta k em vez da categoria adjacente $(k - 1)$ no item i . Cada parâmetro $b_{i,k}$ corresponde ao valor do traço latente no qual o indivíduo tem a mesma probabilidade de escolher as categorias k e $(k - 1)$, isto é, onde $P_{i,k}(\theta_j) = P_{i,k-1}(\theta_j)$.

Para um item com $(m_i + 1)$ categorias de resposta, serão estimados m_i parâmetros.

- Modelo de Crédito Parcial Generalizado (Muraki, 1992)

É baseado no modelo de crédito parcial, mas com a hipótese da inclinação uniforme relaxada. Assim, o modelo de crédito parcial generalizado é dado por:

$$P_{i,k}(\theta_j) = \frac{\exp[\sum_{u=0}^k Da_i(\theta_j - b_{i,u})]}{\sum_{u=0}^{m_i} \exp[\sum_{v=0}^k Da_i(\theta_j - b_{i,v})]} \quad (10)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, e $k = 0, 1, \dots, m_i$, e $b_{i,0} \equiv 0$ onde:

Para um item com $(m_i + 1)$ categorias de resposta, serão estimados m_i parâmetros mais sua inclinação.

- Modelo de Escala Gradual (Andrich, 1978)

Também é uma extensão do modelo de Rasch para itens dicotômicos e pode ser derivado do modelo de crédito parcial. A principal distinção entre eles é que no modelo de escala gradual todos os itens devem ter as mesmas categorias de respostas, igualmente espaçadas entre si.

E, a relação entre eles é que o modelo de escala gradual decompõem o parâmetro $b_{i,k}$ do modelo de crédito parcial em b_i , um parâmetro de posição que representa a dificuldade do item, e em d_k , um parâmetro de categoria, comum a todos os itens, sendo $b_{i,k} = b_i + d_k$. Como no modelo de crédito parcial, a discriminação é assumida igual para todos os itens.

$$P_{i,k}(\theta_j) = \frac{\exp[\sum_{u=0}^k (\theta_j - (b_i + d_u))]}{\sum_{u=0}^{m_i} \exp[\sum_{v=0}^k (\theta_j - (b_i + d_v))]} \quad (11)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, e $k = 0, 1, \dots, m_i$, e $b_{i,0} \equiv 0$.

Em um teste de itens com $(m_i + 1)$ categorias, m_i parâmetros de categoria deverão ser estimados, além dos parâmetros de posição de cada item.

Como nos modelos dicotômicos, para visualizar o ajuste do modelo TRI aos itens utiliza-se a curva característica do item (CCI), mas agora há uma curva para cada categoria de resposta. De acordo com Castro (2008), as curvas de categoria de

resposta mostram a relação entre as probabilidades de respostas nas categorias de cada item, o nível do traço latente e as características do item. Através delas é possível identificar qual a categoria de resposta mais provável de ser escolhida para cada nível de traço latente. A Figura 19 mostra um exemplo de curva característica do item (CCI) para um modelo com respostas politômicas.

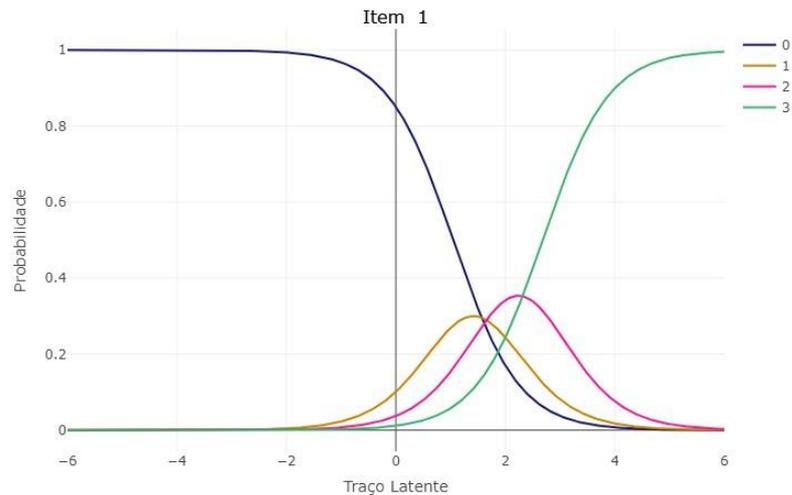


Figura 19: Exemplo de curva característica do item para itens politômicos

Em modelos politômicos, a informação do item compreende os parâmetros de inclinação e de posição e pode ser descrita pela informação de cada categoria ou pela informação total do item. Segundo Embretson e Reise (2000) as curvas de informação em itens politômicos são utilizadas para avaliar a quantidade de discriminação que os itens fornecem pois, diferente do que ocorre nos modelos dicotômicos, o parâmetro de inclinação não pode ser interpretado diretamente como um índice de discriminação do item. A Figura 20 (a) mostra a curva de informação do item por categoria, e, a Figura 20 (b) mostra a curva de informação total do item, ambas para o mesmo item representado pela curva característica da Figura 19.

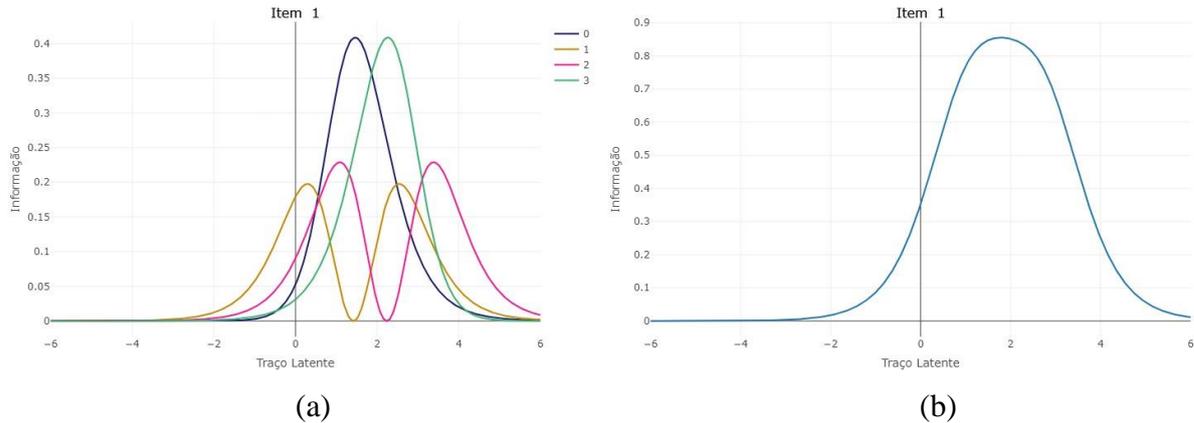


Figura 20: Exemplo de curva de informação de item para itens politômicos

As informações dos itens podem ainda ser somadas para a obtenção da curva de informação do teste, que também pode ser representada através do erro-padrão de estimação, como nos modelos de itens dicotômicos. Essas curvas em modelos politômicos são similares às dos modelos dicotômicos (Figura 18).

