Tópicos em Modelos de Resposta ao Item

Um minicurso no Programa de Verão em Estatística 2017
ICMC/USP e UFSCar
São Carlos
Janeiro 30 a Fevereiro 03

Jorge Luis Bazán
Universidade de São Paulo
jlbazan@icmc.usp.br

http://conteudo.icmc.usp.br/pessoas/jlbazan/

DIA 30/01/2017
SALA: 3-012
HORÁRIO: 14hs a 16hs

AULA 1. INTRODUÇÃO: AVALIAÇÃO, PSICOMETRIA, MODELOS DE MEDIÇÃO (Expositiva)

Conteúdo:
Avaliação. O que é Psicometria?. Principais métodos de medição: Teoria Clássica de Testes (TCT) e Teoria de Resposta ao Item (TRI). Diferencias entre estes modelos de medição e exemplos de avaliações e dados no minicurso

Ministrante: Jorge Luís Bazán

TÓPICOS

1. AVALIAÇÃO
2. PSICOMETRIA
3. MODELOS DE MEDIÇÃO
4. MODELOS DE TESTES CLASSICOS O TEORIA CLÁSSICA DOS TESTES
5. MODELOS DE RESPOSTA AO ITEM
6. MODELOS DE TCT VS MODELOS TRI
7. EXEMPLOS DE AVALIAÇÕES E DADOS NO MINICURSO

1. AVALIAÇÃO

Avaliação é importante para

- Melhora dos processos dos governos federais, estaduais e municipais
- Processos da administração e governabilidade
- Melhora da eficiência e eficácia (serviços, setores produtivos, mercados, competitividade)
- Melhora na qualidade do sistema educacional
- Melhores critérios de seleção e avaliação de estudantes e profissionais,

.....

A avaliação tem a ver com:

- Desenvolvimento de instrumentos (escalas, questionários e suas propriedades)
- Melhor definição dos propósitos de pesquisa (objetivos e resultados)
- Melhores modelos matemáticos e estatísticos
- Melhores sistemas de computo para bases de dados, analises e aplicações de provas
- Desenvolvimento de critérios para propor políticas usando os resultados da avaliação

Isto induz ao desenvolvimento metodológico da avaliação e a um debate ao respeito de sua aplicabilidade.

- No Brasil e no mundo, poucos conhecem todos aspectos técnicos da avaliação pois isto requere conhecimentos de várias profissões.

- Neste minicurso a ênfases é nos modelos estatísticos e então vamos a assumir que existe um conhecimento apropriado das implicâncias da avaliação.

- Nós faremos certo ênfases na avaliação educacional e também na avaliação psicológica porém os modelos que serão apresentados no minicurso não se restringem nestes ambitos.

1.1 O que é avaliação educacional?

O processo de avaliação educacional está relacionado à produção de informações sobre o aprendizado. Isto é algo que está bastante presente no cotidiano escolar e na educação superior: usualmente, os professores aferem o aprendizado dos seus alunos através de diversos instrumentos (observações, questionários, escalas, listas, registros, provas etc.) e indicam, a partir daí, o que precisa ser feito para que seus alunos possam avançar no sistema escolar.

1.2 O que é avaliação em larga escala?

Nas últimas décadas, junto com às avaliações tradicionais na salas de aula, outro tipo de avaliação educacional tem ganhado espaço: são as avaliações externas, geralmente em larga escala, isto é, aplicada simultaneamente a grandes amostras ou censos em forma padronizada incluindo as vezes alunos professores, diretores e coordenadores. Exemplos ENEM, SAEB, SAEM, PISA, Provinha Brasil, Prova ENADE, etc.

Estas avaliações têm objetivos e procedimentos diferenciados das avaliações tradicionais de salas de aula. Por exemplo para certificação, credenciamento, diagnóstico, ranking e rendição de contas.

1.3 O que se avalia nos testes educacionais?

Em geral as avaliações individuais privilegiam o sistema de avaliação cognitiva não em tanto outros sistemas da personalidade podem ser avaliados embora eles sejam menos discutidos na avaliação educacional.

Os sistemas da personalidade são mostrados na figura 1

Ortiz P.El Sistema de la Personalidad. Lima: Orión; 1994.
Ortiz P.La formación de la Personalidad. Lima: Dimaso Editores; 1997

Posteriormente proporcionaremos mais exemplos.

