

INTRODUÇÃO A PSICOMETRIA

TÓPICOS

1. Contexto
2. Que é Psicometria
3. Como é a construção de um instrumento psicométrico?
4. Que mede um teste?
5. Como é determinado a qualidade dos instrumentos psicométricos?
6. Quais são os conceitos básicos da medição?
7. Que é um modelo de medição?

1. Contexto

	Profissão A	Ciência B
Metodologias	Enfoque Quantitativo	Paradigma quantitativo
Profissão	Consultoria Estatística	Psicometria
Ciência	Novos paradigmas quantitativos	Psicologia Matemática

Cuadro 1. Aplicações da Estatística em Psicologia

Que é Psicometria?

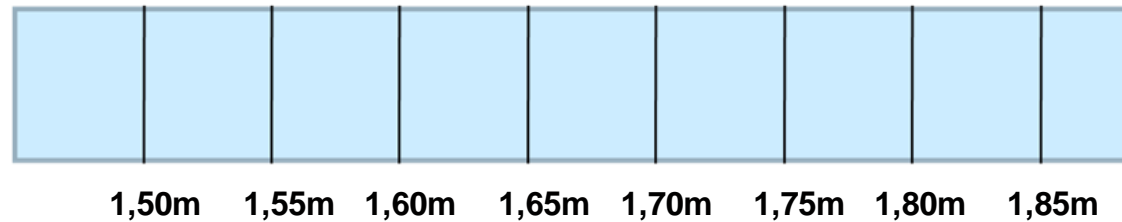
- É o campo de estudo relacionado com a teoria e técnica da medição psicológica, incluindo a medição de conhecimentos, habilidades, atitudes e traços de personalidade e a medição educacional.
- A área está principalmente associada com a construção e validação de instrumentos de medição, como questionários, provas, escalas, inventários e testes, entre outros.
- A psicometria tem dois tarefas de pesquisa principais:
 - a) A construção de instrumentos e procedimentos de medição, e
 - b) O desenvolvimento e aperfeiçoamento de abordagens teóricas para a medição.

- Profissionais da psicologia, psicometristas são científicos envolvidos no planejamento do teste para tentar medir diferentes características humanas.
- A área sofreu um rápido crescimento desde a sua criação. Os testes psicométricos são utilizados em escolas, organizações, empresas, governos, forças armadas, e, claro, em ambientes hospitalares e clínicos.
- Todos os expertos em psicometria devem ter pelo menos um Mestrado, e a maioria tem Doutorado.
- Por causa de que a psicometria é considerada uma área da psicologia, uma licenciatura em Psicologia não é incomum como formação previa.

- Os graduados em Psicometria costumam trabalhar nos departamentos de Psicologia, mas não é infrequente encontrar muitos especialistas com uma graduação em Estatística.
- De acordo com um recente artigo no Journal Washington Monthly, psicometristas (muitas vezes chamado de "test makers") estão em grande demanda.
- Cada vez são mais requeridos testes, e não há especialistas suficientes em psicometria para atender a demanda.
- Em uma sociedade com uma cultura de medição, qualquer psicólogo especialista em psicometria não deve ter dificuldade em encontrar emprego.

3. Como é a construção de um instrumento psicométrico?

"Brincando com a altura" (*)



(*) Prof CAW Glas - University of Twente - Holanda
ABE - SINAPE 2006.

Questionário para medir altura: alguns itens

1. Na cama, eu frequentemente sinto frio nos pés.
2. Eu frequentemente desço as escadas de dois em dois degraus.
3. Eu acho que me daria bem em um time de basquete.
4. Como policial, eu impressionaria muito.
5. Na maioria dos carros eu me sinto desconfortável.
6. Eu literalmente olho para meus colegas de cima para baixo.
7. Você é capaz de pegar um objeto no alto de um armário, sem usar escada?
8. Você abaixa quando vai passar por uma porta?
9. Você consegue guardar a bagagem no porta-malas do avião?
10. Você regulava o banco do carro para trás?
11. Normalmente quando você está andando de carona lhe oferecem o banco da frente?

