XV MGEST
ENCONTRO MINEIRO
DE ESTATÍSTICA
BELO HORIZONTE

Teoria de Resposta ao Item: Modelos para Respostas não Dicotômicas com uso de Software Livre

Marcos Antonio da Cunha Santos
Departamento de Estatística UFMG
msantos@est.ufmg.br

Objetivos:

- Descrição introdutória dos principais modelos da TRI e das implementações disponíveis em software livre

- Apresentar algumas implementações no R e procedimentos para algumas análises, como unidimensionalidade e comportamento diferencial do item

MINICURSO

PARTE 1

1. Introdução
2. Modelos para respostas dicotômicas
3. Modelos para respostas não dicotômicas:
    - Modelo de Resposta Gradual
    - Modelo de Resposta Contínua (Samejima)
    - Modelos de desdobramento: modelo GGUM

PARTE 2

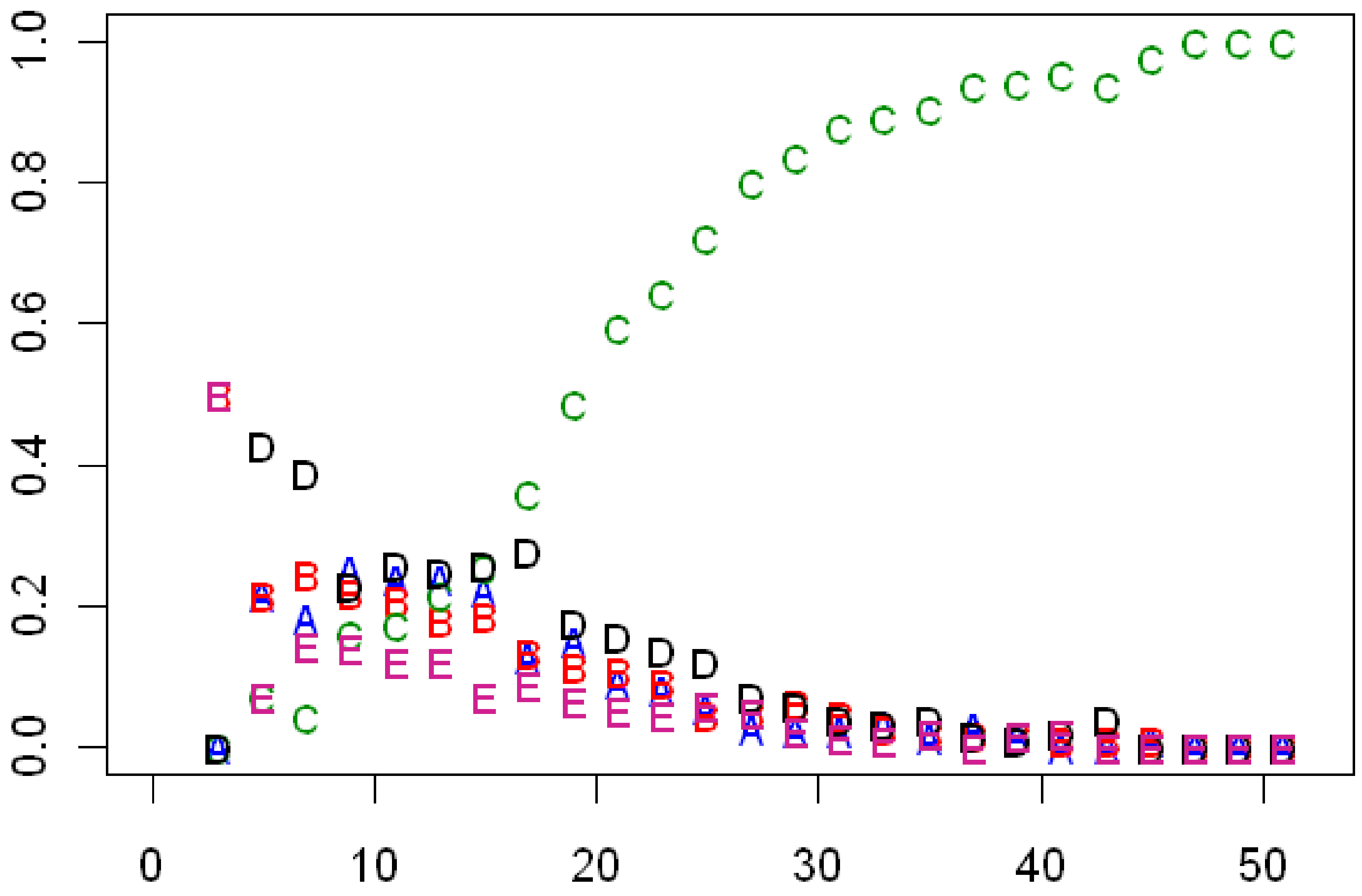
- Comportmento diferencial do item

Parte 1 – Alguns modelos da TRI e suas implementações

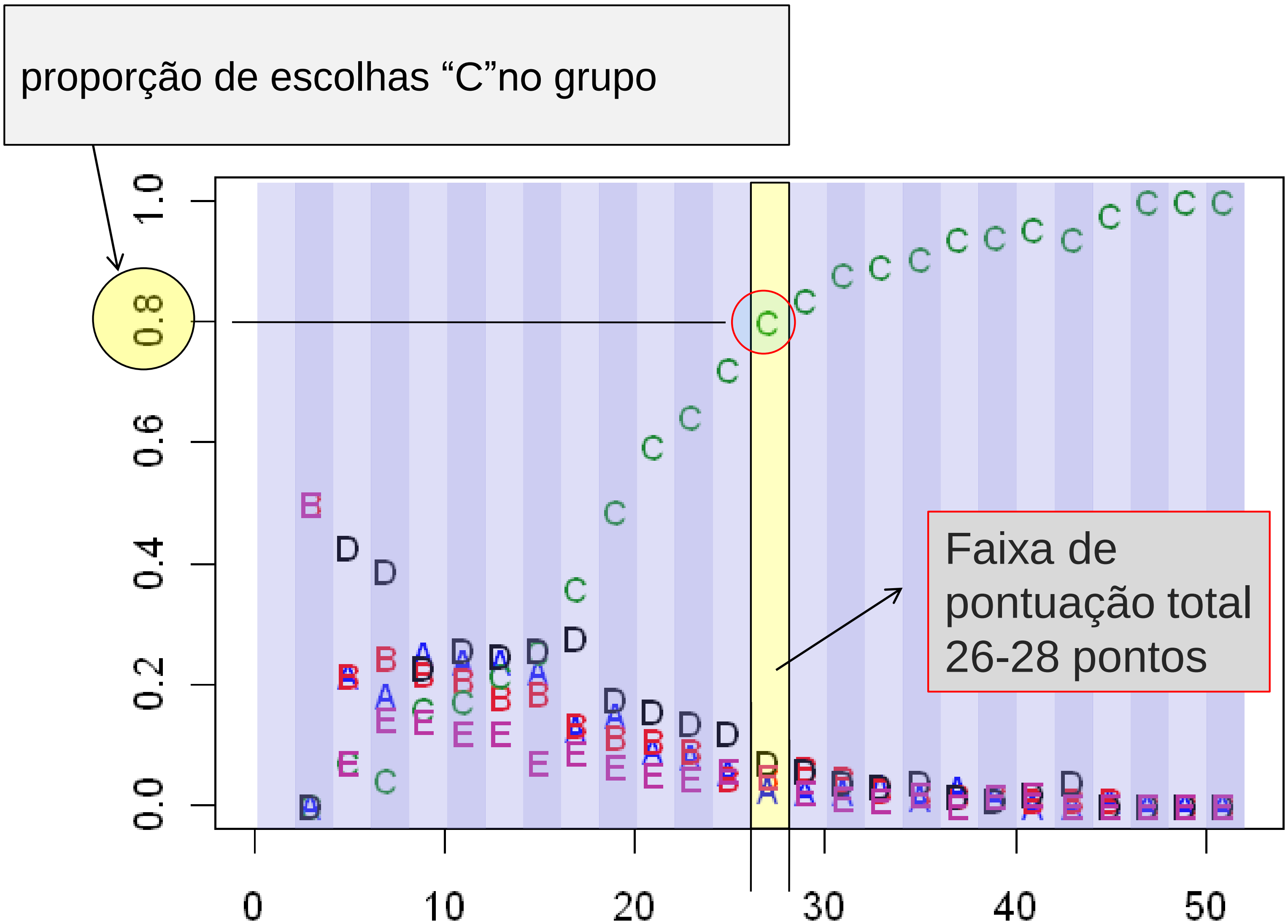
1. Introdução
2. Modelos para respostas dicotômicas
3. Modelos para respostas não dicotômicas:
    - Modelo de Resposta Gradual
    - Modelo de Resposta Contínua (Samejima)
    - Modelos de desdobramento: modelo GGUM

Introdução

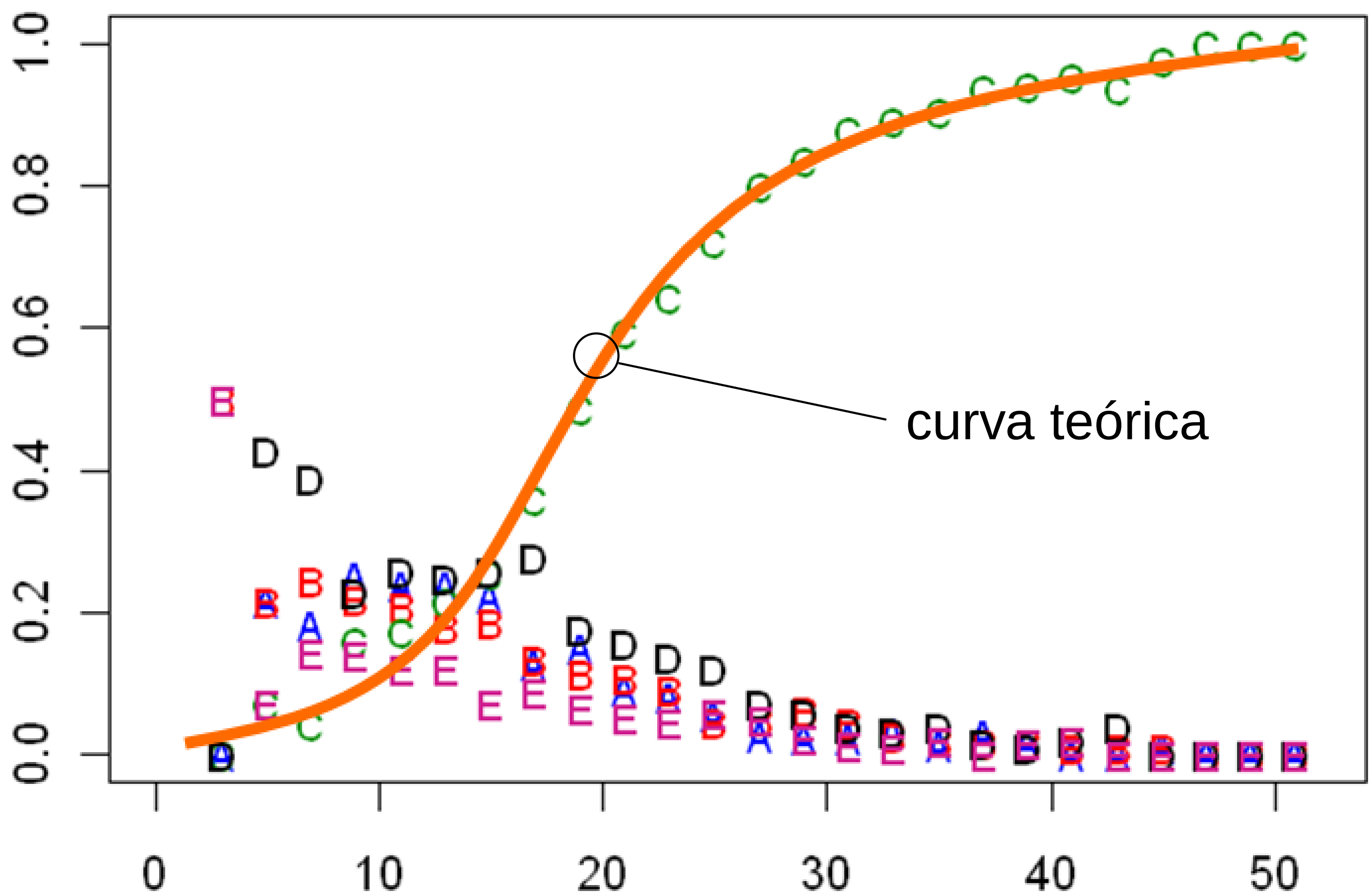
Exemplo: proporção de acertos
Item de uma prova com 54 questões (Antigo vestibular UFMG)



Introdução – Curva característica do item



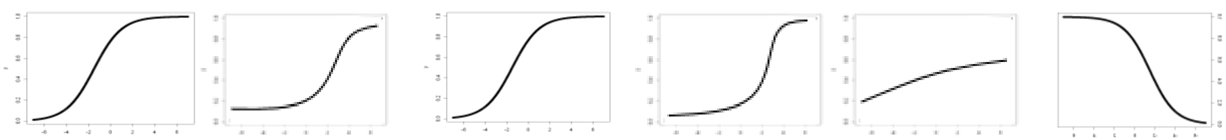
Exemplo: proporção de acertos



Em teste tipo certo/errado: a cada item corresponde uma curva característica, definida por um conjunto de parâmetros

Exemplos de curvas associadas aos itens:

Comportamento atípico



	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Indivíduo 1	1	1	1	1	1	**0**
Indivíduo 2	0	1	1	1	1	**0**
Indivíduo 3	0	0	1	1	1	**0**
Indivíduo 4	0	0	0	1	1	**0**
Indivíduo 5	0	0	0	0	1	**1**
...	...	...	...	...	...	**...**

Modelos para descrever o comportamento de um item:

- Desejável clareza na interpretação dos parâmetros do modelo adotado - desafio na modelagem matemática.

- Alguns modelos se tornaram bem sucedidos na descrição dos resultados e aplicação de testes ou questionários (modelos 1, 2 e 3 parâmetros).

Nesta breve introdução apresentaremos alguns modelos mais usuais da TRI e sugestões de como implementá-los.

Modelos para descrever o comportamento de um item:

- Nos modelos da TRI, as estimativas dos parâmetros dos itens e das habilidades não são triviais e exigem métodos computacionais. Métodos usuais: algoritmo EM, MCMC

O modelo 1P (modelo de Rasch)

Rasch propôs um modelo que atende às seguintes propriedades:

- a curva de resposta é da família das funções logísticas
- A *habilidade do indivíduo j* e o *nível de dificuldade do item i* são medidos em uma mesma escala. É possível portanto:

$$A_j = D_i$$

A probabilidade de acerto do item é

$$P(acerto\ do\ item\ i) = \frac{\text{habilidade do indivíduo}}{\text{habilidade do indivíduo} + \text{dificuldade do item}}$$

A seguinte transformação é introduzida na equação para simplificar o tratamento matemático:

$$A_j = e^{\alpha_j}$$
$$D_i = e^{\beta_i}$$

Deste modo a relação entre a probabilidade de acerto de um *item i* (escala de 0 a 1), para o *indivíduo j* com habilidade θ_j é

$$P(U_j=1|\theta_j) = \frac{e^{(\theta_j-\beta_i)}}{1+e^{(\theta_j-\beta_i)}}$$

$$\frac{\text{habilidade do indivíduo}}{\text{habilidade do indivíduo + dificuldade do item}}$$

Características do modelo:

- atribui a cada indivíduo uma medida de habilidade θ
- atribui a cada item um único parâmetro β (dificuldade)
- separa conceitualmente:

habilidade do indivíduo ⟺ dificuldade do item

Os parâmetros de dificuldade e habilidade são medidos em uma mesma escala padronizada. Na maioria das aplicações práticas, a escala N(0,1)

Outros modelos usuais para respostas dicotômicas acrecentam mais parâmetros ao modelo de Rasch:

- modelo 2P:
  - β, parâmetro de dificuldade do item
  - α, parâmetro de discriminação
- modelo 3P
  - β, parâmetro de dificuldade do item
  - α, parâmetro de discriminação
  - *c*, parâmetro de probabilidade de acerto ao acaso.

