



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



# DETECÇÃO DE FUNCIONAMENTO DIFERENCIAL DO ITEM ATRAVÉS DA REGRESSÃO LOGÍSTICA E DA TEORIA DA RESPOSTA AO ITEM – UMA INTERFACE GRÁFICA

Autor: Ruben Ladwig

Orientador: Professora Dr<sup>a</sup>. Stela Maris de Jesus Castro

Porto Alegre, 02 de Julho de 2012.

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática  
Departamento de Estatística

DETECÇÃO DE FUNCIONAMENTO  
DIFERENCIAL DO ITEM ATRAVÉS DA  
REGRESSÃO LOGÍSTICA E DA TEORIA  
DA RESPOSTA AO ITEM - UMA  
INTERFACE GRÁFICA

Autor: Ruben Ladwig

Monografia apresentada para obtenção  
do grau de Bacharel em Estatística.

Banca Examinadora:  
Professora Dr<sup>a</sup>. Stela Maris de Jesus Castro  
Professora MSc. Vanessa Bielefeldt Leotti Torman

Porto Alegre, 02 de Julho de 2012.

Dedico este meu trabalho à minha família, que é minha base e meu norte.

*"O mestre disse a um dos seus alunos: Queres saber em que consiste o conhecimento? Consiste em ter consciência tanto de conhecer uma coisa quanto de não a conhecer. Este é o conhecimento."*

-Confúcio

## **AGRADECIMENTOS**

Agradeço à minha professora orientadora Stela Maris de Jezus Castro pelo apoio que demonstrou durante a elaboração deste trabalho e por me apresentar o fascinante mundo da teoria da resposta ao item.

Agradeço à Professora Suzi Camey e ao colega Isafias Prestes, por terem me dado a primeira oportunidade dentro do curso de estatística, onde aprendi muito sobre programação e estatística como um todo. À Professora Marilene Bandeira, que com sua disponibilidade me ajudou a tomar a difícil decisão de mudar o meu rumo profissional. A todo departamento de estatística, que com seus professores e funcionários me fizeram perceber o mundo de outra forma.

Agradeço também aos colegas, que com sua amizade tornaram toda esta jornada muito mais bonita. Em especial, agradeço à colega Greice Helen da Costa Laureano, cujo companheirismo e bondade jamais serão esquecidos.

Finalmente, agradeço à minha família, cujo suporte tornou possível essa difícil jornada. A meu pai, Luiz Carlos, que sempre me mostrou o que era correto e pra sempre será o meu modelo. A minha mãe, Luciane, por seu apoio incondicional. Aos meus irmãos, Lucas, Luiza e Bruna, por sempre colocar um sorriso no meu rosto. Amo a todos vocês, para sempre.

## RESUMO

Neste trabalho revisamos o método de detecção de funcionamento diferencial do item (DIF) através do uso da regressão logística ordinal, usando como critério de correspondência as estimativas de traço latente obtidas pelo modelo da teoria da resposta ao item, utilizando o modelo de resposta gradual de Samejima. A este método aplicamos um processo iterativo de purificação das estimativas do traço latente, utilizando parâmetros específicos dos subgrupos em itens que apresentem DIF. Este método todo está implementado ou em softwares proprietários ou no software R, que apresenta uma interface pouco amigável. Como alternativa, propomos uma interface gráfica que funciona sob o R, a fim de tornar o método acessível a um público maior. Como ilustração, aplicamos o método para identificar DIF em um banco de dados simulado, onde encontramos DIF em 3 dos 10 itens.

**Palavras-chave:** Funcionamento diferencial do item, DIF, Regressão Logística Ordinal, Teoria da resposta ao Item, R.

## SUMÁRIO

Agradecimentos .....	5
Resumo .....	6
Sumário.....	7
1. Introdução .....	8
2. Funcionamento diferencial do item (DIF) .....	9
2.1. Teoria da resposta ao item (TRI) .....	9
3. Método (regressão logística – TRI) para detecção de DIF .....	12
3.1. Definição do modelo de Regressão logística ordinal .....	12
3.2. Critério de correspondência.....	14
3.3. Método iterativo de purificação.....	15
4. Implementação em interface gráfica.....	16
4.1. Instruções de uso da interface.....	17
5. Aplicação a dados simulados.....	23
5.1. Resultados.....	23
6. Conclusão .....	29
Referências bibliográficas .....	30

## 1. INTRODUÇÃO

Frequentemente, deseja-se conhecer ou medir variáveis que não podem ser medidas diretamente. Estas variáveis são conhecidas como traços latentes, os quais são obtidos através de questionários, isto é, instrumentos compostos por itens que medem indiretamente a característica de interesse, como por exemplo, níveis de depressão (Castro 2008), proficiência em alguma área do conhecimento, etc. Entretanto, estes itens podem ter um desempenho distinto (funcionamento diferencial do item – DIF) entre subgrupos irrelevantes para o traço latente. Isto é, entre os subgrupos há uma probabilidade distinta de endossar uma resposta em específico, para o mesmo nível de traço latente (Camilli e Shepard 1994).

Portanto, a identificação de tais itens é de grande importância para o pesquisador, pois essa diferença sistemática na medida pode levar a erros no teste de hipóteses, planejamento de políticas e implementações, e levar à pesquisa equivocada de diferenças no traço latente que na verdade não existem (Perkins *et al.* 2006).

Existem muitos métodos para a detecção de DIF (De Boeck e Wilson 2004, Jones 2006, Shealy e Stout 1993). Este trabalho se focará na apresentação do método de regressão logística ordinal, que permite detectar vários tipos de DIF (Swaminathan e Rogers 1990; Zumbo 1999). Este método é aprimorado com o uso das estimativas dos traços latentes obtidas através da teoria da resposta ao item (TRI) como critério de correspondência, ao invés do escore total, e também do uso de um processo iterativo que leva em consideração o DIF nas estimativas dos traços latentes em cada passo, eliminando a identificação errônea ou a não identificação em virtude da presença de DIF (Crane *et al.* 2006). A combinação destes métodos permite identificar múltiplas fontes de DIF e levar em consideração também o impacto do DIF na medida do traço latente.

Esta metodologia está implementada no software livre R (<http://www.r-project.org>) através do pacote **lordif** (Choi *et al.* 2011), e este trabalho propõe uma interface gráfica afim de facilitar o uso do método.



## 2. FUNCIONAMENTO DIFERENCIAL DO ITEM (DIF)

Citando Steinberg e Thissen (2006 p. 404):

*(...) DIF significa que um item tem desempenho diferente, ou mede de forma diferente, para um subgrupo da população em relação a outro subgrupo para o mesmo nível do construto; um item que apresenta DIF necessariamente é menos válido para pelo menos um dos subgrupos da população.*

A existência de DIF em um item significa que dois indivíduos, pertencentes a subgrupos distintos da população estudada, que tenham o mesmo nível de traço latente, tem probabilidades distintas para dar certa resposta para este item, quando seria esperado que a probabilidade fosse igual.

O DIF independe do modelo utilizado para a análise do questionário, entretanto, a sua definição geralmente se dá, em termos da teoria da resposta ao item, através da sua função de resposta ao item. Como a função de resposta ao item tem correspondência direta com os parâmetros do mesmo, a detecção de DIF basicamente se dá através do teste da hipótese nula de que um item tem os mesmos parâmetros para todos os subgrupos da população, ou seja, de que o item apresenta o mesmo comportamento para todos os subgrupos da população.