12. Quando você e várias pessoas vão tirar fotos, formando-se três fileiras, onde ninguém ficará agachado, você costuma ficar atrás?
13. Você tem dificuldade para se acomodar no ônibus?
14. Em uma fila, por ordem de tamanho, você é sempre colocado atrás?

Dicotômica: sim - não, verdadeiro ou falso, certo ou errado.

Politômicos: nunca, raramente, a metade do tempo, muitas vezes, sempre.



Posição de examinados e ítems na mesma escala

O que mede um teste?

- Um teste ou medida pode ser visto com um conjunto de questões de auto-relato (também chamado de "ítems"), cujas respostas são pontuadas e de alguma forma agregadas para obter uma pontuação composta.

- As características essenciais são: a) Uma série de perguntas as quais os indivíduos respondem; b) Um escore composto que surge a partir da pontuação das respostas para as perguntas.
- O conjunto resultante de perguntas é referido como uma "escala", "teste" ou "medida". Em geral, um instrumento psicométrico.
- Dois tipos de resultados estão disponíveis a partir dos itens, mas precisasse notar que o importante não é tanto o formato da pergunta se não o formato da resposta ou pontuação.
- Pontuações binárias (resposta dicotômicas), (a) os itens que estão qualificados como resposta *correta* ou *incorreta* em teste de desempenho (por exemplo, no caso de múltipla escolha), ou (b) itens que são

classificados dicotomicamente de acordo com um tipo de pontuação, ou escala de personalidade (*verdadeiro - falso, de acordo, em desacordo*).

- Respostas ao Item ordinais (respostas graduadas, Likert, tipo Likert, ou item politomicos); envolvendo mais de duas opções de pontuação tais, como uma escala de 5 pontos, *em total acordo até em total discordo* ou em uma escala de personalidade ou medida da atitude.
- As considerações de validade e confiabilidade dos instrumentos psicométricos pelo geral são vistas como elementos essenciais para determinar a qualidade de qualquer teste.

5. Como é determinada a qualidade dos instrumentos psicométricos?

- Associações Profissionais e usuários muitas vezes têm estas preocupações dentro de contextos mais amplos no desenvolvimento de critérios para avaliar a qualidade de qualquer teste num determinado contexto.
- *The Standards for Educational and Psychological Testing* (1999) é um conjunto de criterios de avaliação desenvolvidos pela American Educational Research Association (AERA), American Psychological Association (APA), e o National Council on Measurement in Education (NCME).

Construção de Testes, Avaliação e Documentação

1. Validade
2. Erros de medida e confiabilidade
3. Desenvolvimento do teste e revisão
4. Escalas, Normas e comparabilidade dos escores
5. Administração de teste, Qualificação e Relatórios
6. Documentação de apoio para os testes

Equidade dos Testes

1. Teste de equidade e uso do teste
2. Os direitos e as responsabilidades dos examinadores
3. Testes individuais de pessoas de diversa procedência linguística
4. Teste individuais para pessoas deficientes

Parte III: Aplicações de teste

1. As responsabilidades de usuários de teste
2. Avaliação e Medição Psicológica
3. Avaliação e Medição Educacional
4. Avaliação e Certificação do trabalho
5. Teste de Avaliação de Programas e Políticas Públicas

- No país utilizasse as normas de 1954 e não 1999.
- A nova versão está prevista para 2012.
- Há uma necessidade de reformulação de disciplinas em Estatística e Psicologia, por exemplo de Teoria da Resposta ao Item, Variáveis latentes, Estatísticas na Psicologia, Medição Psicológica, Construção de Testes, Psicometria, etc.

Quais são os princípios da medição?

- Se você quiser medir o quanto de habilidade uma pessoa tem, você deve ter uma escala de medição, ou seja, uma regra com uma métrica.
- Esta regra deve ser utilizada para determinar que capacidade uma determinada pessoa tem.
- A aproximação habitual é definir uma medida da capacidade e desenvolver um teste que consiste num determinado número de itens sob a definição (perguntas).
- Cada um desses itens mede alguma faceta de uma particular habilidade de interesse.

- Assumisse que cada examinado que responde a um item de um teste tem certa quantidade da capacidade subjacente.
- Assim, podemos considerar que cada examinando tem um valor numérico, denotado por θ que toma o lugar da sua posição na escala de habilidade.

