Desafios da Estatística na Avaliação Educacional: Enem, Prova Brasil, Saresp etc.

Heliton Ribeiro Tavares

Faculdade de Estatística / Universidade Federal do Pará
heliton@ufpa.br

Etapas da Apresentação

1. Breve introdução
2. Alguns resultados do Enem
3. Teoria da Resposta ao item (TRI)
4. Estimação
5. Principais Aplicativos (Softwares)
6. Algumas especificidades: Duas provas, Redação no Enem, o SiSu etc.
7. Principais Desafios da Estatística para Sistemas de Avaliação

Teoria Clássica de Medidas (TCM)

- Escores brutos ou padronizados
- Resultados dependem do particular conjunto de itens que compõem o instrumento de medida
- Inviável a comparação entre indivíduos que não foram submetidos "aos mesmos instrumentos de medida".
- Impossibilidade de construção de Escalas de Proficiência.

Teoria Clássica de Medidas (TCM)

- Escores brutos ou padronizados
- Resultados dependem do particular conjunto de itens que compõem o instrumento de medida
- Inviável a comparação entre indivíduos que não foram submetidos "aos mesmos instrumentos de medida".
- Impossibilidade de construção de Escalas de Proficiência.

Teoria Clássica de Medidas (TCM)

- Escores brutos ou padronizados
- Resultados dependem do particular conjunto de itens que compõem o instrumento de medida
- Inviável a comparação entre indivíduos que não foram submetidos "aos mesmos instrumentos de medida".
- Impossibilidade de construção de Escalas de Proficiência.

Teoria Clássica de Medidas (TCM)

- Escores brutos ou padronizados
- Resultados dependem do particular conjunto de itens que compõem o instrumento de medida
- Inviável a comparação entre indivíduos que não foram submetidos "aos mesmos instrumentos de medida".
- Impossibilidade de construção de Escalas de Proficiência.

Algumas Avaliações Educacionais no Brasil e no Mundo

- No Brasil vem sendo usada extensamente em avaliação educacional SAEB/Prova Brasil, Provinha Brasil de LP e MT, Encceja, ENADE, ENEM, Saresp, Prova São Paulo ...
- No mundo: TOEFL, GRE, PISA, ...
- O Enem é parte fundamental dos exames vestibulares de algumas universidades
- Decreto do MEC impõe explicitamente no Enem o uso da Teoria da Resposta ao item via "modelo logístico de 3 parâmetros", com vistas a realizar avaliações paralelar ou mais de uma edição por ano.
- Preocupação com Comparabilidade.

Algumas Avaliações Educacionais no Brasil e no Mundo

- No Brasil vem sendo usada extensamente em avaliação educacional SAEB/Prova Brasil, Provinha Brasil de LP e MT, Encceja, ENADE, ENEM, Saresp, Prova São Paulo ...
- No mundo: TOEFL, GRE, PISA, ...
- O Enem é parte fundamental dos exames vestibulares de algumas universidades
- Decreto do MEC impõe explicitamente no Enem o uso da Teoria da Resposta ao item via "modelo logístico de 3 parâmetros", com vistas a realizar avaliações paralelar ou mais de uma edição por ano.
- Preocupação com Comparabilidade.

Algumas Avaliações Educacionais no Brasil e no Mundo

- No Brasil vem sendo usada extensamente em avaliação educacional SAEB/Prova Brasil, Provinha Brasil de LP e MT, Encceja, ENADE, ENEM, Saresp, Prova São Paulo ...
- No mundo: TOEFL, GRE, PISA, ...
- O Enem é parte fundamental dos exames vestibulares de algumas universidades
- Decreto do MEC impõe explicitamente no Enem o uso da Teoria da Resposta ao item via "modelo logístico de 3 parâmetros", com vistas a realizar avaliações paralelar ou mais de uma edição por ano.
- Preocupação com Comparabilidade.

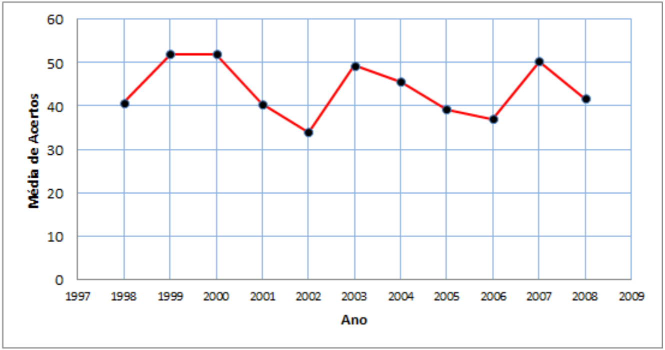
Algumas Avaliações Educacionais no Brasil e no Mundo

- No Brasil vem sendo usada extensamente em avaliação educacional SAEB/Prova Brasil, Provinha Brasil de LP e MT, Encceja, ENADE, ENEM, Saresp, Prova São Paulo ...
- No mundo: TOEFL, GRE, PISA, ...
- O Enem é parte fundamental dos exames vestibulares de algumas universidades
- Decreto do MEC impõe explicitamente no Enem o uso da Teoria da Resposta ao item via "modelo logístico de 3 parâmetros", com vistas a realizar avaliações paralelar ou mais de uma edição por ano.
- Preocupação com Comparabilidade.

Algumas Avaliações Educacionais no Brasil e no Mundo

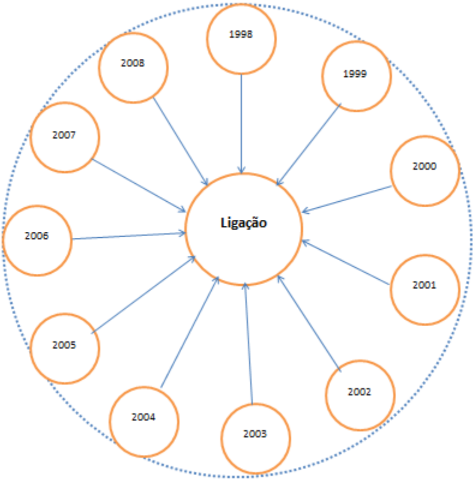
- No Brasil vem sendo usada extensamente em avaliação educacional SAEB/Prova Brasil, Provinha Brasil de LP e MT, Encceja, ENADE, ENEM, Saresp, Prova São Paulo ...
- No mundo: TOEFL, GRE, PISA, ...
- O Enem é parte fundamental dos exames vestibulares de algumas universidades
- Decreto do MEC impõe explicitamente no Enem o uso da Teoria da Resposta ao item via "modelo logístico de 3 parâmetros", com vistas a realizar avaliações paralelar ou mais de uma edição por ano.
- Preocupação com Comparabilidade.