### 2.1. TEORIA DA RESPOSTA AO ITEM (TRI)

A Teoria da Resposta ao Item é uma técnica utilizada tanto para criar e avaliar instrumentos (questionários) que tentam medir traços latentes, como também para estimá-los. O foco da análise se dá em cada item, ao contrário da teoria clássica do teste, que se foca no instrumento como um todo (Andrade *et al.* 2000). A ideia central é a de que as respostas a cada item covariam, em virtude da relação dos itens com o traço latente. Com isso, pode-se estimar a relação entre a probabilidade de um indivíduo dar uma resposta e o seu nível de traço latente que está sendo avaliado.

Os modelos TRI unidimensionais cumulativos podem ser especificados em torno de quantos parâmetros serão estimados (figura 1) e o tipo de resposta dada (dicotômica ou politômica). De particular interesse para este trabalho estão os modelos de dois parâmetros, os quais estimam os parâmetros de discriminação e dificuldade. O parâmetro de

discriminação é uma medida do quanto o item é capaz de discriminar entre indivíduos que tenham níveis distintos do traço latente, enquanto o parâmetro de dificuldade é uma medida do quão alto deve ser o nível do traço latente para um indivíduo endossar aquele item. Estes modelos são definidos a seguir:

a) Modelo de dois parâmetros dicotômico (Birnbaum 1968):

$$P(u_i = 1 | \theta) = \left[ \frac{1}{1 + e^{-a_i(\theta - b_i)}} \right]$$

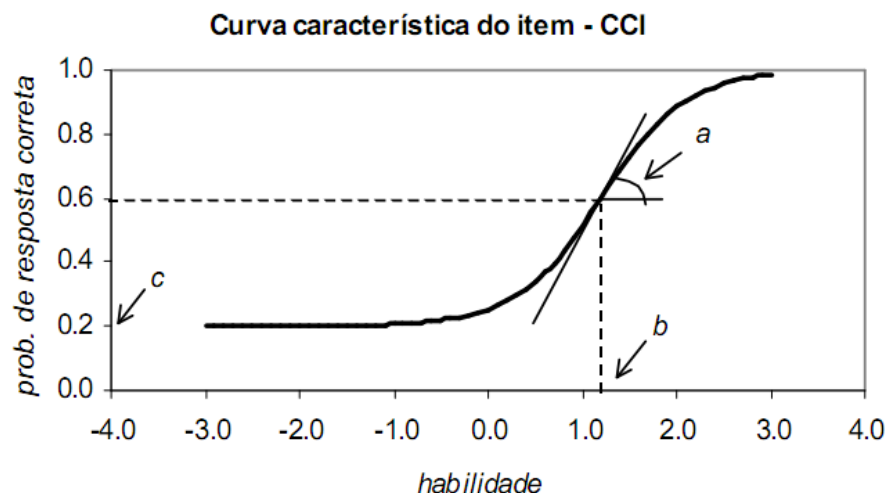
onde  $u_i$  é a resposta ao item  $i$ ,  $a_i$  é o parâmetro de discriminação do item  $i$ ,  $b_i$  é o parâmetro de dificuldade do item  $i$ , e  $\theta$  é o nível do traço latente.

b) Modelo de resposta gradual de Samejima (Samejima 1969):

$$P(u_i \geq k | \theta) = \left[ \frac{1}{1 + e^{-a_i(\theta - b_{ik})}} \right]$$

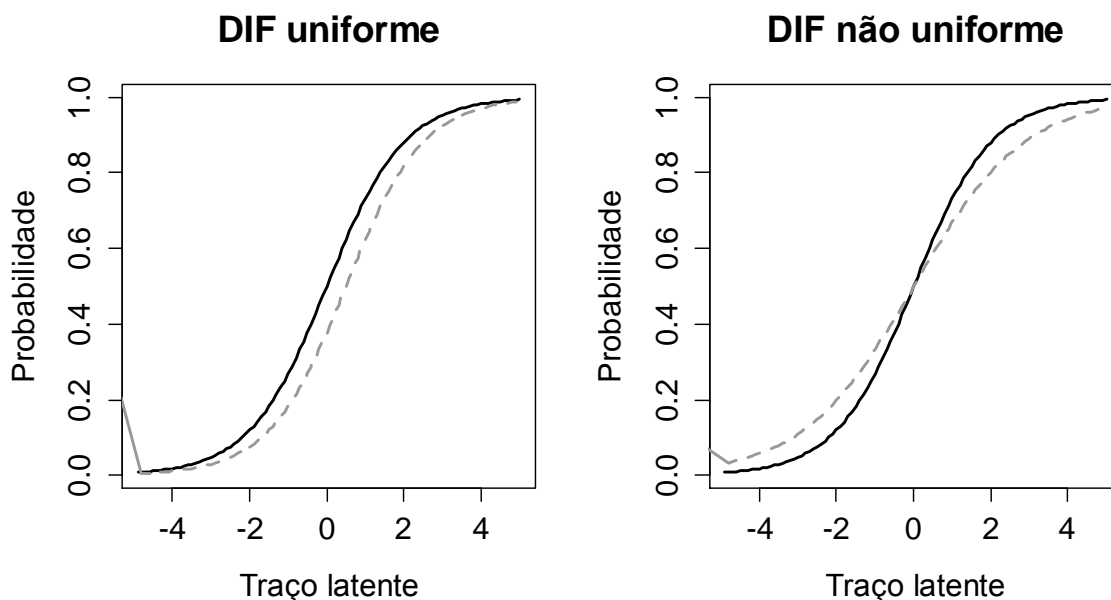
onde  $u_i$  é a resposta ao item  $i$  com níveis  $k = 0, 1, \dots, m-1$ ,  $a_i$  é o parâmetro de discriminação do item  $i$  e  $b_{ik}$  é o parâmetro de dificuldade da  $k$ -ésima categoria do item  $i$ , para um item com  $m$  categorias de resposta.

O modelo de resposta gradual de Samejima nada mais é que uma generalização do modelo de dois parâmetros dicotômico e é utilizado quando as respostas do item podem ser caracterizadas como respostas categóricas ordenadas, como é o caso das escalas Likert.



**Figura 1- Ilustração dos possíveis parâmetros estimados, onde  $a$  é o parâmetro de discriminação do item,  $b$  é o parâmetro de dificuldade do item e  $c$  é o parâmetro de acerto casual do item.**

É possível definir o DIF em torno destes parâmetros. Se há diferenças no parâmetro de dificuldade  $b$ , diz-se que o DIF é uniforme, isto é, há diferença nas probabilidades de resposta entre os subgrupos para aquele item, para qualquer nível do traço. Se há diferenças no parâmetro de discriminação  $a$ , diz-se que o DIF é não uniforme, isto é, há interação entre o nível do traço latente, o subgrupo e a probabilidade de resposta para aquele item latente (figura 2).



**Figura 2 - Ilustração dos dois tipos de funcionamento diferencial do item em um modelo de dois parâmetros. À esquerda temos o DIF uniforme, que nada mais é que um deslocamento da curva característica. À direita, o DIF não uniforme, evidenciando que a discriminação (inclinação) é diferente para os dois grupos.**

### 3. MÉTODO (REGRESSÃO LOGÍSTICA – TRI) PARA DETECÇÃO DE DIF

#### 3.1. DEFINIÇÃO DO MODELO DE REGRESSÃO LOGÍSTICA ORDINAL

Seja  $U_i$  uma variável aleatória discreta ordinal representando a resposta ao item  $i$ , e seja  $u_i = 0, 1, \dots, m_i-1$  a resposta efetiva do item  $i$  com  $m_i$  categorias de resposta. São criados três modelos de regressão logística ordinal para cada item. Cada modelo adiciona um novo termo explicativo.