Que é um modelo de medição?

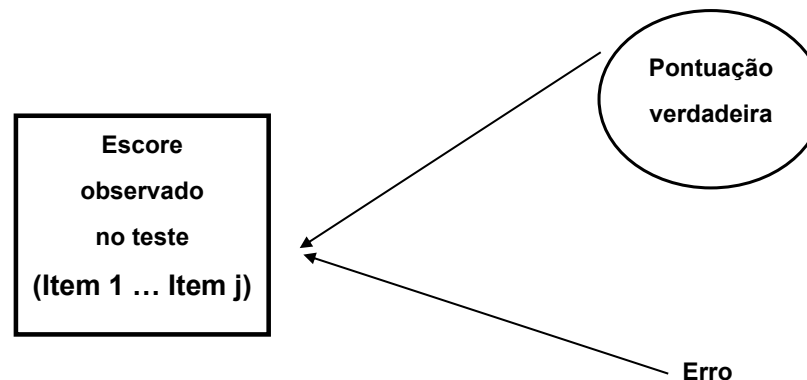
Utilizado para relacionar as variáveis observadas, registradas e medidas (respostas aos itens) com as variáveis latentes (habilidade).

- Modelo de Teoria Clássica
- Modelo de Resposta ao Item

Estes modelos não são os únicos, mas são os mais consolidados.

7.1 Modelo dos Testes clássicos ou Teoria Clássica dos Testes

Ela expressa uma relação linear entre o verdadeiro valor de habilidade e o escore de habilidade observado.



$$\text{Escore observado} = \text{pontuação verdadeira} + \text{erro}$$

- O resultado do teste o escore de linha é a soma das pontuações recebidas sobre os itens do teste.

- Tradicionalmente, a teoria da medição foi estabelecida baseado numa análise de escala- ou de nível do teste baseado em métodos de correlação.
- Os resultados são, é claro, não segmentados (ou seja, você não tem ideia de como uma pessoa com determinado valor no teste executa a um nível particular de habilidade), há uma única e simples medida geral do desempenho.
- A principal ferramenta estatística é o ANOVA dos efeitos aleatórios, ou análise de componentes de variância, cujo principal objetivo é medir a quantidade de erro na medida.
- O Alfa de Cronbach é visto como uma medida apropriada neste contexto.

- Esta teoria é válida para qualquer formato de pontuação dos Ítems. É aplicado tanto para itens dicotômicos quanto para itens politômicos o qualquer subtipo.

altura.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : puntaje 10 Visible: 18 de 18

	sujeto	i01	i02	i03	i04	i05	i06	i07	i08	i09	i10	i11	i12	i13	i14	puntaje
1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	10
2	2	1	1	1	1	1	1	1	1	1	1	1	0	1	1	13
3	3	1	1	1	0	1	0	0	1	0	1	0	0	0	1	7
4	4	1	0	0	1	1	0	1	0	1	1	0	0	1	0	7
5	5	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
6	6	1	0	1	1	1	1	1	1	1	1	1	1	1	1	13
7	7	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
8	8	1	1	0	1	0	0	1	1	1	1	1	1	1	1	11
9	9	0	1	1	1	1	1	1	1	0	1	1	1	1	1	12
10	10	1	0	1	1	1	0	1	0	1	0	1	0	1	1	9
11	11	1	0	0	1	1	0	0	1	1	1	1	0	1	1	9
12	12	1	1	0	1	0	1	1	1	1	1	1	1	1	1	12
13	13	1	1	1	1	1	0	1	1	1	1	1	0	1	1	12
14	14	1	0	0	1	1	1	0	1	0	1	0	0	1	1	8
15	15	1	1	1	1	1	1	1	1	1	0	1	1	1	1	13
16	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
17	17	0	1	1	1	1	0	1	0	1	1	1	1	0	1	10
18	18	1	1	0	1	1	0	1	1	1	0	1	0	1	1	10
19	19	1	0	0	1	1	0	1	1	1	1	1	0	0	1	9