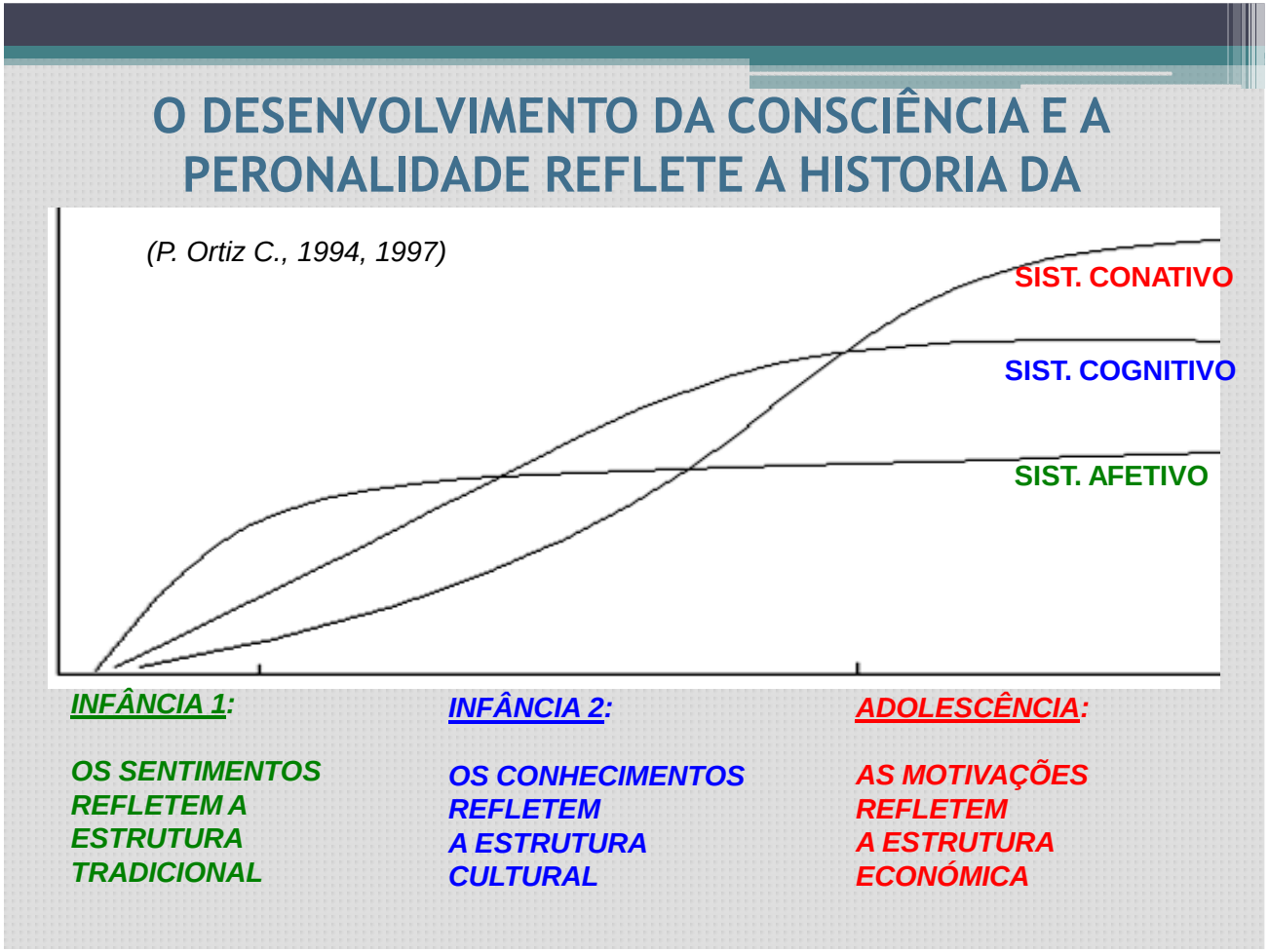


Figura 1. Sistemas da Personalidade (Ortiz, 1994,1997)

2. PSICOMETRIA

2.1 Contexto

	PSICOLOGIA	
ESTATÍSTICA	Profissão A	Ciência B
Metodologias para I	Enfoque Quantitativo	Paradigma quantitativo
Profissão II	Consultoria Estatística	Psicometria
Ciência	Novos paradigmas quantitativos	Psicologia Matemática

Figura 2. Aplicações da Estatística em Psicologia

A Psicometria fica caracterizada como o uso da profissão do Estatística para a Ciencia Psicologica.

2.2 O 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 psicometría tem dois tarefas de pesquisa principais:

(i) A construção de instrumentos e procedimentos de medição, e

(Ii) o desenvolvimento e aperfeiçoamento de abordagens teóricas para a medição.

2.3 O que é Psicometrista?

- Psicométras ou psicometristas (muitas vezes chamado de "test makers") estão em grande demanda.

http://weusemath.org/?career=psychometrician

- Os psicometristas são científicos envolvidos no planejamento do teste para tentar medir diferentes características humanas.

- Todos os expertos em psicometría devem ter pelo menos um mestrado preferentemente em Medição educacional, psicologia organizacional, matemática ou campos relacionados com uma experiência relevante e treinamento. Um Doutorado na área é altamente desejável.

- Por causa de que a Psicometría 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 e infrequente encontrar muitos especialistas trabalhando em departamentos de Matemática, Estatística ou Computação.

- 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.

- Cada vez são mais requeridos testes, e não há especialistas suficientes em psicometría para atender a demanda. Assim, numa sociedade com uma cultura de medição, qualquer psicólogo especialista em psicometria não deve ter dificuldade em encontrar emprego.