6.3 AJUSTE DE MODELOS NO APLICATIVO TRIDIF

O menu “Ajuste do modelo” pode ser visto na Figura 21 com seus *boxes* internos contraídos. Apesar da Figura 21 detalhar os modelos para itens dicotômicos, a estrutura do menu é similar para ambos os tipos de itens. Dentro do *box* externo “Ajuste do modelo”, há duas informações textuais, duas opções de escolha para o pesquisador e dois *boxes* internos onde são apresentados os resultados do ajuste do modelo, são eles: “Traços latentes” e “Itens do modelo”. Cada uma dessas funcionalidades será detalhada a seguir.

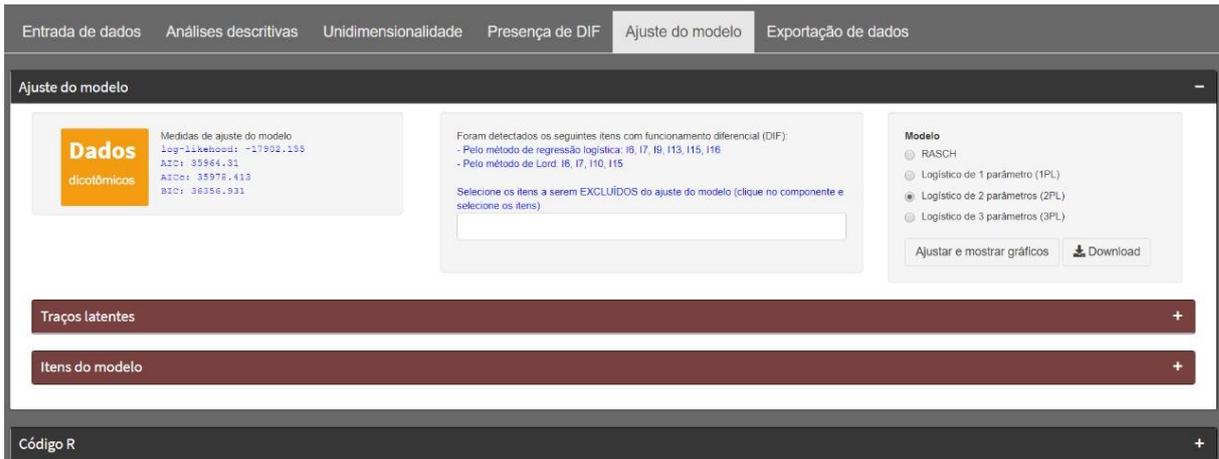


Figura 21: Menu Ajuste do modelo

- Informação textual – medidas de ajuste do modelo: Este texto traz alguns valores úteis caso o pesquisador queira escolher um modelo entre dois ou mais modelos ajustados pelo aplicativo. Esses valores são: o máximo do logaritmo natural da função de verossimilhança (*log-likelihood*) que posteriormente pode ser utilizado para o cálculo do teste da razão de verossimilhança (TRV) entre dois modelos encaixados; o critério de informação proposto por Akaike, 1974 (AIC); este mesmo critério corrigido para pequenas amostras proposto por Bozdogan, 1987 (AICc); e, o critério bayesiano proposto por Schwarz, 1978 (BIC). Tanto o teste da razão de verossimilhança quanto os critérios AIC, AICc e BIC buscam o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que proporcione um melhor ajuste. Pelos critérios AIC, AICc e BIC, o modelo mais satisfatório é aquele que atinge o menor valor no critério.
- Informação textual – itens com funcionamento diferencial (DIF): Este texto traz a informação dos itens detectados com funcionamento diferencial (DIF) para cada método do menu “Presença de DIF”, como mostra a Figura 22. Essa informação é apresentada mesmo que o pesquisador não tenha acessado o menu “Presença de DIF”.

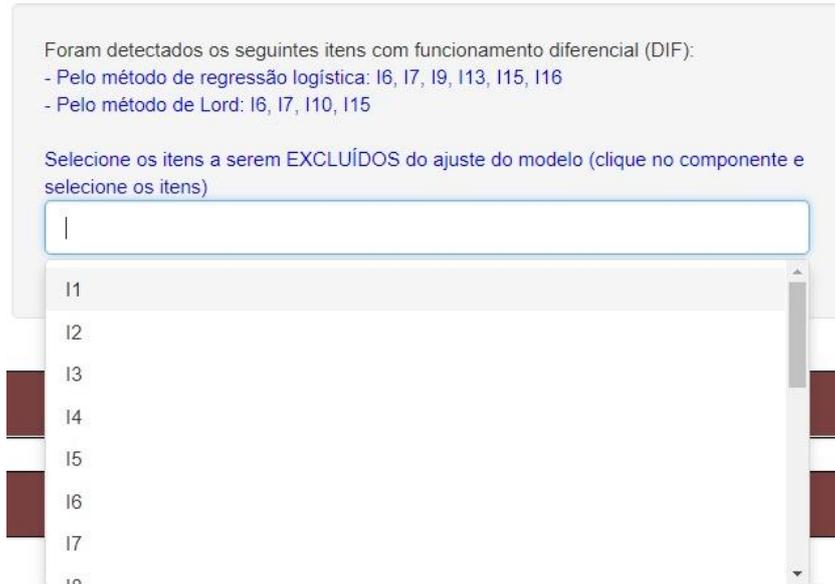


Figura 22: Itens com DIF e escolha de itens para exclusão do modelo

- Opção de Escolha - Itens a serem excluídos do cálculo do modelo: Caso o pesquisador deseje retirar algum item do modelo, basta selecioná-lo nesta lista de opções (Figura 22). Ao clicar na caixa os itens são listados e assim podem ser selecionados.
- Opção de Escolha – “Modelo”: Nesta opção, o pesquisador deve escolher o modelo a ser ajustado, na Figura 21, as opções são para itens dicotômicos.
- *Box* interno – “Traços latentes”: Este *box* é padrão para qualquer modelo e apresenta uma tabela e três gráficos, como mostra a Figura 23. A tabela traz as estimativas dos traços latentes dos respondentes a partir de três estimadores: o escore bruto e o escore padronizado (oriundos da teoria clássica de teste) e o escore TRI, que é o traço latente estimado pelo modelo TRI escolhido na interface. Nesta tabela também é informado o erro padrão do escore TRI. A partir dos escores TRI dos candidatos são feitos dois gráficos, o histograma e o *boxplot*. E, por último, é apresentada a curva de informação do teste junto a curva de erro padrão.