Média de Acertos no Enem: 1998 a 2008



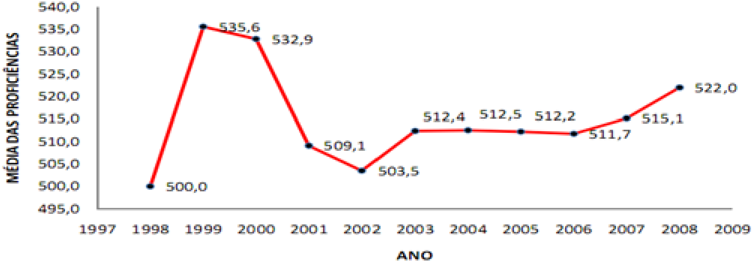
Serão os resultados dessas provas comparáveis?

Prova de Ligação: 1998 a 2008



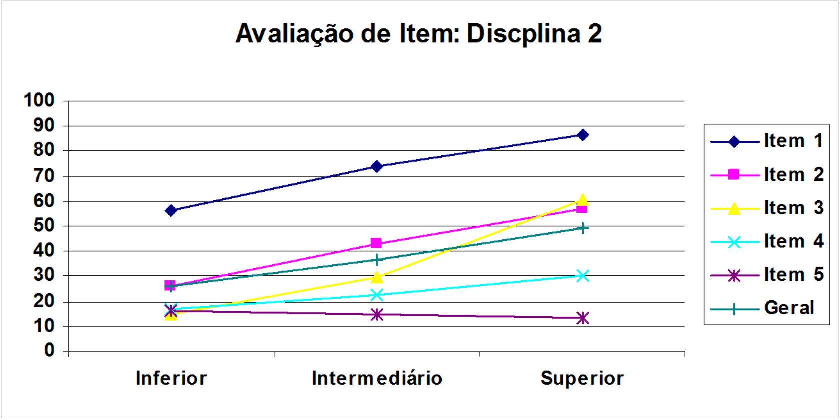
Cerca de 40 mil alunos participaram dessa prova.

Médias do Enem 1998 a 2008 na mesma escala

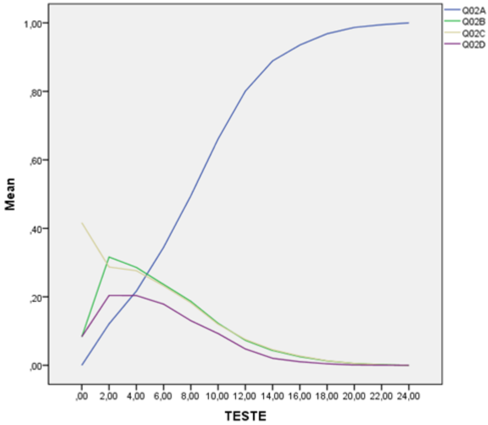


Agora as médias podem ser efetivamente comparadas.

Proporção de acertos em função do escore: 3 grupos



Proporção de acertos em função do escore



Teoria da Resposta ao item (TRI)

- Teoria desenvolvida para suprir necessidades na área psicométrica e, posteriormente, educacional.
- É composta por conjunto de modelos que consideram variáveis latentes.
- Modelos de Resposta ao Item (MRI) : representam o relacionamento entre traços latentes de indivíduos e parâmetros do item de um instrumento de medida (prova, questionário). Modela a probabilidade de responder corretamente ao item.
- Existe um grande número de classes de MRI: dicotômicos e policotômicos, desdobramento, um e múltiplos grupos, multidimensionais, multivariados, longitudinais, dentre outros.

Teoria da Resposta ao item (TRI)

- Teoria desenvolvida para suprir necessidades na área psicométrica e, posteriormente, educacional.
- É composta por conjunto de modelos que consideram variáveis latentes.
- Modelos de Resposta ao Item (MRI) : representam o relacionamento entre traços latentes de indivíduos e parâmetros do item de um instrumento de medida (prova, questionário). Modela a probabilidade de responder corretamente ao item.
- Existe um grande número de classes de MRI: dicotômicos e policotômicos, desdobramento, um e múltiplos grupos, multidimensionais, multivariados, longitudinais, dentre outros.

Teoria da Resposta ao item (TRI)

- Teoria desenvolvida para suprir necessidades na área psicométrica e, posteriormente, educacional.
- É composta por conjunto de modelos que consideram variáveis latentes.
- Modelos de Resposta ao Item (MRI) : representam o relacionamento entre traços latentes de indivíduos e parâmetros do item de um instrumento de medida (prova, questionário). Modela a probabilidade de responder corretamente ao item.
- Existe um grande número de classes de MRI: dicotômicos e policotômicos, desdobramento, um e múltiplos grupos, multidimensionais, multivariados, longitudinais, dentre outros.

Teoria da Resposta ao item (TRI)

- Teoria desenvolvida para suprir necessidades na área psicométrica e, posteriormente, educacional.
- É composta por conjunto de modelos que consideram variáveis latentes.
- Modelos de Resposta ao Item (MRI) : representam o relacionamento entre traços latentes de indivíduos e parâmetros do item de um instrumento de medida (prova, questionário). Modela a probabilidade de responder corretamente ao item.
- Existe um grande número de classes de MRI: dicotômicos e policotômicos, desdobramento, um e múltiplos grupos, multidimensionais, multivariados, longitudinais, dentre outros.