**Modelo 1:**  $\text{logit } P(u_i \geq k) = \alpha_0 + \beta_1 \cdot \theta$

**Modelo 2:**  $\text{logit } P(u_i \geq k) = \alpha_0 + \beta_1 \cdot \theta + \beta_2 \cdot (\text{grupo})$

**Modelo 3:**  $\text{logit } P(u_i \geq k) = \alpha_0 + \beta_1 \cdot \theta + \beta_2 \cdot (\text{grupo}) + \beta_3 \cdot (\theta \cdot \text{grupo})$

onde  $k$  é a categoria da resposta  $u_i$ ;  $P(u_i \geq k)$  é a probabilidade cumulativa que a resposta efetiva do item,  $u_i$ , seja  $k$  ou maior,  $\theta$  é o termo relativo ao traço latente e é usado como critério de correspondência (definido na seção 3.2),  $(\text{grupo})$  é uma variável referente ao subgrupo ao qual o indivíduo pertence e  $(\theta \cdot \text{grupo})$  é a interação entre o traço latente e o grupo.  $\alpha_0$  é o ponto de corte associado a cada categoria da variável resposta no modelo de regressão logística ordinal (McCullagh e Nelder 1989) e os  $\beta$ 's são os parâmetros de regressão associados a cada termo explicativo.

O modelo 1, portanto, supõe que não há a presença de DIF. O modelo 2 supõe a presença de DIF uniforme, e o modelo 3 a presença de DIF uniforme e não uniforme.

### 3.1.1. Critérios para a detecção do DIF

Quando a inserção de um novo termo explicativo melhora significativamente o modelo, é sinal de que o item analisado apresenta DIF.

Podemos comparar as diferenças entre os modelos a partir de testes da razão de verossimilhança, o que possibilita a identificação da presença, ou não, de DIF (Swaminathan e Rogers 1990).

A diferença entre os modelos 1 e 2 é comparada com uma distribuição  $\chi^2$  com 1 grau de liberdade. Caso a estatística para o item seja significativa, o mesmo é identificado como tendo DIF uniforme. Fazendo-se a mesma comparação, considerando os modelos 2 e 3, pode-se identificar a presença do DIF não uniforme. Comparar diretamente os modelos 1 e 3, considerando uma distribuição  $\chi^2$  com 2 graus de liberdade, é uma forma de atestar a presença de DIF como um todo, sem definir se este é uniforme ou não uniforme.

Como o poder estatístico é associado ao tamanho da amostra, uma diferença muito pequena para efeitos práticos pode ser considerada estatisticamente significativa, dada uma amostra grande o suficiente. Assim, outros critérios foram propostos, tentando levar em consideração a magnitude desta diferença. Entre elas, foi proposto o uso de estatísticas pseudo- $R^2$  como uma medida de magnitude e classificação do DIF (Zumbo 1999). São comparadas as estatísticas pseudo- $R^2$  entre os modelos, e a literatura sugere que uma diferença entre elas menor que 0,13 é insignificante, entre 0,13 e 0,26 é considerada moderada e maior do que 0,26 é considerada grande. Outros autores, entretanto, consideram que esses níveis são muito altos e, portanto, podem levar a não identificação de DIF (Jodoin e Gierl 2001; Kim *et al.* 2007). Crane, van Belle e Larson (2004) propuseram o uso da diferença entre o parâmetro  $\beta_1$  dos modelos 1 e 2 para a detecção de DIF uniforme, através da fórmula:

$$\Delta\beta_1 = \left| (\beta_{1Modelo2} - \beta_{1Modelo1}) / \beta_{1Modelo1} \right|$$

Caso  $\Delta\beta_i$  seja maior do que 10%, o item é identificado como tendo DIF. Entretanto, outros estudos usaram pontos de corte diferentes, como 5% ou até mesmo 1%, a fim de detectar efeitos diferenciais menores (Crane *et al.* 2004).

O caso é que não há consenso na definição dos pontos de corte ou *thresholds* para estes critérios de detecção de DIF. Eventualmente, o pesquisador pode desejar utilizar-se de simulação por Monte Carlo para avaliar estes pontos de corte. Vários bancos de dados são gerados, sob a hipótese nula de que não há DIF nos itens e distribuições empíricas são geradas a partir das medidas de magnitude calculadas nos bancos.

O pacote **lordif** apresenta esta possibilidade. Mais detalhes sobre este procedimento podem ser encontrados em Choi (2011, p. 4).

### 3.2. CRITÉRIO DE CORRESPONDÊNCIA

Para que as comparações entre os subgrupos possam ser realizadas, a fim de detectar a presença do DIF, é necessário que os indivíduos destes subgrupos sejam comparados com os aqueles que tenham o mesmo nível de traço latente, logo, é necessário obter uma estimativa deste traço latente para se realizar a análise.

Uma das principais limitações do método de detecção de DIF por regressão logística ordinal é a dependência da soma total dos escores como critério de correspondência, o que é relativamente precário quando se leva em consideração que a soma não necessariamente corresponde ao nível do traço latente, a não ser que as propriedades do modelo estatístico de Rasch (Rasch 1960) estejam asseguradas. E mesmo quando o modelo de Rasch é adequado, a relação entre a soma e o nível do traço latente não é ótima, dada a natureza não linear do traço (Dorans e Holland 1993).

Para evitar este problema, utilizar-se de um método híbrido de regressão logística e teoria da resposta item que usa os escores TRI como critério de correspondência oferece uma escolha mais adequada para a detecção de DIF.

A fim de obter as estimativas do traço latente, é aplicado o modelo de resposta gradual de Samejima apresentado anteriormente, que nada mais é que uma generalização do modelo dicotômico de dois parâmetros. Como o objetivo principal é providenciar um

critério de correspondência, a escolha de um modelo TRI específico não é de grande importância (Cook 1999).

### 3.3. MÉTODO ITERATIVO DE PURIFICAÇÃO

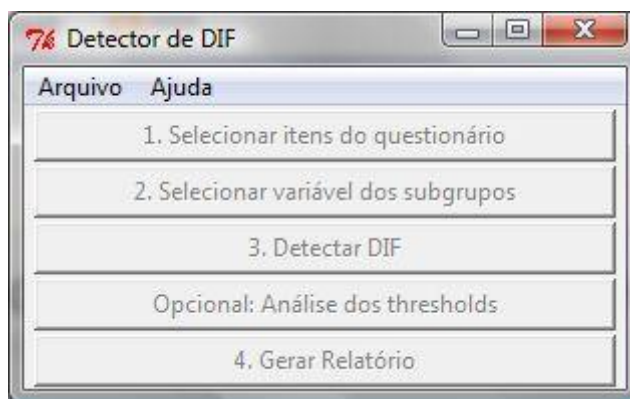
Quando um item é identificado como tendo DIF, pode-se purificar o critério de correspondência, utilizando-se dos parâmetros TRI dos subgrupos nos itens que apresentaram DIF. Tal prática diminui o efeito dos itens com DIF na presença de falsos positivos, isto é, quando o modelo indica que um item apresenta DIF, mas isso não é verdadeiro, e de falsos negativos, isto é, quando o modelo deixa de indicar a presença de um item com DIF, quando este na verdade apresenta o mesmo (Crane *et al.* 2006).