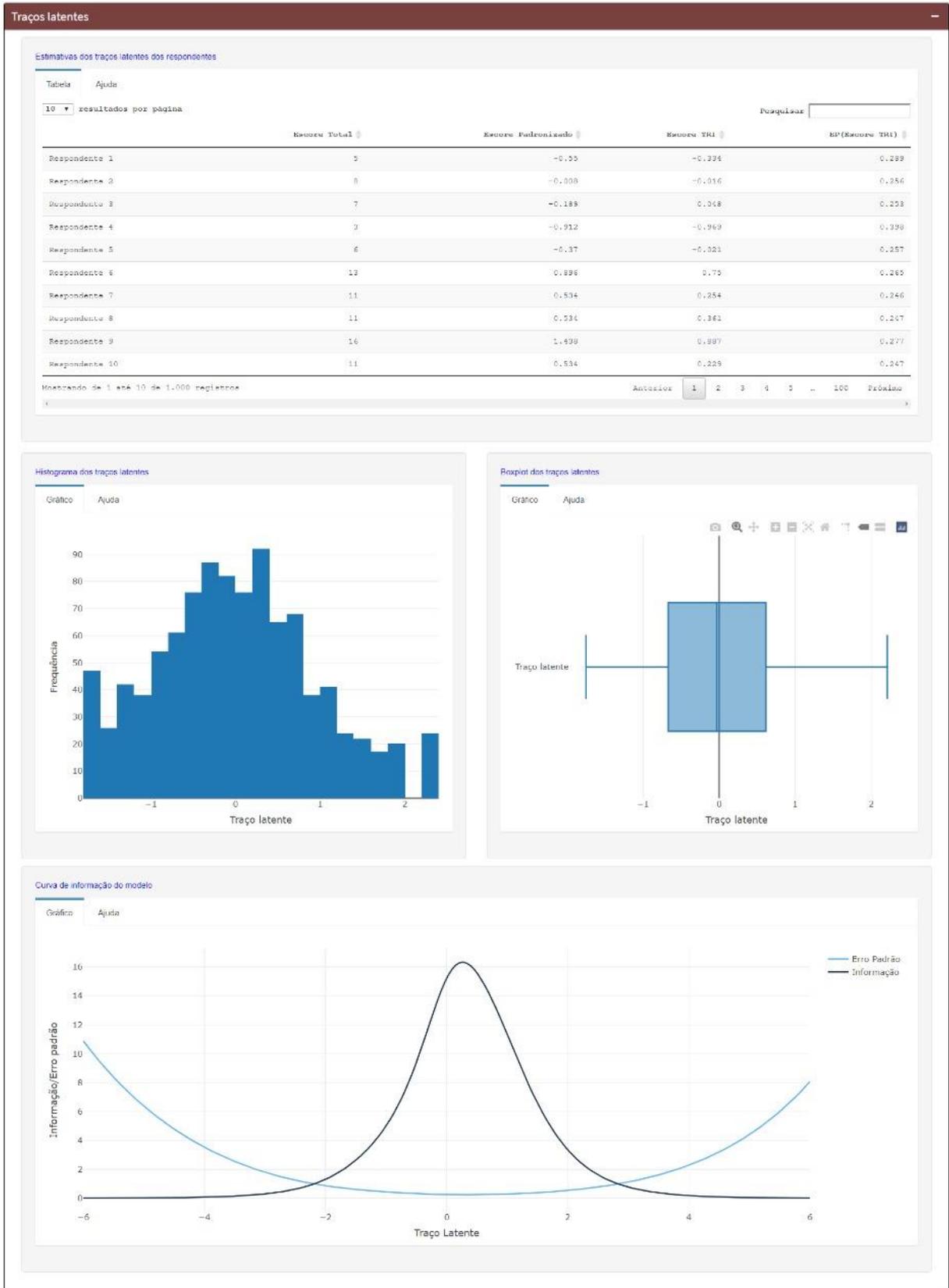


Figura 23: Box Traços latentes

- *Box* interno – “Itens do Modelo”: Este *box* apresenta uma tabela com os coeficientes ajustados pelo modelo para cada item. Esses coeficientes mudam de acordo com o modelo selecionado na interface e com o número de categorias dos itens, no caso de modelos politômicos. Além disso, esse *box* também apresenta gráficos que variam de acordo com o tipo de item do modelo.

A Figura 24 mostra o início deste *box* para um modelo 2PL de itens dicotômicos. Após a tabela, primeiro são apresentadas as curvas características dos itens sobrepostas em um mesmo gráfico. Ao lado, da mesma forma, são apresentadas as curvas de informação dos itens. E, logo abaixo, aparecem duas listas de gráficos, uma com as curvas características dos itens, desenhadas isoladamente e outra com as curvas de informação dos itens, também desenhadas isoladamente.

A Figura 25 mostra o início deste *box* para um modelo de resposta gradual com itens politômicos. Após a tabela, primeiro são apresentadas as curvas de informação dos itens sobrepostas em um mesmo gráfico. Abaixo, aparecem três listas de gráficos, uma com as curvas características dos itens, outra com as curvas de informação de categorias dos itens e a última com as curvas de informação total dos itens.

Para o ajuste dos modelos foi utilizado o pacote *mirt* (Chalmers, 2012). Esse pacote permite ajustes em itens dicotômicos ou politômicos, assim como ajustes uni ou multidimensionais (mais de um traço latente, que não é o caso dos ajustes deste aplicativo). Para a geração das tabelas foi utilizado o pacote *DT* (Xie et al., 2021). E, para a geração dos gráficos, o pacote *plotly* (Sievert, 2020).