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
i01	10.05	3.090	.143	.471
i02	9.98	3.261	.057	.490
i03	10.32	2.804	.240	.442
i04	9.91	3.161	.259	.450
i05	9.97	3.168	.147	.468
i06	10.47	2.990	.141	.475
i07	9.92	3.170	.230	.454
i08	9.96	2.975	.331	.426
i09	10.05	3.236	.036	.500
i10	9.98	3.130	.171	.463
i11	9.91	3.053	.384	.428
i12	10.49	2.929	.183	.461
i13	10.02	3.092	.164	.464
i14	9.89	3.358	.060	.483

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
.481	14

Recursos - TCT

- Softwares
- Pacotes estatísticos (Excel, SPSS, SAS)
- ITEMAN (disponível a partir de <http://www.assess.com/xcart/home.php?cat=18>)

Leitura

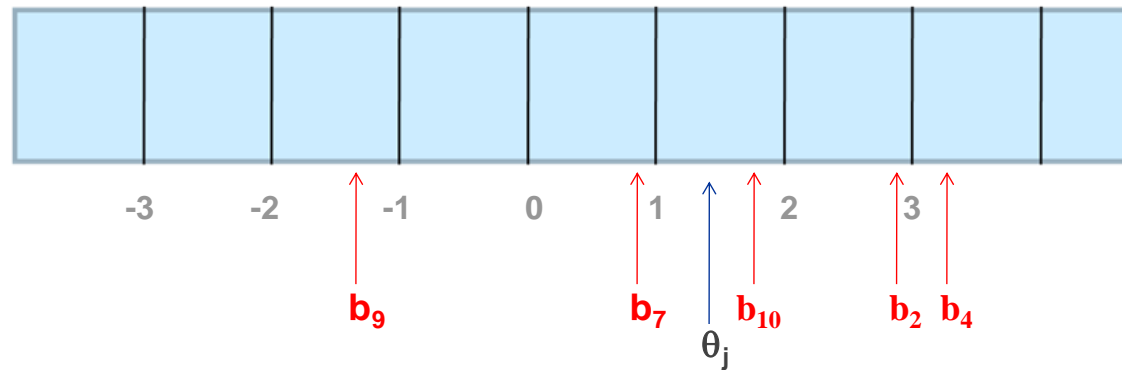
Matlock-Hetzel (1997) *Basic Concepts in Item and Test Analysis* available at www.ericae.net/ft/tamu/Espy.htm

Todos os livros das bibliotecas de psicologia atuais em Espanhol são traduzidas via Mexico, argentina e Colômbia.

7.2 Modelo de Resposta ao item

- De acordo com a teoria de resposta ao item, o interesse primário está em saber se o examinando tem um determinado item correto ou não, ao invés de saber a pontuação total
- Específica como o traço latente e as características do item estão relacionados com as respostas das pessoas aos itens.
- Modelo mais simples: Modelo Rasch

P (resposta correta item) = function {nível de habilidade, dificuldade do item}



Posição de examinados e itens numa mesma escala

θ_j : Traço latente do examinando (parâmetro da pessoa: "habilidade")