Modelos para respostas dicotômicas

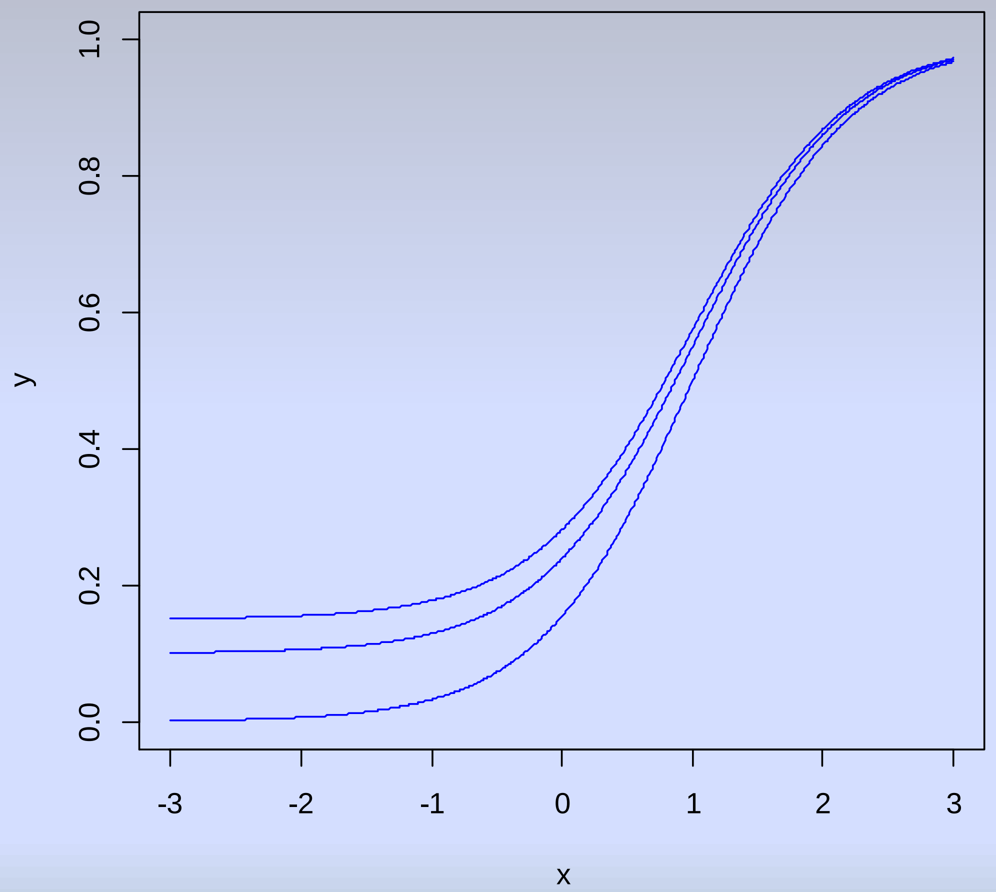
Modelo de três parâmetros (permite obter as expressões para os modelos 1P e 2P):

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

- *a:* parâmetro de discriminação: é o poder de discriminar os respondentes em relação à habilidade exigida pelo item
- *b:* parâmetro de dificuldade: quanto maior seu valor, mais difícil é a questão
- *c:* parâmetro de acerto ao acaso: probabilidade de um item ser respondido corretamente ao acaso

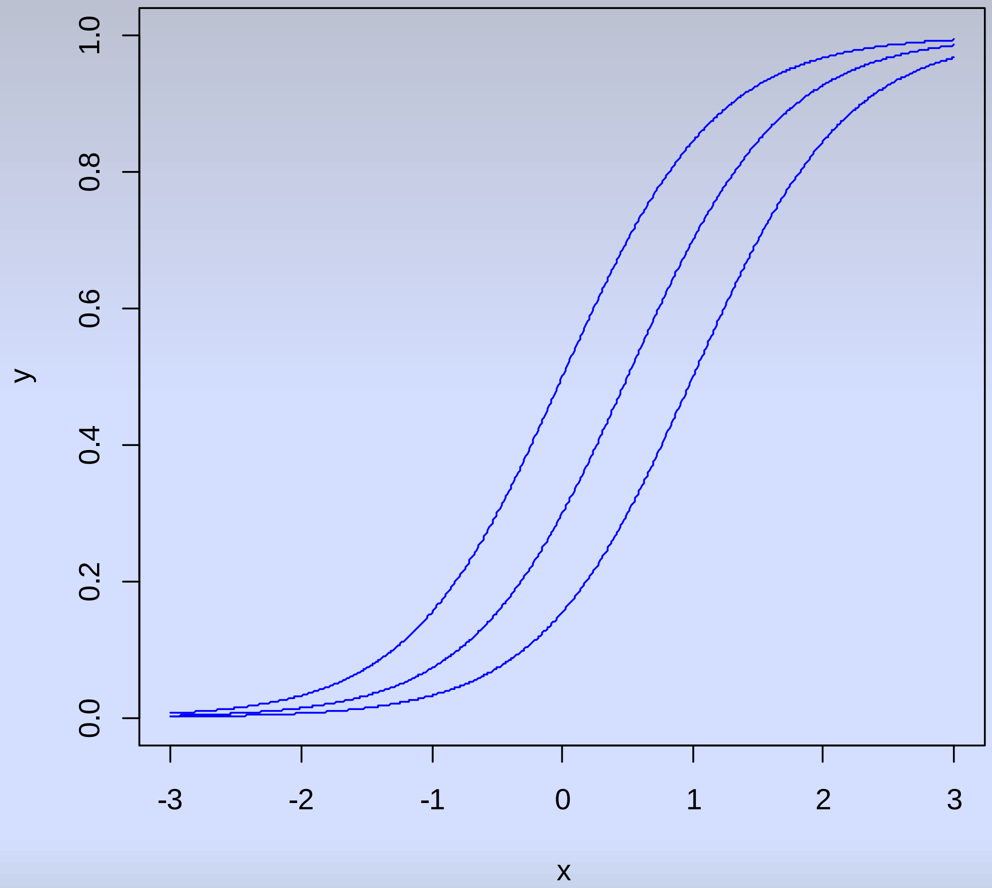
Modelos para respostas dicotômicas

Modelo 3P**:** (a=1, b=0, c=0) (a=1, b=0.5, c=0.1) (a=1, b=1, c=0.15)



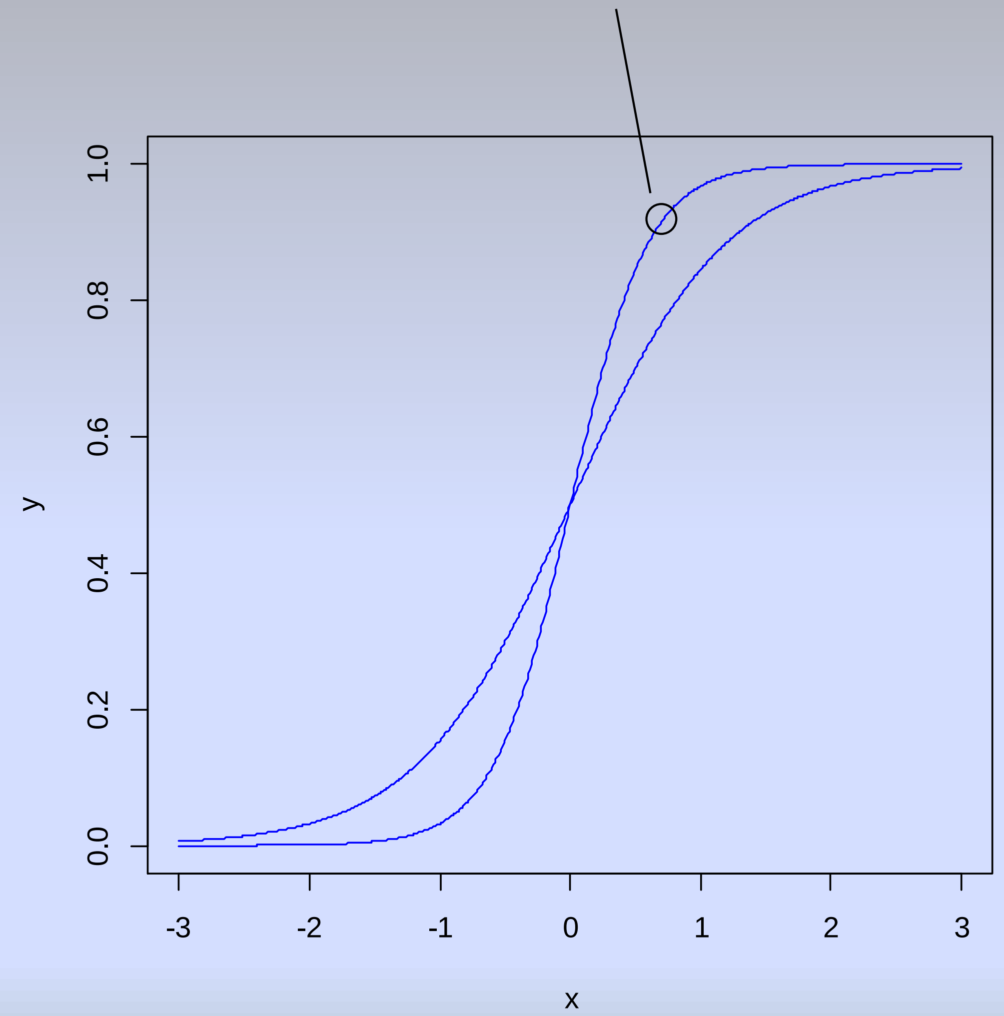
Modelos para respostas dicotômicas

Modelo 3P**:** (a=1, b=0, c=0) (a=1, b=0.5, c=0) (a=1, b=1.0, c=0)



Modelos para respostas dicotômicas

Modelo 3P **(**a=1, b=0, c=0) e (a=2, b=0, c=0)



Implementação dos modelos 1P, 2P e 3P e outros no R:

- pacote "MIRT"
- pacote "ltm"

Funções implementadas no "ltm":

rasch( )	modelo 1P
ltm( )	modelo 2P
tpm( )	modelo 3P
grm ( )	modelo de resposta gradual
outras	

Exemplo pacote ltm - Law School Admission Test (LAST)

1000 individuos, 5 questoes

```
library(ltm)      #carrega o pacote ltm
head(LSAT)        #primeiras linhas do banco
```

	Item 1	Item 2	Item 3	Item 4	Item 5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	1
5	0	0	0	0	1
6	0	0	0	0	1

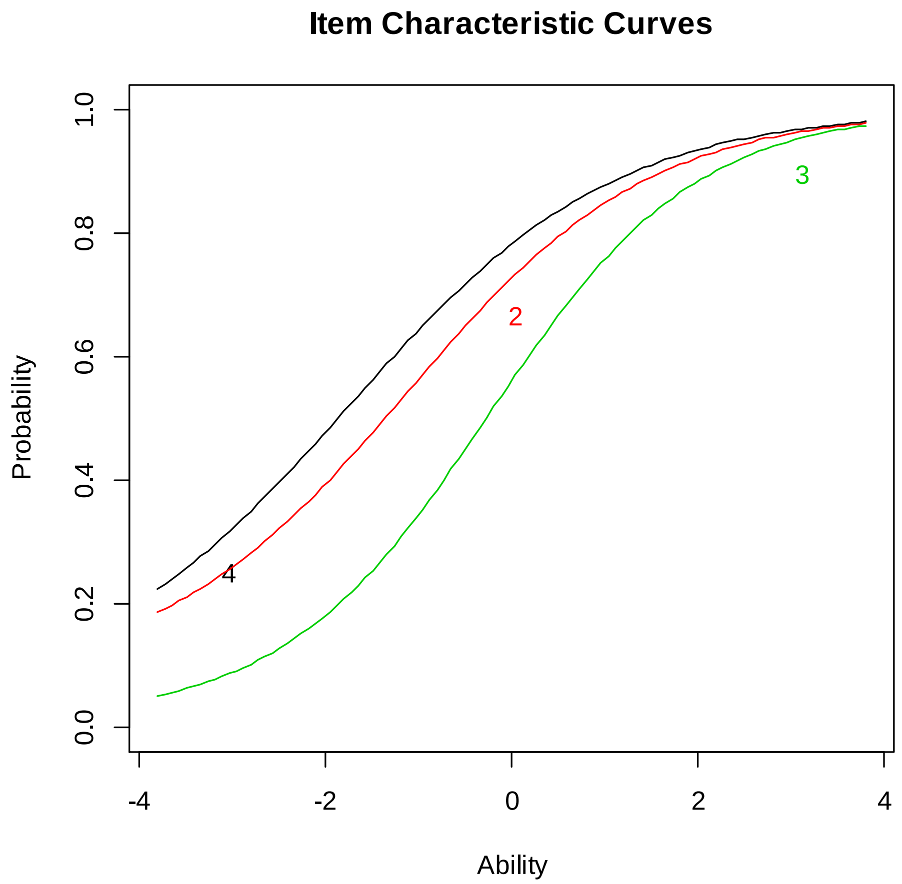
Pacote ltm – modelo Rasch

```
# modelo Rasch sem restrições na discriminação
modelo1=rasch(LSAT)

# valores ajustados dos parâmetros
coef(modelo1, prob=TRUE)
```

```
            Dffclt     Dscrmn     P(x=1|z=0)
Item 1   -3.6152665  0.7551347  0.9387746
Item 5   -2.7801716  0.7551347  0.8908453
Item 4   -1.7300903  0.7551347  0.7869187
Item 2   -1.3224208  0.7551347  0.7307844
Item 3   -0.3176306  0.7551347  0.5596777
```

Modelos para respostas dicotômicas

```
# Curva característica - ajuste ao modelo 3P
modelo3=tpm(LSAT)
plot(modelo3, items=c(4,2,3))
```



Curva de Informação

- Derivando a função log-verossimilhança em relação ao parâmetro θ temos a função escore:

$$S(\theta) \equiv \frac{\partial}{\partial \theta} \log L(\theta)$$

- O EMV é a solução da equação $S(\theta) = 0$

- A derivada segunda da função log-verossimilhança no ponto que a função atinge o valor máximo define a Informação de Fisher.

Modelos para respostas dicotômicas

- A informação esperada de Fisher é

$$I(\theta) = -E_\theta\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right] = -\int\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right] f(x;\theta)\,dx$$

- Sob condições apropriadas de regularidade,

$$\sigma^2(\theta) = \frac{1}{I(\theta)}$$

Modelos para respostas dicotômicas

- O termo informação foi introduzido em estatística por Fisher, que a definiu como o inverso da precisão com que se estima o parâmetro

- A quantidade de informação é calculada para cada valor de θ, sendo uma função definida de $-\infty$ a $+\infty$

- A função de informação não depende da distribuição dos examinandos ao longo da distribuição da escala de habilidades

- Quanto maior o valor da informação $\rightarrow$ menos incerteza sobre os parâmetros

Modelos para respostas dicotômicas

Como exemplo, no modelo 3P:

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

Neste modelo, a informação é maior quando:

- b_i é próximo de θ,
- maior o parâmetro a_i
- menor o valor de c_i

- A utilidade da função de informação do item na avaliação e desenvolvimento de testes depende do ajuste da CCI

- um item pode apresentar um valor alto da função de informação em uma região da escala de habilidades, mas o item pode não ter valor se o interesse é medir o construto em outra posição da escala.

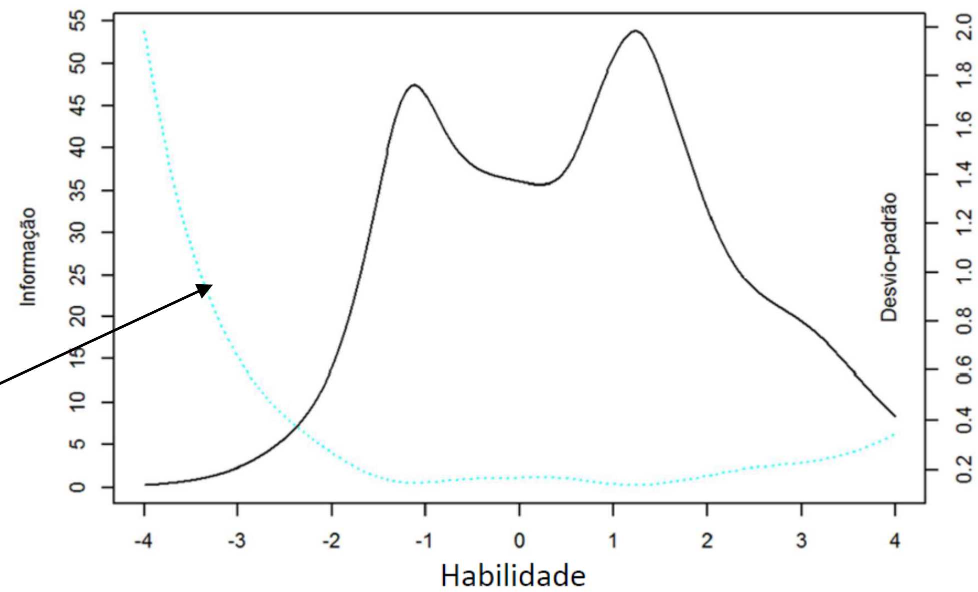
- Função de informação do teste: $I(\theta) = \sum_{i=1}^{N} I_i(\theta)$

Exemplo real de função de informação do teste $I(\theta) = \sum_{i=1}^{N} I_i(\theta)$

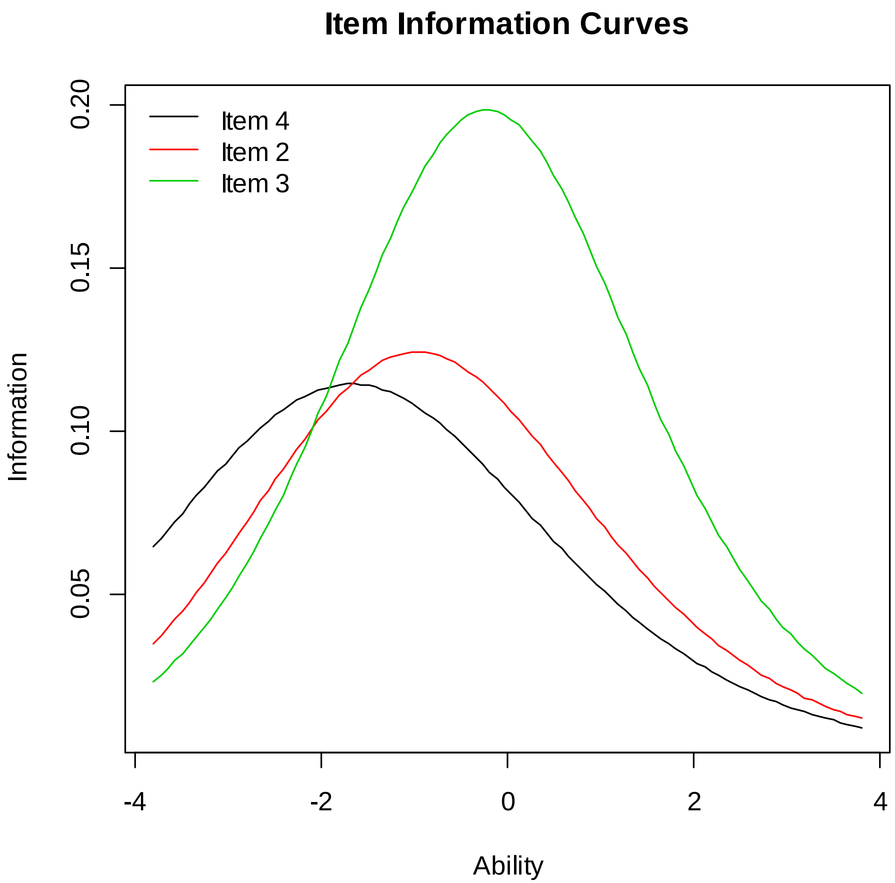
$$\text{Var}(\hat{\theta}) = \frac{1}{I(\hat{\theta})}$$