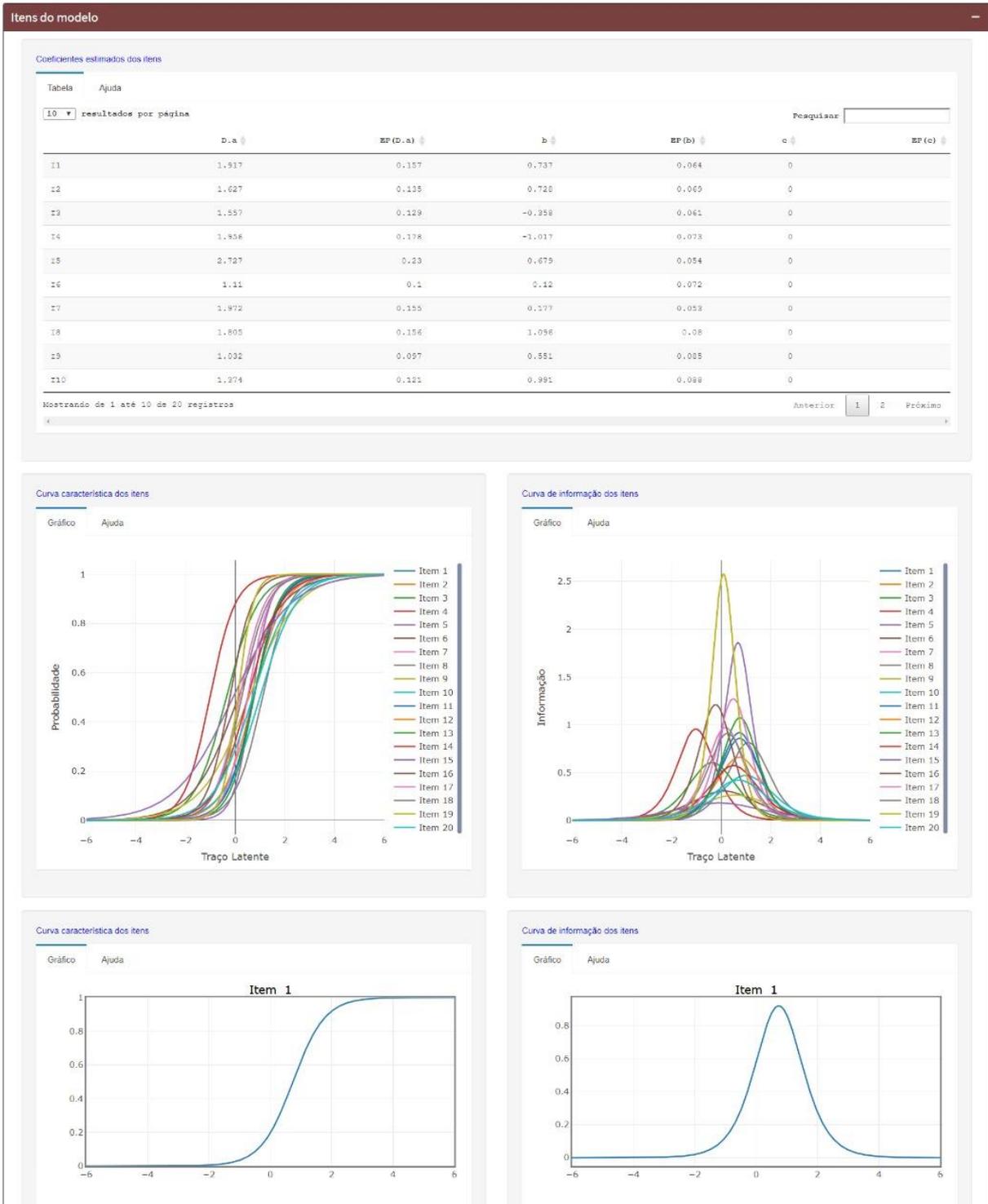


Figura 24: Box “Itens do modelo” para modelos dicotômicos



Figura 25: Box “Itens do modelo” para modelos politômicos

7. MENU PRESENÇA DE DIF

De acordo com Castro (2008), o funcionamento diferencial do item (diferencial item functioning – DIF) ocorre quando um item do instrumento não tem a mesma relação com o traço latente entre dois ou mais grupos de respondentes. Esse comportamento fica claro quando um item apresenta uma curva de resposta diferente para cada grupo ou, equivalentemente, qualquer dos parâmetros do item difiram entre os grupos.

A identificação de itens com funcionamento diferencial deixa a cargo do pesquisador a avaliação da permanência ou não destes itens para o ajuste do modelo. Nessa decisão deve ser levado em consideração que a presença de um grande número de itens com DIF pode causar até mesmo o funcionamento diferencial do teste (FDT), invalidando o instrumento de medida (Sisto, 2006).

Há duas formas distintas de DIF, DIF uniforme e DIF não uniforme. O DIF é considerado uniforme quando não há interação entre o traço latente e o fato de pertencer a um determinado grupo, ou seja, quando diferenças entre grupos em respostas ao item são encontradas em todos os níveis do traço latente. E, é considerado não uniforme quando há interação entre o traço latente e grupo, ou seja, diferenças entre grupos em respostas ao item dependem dos níveis de traço latente. (Crane et al., (2004) e Mellenbergh (1982)).

O aplicativo TRIDIF apresenta quatro métodos de identificação que detectam funcionamentos diferenciais uniformes ou não uniformes. Para esta detecção, podem ser utilizados apenas dois grupos, como por exemplo, sexo masculino ou feminino. Essas duas possibilidades são denominadas grupo focal (minoritário) e grupo referência (majoritário).

Para testes com itens dicotômicos, é possível realizar a identificação através do método de Regressão Logística (Swaminathan and Rogers, 1990) e do método de Lord (1980). Para testes com itens politômicos, é possível realizar a identificação através do método Logito Cumulativo (French & Miller 1996) e do método Logito de Categoria Adjacente (Zumbo 1999).

Dessa forma, este capítulo será dividido em duas seções, a primeira com definições e representações dos métodos de detecção disponíveis no aplicativo, e, a segunda com a demonstração de como realizar essa detecção no aplicativo TRIDIF.

7.1 MÉTODOS DE DETECÇÃO DE DIF

O funcionamento diferencial do item (DIF) fica claro quando um item apresenta uma curva de resposta diferente para cada grupo ou, equivalentemente, qualquer dos parâmetros do item difiram entre os grupos. Assim, os métodos de detecção se baseiam na construção de modelos lineares generalizados e de testes de hipótese cuja hipótese nula é de que um item tem os mesmos parâmetros para ambos os grupos. As definições dos métodos de detecção presentes no aplicativo TRIDIF estão a seguir e foram baseadas em Andriola (2001) e Hladká(2021).

- Método de Regressão Logística (Swaminathan and Rogers, 1990)

O método de Regressão Logística ajusta três modelos logísticos aos itens:

- Modelo 1: sem considerar a variável grupo, supondo que não haja a presença de DIF.
- Modelo 2: considerando a variável grupo, supondo a presença de DIF uniforme.
- Modelo 3: considerando a variável grupo e a interação entre grupo e traço latente, supondo a presença de ambos tipos de DIFs.

A seguir está definido apenas o modelo logístico completo, considerando a variável grupo e a interação entre grupo e traço latente, supondo a presença de ambos tipos de DIF. Os demais modelos são derivados diretamente deste.

$$\text{logit } P(X_{ij} = 1) = b_0 + b_1\theta_j + b_2(\text{grupo}_j) + b_3(\theta_j \text{ grupo}_j) \quad (12)$$

onde:

$P(X_{ij} = 1)$ é a probabilidade de acerto ou de resposta afirmativa do respondente j no item i

b 's são os parâmetros da regressão logística

θ_j é o traço latente do respondente j , estimado através do score total do indivíduo

$grupo_j$ é o grupo ao qual pertence o indivíduo j , 0 se for o grupo referência e 1 se for o grupo focal

Com os modelos ajustados, realiza-se um teste de razão de verossimilhança para compará-los. Se o tipo de DIF a ser identificado é o uniforme, a diferença entre os modelos 1 e 2 é comparada com uma distribuição χ^2 com um grau de liberdade. Se o tipo de DIF a ser identificado é o não uniforme, essa mesma comparação é feita para a diferença entre os modelos 2 e 3. Caso a identificação seja para ambos tipos de DIF, a diferença entre os modelos 1 e 3 é comparada com uma distribuição χ^2 com dois graus de liberdade. O resultado do teste diz se o item possui ou não aquele tipo de DIF.

- Método de Lord (1980)

O método de Lord, ajusta dois modelos TRI com o número de coeficientes determinado pelo pesquisador, um para cada grupo de respondentes. Nestes ajustes as Equações (2), (3) ou (4) são utilizadas, de acordo com o número de parâmetros escolhido. Um item terá DIF se os parâmetros estimados nos modelos dos dois grupos não coincidirem, isto é, se têm diferenças significativas.

Os coeficientes dos dois modelos são comparados através do teste qui-quadrado de Lord, cuja estatística é a seguinte:

$$\chi_i^2 = V_i \Sigma^{-1} V_i' \quad (13)$$

Onde:

V_i é um vetor linha cujas colunas são as diferenças entre as estimativas dos parâmetros dos modelos TRI ajustados nos grupos de referência e focal

V_i' é o vetor V_i transposto

Σ^{-1} é a inversa da matriz de covariância das diferenças nas estimativas dos parâmetros dos itens.