O método proposto por Crane recalcula os níveis do traço latente usando a informação do cálculo anterior. Itens que não tiveram DIF detectado usam dados da amostra inteira. Itens que apresentaram DIF são divididos em grupos referentes aos subgrupos da população. Estes itens são tratados de forma distinta nos subgrupos, tendo seus parâmetros calculados de maneira específica para cada subgrupo. Após isso, o processo de detecção é reiniciado. Se os itens que apresentam DIF são diferentes entre os dois cálculos, os falsos positivos e falsos negativos encontrados são atribuídos ao DIF. Esse processo é realizado sistematicamente, até que os itens que apresentam DIF sejam os mesmos em dois cálculos consecutivos.

#### 4. IMPLEMENTAÇÃO EM INTERFACE GRÁFICA

É de interesse prover uma forma amigável e gratuita para aplicar os métodos anteriormente abordados. Atualmente os métodos apresentados podem ser usados através do software **difwithpar**, que depende dos softwares proprietários **Stata** (StataCorp. 2007) e **PARSCALE** (Muraki e Bock 2005). O método foi estendido ao software livre **R** (R Development Core Team 2010) usando o pacote **lordif** (Choi *et al.* 2011).

Entretanto, o uso do software R não é trivial, e requer conhecimentos de comandos e o uso de uma interface por linha de comando. Em virtude disto, neste trabalho desenvolvemos uma interface gráfica em TCL/TK que permite o uso do pacote **lordif** de maneira intuitiva, sem a necessidade do aprendizado anterior da linguagem R. O usuário terá acesso à interface gráfica, sem ter que se preocupar com a manipulação e indexação de seus dados.



**Figura 3 - Janela principal da interface de detecção de DIF. Os botões são liberados de acordo com as definições que são dadas pelo usuário.**

Os principais comandos e opções referentes à metodologia anteriormente apresentada estão disponíveis em menus e botões de simples acesso, e a interface está estruturada num fluxo passo a passo (figura 3). Ao final da análise, o usuário gera um relatório que é salvo em seu computador, onde estão as saídas gráficas do método, permitindo a interpretação dos resultados. Além do pacote **lordif** e suas dependências, a interface utiliza alguns outros pacotes que permitem a leitura de bancos de dados de alguns softwares proprietários.



## 4.1. INSTRUÇÕES DE USO DA INTERFACE

Toda a documentação da interface aqui apresentada, bem como um exemplo de uso da mesma pode ser encontrada em: <https://sites.google.com/site/detectadif/documentacao>

### 4.1.1. Acesso à interface gráfica

Tendo acesso a um computador com sistema operacional Windows, onde esteja instalado o software R (na data da elaboração deste trabalho, a versão atual é a 2.15.0), o usuário deve executar o programa (o arquivo de instalação do mesmo pode ser obtido em <http://www.r-project.org/>) e entrar com as seguintes linhas de comando, observando o uso de letras maiúsculas e minúsculas (é necessário ter acesso à internet):

```
setInternet2(use = TRUE)
source("https://sites.google.com/site/detectadif/home/DETECTADIF.txt")
```

### 4.1.2. Instalação dos pacotes

Na primeira vez que a interface é executada, antes de proceder qualquer leitura de banco de dados, o usuário deve acessar o menu “Ajuda” e clicar na opção “Instalar pacotes necessários”. Isso instalará pacotes que contém os procedimentos necessários para a execução da interface e suas sub-rotinas internas. Este procedimento só precisa ser realizado uma única vez.

### 4.1.3 Abertura do banco de dados

Para abrir o banco de dados que será analisado, o usuário deverá clicar no menu “Arquivo” e selecionar a opção “Abrir banco de dados”. Os formatos atualmente aceitos pelo programa são: CSV (bancos em formato de texto) no padrão brasileiro, SAV (bancos

do SPSS) e XLS e XLSX (bancos do Excel)<sup>1</sup>. Bancos que estão em planilhas do Excel necessariamente precisam estar em uma planilha nomeada “Plan1” (nome da planilha, não do arquivo Excel).

Dado que os bancos em formato CSV são mais simples e a grande maioria dos softwares de bancos de dados tem a opção de exportar para este formato, sugerimos o uso deste formato para o usuário.

#### 4.1.3.1. Sobre a estrutura do banco de dados

Cada banco deve conter os dados referentes às respostas dos indivíduos pesquisados a cada um dos itens do instrumento e pelo menos uma variável referente aos subgrupos. Bancos no formato CSV ou Excel devem conter nomes das variáveis na primeira linha e os dados nas linhas subsequentes, como mostra o exemplo (figura 4).

	A	B	C	D	E	F	G	H	I
1	Sexo	Raça	Questão 1	Questão 2	Questão 3	Questão 4	Questão 5	Questão 6	Questão 7
2	Masculino	Branca	0	3	0	0	1	1	0
3	Feminino	Parda	3	1	1	0	1	2	3
4	Masculino	Branca	0	2	2	2	0	1	3
5	Masculino	Parda	0	3	3	3	2	1	0
6	Masculino	Negra	1	2	1	2	1	2	2
7	Feminino	Branca	1	0	3	2	3	3	3
8	Feminino	Parda	2	3	2	0	3	1	2
9	Masculino	Negra	2	2	3	2	2	3	0
10	Feminino	Branca	3	3	2	1	0	2	1
11	Masculino	Branca	2	0	2	1	2	2	1

**Figura 4 - Exemplo de banco de dados. No caso há duas variáveis referentes aos subgrupos (o mínimo é uma) e à direita estão as colunas referentes às respostas dos indivíduos.**

Não é permitido que nenhuma variável apresente dados faltantes, de outra forma, os parâmetros do modelo TRI não poderão ser calculados.

#### 4.1.4. Seleção das variáveis do instrumento e da variável que indica o subgrupo

Uma vez realizada a leitura do banco, o usuário deve clicar no botão “1. Selecionar itens do questionário”. Isso abrirá uma janela onde estão listados todos os nomes das

<sup>1</sup> Apenas em versões 32 bits do R

variáveis do banco. O usuário deve então selecionar todas as variáveis referentes às questões que compõem o seu instrumento. Isso pode ser feito pressionando shift+clique do botão esquerdo do mouse, para selecionar grupos de variáveis ou ctrl+clique do botão esquerdo do mouse, para selecionar as variáveis individualmente. Após feita a seleção, clique em OK.

Na janela principal, clique no botão “2. Selecionar variável dos subgrupos” para selecionar a variável que indica a qual subgrupo pertence cada indivíduo. Uma janela semelhante a anterior é aberta, mas aqui só é permitido selecionar uma única variável. Caso o usuário deseje realizar a detecção de DIF em relação à outra família de subgrupos, é só repetir o procedimento.