Teste de matemática do 5º e 9º ano do EF (166 itens calibrados)

$\text{Var}(\hat{\theta})$



```
# curvas de informação no ltm
plot(modelo3, type="IIC", items=c(4,2,3),legend =TRUE)
```



Modelos para respostas dicotômicas

```
# habilidades estimadas por padrão de
# respostas
factor.scores(modelo3)
```

```
Call:
tpm(data = LSAT)

Scoring Method: Empirical Bayes

Factor-Scores for observed response patterns:
   Item 1 Item 2 Item 3 Item 4 Item 5 Obs     Exp     z1 se.z1
1       0      0      0      0      0   3   2.230 -1.871 0.790
2       0      0      0      0      1   6   5.820 -1.463 0.792
3       0      0      0      1      0   2   2.583 -1.446 0.793
4       0      0      0      1      1  11   8.944 -1.032 0.796
5       0      0      1      0      0   1   0.701 -1.325 0.796
6       0      0      1      0      1   1   2.628 -0.908 0.800
7       0      0      1      1      0   3   1.185 -0.889 0.801
8       0      0      1      1      1   4   5.969 -0.463 0.809
```

Modelo de Resposta Gradual

- proposto por Samejima, 1969

- No Modelo de Resposta Gradual, itens são formados por um conjunto de respostas *ordenadas*.

ordenação

Exemplo (uma possibilidade)

a) Concordo fortemente
b) Concordo
c) Discordo
d) Discordo fortemente

Exemplo: item com 4 categorias de respostas

Respostas:

a) Concordo fortemente	k=3	categorias
b) Concordo	k=2	
c) Discordo	k=1	
d) Discordo fortemente	k=0	

Exemplo: item com 4 categorias de respostas

Respostas:

a) Concordo fortemente	k=3
b) Concordo	k=2
c) Discordo	k=1
d) Discordo fortemente	k=0

conjunto escolha 1 = *conjunto escolha 1 ou superior* - *conjunto escolha 2 ou superior*

P(escolha 1) = P(escolha 1 ou superior) - P(escolha 2 ou superior)

P(escolha k ou superior) é dada por

$$P(U_{ij} \geq k|\theta) = \frac{1}{1 + exp[-a_i(\theta_j - b_{i,k})]}$$

expressão proposta pelo modelo

De acordo com o modelo, probabilidade do j-ésimo respondente escolher a *categoria k* do *item i* é

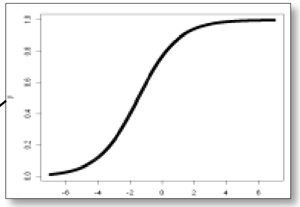
$$P(U_{ij} = k|\theta) = \frac{1}{1 + exp[-a_i(\theta_j - b_{i,k})]} - \frac{1}{1 + exp[-a_i(\theta_j - b_{i,k+1})]}$$

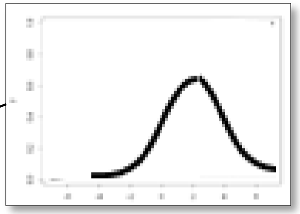
Modelo de Resposta Gradual

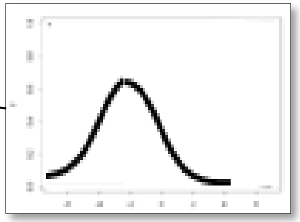
categoria de resposta no item ⟺ curva característica

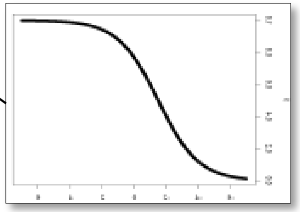
Respostas:

a) Concordo fortemente k=3
b) Concordo k=2
c) Discordo k=1
d) Discordo fortemente k=0









Modelo de Resposta Gradual

Os parâmetros indicam a forma e a localização das posições centrais das curvas de resposta das categorias:

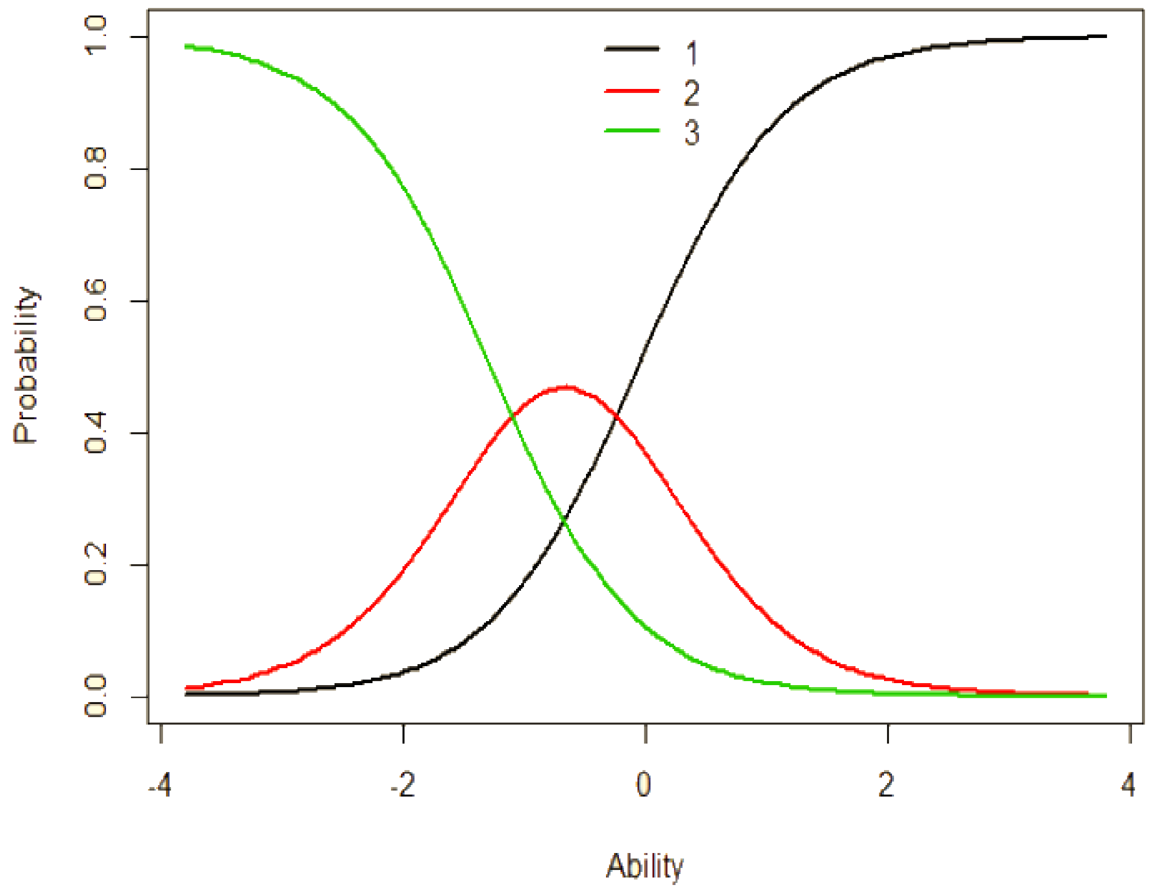
- O parâmetro a_i no MRG não pode ser interpretado como discriminação

- O parâmetro b_{ij} - limite das categorias - representa a habilidade necessária para um indivíduo responder acima do limiar j com 50% de probabilidade.

- Nos modelos de resposta graduadas, os b_{ij} de um mesmo item são ordenados

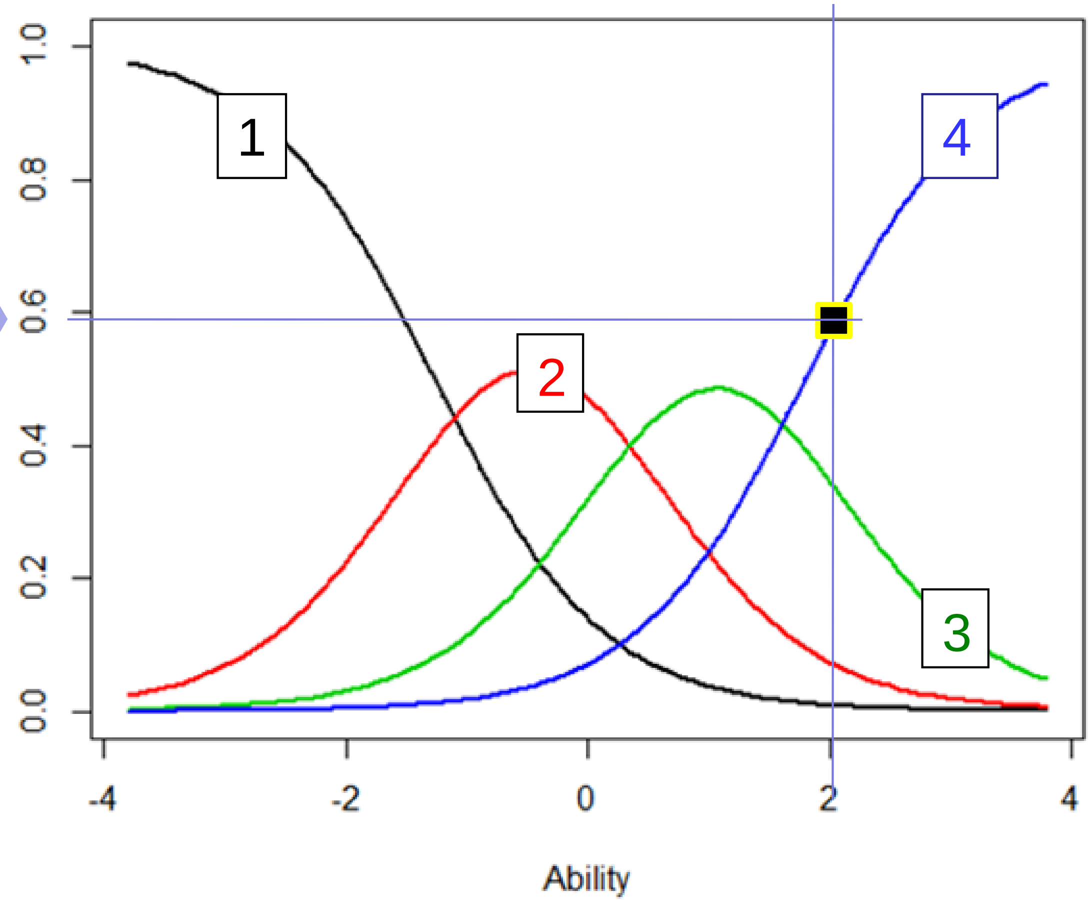
- b_{ij} é dito parâmetro de locação

Curvas características - item com 3 categorias de respostas

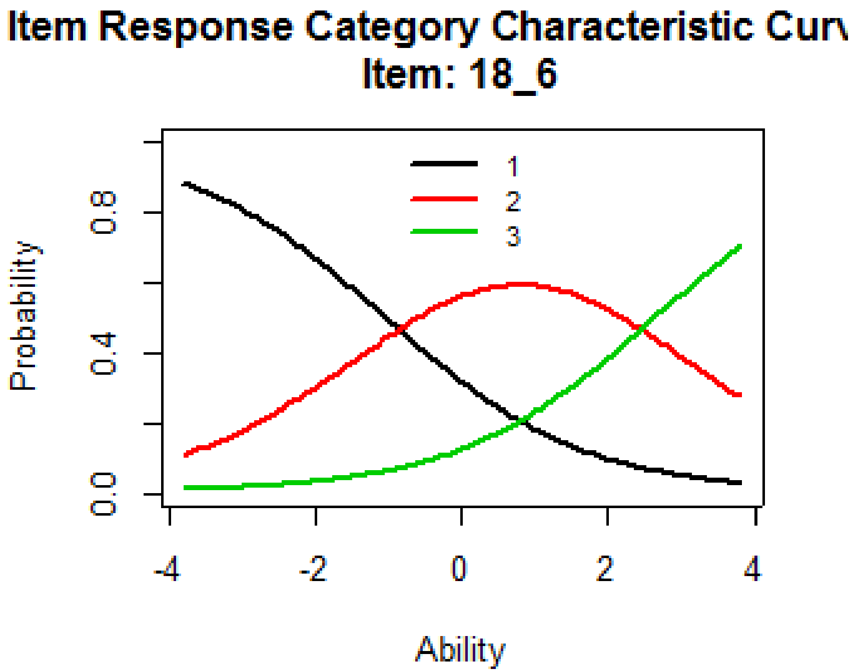


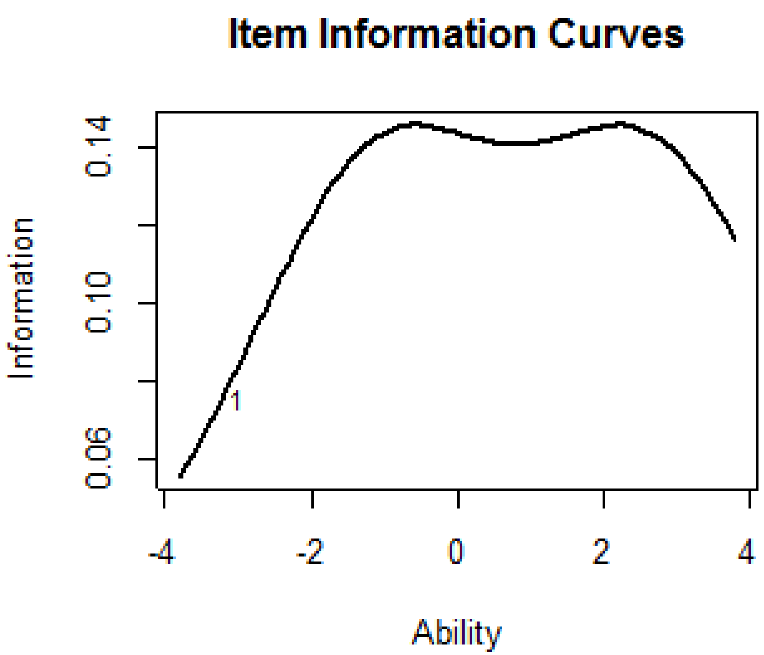
Curvas características - item com 4 categorias de respostas

probabilidade de escolha da categoria 4 para um indivíduo com nível 2



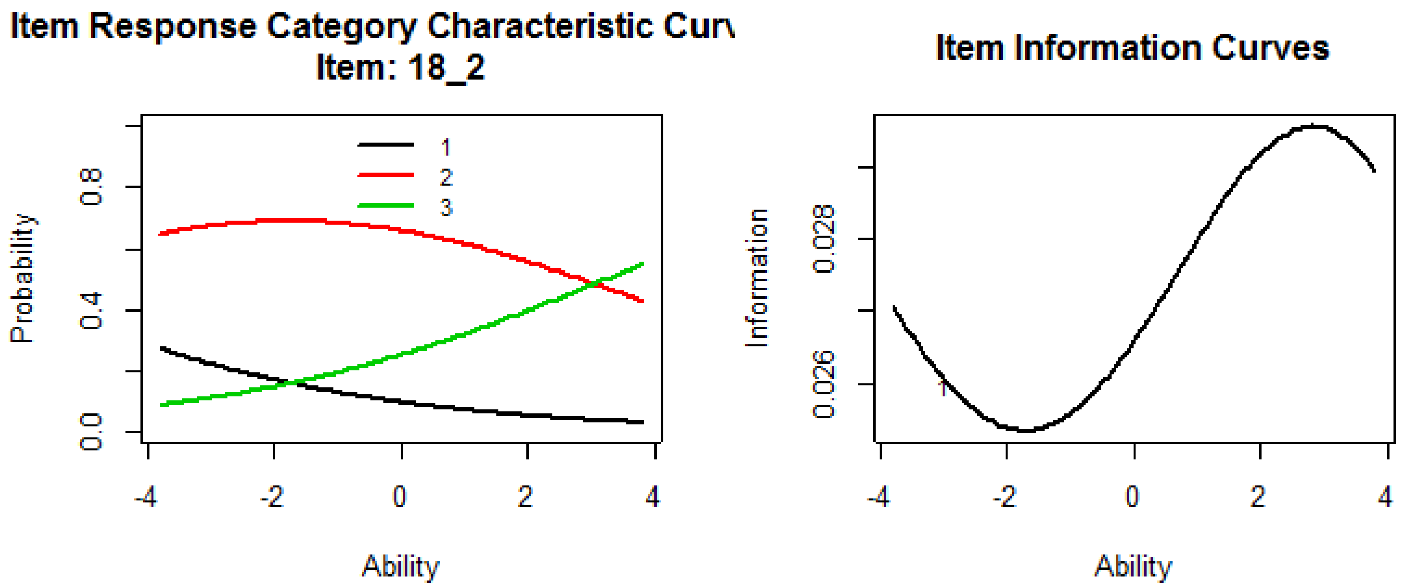
Curva Característica do Item e Curva de Informação do Item





Análise da curva característica do item

Exemplo de item a ser recodificado, revisado ou retirado (pouca discriminação)



Implementação para o Modelo Resposta Gradual: pacotes “ltm” ou “MIRT”

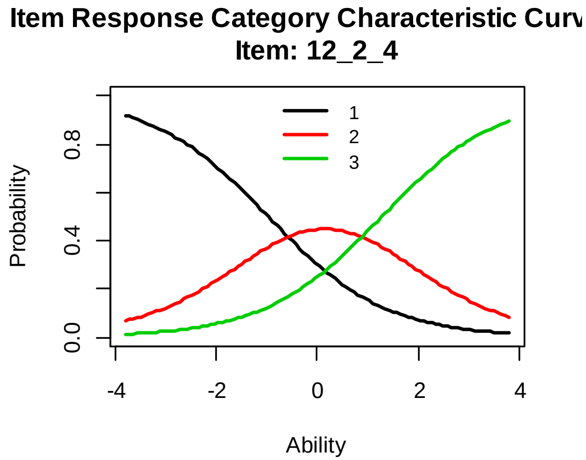
No ltm:

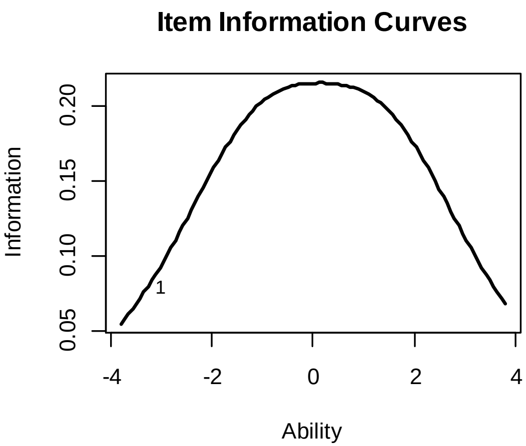
```
# ajuste com o modelo de Samejima
fit=grm(matriz.resp)

# curva característica para item 3
plot(fit, item=3, lwd = 2, legend = TRUE)
```

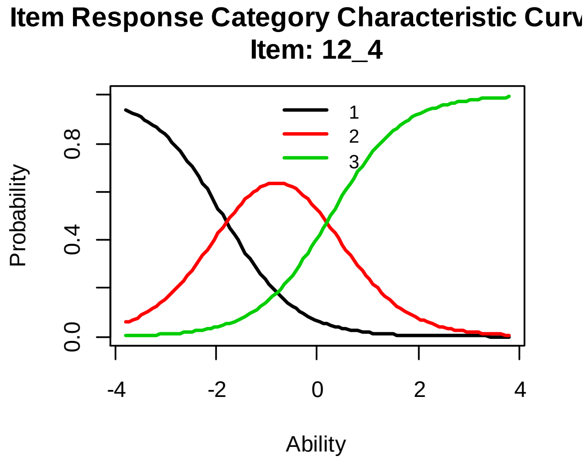
pacote ltm

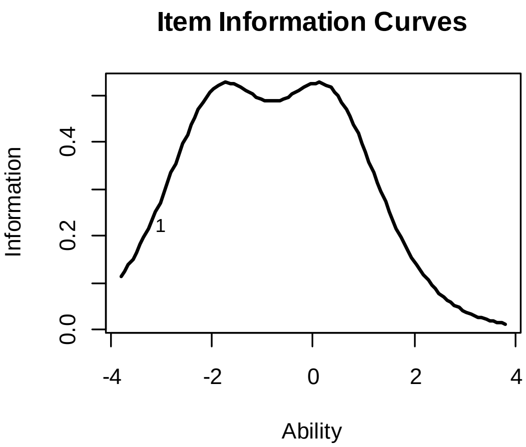
```
# curva de informação do item 2
plot(fit, type = "IIC", items = 2, lwd = 2)

# curva de informação do item 3 e 4
plot(fit, type = "IIC", items = c(3,4), lwd = 2)

# curva de informação do teste
plot(fit, type = "IIC", items = 0, lwd = 2)

# escores
factor.scores(fit, resp.patterns = ...)
```

Curva Característica do Item e curva de informação do item









Verificação de unidimensionalidade

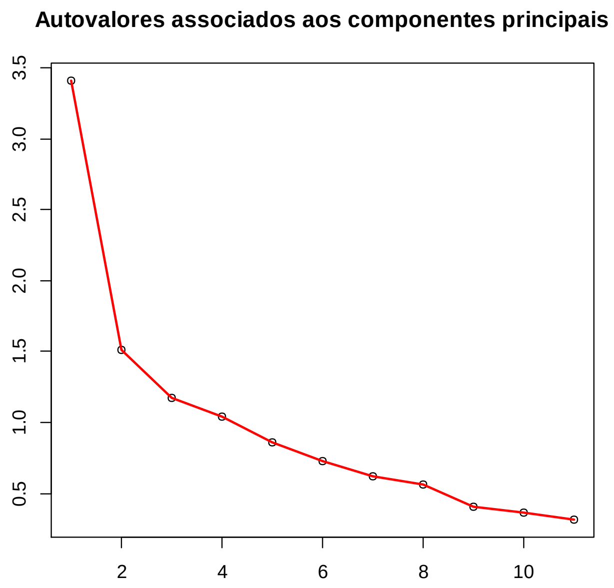
```
# matriz policórica (pacote minicurso)
mpolicorica=poli(matriz.resp)
Mpolicorica
```

matriz correlação policórica

	qn35	qn55	qn56	qn57	qn58	qn59	qn60	qn61	qn62	qn63	qn64
qn35	1.000	0.042	0.012	0.132	0.007	0.029	0.025	0.041	0.097	0.201	-0.020
qn55	0.042	1.000	0.156	0.171	0.193	0.610	0.216	0.180	0.173	0.083	0.210
qn56	0.012	0.156	1.000	0.400	0.340	0.125	0.296	0.425	0.238	0.522	0.297
qn57	0.132	0.171	0.400	1.000	0.398	0.119	0.156	0.208	0.145	0.270	0.258
qn58	0.007	0.193	0.340	0.398	1.000	0.327	0.229	0.197	0.148	0.170	0.175
qn59	0.029	0.610	0.125	0.119	0.327	1.000	0.343	0.181	0.113	0.080	0.162
qn60	0.025	0.216	0.296	0.156	0.229	0.343	1.000	0.373	0.220	0.208	0.191
qn61	0.041	0.180	0.425	0.208	0.197	0.181	0.373	1.000	0.566	0.447	0.362
qn62	0.097	0.173	0.238	0.145	0.148	0.113	0.220	0.566	1.000	0.326	0.416
qn63	0.201	0.083	0.522	0.270	0.170	0.080	0.208	0.447	0.326	1.000	0.353
qn64	-0.020	0.210	0.297	0.258	0.175	0.162	0.191	0.362	0.416	0.353	1.000

Verificação de unidimensionalidade

```
A=plot.autov(matriz.resp)    #pacote minicurso
```



Modelo Multidimensional - R package “MIRT”

O modelo muldimensional:

$$\Phi(x_{ij}=1|\boldsymbol{\theta}_i,\boldsymbol{\alpha}_j,d_j,\gamma_j)=\gamma_j+\frac{(1-\gamma_j)}{1+\exp\left[-D(\boldsymbol{\alpha}_j^{\top}\boldsymbol{\theta}_i+d_j)\right]}$$

- $j=1,\ldots,n$ itens
- *m* variáveis latentes $\boldsymbol{\theta}_i=(\theta_{i1},\ldots,\theta_{im})$
- *m* níveis $\boldsymbol{\alpha}_j=(\alpha_1,\ldots,\alpha_m)$ associados às variáveis latentes

Modelo Multidimensional

```
library(mirt)

#Desagrupar os dados
dados<-expand.table(LSAT7)# Ex. com 1000 respondentes e 5 itens

#Ajuste modelos de 1,2 e 3 parâmetros

fit.1PL<- mirt(dados, 1, itemtype ="1PL")

fit.2PL<- mirt(dados, 1, itemtype ="2PL")

fit.3PL<- mirt(dados, 1, itemtype "3PL", guess=.2)

#Parâmetros dos modelos
coef(fit.1PL)
coef(fit.2PL)
coef(fit.3PL)

#a=discriminação, d=dificuldade, g=acerto ao acaso
```

Modelo Multidimensional

```
library(mirt)

#Desagrupar os dados
dados<-expand.table(LSAT7)# Ex. com 1000 respondentes e 5 itens

#Ajuste modelos de 1,2 e 3 parâmetros

fit.1PL<- mirt(dados, 1, itemtype ="1PL")

fit.2PL<- mirt(dados, 1, itemtype ="2PL")

fit.3PL<- mirt(dados, 1, itemtype "3PL", guess=.2)

#Parâmetros dos modelos
coef(fit.1PL)
coef(fit.2PL)
coef(fit.3PL)

#a=discriminação, d=dificuldade, g=acerto ao acaso
```

número de dimensões

Modelo Multidimensional

```
residuals(fit.1PL)

#Curva de informação total do teste
plot(fit.1PL)
plot(fit.2PL)
plot(fit.3PL)

#Curva de informação do item
itemplot(fit.3PL,1,type="info")

#Comparação de modelos etc
```

Modelo Multidimensional

Material adicional – Modelo de Resposta Gradual

Roteiro com comandos

http://www.est.ufmg.br/~msantos/script_minicurso.txt

dados de exemplo (banco de treino):

http://www.est.ufmg.br/~msantos/fcomplex_exemplo.txt

Modelo TRI para resposta contínua

$$P(X_{ij} \geq x|\theta_i, a_j, b_j, \alpha_j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} e^{\frac{-t^2}{2}} dt$$

$$v = a_j(\theta_i - b_j - \frac{1}{\alpha_j} ln \frac{x_{ij}}{k_j - x_{ij}})$$

a_j é o parâmetro de discriminação
b_j é o parâmetro de dificuldade
α_j parâmetro de escala
k = valor máximo da escala para o item

Modelo de Resposta Contínua (Samejima, 1973).

- O modelo é uma generalização do modelo de resposta gradual
- A cada item são atribuídos três parâmetros: a, b e α.
- Os parâmetros com interpretação semelhante ao do modelo de resposta gradual
- o parâmetro α é um parâmetro geométrico, sem uma interpretação prática.

Modelo para resposta contínua

Modelo de Resposta Contínua (Samejima, 1973).

- Este modelo aplicável quando a resposta a um item pode ser situada em uma escala contínua, entre um valor mínimo e um valor máximo.
- Diferentes itens podem ter diferentes valores mín/máx
- O modelo foi pouco utilizado, após a sua proposição, devido a dificuldades de implementação computacional. Uma implementação foi viabilizada após a contribuição de Shojima (2005).
- Implementação: pacote R "EstCRM"

Modelo para resposta contínua

Modelo de resposta contínua como generalização do modelo de resposta gradual

Resposta Gradual		Resposta Contínua
a) Concordo fortemente	$k=3$	K máximo
b) Concordo	$k=2$	
c) Discordo	$k=1$	
d) Discordo fortemente	$k=0$	K mínimo

Funções implementadas no pacote "EstCRM"

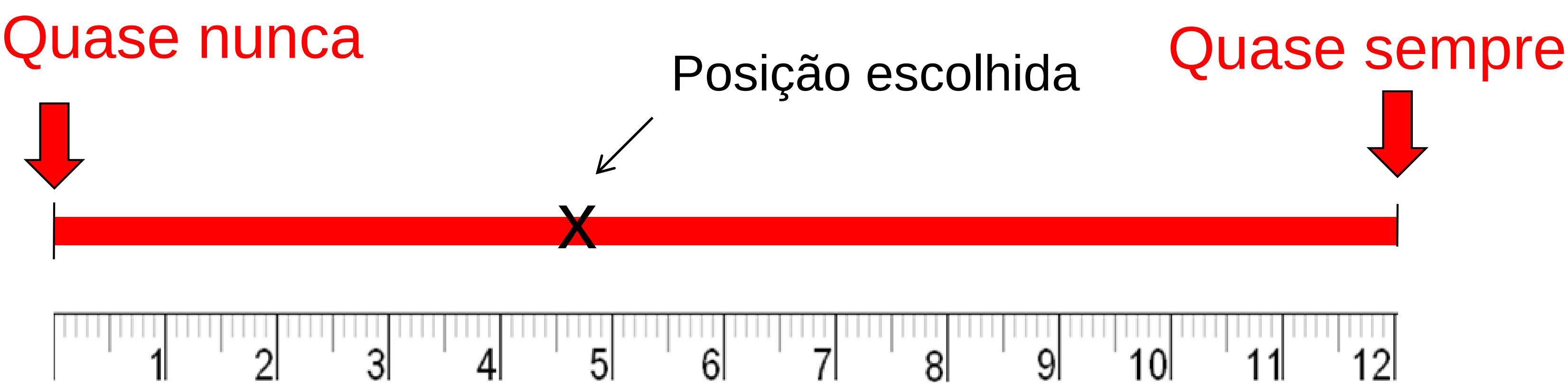
Objetivo	Função
Estimação dos parâmetros dos itens	EstCRMitem(…)
Estimação das habilidades individuais	EstCRMperson (…)
Gráfico 3D das curvas teóricas de resposta	plotCRM(…)
Dados simulados	simCRM(…)
Outras	…

Modelo para resposta contínua

Dados de exemplo que acompanha o pacote: planilha “EPIA”

- 1033 estudantes marcaram em um segmento de reta (112 mm) uma posição entre dois pontos de referência, segundo seu julgamento para cinco itens
- Os itens avaliam uma escala de impulsividade.
- O escore é a distância em milímetros do ponto inicial da escala

Item 5: Tende a fazer muitas coisas ao mesmo tempo



item 1 Longs for excitement
Item 2 Does not stop and think things over before doing anything
Item 3 Often shouts back when shouted at
Item 4 Likes doing things in which he/she has to act quickly
Item 5 Tends to do many things at the same time

Modelo para resposta contínua

Visualização do banco de dados “*EPIA*”

```
library(EstCRM)    #carrega o pacote EstCRM
fix(EPIA)          #abre janela para edição de dados
```

dados do exemplo

Editor de dados

	Item 1	Item 2	Item 3	Item 4	Item 5
1	96	36	80	78	79
2	42	2	1	1	1
3	64	23	36	13	38
4	100	51	106	57	100
5	10	5	65	4	5
6	59	24	94	98	8
7	97	58	103	59	14
8	7	9	106	79	8
9	110	20	110	1	110

Calibração

Os vetores “max.item” e “min.item” definem uma mesma escala de 0 a 112 para todos os itens

```
max.item = c(112,112,112,112,112)
min.item = c(0,0,0,0,0)
```

Estimando os parâmetros dos itens

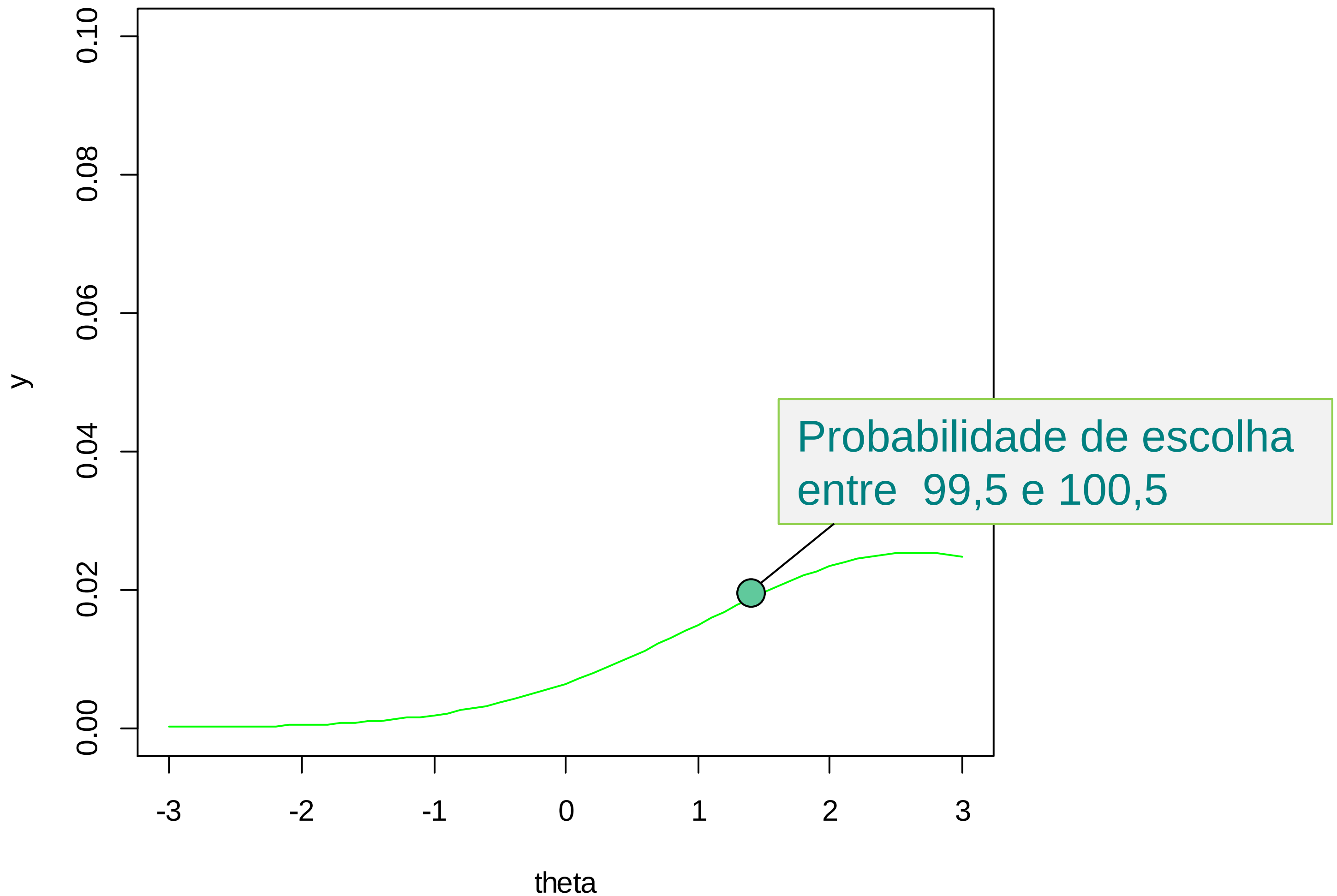
```
CRM = EstCRMitem(EPIA, max.item, min.item, max.EMCycle =
500, converge = 0.01)
par <- CRM$param
```