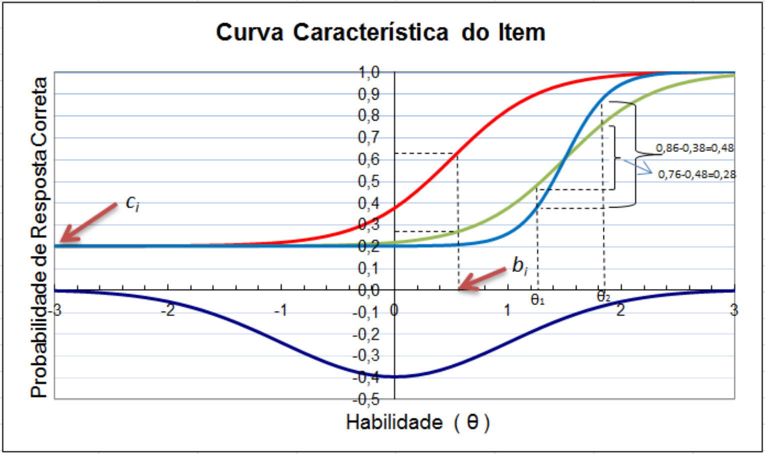
O modelo logístico de 3 parâmetros (ML3)

Dos modelos propostos pela TRI, o *modelo logístico unidimensional de 3 parâmetros (ML3)* é atualmente o mais utilizado. Baseia-se na suposição de que quanto maior o conhecimento (habilidadde) do indivíduo na área avaliada, maior é a probabilidade de ele responder corretamente ao item. Assim, podemos esperar o seguinte comportamento para a probabilidade de resposta correta em função da habilidade:

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

com $i = 1, 2, \cdots, n$, e $j = 1, 2, \cdots, n$ e $\zeta_i = (a_i, b_i, c_i)$.

Exemplo de uma Curva Característica do Item – CCI



Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala

Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala

Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala

Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala.

Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala.

Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala

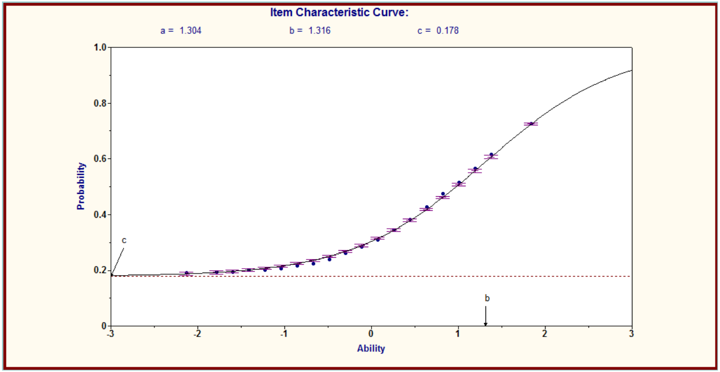
Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala.

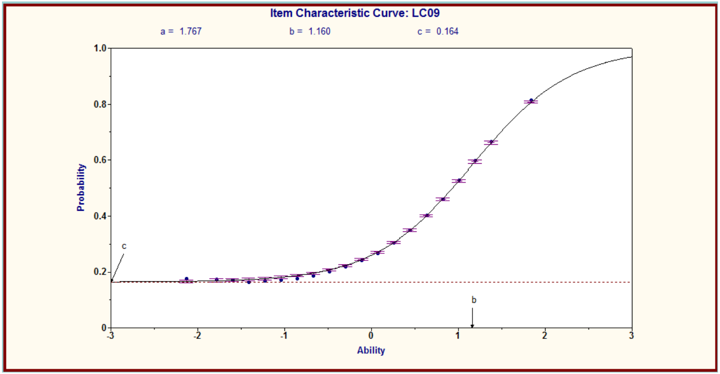
Interpretação:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ij} = 1|\theta_j, \zeta_i)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Modelo (ou Função) de Resposta do Item – MRI.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação do item i.
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i.
- D é um fator de escala.

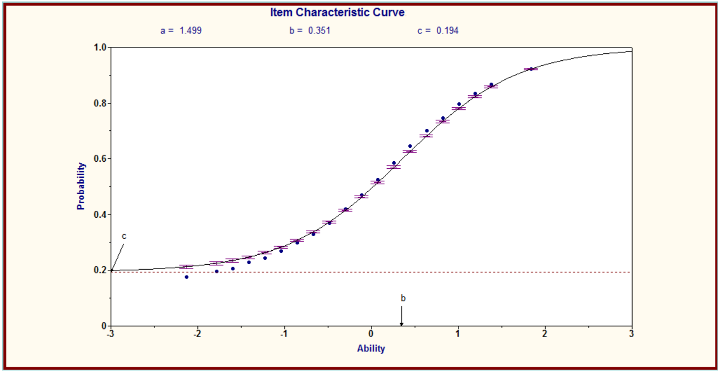
Exemplo de Item na PRÁTICA



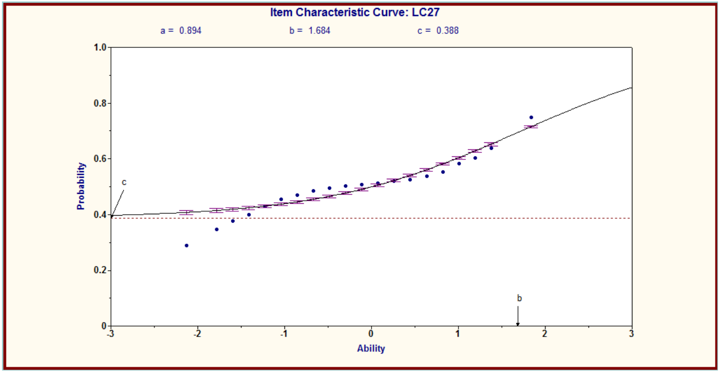
Exemplo de Item na PRÁTICA



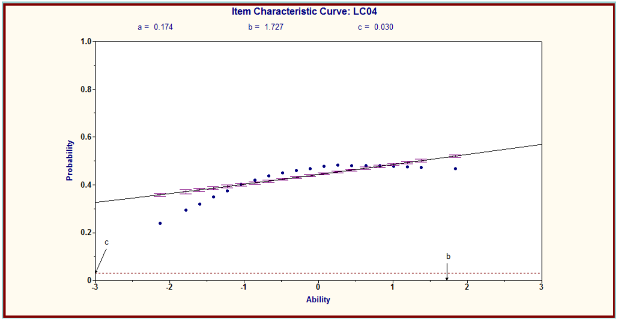
Exemplo de Item na PRÁTICA. Possível melhoria