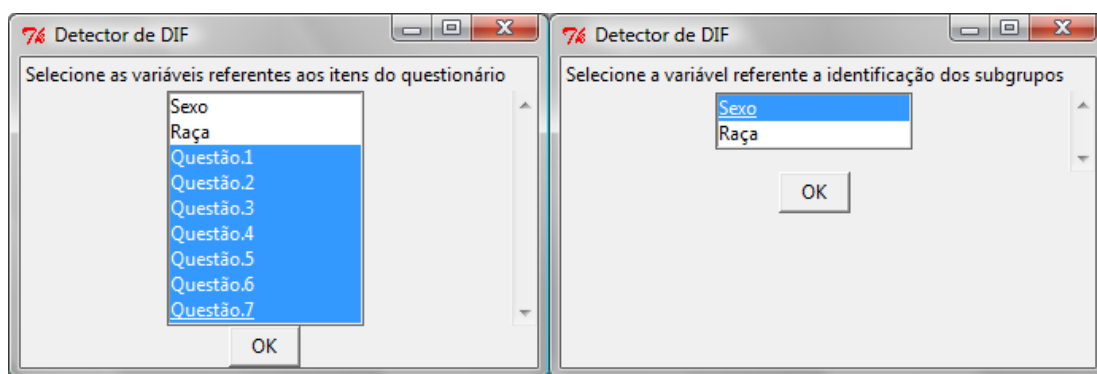


Figura 5 - Janelas de seleção de variáveis do questionário (esquerda) e subgrupos (direita)

#### 4.1.5 Detecção da presença de DIF

Uma vez selecionadas as variáveis, o usuário pode realizar a análise clicando no botão “3. Detectar DIF”. Ao fazê-lo, uma nova janela será aberta, e o usuário poderá selecionar qual método para a detecção de DIF deseja usar: “Qui-quadrado” para o teste da razão de verossimilhanças, “Pseudo-R<sup>2</sup>” para o uso da comparação das estatísticas Pseudo-R<sup>2</sup> dos modelos de regressão logística ordinal e “Diferença nos betas” para o uso das diferenças nos parâmetros referentes ao traço latente na regressão logística ordinal.

Os valores comumente utilizados como níveis de corte já estão especificados nas caixas de texto, mas se desejar, o usuário pode alterá-los, tendo o cuidado de usar o ponto (.) como separador decimal. Selecionado o critério desejado, pode-se clicar em “Calcular DIF” para iniciar o procedimento.

É importante salientar que, dependendo do tamanho do banco de dados e do método utilizado, este procedimento pode demorar. É importante aguardar a mensagem de finalização do procedimento que surgirá antes de partir para o próximo passo. Nessa mensagem estarão indicados quais itens apresentaram DIF.

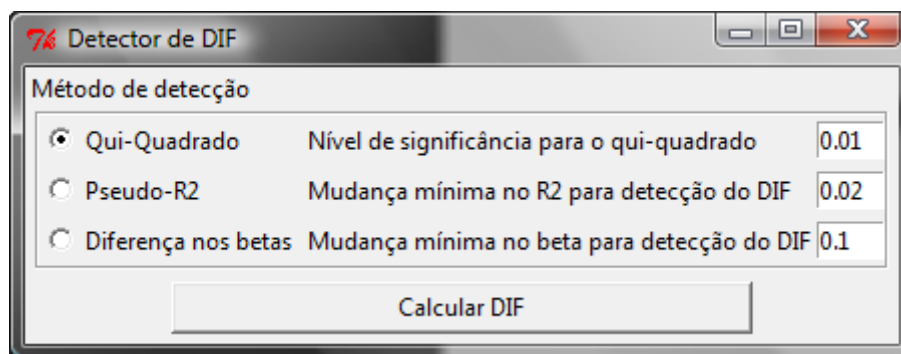


Figura 6 - Janela de seleção do método de detecção do DIF

#### 4.1.5 Geração do relatório

Finalizado o procedimento de detecção, o usuário pode clicar em “4. Gerar Relatório”, para visualizar os resultados.

Uma janela será aberta, perguntando onde deve ser salvo o relatório. O usuário deve especificar o local e o nome do arquivo, que deve ser terminado em “.PDF”. Na mesma pasta onde será salvo o arquivo PDF, será salvo um arquivo TXT contendo as estatísticas calculadas. Feito isso, o usuário deve fechar a interface e o R, para então poder ler o relatório.

##### 4.1.5.1. – Estatísticas encontradas no TXT

No relatório em TXT estão listadas as seguintes estatísticas, em colunas:

**item** – refere-se ao item analisado, na ordem em que aparece no banco de dados, da esquerda para a direita.

**ncat** – número de categorias encontradas naquele item, após possível agrupamento de categorias

**chi12** – p-valor do teste da razão de verossimilhanças entre o modelo 1 e 2

**chi13** – p-valor do teste da razão de verossimilhanças entre o modelo 1 e 3

**chi23** – p-valor do teste da razão de verossimilhanças entre o modelo 2 e 3

**beta12** – diferença nos betas, de acordo com a fórmula apresentada na seção 3.1.1.

**pseudo12.McFadden** – diferenças nas estatísticas pseudo-R2 de McFadden entre os modelos 1 e 2

**pseudo13.McFadden** – diferenças nas estatísticas pseudo-R2 de McFadden entre os modelos 1 e 3

**pseudo23.McFadden** – diferenças nas estatísticas pseudo-R2 de McFadden entre os modelos 2 e 3

**pseudo12.Nagelkerke** – diferenças nas estatísticas pseudo-R2 de Nagelkerke entre os modelos 1 e 2

**pseudo13.Nagelkerke** – diferenças nas estatísticas pseudo-R2 de Nagelkerke entre os modelos 1 e 3

**pseudo23.Nagelkerke** – diferenças nas estatísticas pseudo-R2 de Nagelkerke entre os modelos 2 e 3

**pseudo12.CoxSnell** – diferenças nas estatísticas pseudo-R2 de CoxSnell entre os modelos 1 e 2

**pseudo13.CoxSnell** – diferenças nas estatísticas pseudo-R2 de CoxSnell entre os modelos 1 e 3

**pseudo23.CoxSnell** – diferenças nas estatísticas pseudo-R2 de CoxSnell entre os modelos 2 e 3

**df12** – graus de liberdade para o teste da razão de verossimilhanças entre o modelo 1 e 2

**df13** – graus de liberdade para o teste da razão de verossimilhanças entre o modelo 1 e 3

**df23** – graus de liberdade para o teste da razão de verossimilhanças entre o modelo 2 e 3

#### 4.1.6 Análise dos thresholds

Opcionalmente, o usuário pode realizar uma análise dos *thresholds* clicando no botão “Opcional: Análise dos thresholds”. Este comando abre uma janela onde serão especificados o número de replicações do processo de Monte Carlo (recomenda-se que o número não seja menor do que 500 replicações) e o nível de significância, onde se deve ter o cuidado de usar o ponto (.) como separador decimal. Este é um processo bastante demorado, que depende muito do poder de processamento do computador, tamanho do banco de dados e número de replicações. Os resultados dessa análise são salvos no mesmo relatório que a detecção de DIF, de forma que este passo deve ser realizado antes da geração do relatório.