Visualizando os parâmetros

CRM

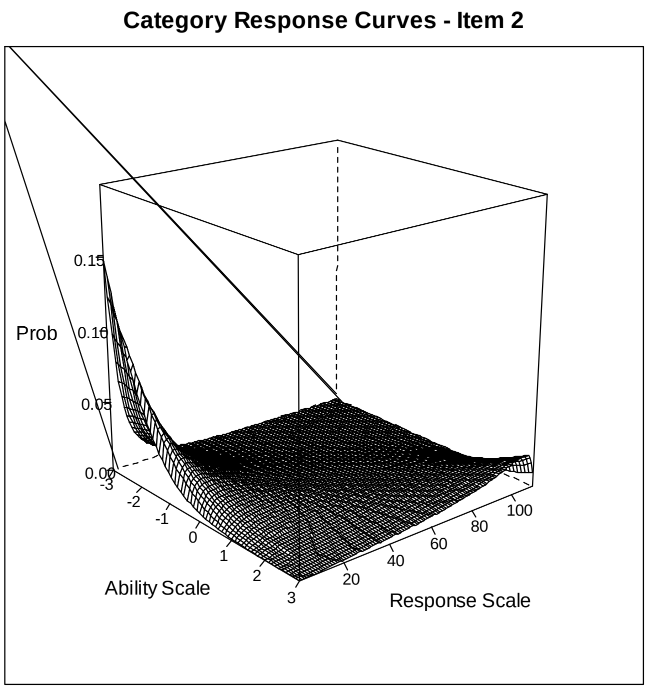
```
Largest Parameter change at the last EM Cycle is 1.386805e-05

Final Paramater Estimates:

               a      b    alpha
      Item 1  0.394 -0.623  0.591
      Item 2  0.549  1.299  0.731
      Item 3  0.324  0.235  0.645
      Item 4  0.672  0.394  0.943
      Item 5  0.626  0.343  0.921
```

Modelo para resposta contínua

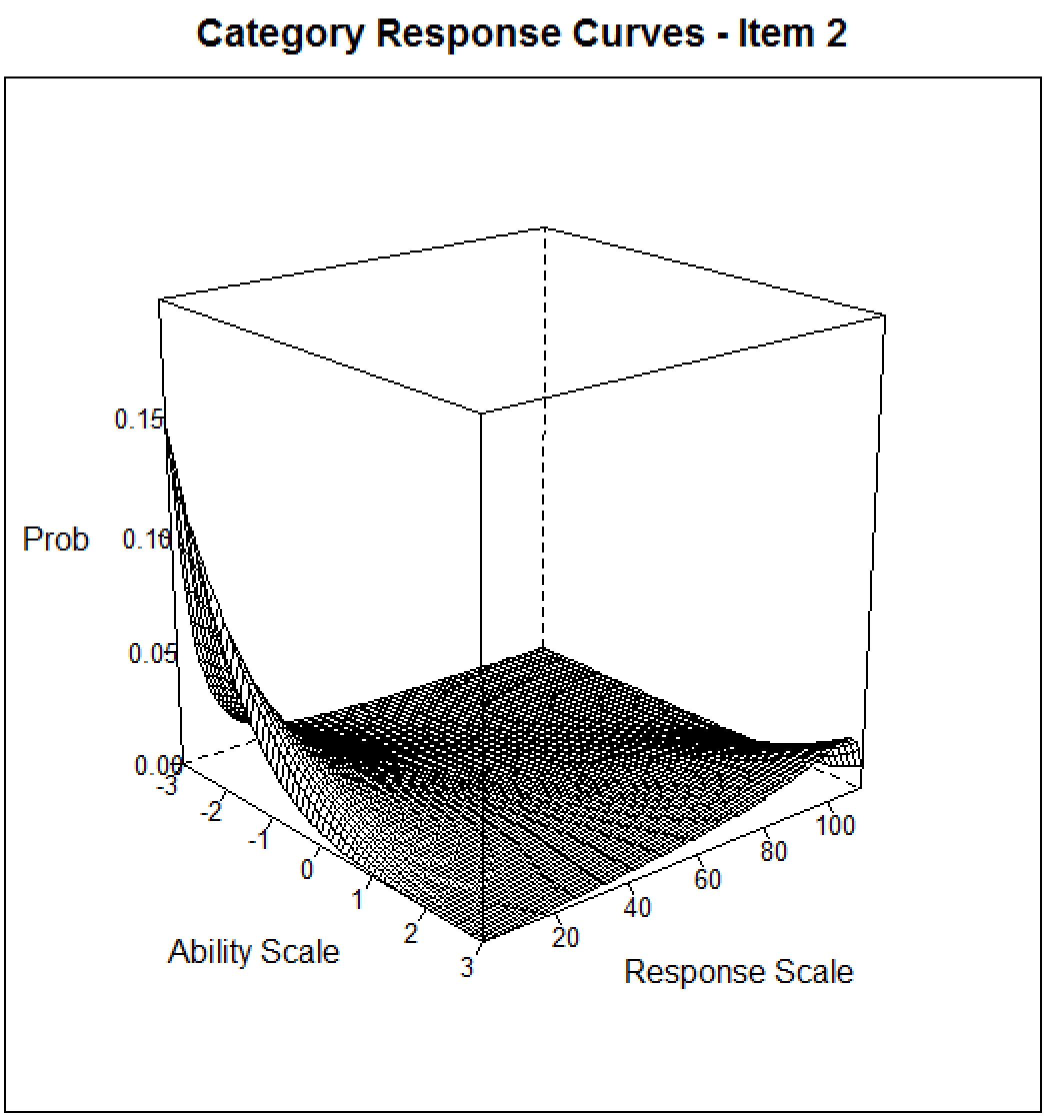


Obtendo as curvas para o item 2

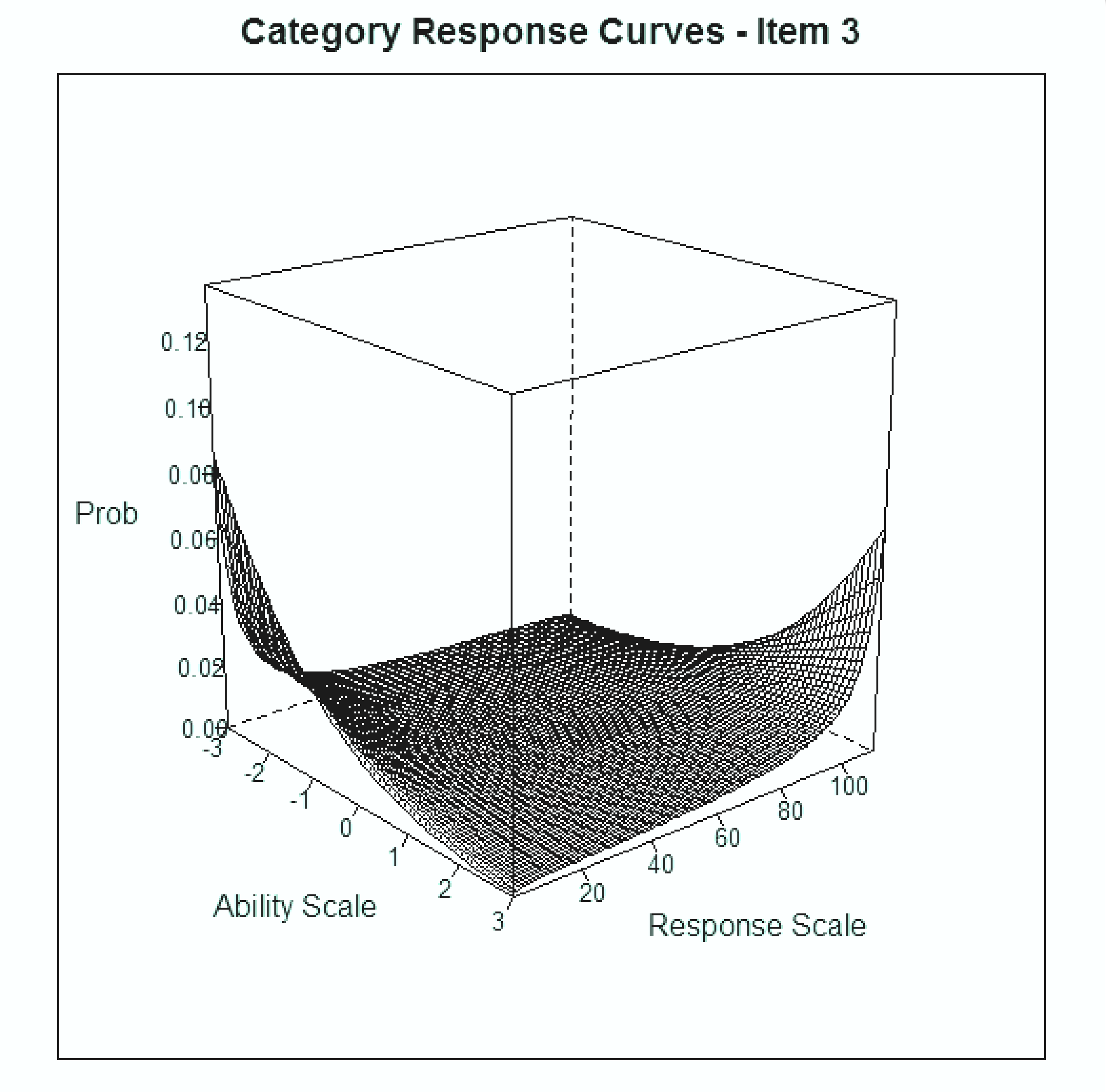
```
plotCRM(par,item=2,min.item, max.item)
```



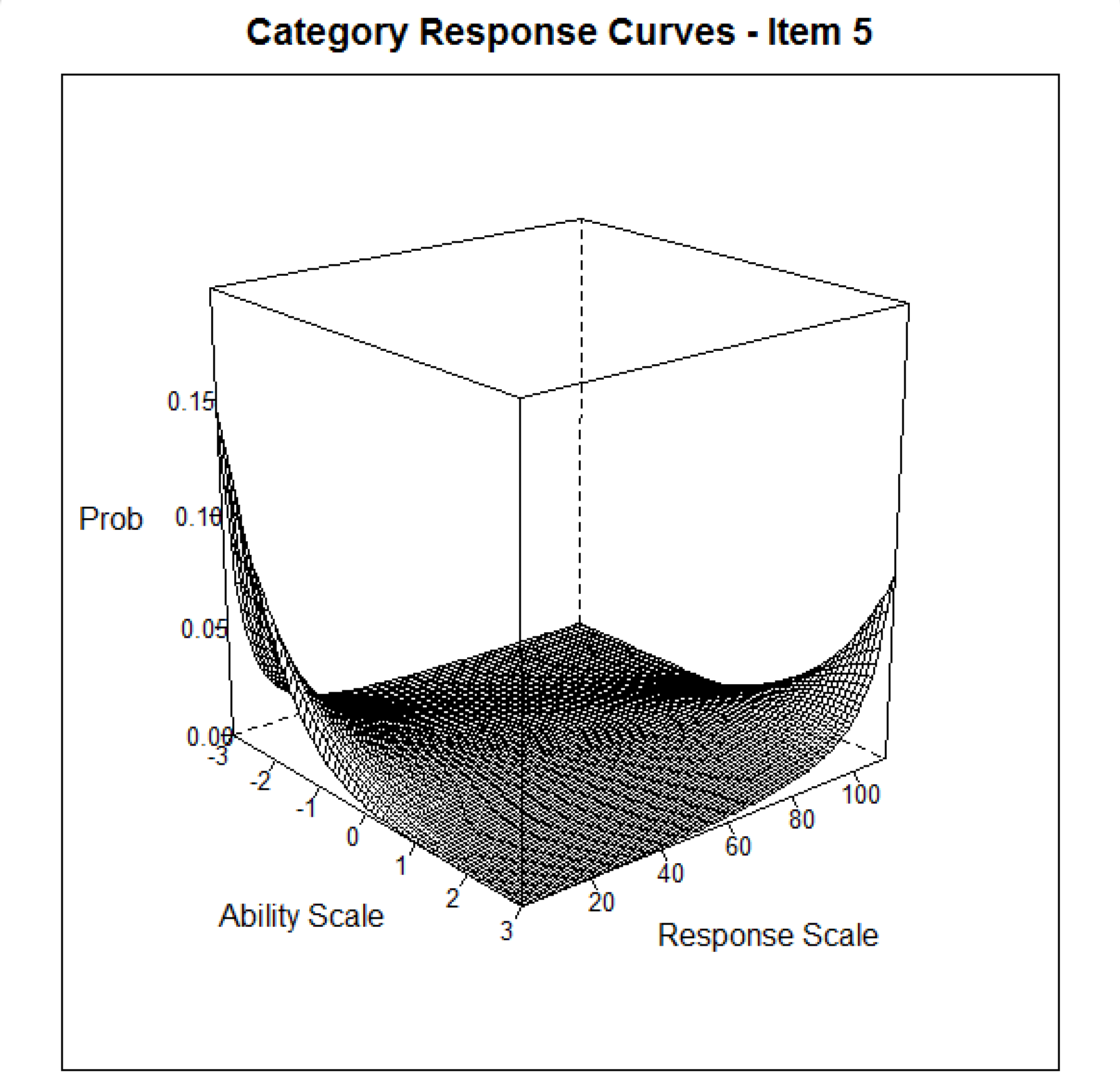
Modelo para resposta contínua



Modelo para resposta contínua



Modelo para resposta contínua



valores individuais de θ

```
CRMthetas = EstCRMperson(EPIA,par,min.item,max.item)
CRMthetas
```

```
$thetas
         ID    Theta Est.        SE
   [1,]   1  0.9312837781 0.8440008
   [2,]   2 -3.3614833094 0.8440008
   [3,]   3 -0.6678324276 0.8440008
   [4,]   4  1.3603595062 0.8440008
   [5,]   5 -2.2138970848 0.8440008
   [6,]   6  0.1266999241 0.8440008
   [7,]   7  0.4191604986 0.8440008
   [8,]   8 -0.7021624359 0.8440008
   [9,]   9  0.6289758194 0.8440008
  [10,]  10  4.2980966234 0.8440008
  [11,]  11  0.2372013014 0.8440008
  [12,]  ....
```

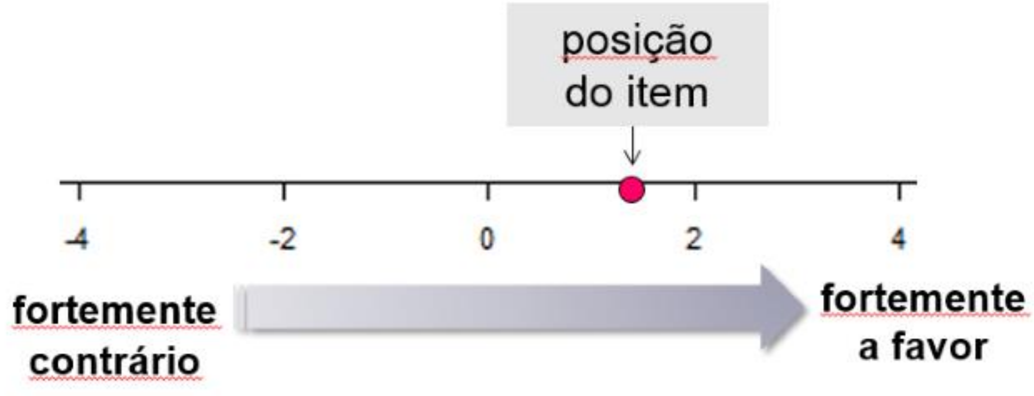
Generalized Graded Unfolding Model (GGUM) (Modelo de Desdobramento Graduado Generalizado)

- Modelo da Teoria de Resposta ao Item, proposto por Roberts,J.S., Donoghue,J.R., Laughlin,J.E.(2000)
- Finalidade: escala de atitudes
- Métodos de implementação: algoritmo EM (Expectation/Maximization)

Modelo GGUM

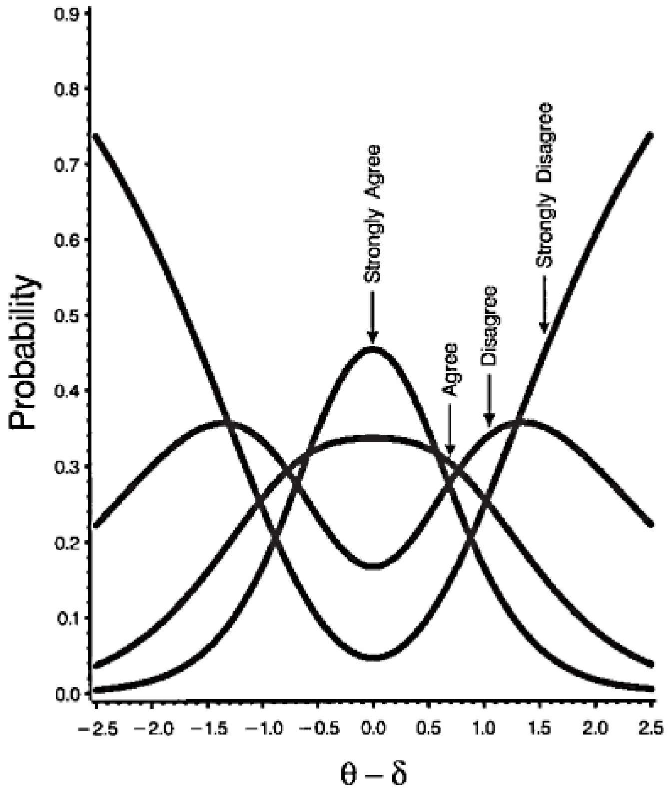
- Indivíduos são posicionados em uma escala contínua de atitude

- Indivíduos podem escolher uma categoria por duas razões: por estar posicionada “abaixo” ou “acima” da sua posição na escala