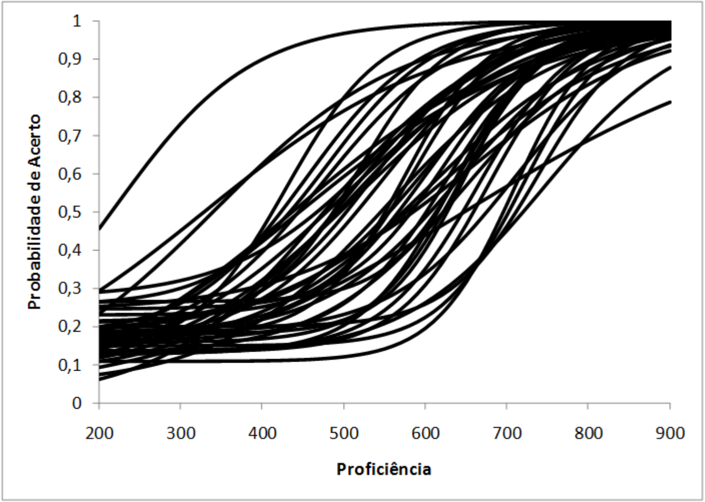
Exemplo de Item com comportamento inesperado



Da Teoria Clássica



Enem 2009: LCT



Processo de Estimação

Geral: Estimação por máxima verossimilhança $\Rightarrow$ contruir a verossimilhança e encontrar os valores de ζ que maximizam essa função.

Notação:
u_{ji}: resposta do indivíduo j ao item i
$\mathbf{u}_{j.}$: vetor de respostas do indivíduo j a todos os itens
$\mathbf{u}_{..}$: conjunto total de observações

$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_N)$: vetor de habilidades dos N indivíduos

$\zeta = (\zeta_1, \cdots, \zeta_n)$: conjunto de parâmetros dos itens.

Suposições:

As duas principais são:

(S1) as respostas oriundas de indivíduos diferentes são independentes,

(S2) os itens são respondidos de forma independente por cada indivíduo (Independência Condicional), fixada sua habilidade. Ou seja, a habilidade (θ) é a única informação necessária para determinar se o indivíduo acerta ou erra a questão. Ou ainda, o idivíduo não aprende (altera sua habilidade) no momento do teste.

Construção da verossimilhança

O processo é construído em duas etapas, primeiro supõe-se alguma distribuição para as habilidades da população, tal como *Normal*(μ, σ^2). Na prática podemos fixar $\mu = 0$ e $\sigma^2 = 1$. A equação de verossimilhança fica em função apenas dos parâmetros dos itens ζ.

Geral: Supor distribuição para a habilidade: fdp é $g(\theta|\boldsymbol{\eta})$, $\boldsymbol{\eta} = (\mu, \sigma^2)'$

A verossimilhança individual pode ser escrita como

$$P(\mathbf{u}_{j.}|\zeta, \boldsymbol{\eta}) = \int_{\Re} P(\mathbf{u}_{j.}|\theta, \zeta) g(\theta|\boldsymbol{\eta}) d\theta.$$

Verossimilhança marginal

- A função de verossimilhança marginal será:

$$L(\zeta) = P(\mathbf{U}_{..} = \mathbf{u}_{..}|\zeta) \overset{(S1)}{=} \prod_{j=1}^{N} P(\mathbf{U}_{j.} = \mathbf{u}_{j.}|\zeta)$$

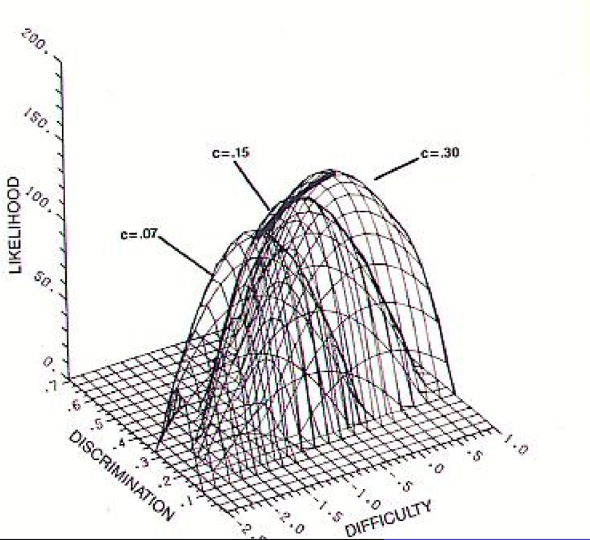
- Encontrar o ponto de máximo da verossimilhança é equivalente a encontrar o ponto que maximiza a log-verossimilhança;

$$\log L(\zeta) = \sum_{j=1}^{N} P(\mathbf{U}_{j.} = \mathbf{u}_{j.}|\zeta).$$

- Para encontrar o ponto de máximo de uma função, derivamos e igualamos a zero:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \mathbf{0}, \quad i = 1, \cdots, n.$$

Visualização da Verossimilhança



Equação de Verossimilhança

Podemos escrever a log-verossimilhança como

$$\begin{aligned}
\frac{\partial \log L(\zeta, \boldsymbol{\eta})}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \left\{ \sum_{j=1}^{N} \log P(\mathbf{u}_{j.} | \zeta, \boldsymbol{\eta}) \right\} \\
&= \sum_{j=1}^{N} \frac{1}{P(\mathbf{u}_{j.} | \zeta, \boldsymbol{\eta})} \frac{\partial P(\mathbf{u}_{j.} | \zeta, \boldsymbol{\eta})}{\partial \zeta_i}.
\end{aligned}$$

Depois de algumas páginas de desenvolvimento, obtemos

$$\frac{\partial P(\mathbf{u}_{j.} | \zeta, \boldsymbol{\eta})}{\partial \zeta_i} = \int_{\Re} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial \zeta_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] P(\mathbf{u}_{j.} | \theta, \zeta) g(\theta | \boldsymbol{\eta}) d\theta$$