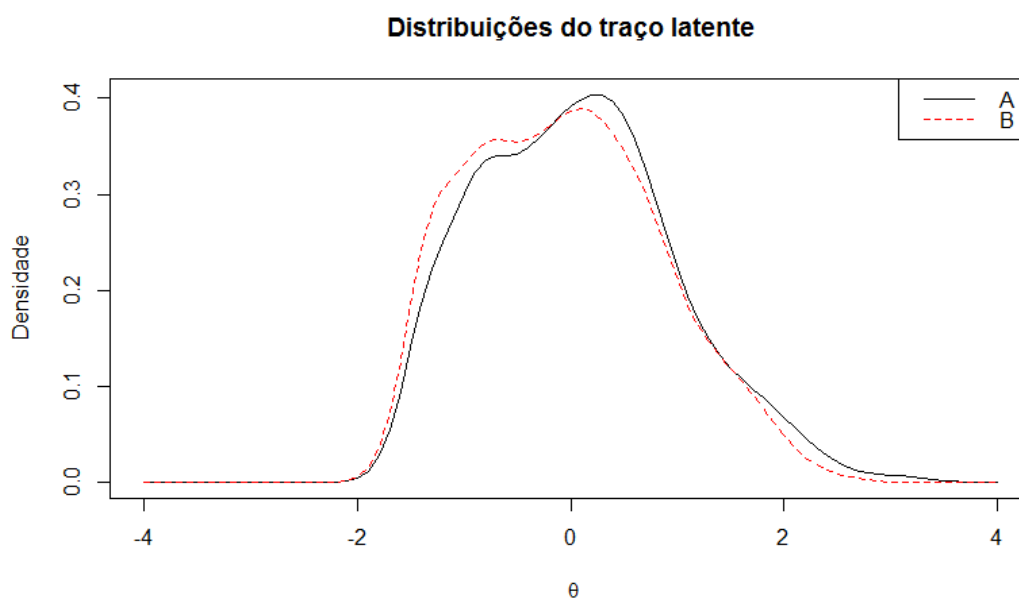
## 5. APLICAÇÃO A DADOS SIMULADOS

A fim de ilustrar o procedimento, aplicamos o mesmo em um banco de dados simulado, referente a um instrumento idealizado. Responderam ao questionário 1400 pessoas, que foram divididas entre pessoas do grupo A ( $n= 700$ ) e grupo B ( $n= 700$ ). O questionário é composto por politômica, de 10 itens, cada um com quatro alternativas com escores variando de 0 a 3, subentendendo graus crescentes de intensidade de um traço latente.

Com o banco resultante, podemos aplicar o método apresentado, utilizando como critério de detecção o teste da razão das verossimilhanças, ao nível de significância de 1% ( $\alpha=0,01$ ). Para a medida da magnitude, usamos o pseudo- $R^2$ .

### 5.1. RESULTADOS

Analisadas as distribuições do traço latente em cada subgrupo (figura 7), o método foi aplicado, e usou 2 iterações para purificar o modelo, encontrando DIF em 3 itens: Item 2, Item 3 e Item 10.



**Figura 7 - Distribuição do traço latente para os grupos A e B. A média do traço latente para o grupo A é de 0,065 (erro padrão 0,425) e para o grupo B é de -0,064 (erro padrão 0,437).**

Para cada item identificado com DIF, foram gerados 4 diagnósticos gráficos.

### 5.1.1. Item 2

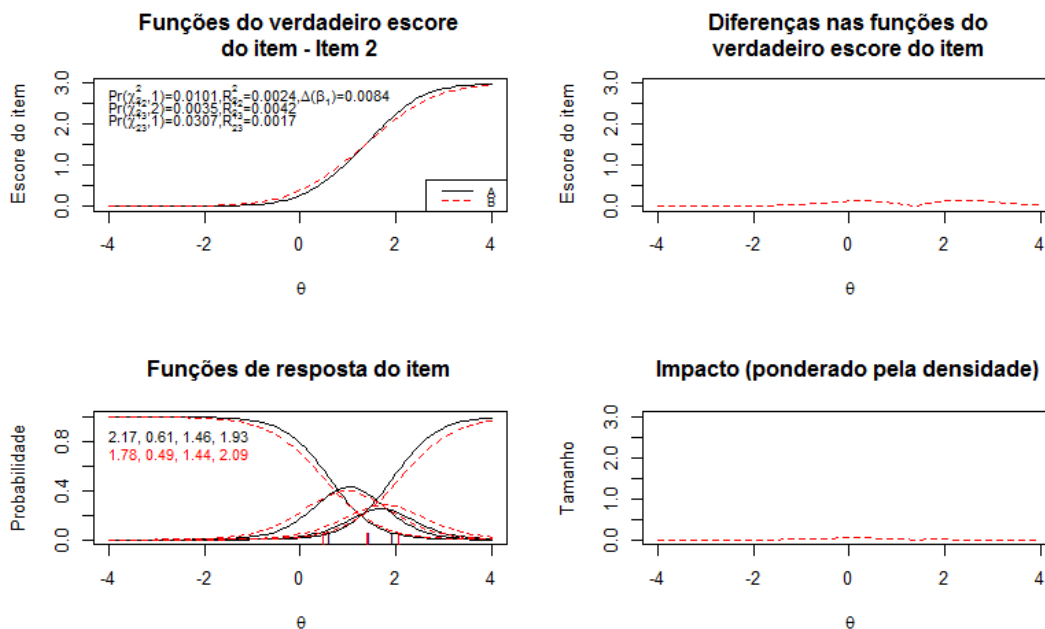


Figura 8 - Diagnósticos do item 2

O gráfico superior da esquerda (figura 8) mostra as funções do verdadeiro escore do item, baseadas nos parâmetros estimados para cada grupo. Os valores presentes neste gráfico representam os testes de razão de verossimilhança, as diferenças nos valores pseudo- $R^2$  e a diferença entre os parâmetros  $\beta_1$ . As inclinações das sigmóides são relativamente diferentes, entretanto, o teste  $\chi^2$  que compara os modelos 2 e 3 não se mostrou significativa (p-valor 0,0307), o que indica que não há presença DIF não uniforme. O teste que detecta a presença de DIF uniforme também não foi significativo, como mostra o teste  $\chi^2$  que compara os modelos 1 e 2 (p-valor 0,0101). Entretanto, se observarmos o teste que compara os modelos 1 e 3, encontramos um valor significativo (p-valor 0,0035), indicando que há a presença de DIF neste item, mas não atribuindo a um determinado tipo. O gráfico inferior da esquerda sobrepõe as funções de resposta do item para ambos os grupos. Os valores presentes no gráfico se referem, em ordem, a discriminação do item e a dificuldade para cada categoria do item em relação à primeira categoria.

O gráfico superior da direita mostra as diferenças entre as duas funções devidas ao DIF.



O gráfico inferior da direita mostra o impacto do DIF, que é a diferença encontrada ponderada pela densidade da amostra. No caso, o impacto parece ser negligenciável. Isso é evidenciado pelas estatísticas pseudo- $R^2$  (0,0024; 0,0042; 0,0017) apresentadas no primeiro gráfico, que estão muito abaixo do valor que a literatura indica como DIF relevante.