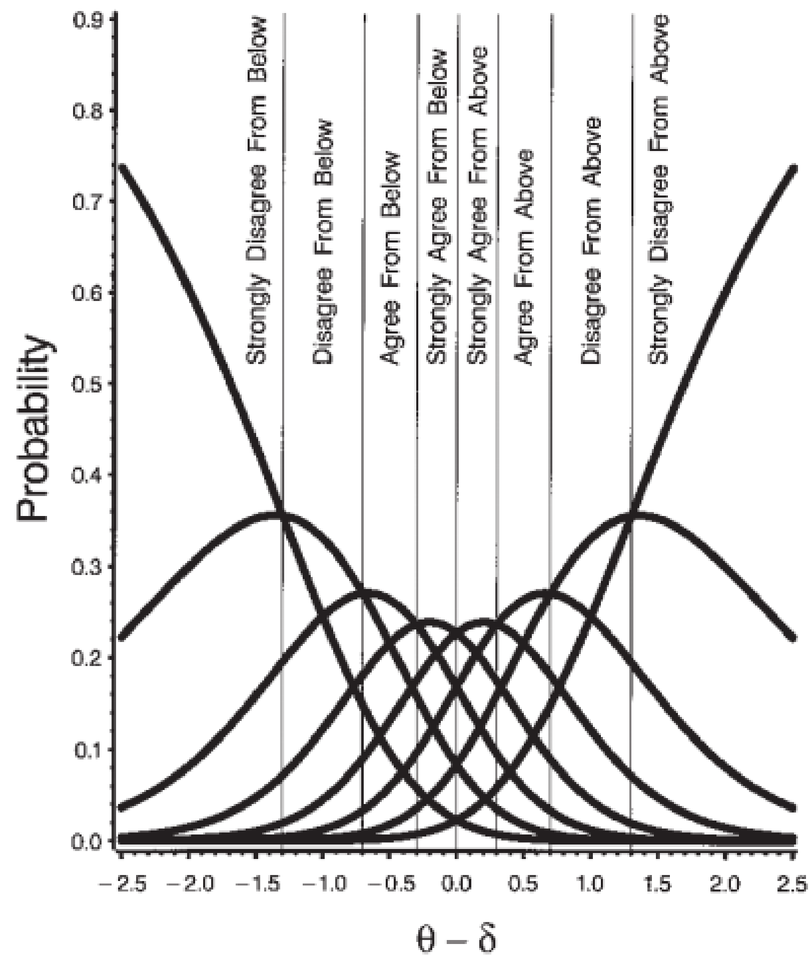
Modelo GGUM

Exemplo de curvas de probabilidade
para item com 4 categorias objetivas



Modelo GGUM

Exemplo de curvas de probabilidades *subjetivas* para item com 4 categorias



Modelo GGUM

A probabilidade associada a uma categoria subjetiva é

$$P(Y_i = y|\theta_j) = \frac{\exp\left\{\alpha_i\left[y(\theta_j - \delta_i) - \sum_{k=0}^{y}\tau_{ik}\right]\right\}}{\sum_{w=0}^{M}\left\{\exp\left\{\alpha_i\left[w(\theta_j - \delta_i) - \sum_{k=0}^{w}\tau_{ik}\right]\right\}\right\}}$$

- $y = 0 \rightarrow$ nível mais forte de discordância abaixo do item
- $y = M \rightarrow$ nível mais forte de discordância acima do item
- M é o número de respostas subjetivas menos 1
- α_i é o parâmetro de discriminação
- θ_j é a habilidade do j-ésimo indivíduo
- $\tau_{i,k}$ é a localização da k-ésima liminar no contínuo de atitute relativo à posição do item i

Modelo GGUM

A probabilidade de escolha de uma categoria do item

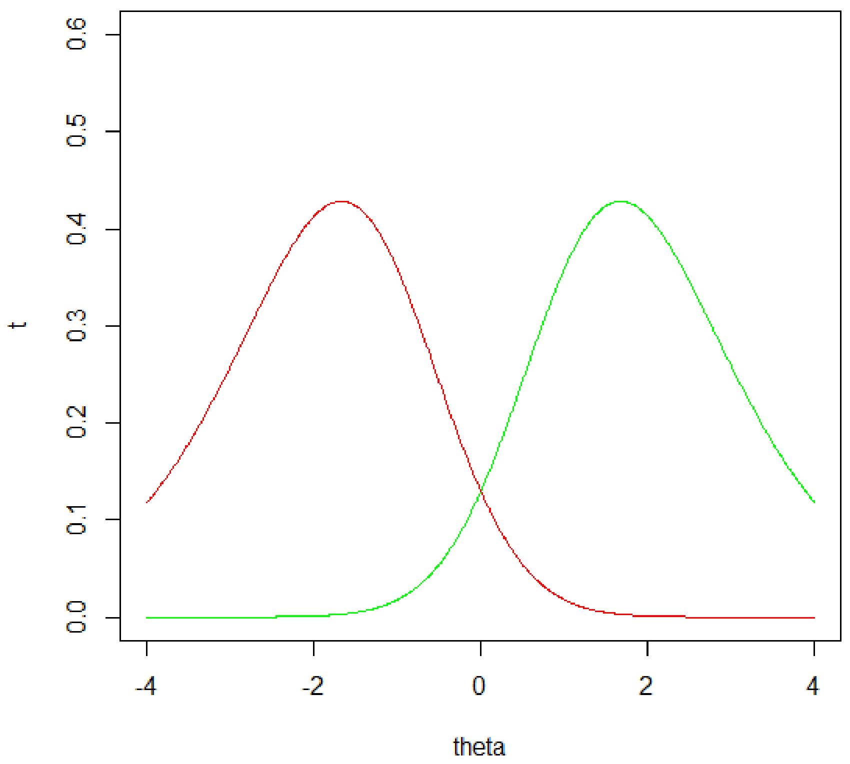
$$P(Z_i = z|\theta_j) = P(Y_i = z|\theta_j) + P[Y_i = (M - z)|\theta_j]$$

$$P(Z_i = z|\theta_j) = \frac{\exp\left\{\alpha_i\left[z(\theta_j - \delta_i) - \sum_{k=0}^{z}\tau_{ik}\right]\right\} + \exp\left\{\alpha_i\left[(M - z)(\theta_j - \delta_i) - \sum_{k=0}^{z}\tau_{ik}\right]\right\}}{\sum_{w=0}^{C}\left\{\exp\left\{\alpha_i\left[w(\theta_j - \delta_i) - \sum_{k=0}^{w}\tau_{ik}\right]\right\} + \exp\left\{\alpha_i\left[(M - w)(\theta_j - \delta_i) - \sum_{k=0}^{w}\tau_{ik}\right]\right\}\right\}}.$$

Modelo GGUM

$$P(Z_i = 1|\theta_j) = P(Y_i = 1|\theta_j) + P(Y_i = 4|\theta_j)$$

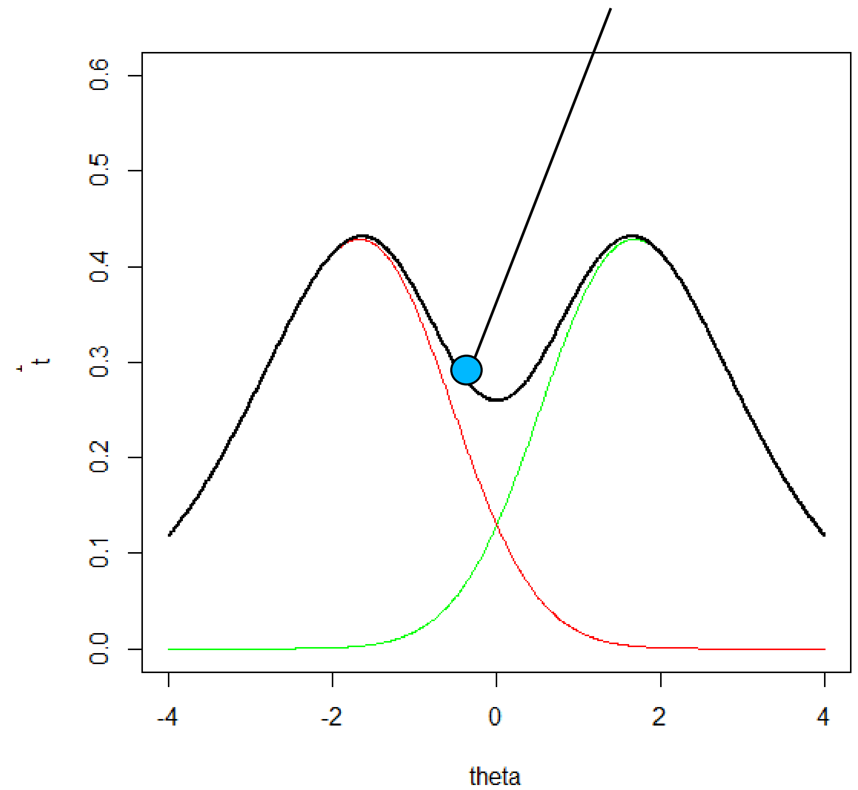
curvas subjetivas



Modelo GGUM

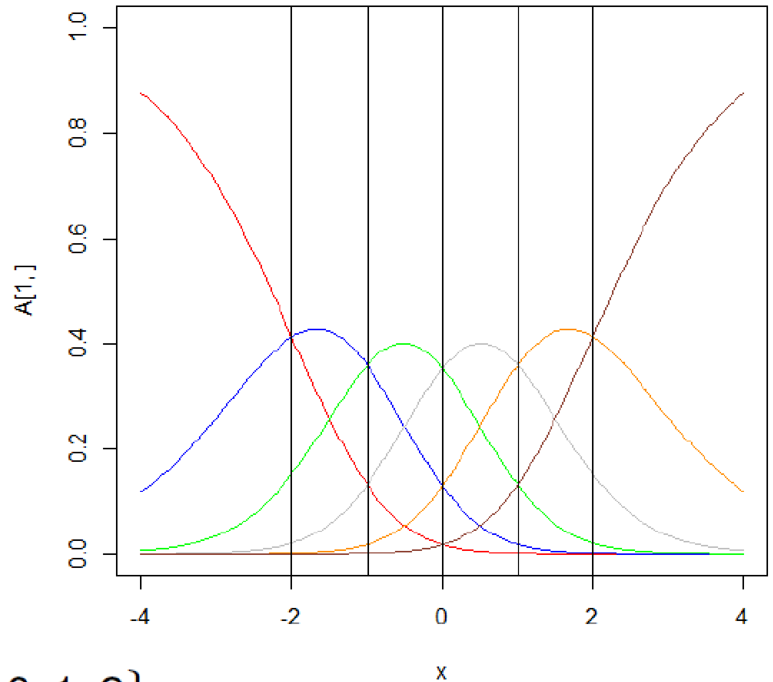
$$P(Z_i = 1|\theta_j) = P(Y_i = 1|\theta_j) + P(Y_i = 4|\theta_j)$$

$$P(Z_i = 1|\theta_j)$$



Interpretação dos parâmetros tau

plotGGUM.y(a=1,d=2,vtau=c(-2,-1))

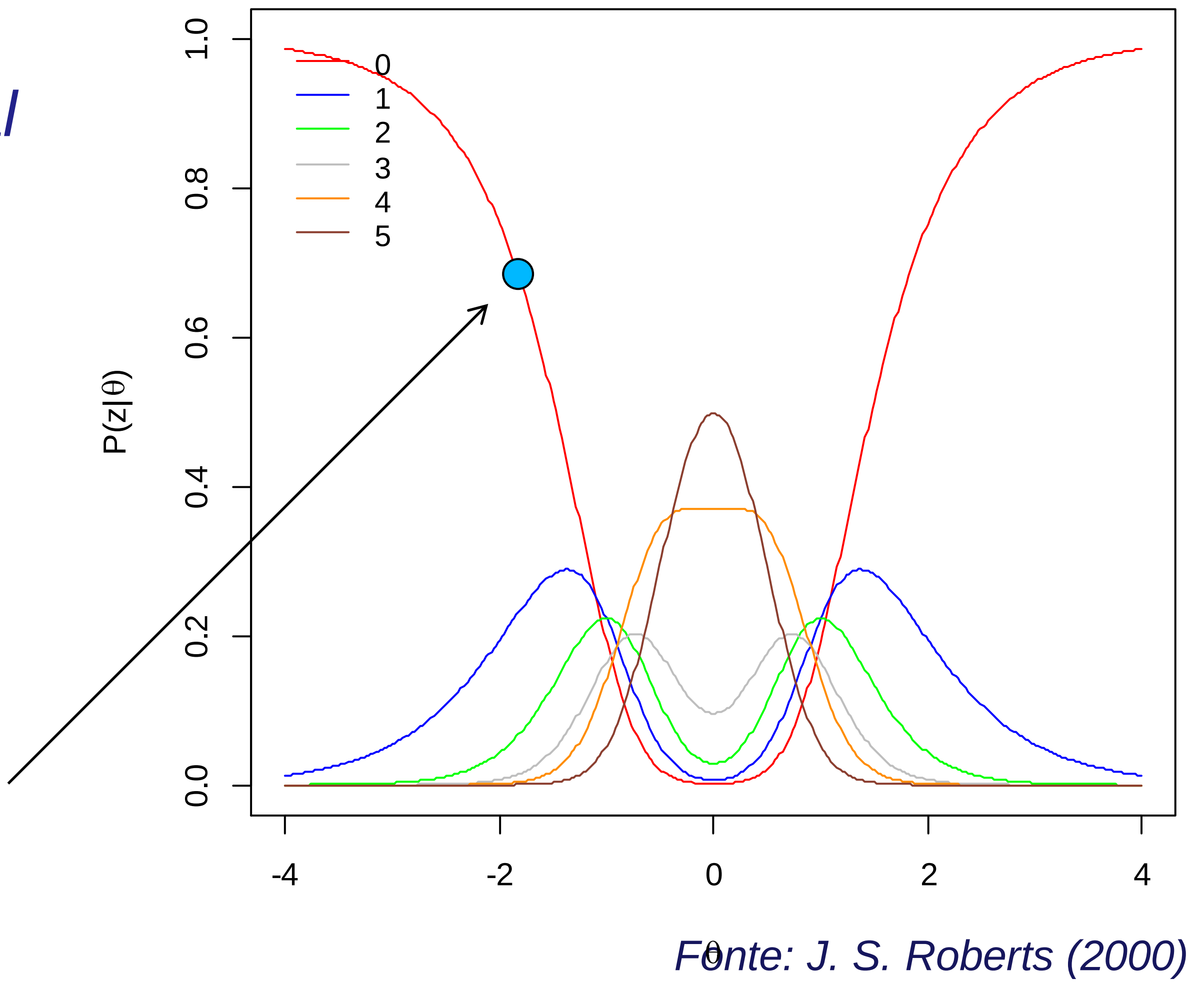


$$\tau = \{0, -2, -1, 0, 1, 2\}$$

Um item aplicado a 750 alunos - Universidade Carolina do Sul

Aborto deve ser ilegal exceto em casos extremos de incesto ou estupro

A. Concordo fortemente
B. Concordo
C. Concordo levemente
D. Discordo fortemente
E. Discordo
F. Discordo fortemente



Fonte: J. S. Roberts (2000)

Modelo GGUM

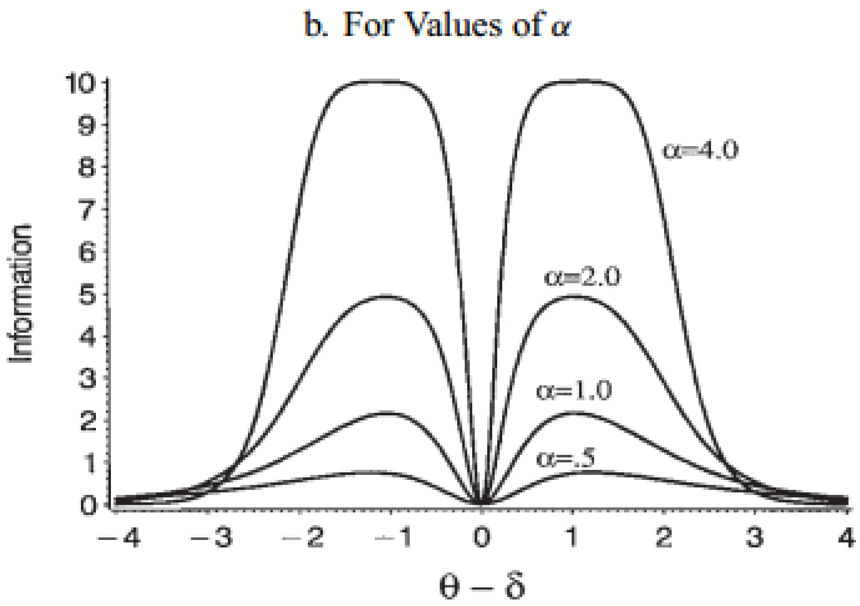
A função de informação de um item para o GGUM

$$I_i(\theta_j) = -\mathrm{E}\left[\frac{\partial^2 \ln(L)}{\partial\theta_j^2}\right] = \alpha_i^2 \left\{ \left\{ \sum_{z=0}^{C} \left[P(Z_i = z)\sigma^2_{Y_i|\theta_j,z} \right] \right\} - \sigma^2_{Y_i|\theta_j} \right\}$$

A função de informação do teste é

$$I(\theta_j) = \sum_{i=1}^{I} I_i(\theta_j) = \sum_{i=1}^{I} \alpha_i^2 \left\{ \left\{ \sum_{z=0}^{C} \left[P(Z_i = z)\sigma^2_{Y_i|\theta_j,z} \right] \right\} - \sigma^2_{Y_i|\theta_j} \right\}$$

Curvas de informação - GGUM



Implementação

Estimação dos parâmetros e escores:

- software GGUM2004 (gratuito)