Desenvolvimento

onde,

$$W_{ji} = \frac{P^*_{ji} Q^*_{ji}}{P_{ji} Q_{ji}}, \quad \text{onde} \quad P^*_{ji} \text{ é o ML2} \quad \text{e} \quad Q^*_{ji} = 1 - P^*_{ji}$$

Usando a notação

$$g^*_j(\theta) \equiv g(\theta | \mathbf{u}_{j.}, \zeta, \boldsymbol{\eta}) = \frac{P(\mathbf{u}_{j.} | \theta, \zeta) g(\theta | \boldsymbol{\eta})}{P(\mathbf{u}_{j.} | \zeta, \boldsymbol{\eta})}, \tag{2}$$

teremos que a função de log-verossimilhança será

$$\frac{\partial \log L(\zeta, \boldsymbol{\eta})}{\partial \zeta_i} = \sum_{j=1}^{N} \int_{\Re} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial \zeta_i} \right) \frac{W_i}{P^*_i Q^*_i} \right] g^*_j(\theta) d\theta.$$

Derivadas de P_{ji}

Finalmente, basta obter as derivadas para a FRI escolhida para cada item. Adotando o ML3 para todos os itens, teremos

$$
\begin{aligned}
\frac{\partial P_{ji}}{\partial a_i} &= D(1-c_i)(\theta_j - b_i)P^*_{ji}Q^*_{ji}, \\
\frac{\partial P_{ji}}{\partial b_i} &= -Da_i(1-c_i)P^*_{ji}Q^*_{ji}, \\
\frac{\partial P_{ji}}{\partial c_i} &= Q^*_{ji}.
\end{aligned}
$$

Agora é só substituir na equação anterior.

Equações de Estimação Finais

Em resumo, as equações de estimação para os parâmetros a_i, b_i e c_i são, respectivamente,

$$
\begin{aligned}
a_i \quad &: \quad D(1-c_i)\sum_{j=1}^{N}\int_{\Re}\left[(u_{ji}-P_i)(\theta-b_i)W_i\right]g_j^*(\theta)d\theta = 0,\\
b_i \quad &: \quad -Da_i(1-c_i)\sum_{j=1}^{N}\int_{\Re}\left[(u_{ji}-P_i)W_i\right]g_j^*(\theta)d\theta = 0,\\
c_i \quad &: \quad \sum_{j=1}^{N}\int_{\Re}\left[(u_{ji}-P_i)\frac{W_i}{P_i^*}\right]g_j^*(\theta)d\theta = 0,
\end{aligned}
$$

Métodos Iterativos

As equações não apresentam solução explícita. Com isso, deve ser adotado algum método iterativo para obtenção das estimativas de máxima verossimilhança, tal como o Método de Newton-Raphson ou Scoring de Fisher. Para isso, devem ser calculadas ainda a derivadas segundas,

$$\frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_l} \quad \forall (i, l). \tag{3}$$

Naturalmente, quando $i = l$ (mesmo item) essas derivadas não serão nulas, mas o desejável seria que as derivadas cruzadas ($i \neq l$) fossem nulas, mas não são. Isso implica que deveremos estimar todos os itens conjuntamente, tendo que inverter matrizes da ordem $3n$ inúmeras vezes, por exemplo.

Métodos Iterativos

Usando

$$r_i(\theta) = \sum_{j=1}^{N} u_{ji} g_j^*(\theta), \quad f_i(\theta) = \sum_{j=1}^{N} g_j^*(\theta),$$

As equações de estimação ficam

$$\begin{aligned}
\frac{\partial \log L(\zeta,\eta)}{\partial a_i} &= D(1-c_i)\int_{\Re}(\theta-b_i)\left[r_i(\theta)-P_i f_i(\theta)\right] W_i d\theta,\\
\frac{\partial \log L(\zeta,\eta)}{\partial b_i} &= -Da_i(1-c_i)\int_{\Re}\left[r_i(\theta)-P_i f_i(\theta)\right] W_i d\theta,\\
\frac{\partial \log L(\zeta,\eta)}{\partial c_i} &= \int_{\Re}\left[r_i(\theta)-P_i f_i(\theta)\right] W_i d\theta.
\end{aligned}$$

Pontos de Quadratura

Na prática, a integral é aproximada usando-se q pontos de quadratura $\overline{\theta}_k$ e respectivos pesos A_k, $k = 1, \cdots q$. As Equações ficam

$$\begin{aligned}
a_i &: \quad D(1-c_i)\sum_{k=1}^{q}(\overline{\theta}_k - b_i)\left[r_{ki} - P_{ki}f_k\right]W_{ki} = 0, \\
b_i &: \quad -Da_i(1-c_i)\sum_{k=1}^{q}\left[r_{ki} - P_{ki}f_k\right]W_{ki} = 0, \\
c_i &: \quad \sum_{k=1}^{q}\left[r_{ki} - P_{ki}f_k\right]\frac{W_{ki}}{P^*_{ki}} = 0.
\end{aligned}$$

Etapas do Algoritmo EM

Mais especificamente, os passos E e M são

Passo E Usar os $\bar{\theta}_k$, os pesos A_k, $k = 1, \cdots, q$ e estimativas iniciais $\widehat{\zeta}_i$ para gerar $g_j^*(\bar{\theta}_k)$ e, posteriormente, r_{ki} e f_k.

Passo M Com $\mathbf{r}$ e $\mathbf{f}$ obtidos no Passo E, resolver as equações de estimação para ζ_i usando o algoritmo Newton-Raphson ou "Scoring" de Fisher.

Vantagem: tratamento de matrizes 3×3 para o ML3, convergência.

Desvantagem: Velocidade de convergência

Método de Newton-Raphson

- Inicia-se com uma estimativa inicial de ζ_i, representada por $\widehat{\zeta}_i^{(0)}$. A cada iteração $1, 2, \cdots$, as estimativas são melhoradas.
- A expressão para a estimativa de ζ_i na iteração $t+1$ será

$$\widehat{\zeta}_i^{(t+1)} = \widehat{\zeta}_i^{(t)} - [\boldsymbol{\Delta}(\widehat{\zeta}_i^{(t)})]^{-1}\mathbf{h}(\widehat{\zeta}_i^{(t)}).$$

- O processo para quando algum critério de parada for alcançado.