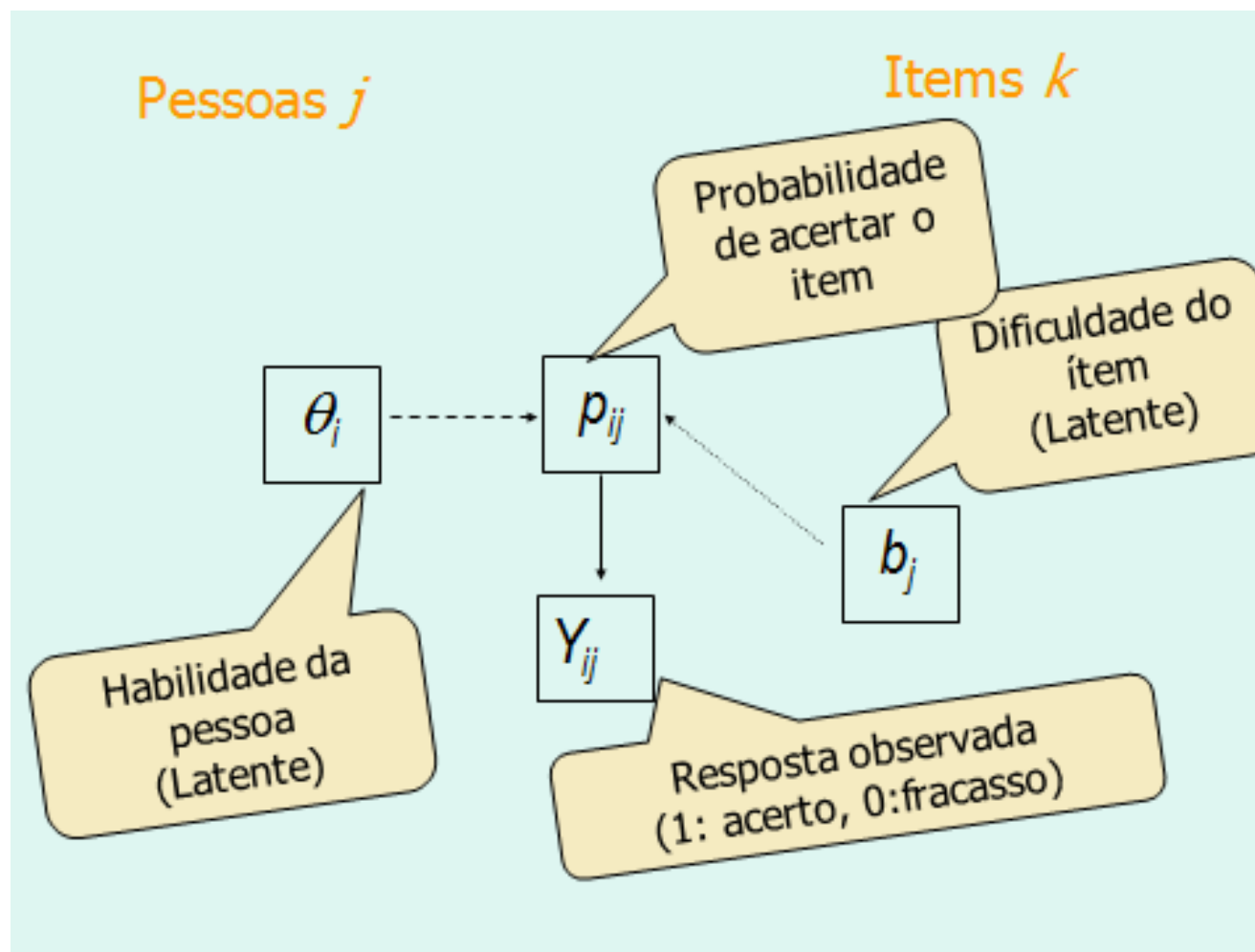
b_i : Dificuldade do item "traço latente" (parâmetro do item)