  Psychometric Research and Development Lab
  Georgia Tech - School of Psychology

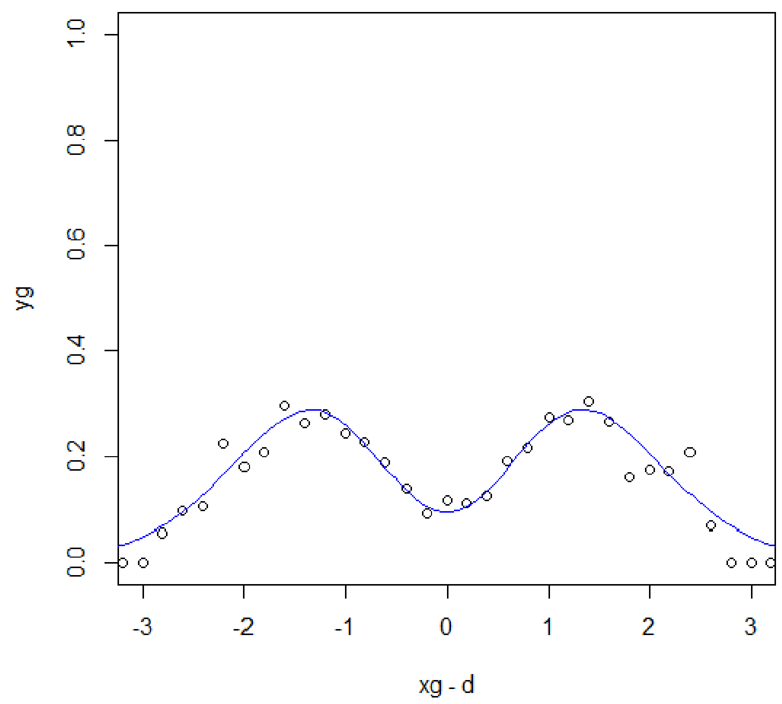
  http://prdlab.gatech.edu/unfolding/freesoftware/

Estimação dos escores:

- R package “ScoreGGUM”
  David R. King and James S. Roberts

Modelo GGUM

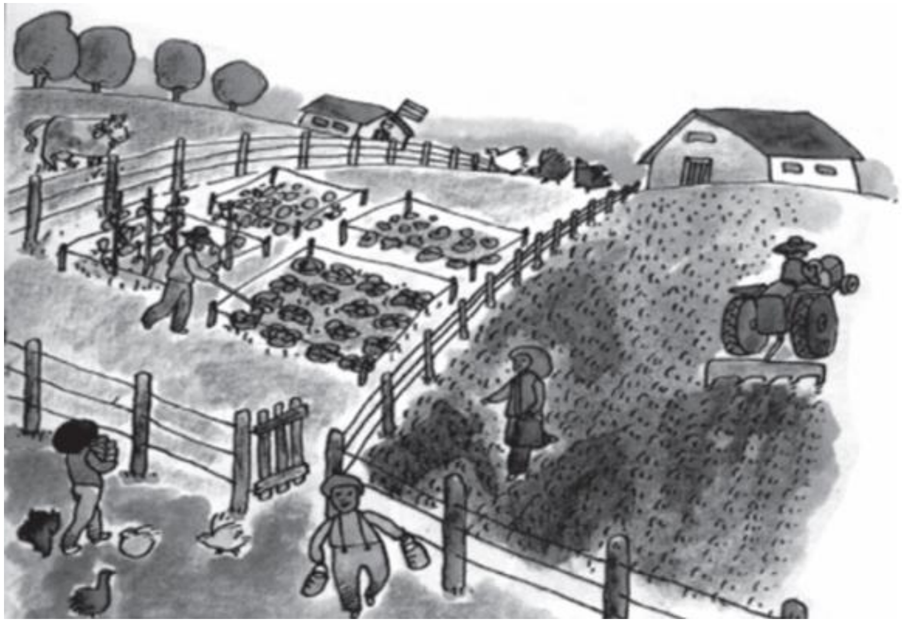
```
plot.fit(item=5,z=2,Mpar,MR,theta.all)
```



Exemplo 1 - Item aplicado aos alunos da 4ª série do ensino fundamental (PROEB-2001)

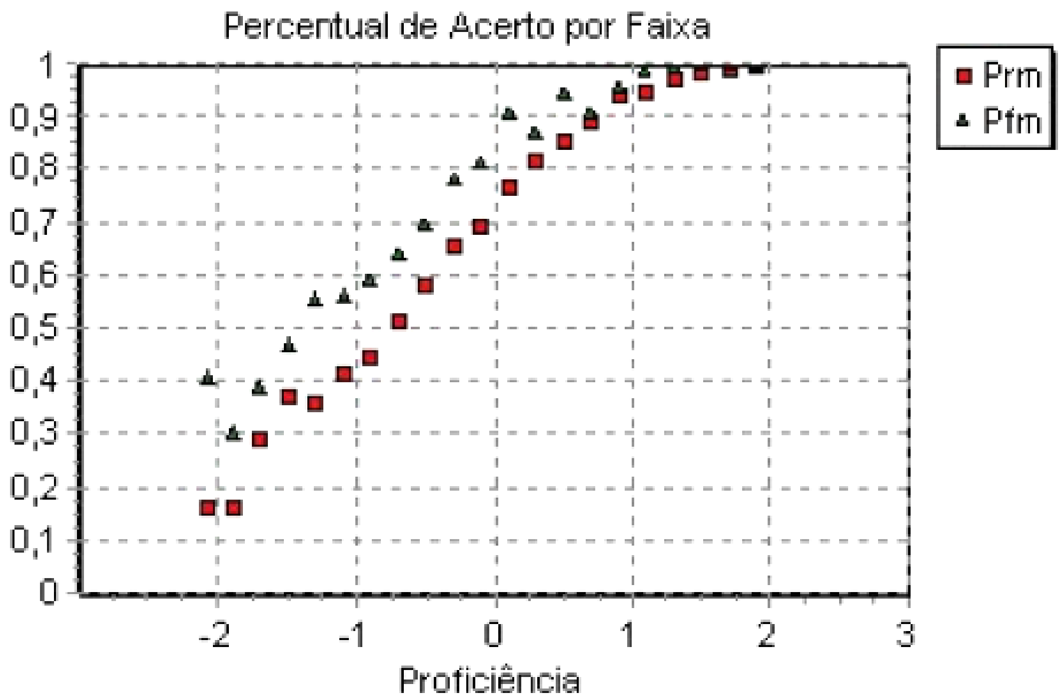
O desenho mostra uma realidade que se relaciona a um espaço

A) urbano
B) industrial
C) rural
D) metropolitano



Fonte: Soares T.,Gamerman, D.,Gonçalves, F.B (2007)

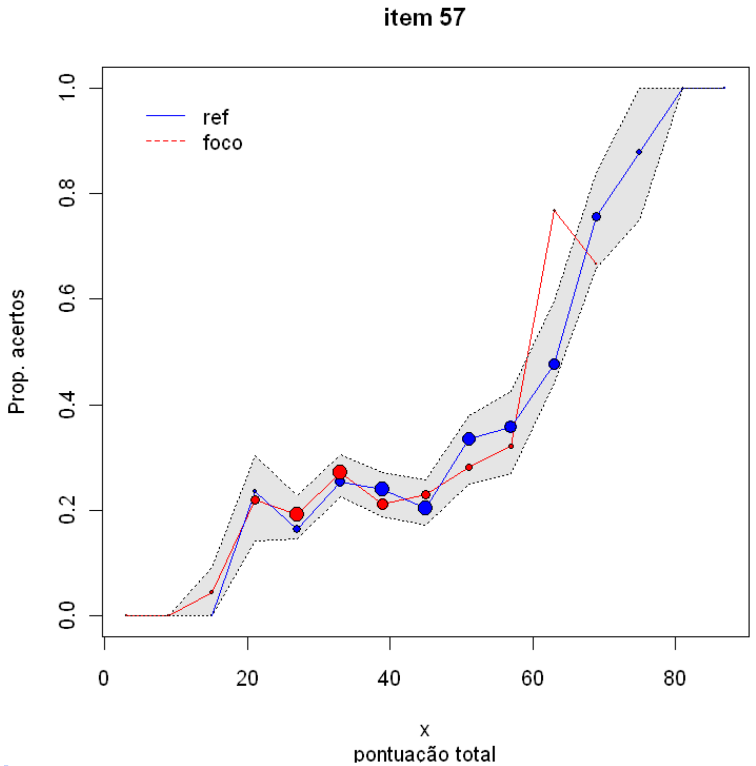
Exemplo 1 - Item aplicado aos alunos da 4ª série do ensino fundamental (PROEB-2001)



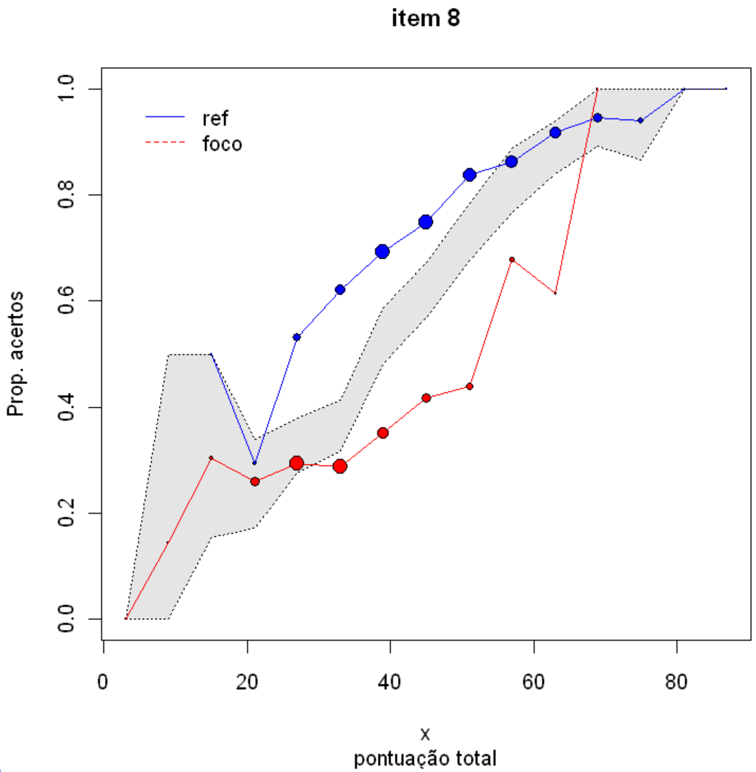
Fonte: Soares T.,Gamerman, D.,Gonçalves, F.B (2007)

Exemplo 2 comparação de dois itens (vestibular UMG 2004)

Um item de prova de História



Um item da prova de Língua Estrangeira



Um comportamento diferencial do item pode ocorrer na presença de:

- Indivíduos com mesma habilidade e diferenças relacionadas ao sexo, origem étnica ou grupos culturais;
- habilidade média menor de um grupo em particular
- fontes de dificuldade que são irrelevantes ou alheios ao construto que está sendo avaliado
- solicitação de informações que examinandos não tiveram igual exposição
- outros

Comportamento Diferencial do Item - DIF

Podemos caracterizar DIF como

- DIF uniforme → a magnitude do DIF é a mesma para todos os níveis de habilidades

- DIF não uniforme → magnitute do DIF varia de acordo com o nível de habilidade.

O pacote difR

- entrada de dados: matriz usual, com indivíduos nas linhas e itens nas colunas, entradas 0 e 1

- a matriz deve conter cabeçalho (primeira linha) com nomes dos itens

- a matriz deve ter duas colunas adicionais, preenchida com 0 ou 1 identificando *o grupo foco* e o *grupo referência*.

Alguns métodos para identificação de DIF

- Método das Áreas
- Método Padronizado
- Mantel-Haenszel
- Teste da Razão de Verossimilhança
- Teste qui-quadrado de Lord
- ... (vários outros)

difR

- Entrada de dados: matriz usual, com indivíduos nas linhas e itens nas colunas, entradas 0 e 1

- A matriz deve conter cabeçalho (primeira linha) com nomes dos itens

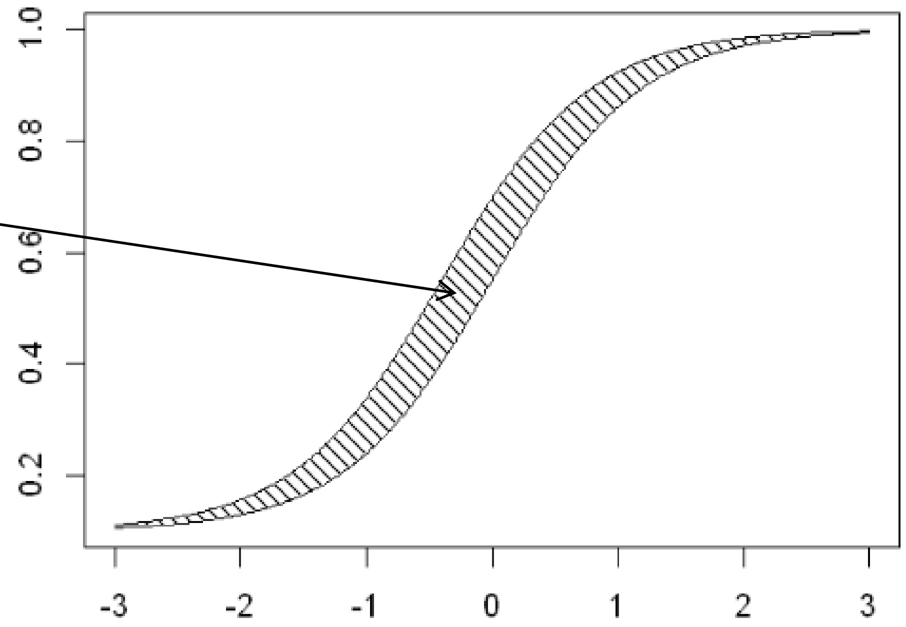
- A matriz deve ter duas colunas adicionais, preenchida com 0 ou 1 identificando *o grupo foco* e o *grupo referência*.

difR
métodos implementados:

R Command	Method
`difBD`	Breslow–Day
`difGenLord`	Generalized Lord
`difGMH`	Generalized Mantel–Haenszel
`difLogistic`	Logistic regression
`difLord`	Lord's chi-square test
`difLRT`	Likelihood ratio test
`difMH`	Mantel–Haenszel
`difRaju`	Raju's area
`difStd`	Standardization

- Método baseados em áreas entre as curvas
    - Rudner, Getson e Knight (1980)
    - Linn e Harnisch (1981)

Computado a área entre as curvas características dos grupos focal e referência



Método Padronizado (Standardized P-Difference):

$$ST_{pDIF} = \frac{\sum_m w_m (P_{f,m} - P_{r,m})}{\sum_m w_m}$$

onde:

w_m = pesos dos grupos

$P_{f,m}$= Proporção de indivíduos no grupo focal que acertaram o item

$P_{r,m}$= Proporção de indivíduos no grupo referência que acertaram

Obs 1: aqui a proporção é em relação ao total de respondentes ao item

Obs 2: índice assume valores em $[-1, 1]$.

Método Padronizado – maneiras de definir os pesos w_m

- “total”: total de sujeitos no nível m de habilidade,
- “reference” = total de sujeitos no grupo de referência com nível m de habilidade;
- “focal” (default): total de sujeitos no grupo focal com nível m de habilidade

Valores de referência para o método padronizado: ver difR manual

O Método Mantel-Haenszel: teste não paramétrico para DIF uniforme, a partir de tabela de contingência:

Tabela : Frequências observadas de respostas a um item

Grupo	Acertos (1)	Erros (0)	Total
Referência	A_j	B_j	n_{Rj}
Focal	C_j	D_j	n_{Fj}
Total	m_{1j}	m_{0j}	T_j

Método Mantel-Haenszel

$$\alpha_{MH} = \frac{\sum_{j=1}^{S} A_j D_j}{\sum_{j=1}^{S} B_j C_j}$$

A_j = frequência de resposta 1 do grupo referência no nível j

B_j = frequência de respostas 0 do grupo referência no nível j

C_j = frequência de respostas 1 do grupo foco no nível j

D_j = frequência de respostas 0 do grupo foco no nível j

Escala *Alfa Delta MH*, escala *MH D-DIF* ou escala Δ_{MH}:

Para itens favorecendo o grupo Foco: $\alpha_{MH} \in [1, +\infty]$
Para itens favorecendo o grupo Referencia:: $\alpha_{MH} \in [0, 1]$
Para evitar esta assimetria na escala Holland and Thayer (1988) propuseram a escala:

$$\Delta_{MH} = -2,35 log(\alpha_{MH})$$

Critério adotado pelo ETS - Educational Testing Service:

- $|\Delta_{MH}| \leq 1 \rightarrow$ negligenciável
- $1 < |\Delta_{MH}| \leq 1.5 \rightarrow$ moderado
- $|\Delta_{MH}| > 1.5 \rightarrow$ forte presença de DIF

Fonte: difR Reference Manual

Exemplo: banco de dados *"verbal"* (pacote difR)

- Respostas de 243 mulheres e 73 homens a um questionário sobre agressão verbal.