O valor desta estatística então é comparada com uma distribuição χ^2 cujo o número de graus de liberdade é o número de parâmetros escolhido

para o ajuste dos modelos. Este método detecta ambos os tipos de DIF simultaneamente.

- Método Logito Cumulativo (French & Miller 1996)

O método Logito Cumulativo é uma extensão do método da regressão logística para modelos ordinais. Esse método também ajusta três modelos aos itens na busca de detecção de DIFs, mas os modelos ajustados são do tipo logitos cumulativos.

A seguir será definido apenas o modelo Logito cumulativo completo, considerando a variável grupo e a interação entre grupo e traço latente, supondo a presença de ambos tipos de DIF. Os demais modelos são derivados diretamente deste.

Considerando $m_i + 1$ categorias ordenadas para o item i , sendo X_{ij} o escore do candidato j no item i que pode variar entre $\{0, 1, \dots, m_i\}$, a probabilidade do indivíduo j obter o escore de pelo menos $m = 1, \dots, m_i$ no item i é dada por:

$$\text{logit } P(X_{ij} \geq m) = b_{0im} + b_{1i}\theta_j + b_{2i}grupo_j + b_{3i}\theta_j grupo_j \quad (14)$$

onde:

$P(X_{ij} \geq m)$ é a probabilidade do respondente j obter pelo menos escore m no item i

b_{0im} são os m interceptos de categoria do item i , a primeira categoria não recebe um intercepto

b_{1i}, b_{2i}, b_{3i} são os demais parâmetros do modelo

θ_j é o traço latente do respondente j , estimado através do escore total do indivíduo

$grupo_j$ é o grupo ao qual pertence o indivíduo j , 0 se for o grupo referência e 1 se for o grupo focal

A probabilidade da categoria, ou seja, a probabilidade do escore ser exatamente m no item i , será a diferença entre as probabilidades cumulativas das duas categorias adjacentes.

$$P(X_{ij} = m) = P(X_{ij} \geq m) - P(X_{ij} \geq m + 1) \quad (15)$$

Como no método de regressão logística, os três modelos são ajustados e comparados através do teste da razão de verossimilhança (modelo 1: sem considerar a variável grupo, supondo que não haja a presença de DIF; modelo 2: considerando a variável grupo, supondo a presença de DIF uniforme; modelo 3: considerando a variável grupo e a interação entre grupo e traço latente, supondo a presença de ambos tipos de DIFs). Os modelos são comparados de acordo com o tipo de DIF a ser identificado e o resultado do teste diz se o item possui ou não aquele tipo de DIF.

- Método Logito de Categoria Adjacente (Zumbo 1999)

O método Logito de Categoria Adjacente também ajusta três modelos aos itens na busca de detecção de DIFs. A seguir será definido o modelo Logito de Categoria Adjacente completo, considerando a variável grupo e a interação entre grupo e traço latente, supondo a presença de ambos tipos de DIF. Os demais modelos são derivados diretamente deste.

$$P(X_{ij} = m) = \frac{e^{\sum_{k=0}^m (b_{0im} + b_{1i}\theta_j + b_{2i}grupo_j + b_{3i}\theta_j grupo_j)}}{\sum_{j=0}^{m_i} e^{\sum_{k=0}^m (b_{0im} + b_{1i}\theta_j + b_{2i}grupo_j + b_{3i}\theta_j grupo_j)}} \quad (16)$$

os termos da expressão são os mesmos já definidos no método logito cumulativo.

Como nos métodos de regressão logística e logito cumulativo, os três modelos são ajustados e comparados através do teste da razão de verossimilhança, o resultado do teste diz se o item possui ou não o tipo de DIF requerido pelo pesquisador.

Em qualquer método de detecção de itens DIF, são realizados vários testes de hipóteses. Estas comparações múltiplas aumentam a chance de rejeitar ou aceitar incorretamente uma hipótese. Para contornar este problema são utilizados métodos de ajuste de p-valor para comparações múltiplas. Há muitos métodos descritos na literatura, a seleção deste método deve sempre levar em consideração a inferência

desejada. De acordo com Kim e Oshima (2013), ajustes de Benjamini & Hochberg (1995) e Holm (1979) costumam ter melhor desempenho para fins de DIF.

Independentemente do método de detecção utilizado, a presença de DIF pode ser visualizada através do gráfico da curva de resposta para cada grupo no item. A Figura 26 (a) mostra esse gráfico na detecção de DIF uniforme em um item de teste com resposta dicotômica. A Figura 26 (b) apresenta o gráfico na detecção de DIF não uniforme também em um item de teste com resposta dicotômica. E, a Figura 26 (c) mostra esse gráfico na detecção de ambos tipos de DIF em um item de teste com resposta politômica.