Método de Newton-Raphson

- Inicia-se com uma estimativa inicial de ζ_i, representada por $\widehat{\zeta}_i^{(0)}$. A cada iteração $1, 2, \cdots$, as estimativas são melhoradas.
- A expressão para a estimativa de ζ_i na iteração $t+1$ será

$$\widehat{\zeta}_i^{(t+1)} = \widehat{\zeta}_i^{(t)} - [\boldsymbol{\Delta}(\widehat{\zeta}_i^{(t)})]^{-1}\mathbf{h}(\widehat{\zeta}_i^{(t)}).$$

- O processo para quando algum critério de parada for alcançado.

Método de Newton-Raphson

- Inicia-se com uma estimativa inicial de ζ_i, representada por $\widehat{\zeta}_i^{(0)}$. A cada iteração $1, 2, \cdots$, as estimativas são melhoradas.
- A expressão para a estimativa de ζ_i na iteração $t+1$ será

$$\widehat{\zeta}_i^{(t+1)} = \widehat{\zeta}_i^{(t)} - [\boldsymbol{\Delta}(\widehat{\zeta}_i^{(t)})]^{-1}\mathbf{h}(\widehat{\zeta}_i^{(t)}).$$

- O processo para quando algum critério de parada for alcançado.

Erro-Padrão

Os estimadores de máxima verossimilhança gozam de propriedades assintóticas conhecidas, tais como vício nulo e eficiência. Sob algumas condições de regularidade (ver Sen & Singer (1993), por exemplo) a distribuição assintótica do estimador de máxima verossimilhança, $\widehat{\zeta}_i$, é normal com vetor de média ζ_i e matriz de covariâncias dada pela inversa da matriz de informação

$$\mathbf{I}(\zeta_i) = -E\left(\frac{\partial^2 \log L(\zeta)}{\partial\zeta_i \partial\zeta_i'}\right) = -\boldsymbol{\Delta}(\zeta_i),$$

As raízes quadradas dos elementos diagonais de $[\mathbf{I}(\zeta_i)]^{-1}$ fornecem os erros-padrão dos estimadores $\widehat{a}_i$, $\widehat{b}_i$ e $\widehat{c}_i$.

Estimação das Habilidades

- Nesta etapa consideramos os parâmetros dos itens conhecidos.
- A habilidade do indivíduo j será estimada com base na distribuiçãoda habilidade, condicionada ao vetor de respostas do indivíduo j, ou seja, alguma característica de $g_j^*(\theta)$:

$$g_j^*(\theta) \equiv g(\theta|\mathbf{u}_{j.}, \zeta, \eta) = \frac{P(\mathbf{u}_{j.}|\theta, \zeta)g(\theta|\eta)}{P(\mathbf{u}_{j.}|\zeta, \eta)}$$

- Podemos usar a Média desta distribuição (chamado de EAP - expected a posteriori), ou o máximo (Moda) da distribuição (chamado de MAP - maximum a posteriori)

Estimação das Habilidades

- Nesta etapa consideramos os parâmetros dos itens conhecidos.
- A habilidade do indivíduo j será estimada com base na distribuiçãoda habilidade, condicionada ao vetor de respostas do indivíduo j, ou seja, alguma característica de $g_j^*(\theta)$:

$$g_j^*(\theta) \equiv g(\theta|\mathbf{u}_{j.}, \zeta, \boldsymbol{\eta}) = \frac{P(\mathbf{u}_{j.}|\theta, \zeta)g(\theta|\boldsymbol{\eta})}{P(\mathbf{u}_{j.}|\zeta, \boldsymbol{\eta})}$$

- Podemos usar a Média desta distribuição (chamado de EAP - expected a posteriori), ou o máximo (Moda) da distribuição (chamado de MAP - maximum a posteriori)

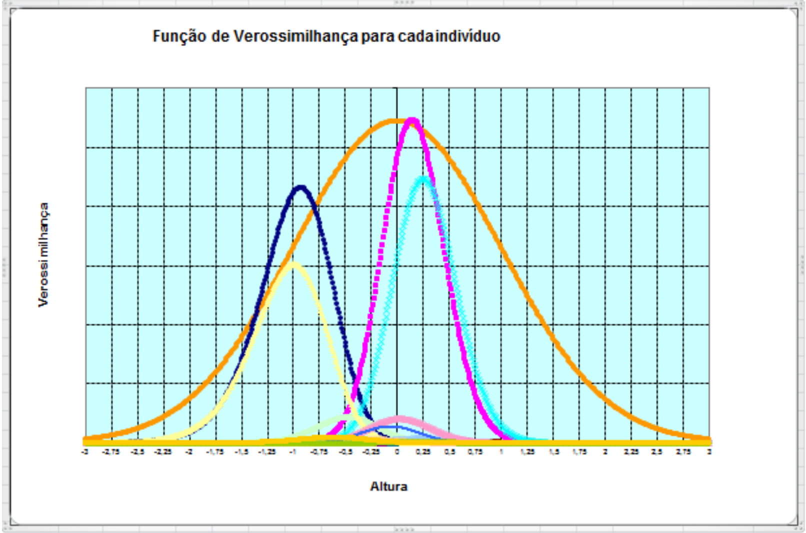
Estimação das Habilidades

- Nesta etapa consideramos os parâmetros dos itens conhecidos.
- A habilidade do indivíduo j será estimada com base na distribuiçãoda habilidade, condicionada ao vetor de respostas do indivíduo j, ou seja, alguma característica de $g_j^*(\theta)$:

$$g_j^*(\theta) \equiv g(\theta|\mathbf{u}_{j.}, \zeta, \boldsymbol{\eta}) = \frac{P(\mathbf{u}_{j.}|\theta, \zeta)g(\theta|\boldsymbol{\eta})}{P(\mathbf{u}_{j.}|\zeta, \boldsymbol{\eta})}$$

- Podemos usar a Média desta distribuição (chamado de EAP - expected a posteriori), ou o máximo (Moda) da distribuição (chamado de MAP - maximum a posteriori)