- Os itens descrevem uma situação frustrante. O valor 1 significa que o sujeito responderia à situação frustrante de forma agressiva; 0 caso contrário.

```
library(difR)      # carrega o pacote difR
data(verbal)       # para carregar os dados
```

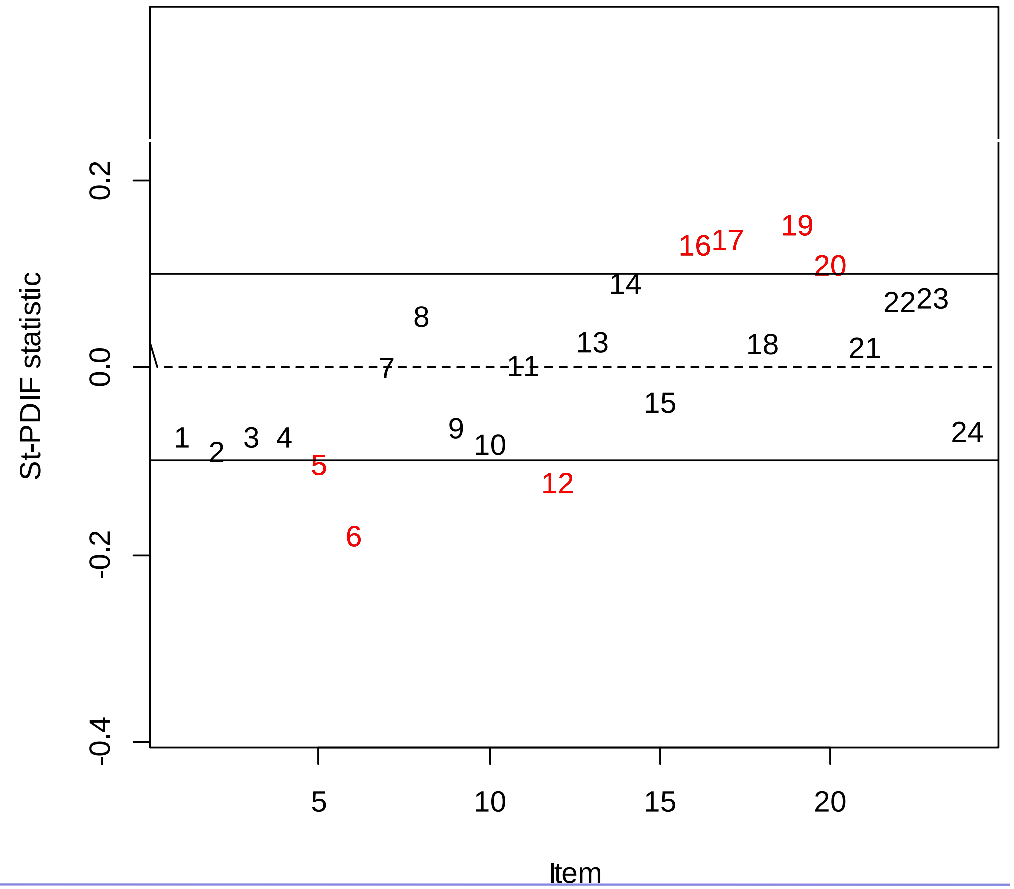
Método Padronizado

```
difStd(verbal2, group = "Gender", focal.name = 1)
```

```
Standardized P-DIF statistic:

            Stat.
S1wantCurse -0.0715 *
S1WantScold -0.0857 *
S1WantShout -0.0718 *
S2WantCurse -0.0719 *
S2WantScold -0.1001 **
S2WantShout -0.1759 **
S3WantCurse  0.0025
S3WantScold  0.0586 *
S3WantShout -0.0619 *
S4WantCurse -0.0787 *
…
```

Comportamento Diferencial do Item - DIF

```
# visualização
res= difStd(verbal2, group="Gender",focal.name=1)
plot(res)
```



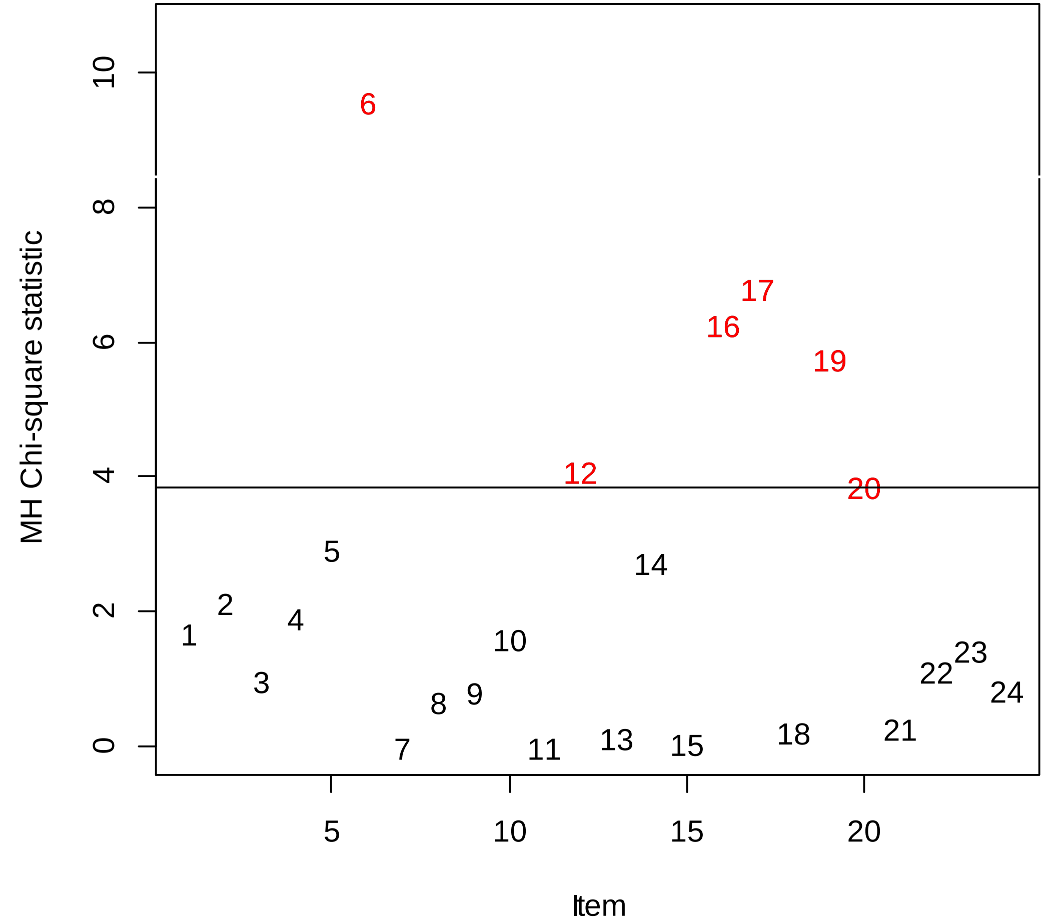
Método Mantel-Haenszel

```
difMH(verbal2, group = "Gender", focal.name = 1)
```

```
…
 Mantel-Haenszel Chi-square statistic:

             Stat.  P-value
S1wantCurse 1.7076 0.1913
S1WantScold 2.1486 0.1427
S1WantShout 0.9926 0.3191
S2WantCurse 1.9302 0.1647
S2WantScold 2.9540 0.0857  .
S2WantShout 9.6032 0.0019  **
S3WantCurse 0.0013 0.9711
S3WantScold 0.6752 0.4112
S3WantShout 0.8185 0.3656
Detection threshold: 3.8415 (significance level: 0.05)
```

Comportamento Diferencial do Item - DIF

```
# visualização
res= difMH(verbal2, group="Gender",focal.name=1)
plot(res)
```



estimados simultâneamente

variável
latente θ

parâmetros
dos itens

valores associados ao
indivíduo, equipe, escola, etc

Métodos computacionais usados na TRI:

- intensivos
- metodologia complexa para estimação dos parâmetros e escores
- questão da convergência dos algoritmos utilizados (crítico)
- escolha adequada do modelo

Algoritmo EM

Dado valores dos parâmetros dos itens, estimar as habilidades a partir da matriz de respostas

	Item 1 (a_1,b_1)	Item 2 (a_2,b_2)	Item 3 (a_3,b_3)	Item 4 (a_4,b_4)	
Indivíduo 1	1	0	1	0	θ_1
Indivíduo 2	1	1	1	1	θ_2
Indivíduo 3	1	1	0	1	θ_3
Indivíduo 5	...	...	...	...	...

Algoritmo EM

- fixado os valores das habilidades, estimar os parâmetros dos itens

	Item 1 (a_1,b_1)	Item 2 (a_2,b_2)	Item 3 (a_3,b_3)	Item 4 (a_4,b_4)	
Indivíduo 1	1	0	1	0	θ_1
Indivíduo 2	1	1	1	1	θ_2
Indivíduo 3	1	1	0	1	θ_3
Indivíduo 5	...	...	...	...	...

Procedimento computacional – ciclo EM

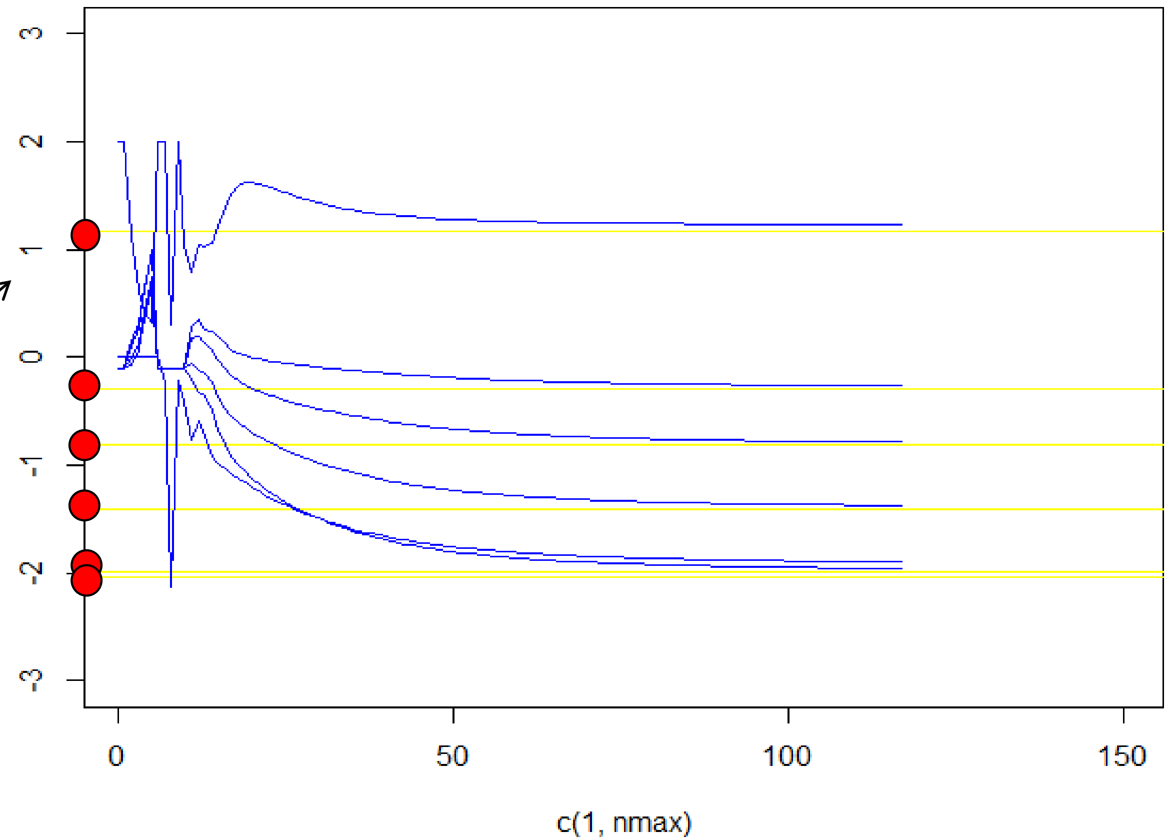
- valores iniciais dos parâmetros
- o ciclo termina quando:
  - é atingido um número máximo de ciclos
  - diferença entre parâmetros ou habilidades em ciclos consecutivos é pequena
  - outros critérios de parada
  - Normalmente critérios já pré-definidos no software

- Outros algoritmos: MCMC, métodos bayesianos

Exemplo de convergência observada para obtenção de parâmetros de um item no modelo GGUM

Valores iniciais dos parâmetros do item

Número de ciclos

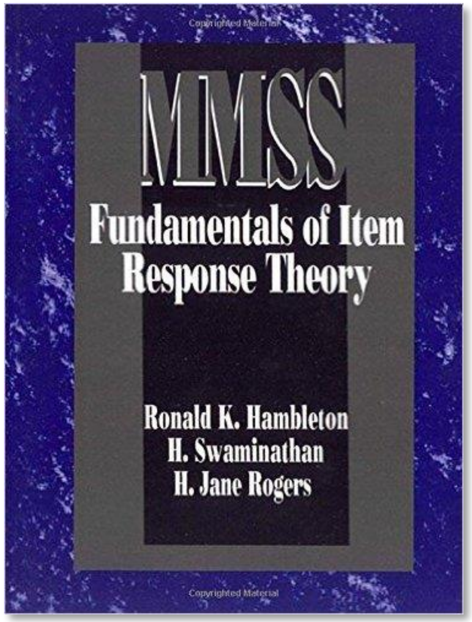


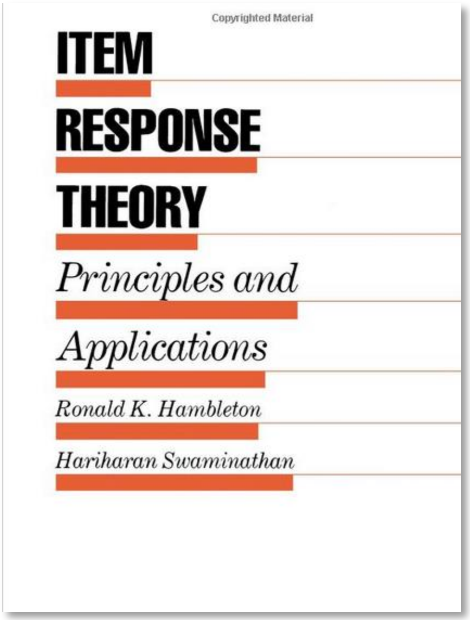
Indicações

Outros pacotes úteis no R

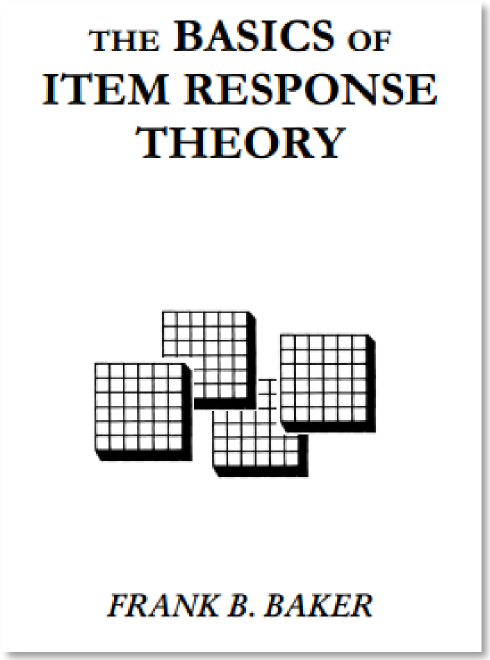
"plink" – para equalização
"irtoys" – modelos 1P, 2P e 3P com plots úteis

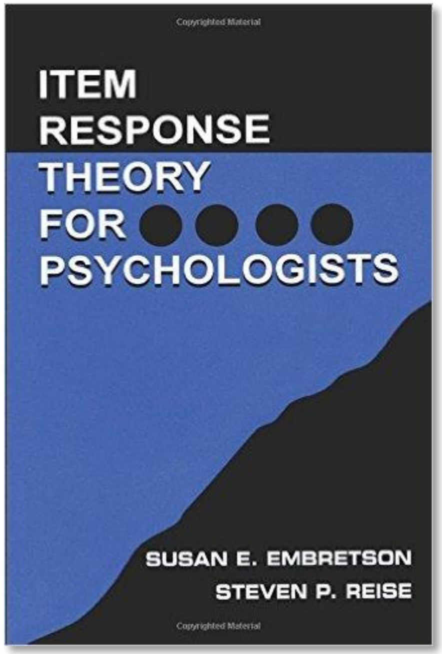
Indicações


Copyrighted Material
MMSS
Fundamentals of Item Response Theory
Ronald K. Hambleton
H. Swaminathan
H. Jane Rogers
Copyrighted Material


Copyrighted Material
ITEM
RESPONSE
THEORY
Principles and
Applications
Ronald K. Hambleton
Hariharan Swaminathan

Indicações


THE BASICS OF
ITEM RESPONSE
THEORY
FRANK B. BAKER


ITEM
RESPONSE
THEORY
FOR
PSYCHOLOGISTS
SUSAN E. EMBRETSON
STEVEN P. REISE

Modelos 1,2 e 3P

i

Teoria da Resposta ao Item:

Conceitos e Aplicações

Dalton Francisco de Andrade[1]

Heliton Ribeiro Tavares[2]

Raquel da Cunha Valle[3]

[1] Professor Titular do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará (UFC). *e-mail: dandrade@ufc.br*

[2] Professor do Departamento de Estatística da Universidade Federal do Pará (UFPA). *e-mail: heliton@ufpa.br*

[3] Estatístico da Fundação Carlos Chagas (FCC). *e-mail: rvalle@fcc.gov.br*

Andrade, Tavares & Valle SINAPE 2000

Referências

- J. S. Roberts, J. R. Donoghue, and J. E. Laughlin (2000), Unfolding Unidimensional Polytomous Responses, Applied Psychological Measurement,Vol. 24 No. 1, March 2000, 3–32

- GGUM2004: A Windows-Based Program to Estimate Parameters in the Generalized Graded Unfolding Model (reference manual)

- Hambleton RK, Swaminathan H, Rogers HJ: Fundamentals of Item Response Theory. Newbury Park, CA: SAGE 1991

- 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 (3), 847-862.

Referências

- Rizopoulos, Dimitris (2006), ltm: An R Package for Latent Variable Modeling and Item Reponse Theory Analyses Journal of Statistical Software, 17 (5)

- Roberts J. S., Donoghue J. R., Laughlin J. E.(2000). A General Model for Unfolding Unidimensional Polytomous Responses Using Item Response Theory. Applied Psychological Measurement, 24,3–32.

- Soares, T.,Gamerman, D.,Gonçalves, F.B (2007) Análise Bayesiana do Funcionamento Diferencial do Item, Pesquisa Operacional,v.27, p.271-291

- Zopluoglu, C. (2012). EstCRM: An R package for Samejima's Continuous IRT Model. Applied Psychological Measurement, 36(2), 149-150.