### 5.1.2. Item 3

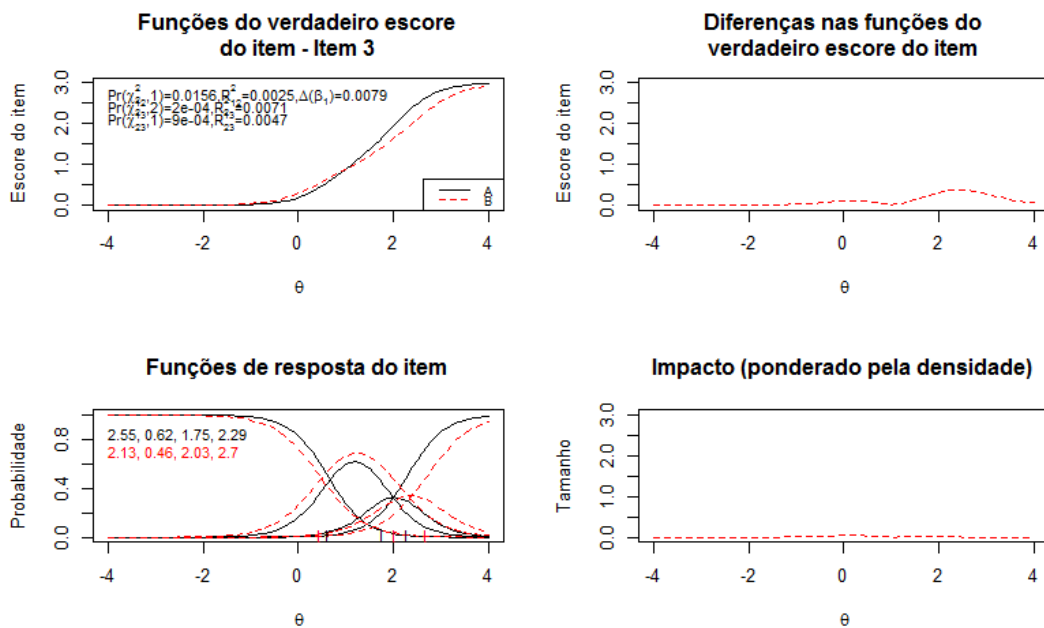


Figura 9 – Diagnósticos do item 3

Neste item (figura 9), as comparações entre os modelos através do teste  $\chi^2$  indicam a presença de DIF não uniforme (p-valor 0,009). As diferenças nos escores são maiores entre aqueles que têm um nível de traço latente aproximado  $\theta = 2,5$ . O impacto ponderado pela densidade se mostra irrisório, dada a pequena proporção de respondentes cujo traço latente se encontra neste nível.

### 5.1.3. Item 10

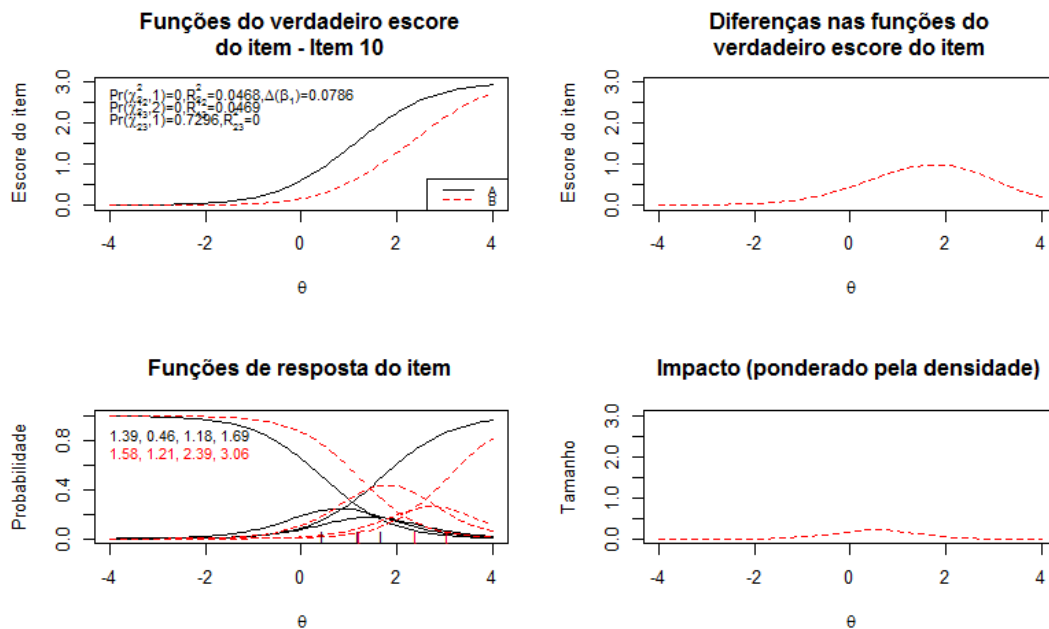


Figura 10 – Diagnósticos do item 10

Neste item (figura 10), as comparações entre os modelos através do teste  $\chi^2$  indicam a presença de DIF uniforme (p-valor 0). Pessoas do grupo A com níveis menores no traço parecem responder este item com níveis mais altos, em comparação com as do grupo B. As diferenças nos escores são maiores entre aqueles que têm um nível de traço latente em torno de  $\theta = 2$ . O impacto ponderado pela densidade neste item é mais evidente, como evidenciam os valores do pseudo- $R^2$ , especialmente pelo fato de uma diferença significativa nos escores se dar em níveis mais comuns do traço latente, na população estudada.

### 5.1.11. Efeito de todos os itens com DIF

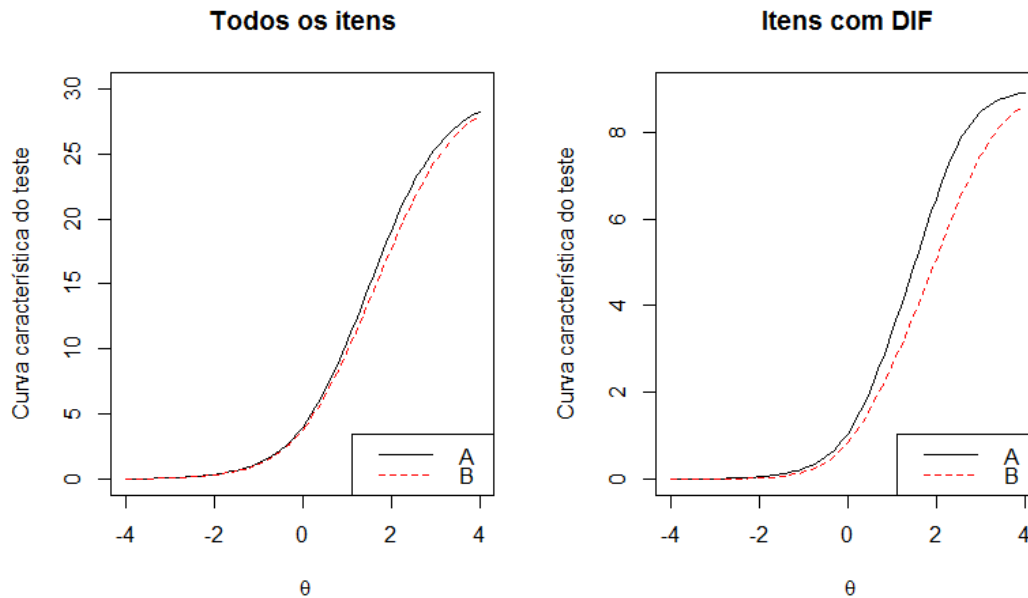
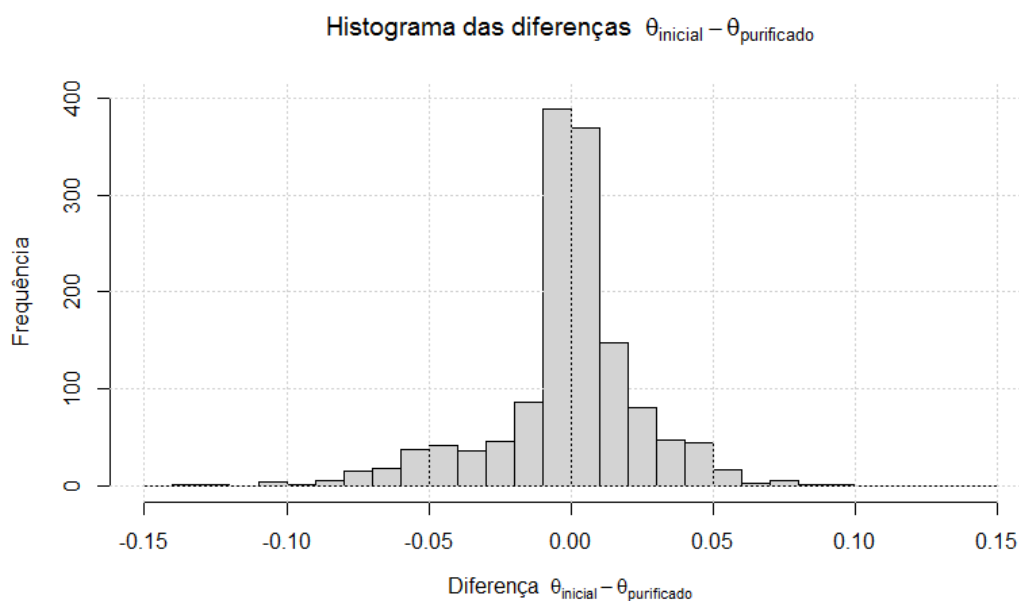