> 0 : Examinado está "acima" do item

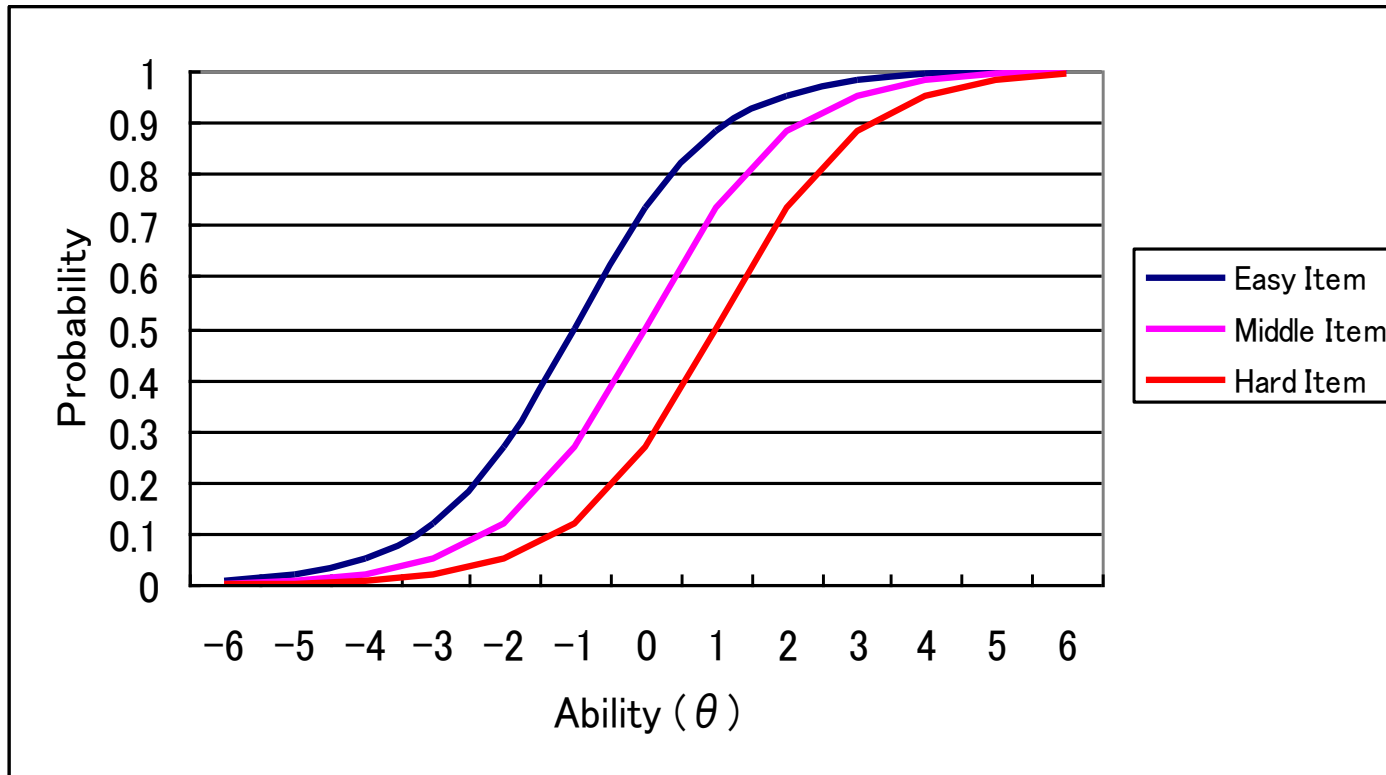
$(\theta_j - b_i) \approx 0$: É considerado "próximo" do item