Estimaçao das Habilidades - Visualização



EAP - expected a posteriori

Este método é bastante conveniente, pois durante o processo de estimação dos parâmetros dos itens as quantidades $g_j^*(\theta)$ são obtidas e guardadas, assim como o vetor de pontos de quadratura $\overline{\boldsymbol{\theta}} = (\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_q)$, usado para aproximar as integrais por somas. Segue que a esperança da posteriori é

$$\widehat{\theta}_j \equiv E[\theta|\mathbf{u}_{j.}, \zeta, \boldsymbol{\eta}] = \int_{\Re} \theta g_j^*(\theta) d\theta.$$

MAP - maximum a posteriori

- Temos que encontrar o ponto de máximo de $g_j^*(\theta)$
- Logaritmizamos, derivamos e igualamos a zero, e aplicamos Newton-Raphson.
- A equação de estimação para θ_j é dada por

$$D\sum_{i=1}^{n} a_i(1-c_i)(u_{ji}-P_{ji})W_{ji} - \frac{(\theta_j-\mu)}{\sigma^2} = 0.$$

- O processo iterativo será feito com base em

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} - [H(\widehat{\theta}_j^{(t)})]^{-1} h(\widehat{\theta}_j^{(t)})$$

- O erro-padrão será dado pelo inverso da informação de Fisher.

MAP - maximum a posteriori

- Temos que encontrar o ponto de máximo de $g_j^*(\theta)$
- Logaritmizamos, derivamos e igualamos a zero, e aplicamos Newton-Raphson.
- A equação de estimação para θ_j é dada por

$$D\sum_{i=1}^{n} a_i(1-c_i)(u_{ji}-P_{ji})W_{ji}-\frac{(\theta_j-\mu)}{\sigma^2}=0.$$

- O processo iterativo será feito com base em

$$\widehat{\theta}_j^{(t+1)}=\widehat{\theta}_j^{(t)}-[H(\widehat{\theta}_j^{(t)})]^{-1}h(\widehat{\theta}_j^{(t)})$$

- O erro-padrão será dado pelo inverso da informação de Fisher.

MAP - maximum a posteriori

- Temos que encontrar o ponto de máximo de $g_j^*(\theta)$
- Logaritmizamos, derivamos e igualamos a zero, e aplicamos Newton-Raphson.
- A equação de estimação para θ_j é dada por

$$D\sum_{i=1}^{n} a_i(1-c_i)(u_{ji}-P_{ji})W_{ji} - \frac{(\theta_j-\mu)}{\sigma^2} = 0.$$

- O processo iterativo será feito com base em

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} - [H(\widehat{\theta}_j^{(t)})]^{-1} h(\widehat{\theta}_j^{(t)})$$

- O erro-padrão será dado pelo inverso da informação de Fisher.

MAP - maximum a posteriori

- Temos que encontrar o ponto de máximo de $g_j^*(\theta)$
- Logaritmizamos, derivamos e igualamos a zero, e aplicamos Newton-Raphson.
- A equação de estimação para θ_j é dada por

$$D\sum_{i=1}^{n} a_i(1-c_i)(u_{ji}-P_{ji})W_{ji} - \frac{(\theta_j-\mu)}{\sigma^2} = 0.$$

- O processo iterativo será feito com base em

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} - [H(\widehat{\theta}_j^{(t)})]^{-1} h(\widehat{\theta}_j^{(t)})$$

- O erro-padrão será dado pelo inverso da informação de Fisher.

MAP - maximum a posteriori

- Temos que encontrar o ponto de máximo de $g_j^*(\theta)$
- Logaritmizamos, derivamos e igualamos a zero, e aplicamos Newton-Raphson.
- A equação de estimação para θ_j é dada por

$$D\sum_{i=1}^{n} a_i(1-c_i)(u_{ji}-P_{ji})W_{ji} - \frac{(\theta_j-\mu)}{\sigma^2} = 0.$$

- O processo iterativo será feito com base em

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} - [H(\widehat{\theta}_j^{(t)})]^{-1} h(\widehat{\theta}_j^{(t)})$$

- O erro-padrão será dado pelo inverso da informação de Fisher.

Principais Aplicativos (Softwares)

- Análise Clássica: ItemAN (www.assess.com)
- Teoria da Resposta ao Item: Bilog-MG, para itens dicotômicos (www.ssicentral.com)
- Teoria da Resposta ao Item: Multilog, para itens policotômicos
- Análise de Fatores Associados: HLM (Hierarquical Linear Models)
- R-Project: ltm, mirt, Lorfif (http://cran.r-project.org/web/packages/ltm/ltm.pdf)
- Dimensionalidade / Análise Fatorial de Informação Plena: TestFact.
- Xcalibre etc.

Principais Aplicativos (Softwares)

- Análise Clássica: ItemAN (www.assess.com)
- Teoria da Resposta ao Item: Bilog-MG, para itens dicotômicos (www.ssicentral.com)
- Teoria da Resposta ao Item: Multilog, para itens policotômicos
- Análise de Fatores Associados: HLM (Hierarquical Linear Models)
- R-Project: ltm, mirt, Lorfif (http://cran.r-project.org/web/packages/ltm/ltm.pdf)
- Dimensionalidade / Análise Fatorial de Informação Plena: TestFact.
- Xcalibre etc.

Principais Aplicativos (Softwares)

- Análise Clássica: ItemAN (www.assess.com)
- Teoria da Resposta ao Item: Bilog-MG, para itens dicotômicos (www.ssicentral.com)
- Teoria da Resposta ao Item: Multilog, para itens policotômicos
- Análise de Fatores Associados: HLM (Hierarquical Linear Models)
- R-Project: ltm, mirt, Lorfif (http://cran.r-project.org/web/packages/ltm/ltm.pdf)
- Dimensionalidade / Análise Fatorial de Informação Plena: TestFact.
- Xcalibre etc.

Principais Aplicativos (Softwares)

- Análise Clássica: ItemAN (www.assess.com)
- Teoria da Resposta ao Item: Bilog-MG, para itens dicotômicos (www.ssicentral.com)
- Teoria da Resposta ao Item: Multilog, para itens policotômicos
- Análise de Fatores Associados: HLM (Hierarquical Linear Models)
- R-Project: ltm, mirt, Lorfif (http://cran.r-project.org/web/packages/ltm/ltm.pdf)
- Dimensionalidade / Análise Fatorial de Informação Plena: TestFact.
- Xcalibre etc.