Figura 11 - Impacto dos itens com DIF nas curvas características do teste

O gráfico apresentado na esquerda (figura 11) é baseado nos parâmetros para todos os 10 itens, incluindo aí os parâmetros específicos de grupo para os itens que apresentaram DIF. O gráfico na direita é baseado apenas nos itens que apresentaram DIF. Percebe-se que as pessoas do grupo B obteriam um escore menor que uma pessoa do grupo A, para o mesmo nível do traço latente, especialmente para níveis relativamente altos do traço latente. E de fato, se analisarmos cada item com DIF em separado, veremos que o grupo B sempre obtém escores menores em níveis altos do traço latente.

Se analisarmos a diferença da estimativa do traço latente antes da purificação com a estimativa pós-purificação para os respondentes do questionário, a média é de -0,001, sendo que as diferenças variaram de -0,13 a 0,10. O intervalo interquartil é de -0,006 até 0,009. O histograma dessas diferenças (figura 12) mostra a distribuição dessas diferenças, de onde concluímos que, para a maioria das pessoas, a diferença nos escores é muito pequena.



**Figura 12 - Histograma das diferenças nas estimativas do traço latente**

Quando repetimos o procedimento, utilizando como critério de detecção as mudanças nos pseudo- $R^2$ , o DIF foi detectado apenas no item 10, justamente aquele onde há o maior impacto ponderado pela densidade. Se utilizarmos o critério de diferença nos coeficientes de regressão, não há DIF detectado.

## 6. CONCLUSÃO

Pelos resultados apresentados, fica evidente a importância de não só identificar o funcionamento diferencial dos itens, mas também estabelecer a magnitude do impacto do mesmo na estimativa dos traços latentes. O pacote lordif, e por extensão, a sua implementação em uma interface gráfica, permitem ao pesquisador explorar o seu instrumento e tomar uma decisão informada sobre a remoção ou não dos itens que o compõe e apresentam DIF.

## REFERÊNCIAS BIBLIOGRÁFICAS

Andrade DF, Tavares HR, Valle RC. Teoria da Resposta ao Item: Conceitos e Aplicações. *SINAPE*. (2000)

Beck AT, Steer RA. Manual for the Beck Depression Inventory. *San Antonio TX: Psychological Corporation*. (1993)

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores 395–479*. Reading, MA: Addison-Wesley (1968)

Camilli G, Shepard LA. Methods for identifying Biased Test Items. *Sage Publications, Thousand Oaks*. (1994)

Castro SMJ. Teoria da Resposta ao Item: Aplicação na Avaliação da Intensidade de Sintomas Depressivos. *Universidade Federal do Rio Grande do Sul, Porto Alegre*. (2008)

Choi SW, Gibbons LE, Crane PK. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *Journal of Statistical Software*, 39, issue 8. (2011)

Cook K. A Comparison of Three Polytomous Item Response Theory Models in the Context of Testlet Scoring. *Journal of Outcome Measurement*, 3, 1-20. (1999)

Crane PK, van Belle G, Larson EB. Test Bias in a Cognitive Test: Differential Item Functioning in the CASI.” *Statistics in Medicine*, 23, 241-256. (2004)

Crane PK, Gibbons LE, Jolley L, van Belle G. Differential Item Functioning Analysis with Ordinal Logistic Regression Techniques: DIFdetect and difwithpar. *Medical Care*, 44, n 11, supl 3, S115-S123. (2006)

De Boeck P, Wilson M. Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach. *Springer-Verlag, New York*. (2004)

Dorans NJ, Holland PW. DIF Detection and Description: Mantel-Haenszel and Standardization. *Differential Item Functioning*, 35-66, Erlbaum, Hillsdale, NJ. (1993)

Jodoin MG, Gierl MJ. Evaluating Type I Error and Power Rates Using an Effect Size Measure with logistic regression Procedure for DIF Detection. *Applied Measurement in Education*, 14, 329-349. (2001)

Jones R. Identification of Measurement Differences Between English and Spanish Language Versions of the Mini-Mental State Examination: Detecting Differential Item Functioning Using MIMIC Modeling. *Medical Care*, 44, S124-S133. (2006)

Kim SH, Cohen AS, Alagoz C, Kim S. DIF Detection Effect Size Measures for Polytomously Scored Items. *Journal of Educational Measurement*, 44(2), 93-116. (2007)

McCullagh P, Nelder JA. Generalized Linear Models. *Chapman and Hall, London*. (1989)

Muraki E, Bock D. PARSCALE 4. *Scientific Software International, Inc., Lincolnwood, IL*. URL <http://www.ssicentral.com/> (2005)

Perkins AJ, Stump TE, Monahan PO, McHorney CA (2006). Assessment of Differential Item Functioning for Demographic Comparisons in the MOS SF-36 Health Survey. *Quality of Life Research*, 15(3), 331-348. (2006)

R Development Core Team. R: A Language and Environment for Statistical Computing. *R foundation for Statistical Computing, Viena, Austria*. URL <http://http://www.r-project.org/> (2010)

Rasch, G. Probabilistic models for some intelligence and attainment tests. *Copenhagen, Danish Institute for Educational Research*. (1960)

Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, 34 (17, Pt. 2). (1969)

Shealy RT, Stout WF. An Item Response Theory Model for Test Bias and Differential Item Functioning. *Differential Item Functioning*, 197-239, *Erlbaum, Hillsdale, NJ*. (1993)

StataCorp. Stata Statistical Software: release 10. *StataCorp LP, College Station, TX*. URL <http://www.stata.com> (2007)

Steinberg L, Thissen D. Using Effect Sizes for Research Reporting: Examples Using Item Response Theory to Analyze Differential Item Functioning. *Psychological Methods*, 11, n 4, 402-415. (2006)

Swaminathan H, Rogers HJ. Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27, 361-370. (1990)

Thissen D. IRTL RDIF: Software for the Computation of the Statistics involved in Item Response Theory Likelihood-Ratio Tests for Differential Item Functioning v.2.0.b. *L.L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill.* URL <http://www.unc.edu/~dthissen/dl.html> (2001)

Zumbo BD. A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. *Directorate of Human Resources Research and Evaluation, Department of national Defense, Ottawa, ON.* (1999)