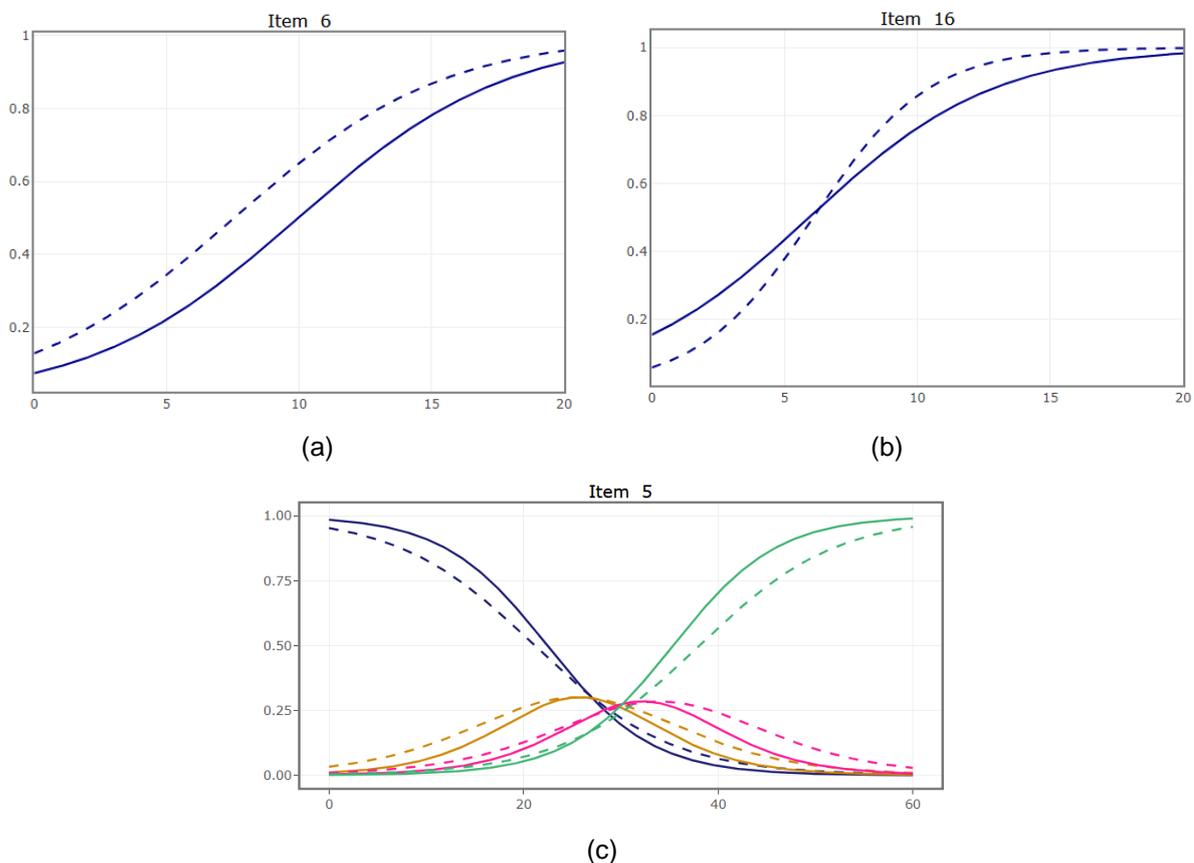


Figura 26: Visualização da detecção de DIF

7.2 DETECÇÃO DE DIF NO APLICATIVO TRIDIF

O menu “Presença de DIF” para testes com respostas dicotômicas pode ser visto na Figura 27 com os *boxes* internos contraídos. Cada um dos métodos de detecção localiza-se em um *box* retrátil externo e dentro de cada um desses *boxes* há

opções de escolha para o pesquisador (opções que variam conforme o método de detecção). Além disso, dentro de cada *box* ainda há mais três *boxes* internos onde são apresentados os resultados da análise DIF, são eles: “Detecção dos Itens”, “Análise gráfica” e “Código R”.

The image shows a software interface with a navigation bar at the top containing: 'Entrada de dados', 'Análises descritivas', 'Unidimensionalidade', 'Presença de DIF', 'Ajuste do modelo', and 'Exportação de dados'. The 'Presença de DIF' tab is active.

Under the 'Presença de DIF' tab, there are two main panels:

- Método de regressão logística:**
 - 'Dados' dropdown: 'dicotômicos'
 - 'Tipo de DIF' radio buttons: 'Uniforme', 'Não uniforme', 'Ambos' (selected)
 - 'Método de Correção' dropdown: 'Nenhum'
 - Three expandable boxes: 'Detecção dos itens', 'Análise gráfica', 'Código R' (each with a '+' icon)
- Método de Lord:**
 - 'Dados' dropdown: 'dicotômicos'
 - 'Modelo' radio buttons: '1PL', '2PL', '3PL' (selected)
 - 'Método de Correção' dropdown: 'Nenhum'
 - Three expandable boxes: 'Detecção dos itens', 'Análise gráfica', 'Código R' (each with a '+' icon)

Figura 27: Menu “Presença de DIF”

A seguir são detalhadas as opções de escolha do pesquisador e os *boxes* internos com os resultados das detecções.

- Opção de escolha – “Tipo de DIF”: Os métodos de Regressão logística, logito cumulativo e logito de categoria adjacente apresentam esta opção onde pode ser selecionado o tipo de DIF a ser identificado. Podendo ser Uniforme, Não Uniforme ou ambos, sendo a opção padrão a identificação de ambos os tipos. Esta opção pode ser vista na Figura 27.
- Opção de escolha – “Modelo”: O método de detecção de Lord apresenta esta opção de entrada. Deve ser selecionado o tipo de modelo TRI que o método irá utilizar para o ajuste dos modelos em ambos os grupos, a comparação dos coeficientes e a identificação das DIFs. Esta opção também pode ser vista na Figura 27.

- Opção de escolha – “Método de Correção”: Esta opção está disponível para todos os métodos de detecção de DIFs e pode ser vista na Figura 28. Caso o pesquisador queira contornar problemas que possam surgir em comparações múltiplas, é possível selecionar um método de ajuste de p-valor. Os métodos disponíveis são: Benjamini & Hochberg (1995), Benjamini & Yekutieli (2001), Bonferroni (Dunn 1961), Holm (1979), Hochberg (1988), Hommel (1988).

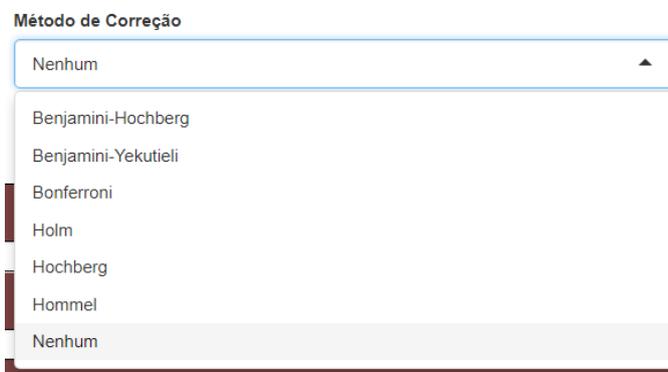


Figura 28: Opção de escolha “Métodos de Correção”

- *Box* interno – “Detecção dos itens”: A Figura 29 mostra esse *box* para o método de Lord. Logo no início do *box* há um texto informando quais os itens foram detectados com DIF. Abaixo é apresentada uma tabela com os resultados do método de detecção. Apesar das colunas dessa tabela modificarem a cada método, todas elas apresentam colunas com uma estatística de teste e seu p-valor, que conclui a detecção ou não de DIF no item.

Detecção dos Itens

Tabela Ajuda

Foram detectados os seguintes itens com funcionamento diferencial (DIF) 16, 17, 110, 115

10 resultados por página

Pesquisar

	Lord Qui	p-valor	aR	EP(aR)	bR	EP(bR)	aF	EP(aF)	bF	EP(bF)
11	3.161	0.206	1.514	0.205	-0.033	0.098	1.14	0.149	1.717	0.193
12	0.981	0.612	1.421	0.191	0.01	0.102	0.929	0.129	1.716	0.222
13	2.289	0.319	1.504	0.229	-1.164	0.15	0.91	0.113	-0.03	0.097
14	0.075	0.963	1.954	0.366	-1.733	0.195	1.412	0.16	-0.949	0.098
15	0.449	0.799	2.231	0.294	-0.029	0.083	1.721	0.206	1.428	0.125
16	0.658	0.035	1.327	0.184	-0.351	0.11	0.698	0.102	0.508	0.143
17	0.41	0.024	1.494	0.213	-0.802	0.122	1.195	0.136	0.853	0.109
18	0.65	0.722	1.678	0.227	0.903	0.103	1.108	0.155	2.058	0.241
19	2.134	0.344	1.112	0.165	-0.019	0.117	0.667	0.104	1.192	0.215
110	14.711	0.001	1.087	0.146	0.524	0.135	1.344	0.161	1.318	0.136

Mostrando de 1 até 10 de 20 registros

Anterior 1 2 Próximo

Figura 29: Box “Detecção dos Itens” com DIF

- **Box interno – “Análise gráfica”:** Na Figura 30 é apresentada a parte inicial deste *box* para o método Logito cumulativo. Este *box* apresenta os gráficos com as curvas para cada categoria de resposta dos itens com DIF considerando os grupos Referência (linha contínua) e Focal (linha tracejada).