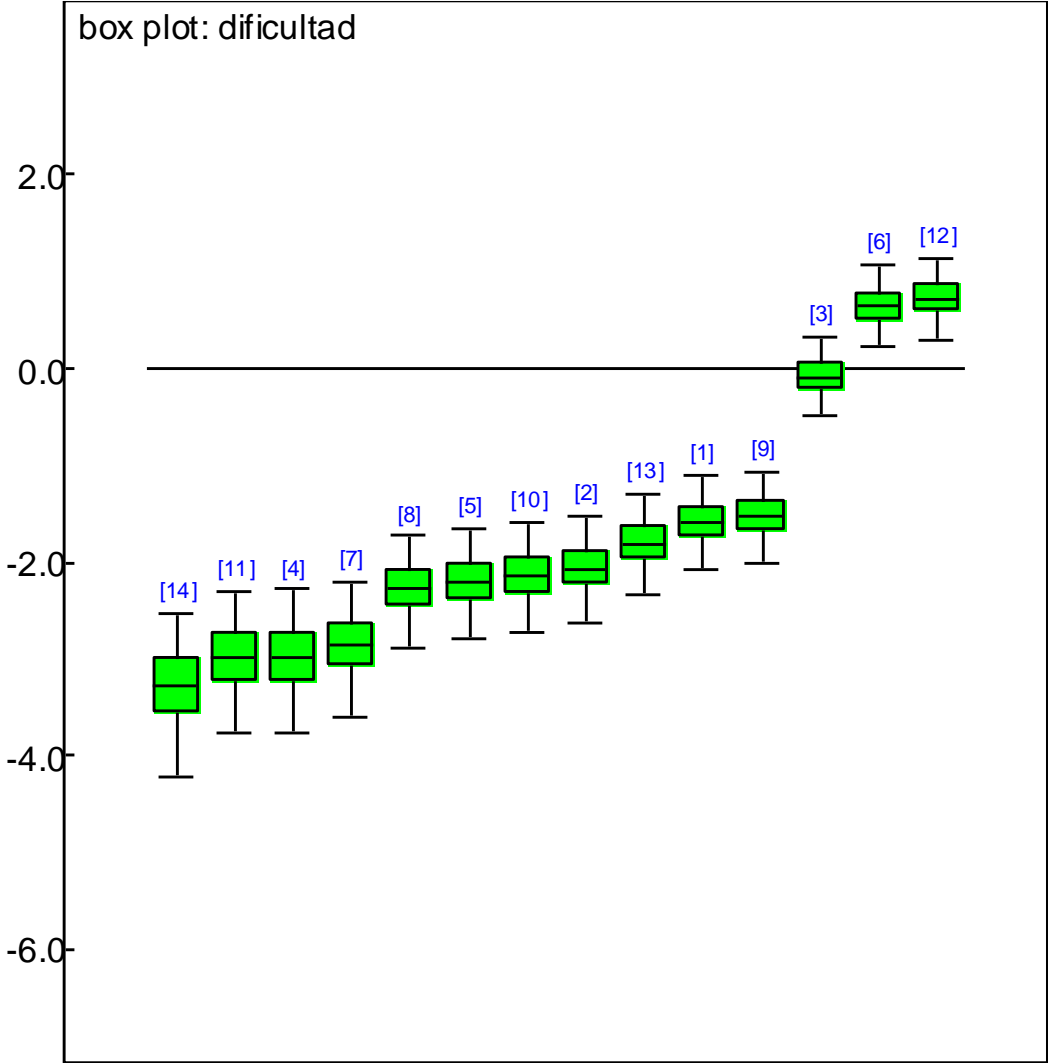
< 0 : É considerado "abaixo" do item

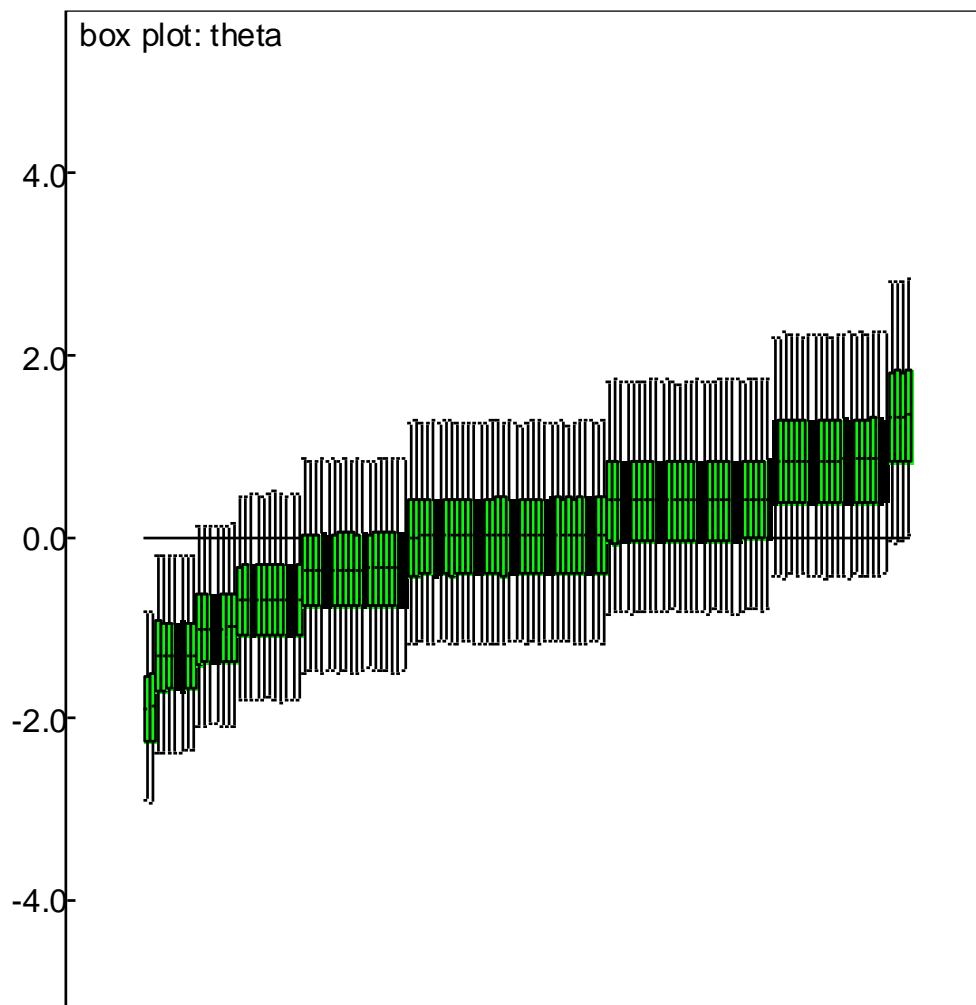
- Baseado nas respostas dos itens de um teste desejasse estimar: parâmetros dos Itens (calibração); Traços latentes dos examinados; Parâmetros da população (distribuição dos traços latentes): média, desvio padrão, etc.
- A probabilidade de uma resposta "correta" para um item é modelada como função da habilidade do examinando e os parâmetros do item.