Principais Aplicativos (Softwares)

- Análise Clássica: ItemAN (www.assess.com)
- Teoria da Resposta ao Item: Bilog-MG, para itens dicotômicos (www.ssicentral.com)
- Teoria da Resposta ao Item: Multilog, para itens policotômicos
- Análise de Fatores Associados: HLM (Hierarquical Linear Models)
- R-Project: ltm, mirt, Lorfif (http://cran.r-project.org/web/packages/ltm/ltm.pdf)
- Dimensionalidade / Análise Fatorial de Informação Plena: TestFact.
- Xcalibre etc.

Principais Aplicativos (Softwares)

- Análise Clássica: ItemAN (www.assess.com)
- Teoria da Resposta ao Item: Bilog-MG, para itens dicotômicos (www.ssicentral.com)
- Teoria da Resposta ao Item: Multilog, para itens policotômicos
- Análise de Fatores Associados: HLM (Hierarquical Linear Models)
- R-Project: ltm, mirt, Lorfif (http://cran.r-project.org/web/packages/ltm/ltm.pdf)
- Dimensionalidade / Análise Fatorial de Informação Plena: TestFact.
- Xcalibre etc.

Principais Aplicativos (Softwares)

- Análise Clássica: ItemAN (www.assess.com)
- Teoria da Resposta ao Item: Bilog-MG, para itens dicotômicos (www.ssicentral.com)
- Teoria da Resposta ao Item: Multilog, para itens policotômicos
- Análise de Fatores Associados: HLM (Hierarquical Linear Models)
- R-Project: ltm, mirt, Lorfif (http://cran.r-project.org/web/packages/ltm/ltm.pdf)
- Dimensionalidade / Análise Fatorial de Informação Plena: TestFact.
- Xcalibre etc.

Algumas especificidades: Duas provas, Redação no Enem, o SiSu etc.

- Na TC, podemos controlar para que a escala dos escores seja a mesma: [0,45] no Enem, por exemplo. Na TRI isso não é realizado propositadamente;
- Provas diferentes levarão a limites (mínimo e máximo) distintos;
- A Nota da Redação tem escala [0,1000], sem média 500, com massa relevante nos extremos;
- O Inep repassa às IES as 5 notas (LCT, MTC, CHT, CNT, Redação). Cada IES decide se usa, e como usa as 5 notas, dando as ponderações que desejar [Autonomia].

Algumas especificidades: Duas provas, Redação no Enem, o SiSu etc.

- Na TC, podemos controlar para que a escala dos escores seja a mesma: [0,45] no Enem, por exemplo. Na TRI isso não é realizado propositadamente;
- Provas diferentes levarão a limites (mínimo e máximo) distintos;
- A Nota da Redação tem escala [0,1000], sem média 500, com massa relevante nos extremos;
- O Inep repassa às IES as 5 notas (LCT, MTC, CHT, CNT, Redação). Cada IES decide se usa, e como usa as 5 notas, dando as ponderações que desejar [Autonomia].

Algumas especificidades: Duas provas, Redação no Enem, o SiSu etc.

- Na TC, podemos controlar para que a escala dos escores seja a mesma: [0,45] no Enem, por exemplo. Na TRI isso não é realizado propositadamente;
- Provas diferentes levarão a limites (mínimo e máximo) distintos;
- A Nota da Redação tem escala [0,1000], sem média 500, com massa relevante nos extremos;
- O Inep repassa às IES as 5 notas (LCT, MTC, CHT, CNT, Redação). Cada IES decide se usa, e como usa as 5 notas, dando as ponderações que desejar [Autonomia].

Algumas especificidades: Duas provas, Redação no Enem, o SiSu etc.

- Na TC, podemos controlar para que a escala dos escores seja a mesma: [0,45] no Enem, por exemplo. Na TRI isso não é realizado propositadamente;
- Provas diferentes levarão a limites (mínimo e máximo) distintos;
- A Nota da Redação tem escala [0,1000], sem média 500, com massa relevante nos extremos;
- O Inep repassa às IES as 5 notas (LCT, MTC, CHT, CNT, Redação). Cada IES decide se usa, e como usa as 5 notas, dando as ponderações que desejar [Autonomia].

Principais Desafios da Estatística para Sistemas de Avaliação

- Métodos e Precisão das Estimativas: correção termos $O(n^{-k})$, uso Multivariado (LC, MT, CNT, CHT).
- Comparabilidade envolvendo Redação
- Análise de Fatores Associados ao desempenho
- ...

Principais Desafios da Estatística para Sistemas de Avaliação

- Métodos e Precisão das Estimativas: correção termos $O(n^{-k})$, uso Multivariado (LC, MT, CNT, CHT).
- Comparabilidade envolvendo Redação
- Análise de Fatores Associados ao desempenho
- ...

Principais Desafios da Estatística para Sistemas de Avaliação

- Métodos e Precisão das Estimativas: correção termos $O(n^{-k})$, uso Multivariado (LC, MT, CNT, CHT).
- Comparabilidade envolvendo Redação
- Análise de Fatores Associados ao desempenho
- ...

Principais Desafios da Estatística para Sistemas de Avaliação

- Métodos e Precisão das Estimativas: correção termos $O(n^{-k})$, uso Multivariado (LC, MT, CNT, CHT).
- Comparabilidade envolvendo Redação
- Análise de Fatores Associados ao desempenho
- ...

Referências Bibliográficas

(1) Andrade, D.F., Tavares, H.R., Valle, R.C. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. Associação Brasileira de Estatística: São Paulo. (disponível em www.ufpa.br/heliton)

(2) Baker, F. B., & Kim, S. H. (2004). Item response theory: Parameter estimation techniques (2nd ed.). New York: Marcel Dekker.

(3) Fox, J.-P. (2010). Bayesian Item Response Modeling: Theory and Applications New York: Springer.

(4) De Boeck, P. Azevedo, C. L. N., Tavares, H. R. (2011). Linear and Nonlinear Generalized Mixed Models: Inference and Applications. Associação Brasileira de Estatística: Fortaleza.

Obrigado!

Héliton Tavares
heliton@ufpa.br

Material disponível em www.ufpa.br/heliton