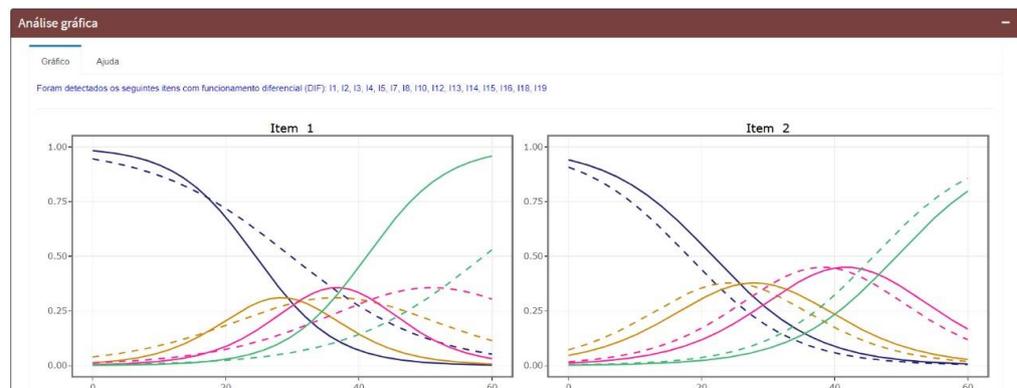


Figura 30: Box “Análise gráfica” DIF em itens politômicos

Para o cálculo dos métodos de identificação em itens dicotômicos (Regressão Logística e Lord) foi utilizado o pacote *difR* (Magis et al., 2010). Já para o cálculo dos métodos de identificação em itens politômicos (Logito Cumulativo e Logito Categoria Adjacente) foi utilizado o pacote *difNLR* (Hladká & Martinkova 2021). Para a geração das tabelas o pacote *DT* (Xie et al., 2021). E, para a geração dos gráficos, o pacote *plotly* (Sievert, 2020).

8. MENU EXPORTAÇÃO DE DADOS

Após o ajuste de modelos, pode ser de interesse do pesquisador salvar os resultados desse ajuste para posteriormente revisá-los ou compará-los com outros ajustes. Também pode ser de interesse editar os nomes de eixos de gráficos e salvá-los para incluí-los em publicações.

O aplicativo TRIDIF traz essas possibilidades no menu “Exportação de dados” que pode ser visto na Figura 31. Este menu possui três *boxes* retráteis, que são: “Exportação de tabelas de resultados”, “Exportação de gráficos” e “Código R”. Os dois primeiros estão detalhados a seguir.

- *Box* “Exportação de tabelas de resultados”: Neste *box* pode ser feito o download de um arquivo *xlsx* com três tabelas de dados. A primeira tabela traz informações gerais do modelo, tipo de dado, itens detectados com funcionamento diferencial, nome e medidas de ajuste do modelo. A segunda tabela traz os dados de entrada com o acréscimo das estimativas de traço latente dos respondentes. E, a terceira tabela traz as estimativas dos coeficientes dos itens ajustados pelo modelo. O download desse arquivo também pode ser feito no menu “Ajuste do modelo”.
- *Box* “Exportação de gráficos”: Neste *box* podem ser salvas imagens com os gráficos gerados pelo modelo. Antes de exportar é possível modificar o nome dos eixos ou até das séries como no caso do gráfico “Curva de informação do modelo”. As imagens são salvas no formato *png*.

Para o download da tabela de dados é utilizado o pacote *xlsx* (Dragulescu & Arendt, 2020). Já o download de imagens é uma funcionalidade disponibilizada pelo pacote *plotly* (Sievert, 2020) que é utilizado para construção dos gráficos do aplicativo.

Entrada de dados Análises descritivas Unidimensionalidade Presença de DIF Ajuste do modelo **Exportação de dados**

Exportação de tabelas de resultados

Após o ajuste do modelo dicotômico Logístico de 2 parâmetros (2PL), tanto os dados da matriz de entrada são juntados às estimativas dos traços latentes dos respondentes quanto os coeficientes estimados dos itens, podem ser exportados num arquivo XLSX.

[Download](#)

Dados
dicotômicos

Exportação de gráficos

Histograma Boxplot Curva de informação do modelo Curva característica do item Curva informação do item

Eixo X

Eixo Y

[Redesenhar](#) [Restaurar](#)

Para download do gráfico, passe o mouse sobre a imagem e pressione a máquina fotográfica ('download plot as a png') que aparece no topo da figura

Código R

```
#tabela informações gerais
ajuste = data.frame(
  campo = c("Modelo", "log-likelihood", "AIC", "AICC", "BIC"),
  valor = c(modelo@Call$itemtype, modelo@Fit$logLik, modelo@Fit$AIC, modelo@Fit$AICC, modelo@Fit$BIC)
)
write.xlsx(ajuste, "resultados.xlsx", sheetName="Ajuste do modelo", append=T, row.names = F)

#tabela traços latentes
tracos = data.frame(DF,
  escore = rowSums(matriz, na.rm = T),
  escorepadrao = scale(rowSums(matriz, na.rm = T)),
  escoretri = fscores(modelo, full.scores.SE = TRUE)[,1],
  sp_escoretri = fscores(modelo, full.scores.SE = TRUE)[,2])
write.xlsx(tracos, "resultados.xlsx", sheetName="Traços latentes", append=T)

#tabela coeficientes dos itens
coeficientes = coef(modelo, IRTpars = TRUE, simplify = TRUE)$items
write.xlsx(coeficientes, "resultados.xlsx", sheetName="Coeficientes", append=T)

#grafico em imagem
jpeg ("nomedoarquivo.jpg")
hist(fscores(modelo, full.scores.SE = TRUE)[,1],
  xlab = "Titulo eixo x",
  ylab = "Titulo eixo y")
dev.off ()
```

Figura 31: Menu “Exportação de dados”

9. CONCLUSÃO

Tendo em vista a ampla utilização da teoria de resposta ao item e a necessidade de várias etapas para um bom ajuste de modelos TRI, o aplicativo TRIDIF foi desenvolvido para facilitar o pesquisador na realização de todas estas etapas. O aplicativo possibilita realizar as análises em uma interface amigável e interativa, com download facilitado dos resultados em tabelas e gráficos, sem a necessidade do conhecimento da linguagem R.

Como apoio ao pesquisador, o aplicativo possui uma ajuda rápida e sucinta para colunas de tabelas ou eixos de gráfico e uma ajuda detalhada, onde são apresentadas as suas funcionalidades e fundamentações teóricas para as análises realizadas.

Caso o pesquisador deseje, poderá utilizar o próprio console do R para reproduzir a análise realizada nos menus do aplicativo, o *box* “Código R” presente em todos os menus permite essa reprodução. Ou, ainda, se o pesquisador for um programador R e queira personalizar ainda mais seus gráficos e tabelas, os dados de exemplo e códigos utilizados no aplicativo TRIDIF estão disponibilizados no GitHub³.

³ <https://github.com/marianalimagarcia/TRIDIF>

REFERÊNCIAS

AKAIKE, Hirotugu. A new look at the statistical model identification. **IEEE transactions on automatic control**, v. 19, n. 6, p. 716-723, 1974.

ANDRADE, Dalton Francisco de; TAVARES, Heliton Ribeiro; DA CUNHA VALLE, Raquel. **Teoria da Resposta ao Item: conceitos e aplicações**. ABE, Sao Paulo, 2000.