Curvas características dos itens







Software Psicométrico

- Winstep
- Rascal
- Bilog
- Conquest
- Quest
- Winmira
- RUMM2020
- Param3PL
- Logimo
- MSP
- LPCM-WIN
- RSP
- T-Rasch

- ICL-WIN
- LEM
- Multilog
- Xcalibret

Estatísticos

- SAS
- R
- Stata
- WinBUGS
- Systat
- OpenStat

Leituras

1. Baker, Frank (2001). The Basics of Item Response Theory available at <http://ericae.net/irt/baker/>
2. Baker, F. and Kim, S. (2004). Item Response Theory: Parameter Estimation Techniques . Marcel Dekker Inc. New York.
3. Bond, T.G and Fox, C.M (2001). Applying the Rasch Model: Fundamental Measurement in the Human Sciences Lawrence Erlbaum Associates.
4. De Boeck, P., & Wilson, M. (Eds.) (2004). Explanatory Item Response Models. A Generalized Linear and Nonlinear Approach. New York: Springer.

5. Embretson, S. and Reise, S. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum.
6. Fox, J.-P. (2010). Bayesian Item Response Modeling: Theory and Applications New York: Springer.
7. Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). Fundamentals of item response theory. Newbury Park: Sage.
8. Lord (1980) Applications of Item Response Theory to Practical Testing Problems.
9. McDonald, R. P. (1999). Test theory: A unified approach. Mahwah, NJ: Lawrence Erlbaum.

10. Thissen, D., & Wainer, H. (Eds.). (2001). Test Scoring. Mahwah, NJ: Lawrence Erlbaum.

11. Van der Linden, W.J. & Hambleton, R.K. (Eds.) (1997). Handbook of modern item response theory. New York: Springer.

IRT ou TRI

- <http://edres.org/irt/>
- <http://work.psych.uiuc.edu/irt/tutorial.asp>
- http://psychcentral.com/psypsych/Item_response_theory

Software e Livros

- <http://www.ssicentral.com>
- <http://www.assess.com>

7.3 TCT vs IRT

Modelo dos testes clássicos	Modelo de resposta ao item
O modelo é expresso a nível de teste	O modelo é expresso a nível do item
As características do Item são dependentes da amostra	As características do item são independentes da amostra (invariância do item)
Estimados da habilidade dependem dos itens	Estimativas da habilidade independente dos itens (Invariância das pessoas)
O mesmo erro de medição para todos examinados	O erro de medição é para cada nível de habilidade
Teste mais longos são mais confiáveis do que os testes mais curtos	Pequenos testes podem ser mais confiáveis do que testes longos

- Embora as bases teóricas da TRI aconteceram entre 1950 e 1960, os métodos não foram amplamente utilizados até os 70, devido à complexidade na estimativa.
- TCT é usado ainda mais a aproximação TRI está cada vez mais predominante.
- A TRI trata novos problemas como multidimensionalidade, diferenciabilidade, testes de adaptação, teste de velocidades (speedeness), testlet, testes longitudinais, equalizações.
- Há uma extensa bibliografia cada vez mais na TRI, como software livre e comercial.