ANDRICH, David. A rating formulation for ordered response categories. **Psychometrika**, v. 43, n. 4, p. 561-573, 1978.

ANDRIOLA, Wagner Bandeira. Descrição dos principais métodos para detectar o funcionamento diferencial dos itens (DIF). **Psicologia: reflexão e crítica**, v. 14, n. 3, p. 643-652, 2001.

BENJAMINI, Yoav; HOCHBERG, Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal statistical society: series B (Methodological)**, v. 57, n. 1, p. 289-300, 1995.

BENJAMINI, Yoav; YEKUTIELI, Daniel. The control of the false discovery rate in multiple testing under dependency. **Annals of statistics**, p. 1165-1188, 2001.

BIRNBAUM, Allan. Lord, Frederic Mather. Some latent trait models and their use in inferring an examinee's ability. **Statistical theories of mental test scores**, 1968.

BOZDOGAN, Hamparsum. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. **Psychometrika**, v. 52, n. 3, p. 345-370, 1987.

CASTRO, Stela Maris de Jesus. **Teoria da resposta ao item: aplicação na avaliação da intensidade de sintomas depressivos**. 2008.

CHALMERS, R. Philip (2012). **mirt: A Multidimensional Item Response Theory Package for the R Environment**. *Journal of Statistical Software*, 48(6), 1-29. doi:10.18637/jss.v048.i06

CHANG, Winston; BORGES, Barbara Ribeiro. (2021). **shinydashboard: Create Dashboards with 'Shiny'**. R package version 0.7.2. <http://rstudio.github.io/shinydashboard>

CHANG, Winston, CHENG, Joe, ALLAIRE, JJ, SIEVERT, Carson, SCHLOERKE, Barret, XIE, Yihui, ALLEN, Jeff, MCPHERSON, Jonathan, DIPERT, Alan e BORGES, Barbara (2021). **shiny: Web Application Framework for R**. R package version 1.6.0. <https://CRAN.R-project.org/package=shiny>

CRANE, Paul K.; BELLE, Gerald van; LARSON, Eric B. Test bias in a cognitive test: differential item functioning in the CASI. **Statistics in Medicine**, v. 23, n. 2, p. 241-256, 2004.

CRONBACH, Lee J. Coefficient alpha and the internal structure of tests. **Psychometrika**, v. 16, n. 3, p. 297-334, 1951.

DRAGULESCU, Adrian & ARENDT, Cole (2020). **xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files**. R package version 0.6.5. <https://CRAN.R-project.org/package=xlsx>.

DUNN, Olive Jean. Multiple comparisons among means. **Journal of the American statistical association**, v. 56, n. 293, p. 52-64, 1961.

EMBRETSON, Susan E.; REISE, Steven. P. **Item response theory for psychologists**. 2000. Lawrence Earlbaum Associates, Mahwah, NJ.

FRENCH, Ann W.; MILLER, Timothy R. Logistic regression and its use in detecting differential item functioning in polytomous items. **Journal of Educational Measurement**, v. 33, n. 3, p. 315-332, 1996.

HAYS, Ron D.; MORALES, Leo S.; REISE, Steve P. Item response theory and health outcomes measurement in the 21st century. **Medical care**, v. 38, n. 9 Suppl, p. II28, 2000.

HLADKÁ, Adéla. **Statistical models for detection of differential item functioning**. 2021.

HLADKA, Adéla & MARTINKOVA, Patricia (2021). **difNLR: DIF and DDF detection by non-linear regression models**. R package version 1.3.7.

HOCHBERG, Yosef. A sharper Bonferroni procedure for multiple tests of significance. **Biometrika**, v. 75, n. 4, p. 800-802, 1988.

HOLM, Sture. A simple sequentially rejective multiple test procedure. **Scandinavian journal of statistics**, p. 65-70, 1979.

HOMMEL, Gerhard. A stagewise rejective multiple test procedure based on a modified Bonferroni test. **Biometrika**, v. 75, n. 2, p. 383-386, 1988.

KIM, Jihye; OSHIMA, T. C. Effect of multiple testing adjustment in differential item functioning detection. **Educational and Psychological Measurement**, v. 73, n. 3, p. 458-470, 2013.

LORD, Frederic Mather. A theory of test scores (Psychometric Monograph No. 7). Iowa City, IA: **Psychometric Society**, v. 35, 1952.

LORD, Frederic Mather. **Applications of Item Response Theory to Practical Testing Problems**. Lawrence Erlbaum Associates. Inc, Hillsdale, NJ, 1980.

MAGIS, D., BÉLAND, S., TUERLINCKX, F., & DE BOECK, P (2010). A general framework and an R package for the detection of dichotomous differential item functioning. **Behavior Research Methods**, 42, 847-862.

MASTERS, Geoff N. A Rasch model for partial credit scoring. **Psychometrika**, v. 47, n. 2, p. 149-174, 1982.

MCHORNEY, Colleen A.; COHEN, Allan S. Equating health status measures with item response theory: illustrations with functional status items. **Medical care**, p. 1143-1159, 2000.

MELLENBERGH, Gideon J. Contingency table models for assessing item bias. **Journal of educational statistics**, v. 7, n. 2, p. 105-118, 1982.

MURAKI, Eiji. A generalized partial credit model: Application of an EM algorithm. **ETS Research Report Series**, v. 1992, n. 1, p. i-30, 1992.

NIJS, Vincent; FANG, Forest, Trestle Technology, LLC and Jeff Allen (2019). **shinyAce: Ace Editor Bindings for Shiny**. R package version 0.4.1. <https://CRAN.R-project.org/package=shinyAce>

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. **Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests**. 1960.

REVELLE, William. (2020). **psych: Procedures for Personality and Psychological Research**. Northwestern University, Evanston, Illinois, USA, R package version 2.0.12 <https://CRAN.R-project.org/package=psych>

RIZOPOULOS, Dimitris (2006). **ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses**, Journal of Statistical Software, 17 (5), 1-25. URL <http://www.jstatsoft.org/v17/i05/>

RStudio Team (2021). **RStudio: Integrated Development Environment for R**. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>

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

SCHWARZ, Gideon. Estimating the dimension of a model. **The annals of statistics**, p. 461-464, 1978.

SIEVERT, Carson. Interactive **Web-Based Data Visualization with R, plotly, and shiny**. Chapman and Hall/CRC Florida, 2020.

SISTO, Fermino Fernandes. O funcionamento diferencial dos itens. **Psico-USf**, v. 11, p. 35-43, 2006.

SWAMINATHAN, Hariharan; ROGERS, H. Jane. Detecting differential item functioning using logistic regression procedures. **Journal of Educational measurement**, v. 27, n. 4, p. 361-370, 1990.

XIE, Yihui, CHENG, Joe e TAN, Xianying (2021). **DT: A Wrapper of the JavaScript Library 'DataTables'**. R package version 0.19. <https://github.com/rstudio/DT>

ZUMBO, Bruno D. **A handbook on the theory and methods of differential item functioning (DIF)**. Ottawa: National Defense Headquarters, v. 160, 1999.