2.4 Como é construído um instrumento psicométrico?

"Brincando com a altura (*)

Figura 3. Uma escala de 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?

Formatos itens:

Dicotômica: Sim - Não, verdadeiro ou falso, certo ou errado.

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

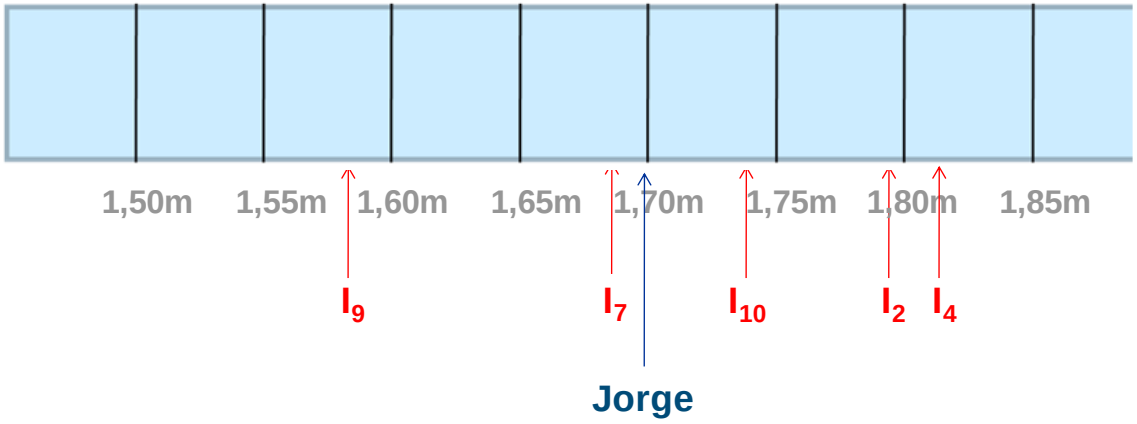


Figura 4. Posição de examinados e itens na mesma escala

2.5 O que mede um teste?

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

- As características essenciais são:
  - Uma série de perguntas ou itens as quais os indivíduos respondem
  - 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, chamamos isto de um instrumento psicométrico.

Podemos obter dois possíveis tipo de resultados dos itens considerando o formato da *resposta da pergunta* (não o formato da pergunta). Estes são chamados de pontuações ou escores:

- Pontuações dicotômicas (binárias), (a) os itens que estão qualificados como resposta *correta* ou *incorreta* em teste de rendimento(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 (ie, *verdadeiro - falso, de acordo com - em desacordo).*

- Pontuações politômicas (múltiplas), respostas ordinais (respostas graduadas, tipo Likert, frequência, intensidade, etc); envolvendo mais de duas opções de pontuação.
  Exemplo: pontuação de 5 pontos,

  (1) Discordo totalmente

  (2) Discordo

  (3) Não concordo nem discordo

  (4) De acordo

  (5) Totalmente de acordo

  relativa a uma questão ao respeito de sua personalidade ou sua atitude frente a determinado objeto.

2.6 Como é determinada a qualidade dos instrumentos psicométricos?

- As considerações de validez e confiabilidade dos instrumentos psicométricos são vistos como elementos essenciais para determinar a qualidade de qualquer teste.

- Associações Profissionais e usuários muitas vezes se preocupam da qualidade dos instrumentos e desenvolvem critérios para avaliar a qualidade de qualquer teste num determinado contexto.

Para dar uma ideia disso, falaremos das Normas para avaliação educacional e psicológica desenvolvida por duas importantes instituições dos Estados Unidos.

- ***The Standards for Educational and Psychological Testing*** 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).** Edições de 1954, 1999, 2014.

Quadro 1. Tópicos para avaliar a qualidade dos instrumentos psicométricos APA, AERA, NCME (1999)

Construção de Testes, Avaliação e Documentação	Validez
	Erros de medida e confiabilidade
	Desenvolvimento de teste e revisão.
	Escalas, Normas e comparabilidade dos escores
	Administração de teste, Qualificação e Relatórios
	Documentação de apoio para os testes
Equivalência dos Testes	Teste de equidade e uso do teste
	Os direitos e as responsabilidades dos examinadores.
	Testes individuais de pessoas de diversa procedência linguística
	Testes individuais para pessoas deficientes
Aplicações dos testes	As responsabilidades de usuários de teste
	Avaliação e Medição Psicológica
	Avaliação e Medição Educacional
	Avaliação e Certificação do trabalho
	Teste de Avaliação de Programas e Políticas Públicas

Quadro 2. Tópicos para avaliar a qualidade dos instrumentos psicométricos APA, AERA, NCME (2014)

Fundamentos	Validez
	Confiabilidade/precisão e erros de medida.
	Equidade em Testes.
Operatividade	Planejamento de testes e desenvolvimento
	Escores, escalas, normas, escores de ligação e escores de corte.
	Administração dos testes, qualificação, reporte e interpretação.
	Documentação de suporte para teste
	Os direitos e responsabilidades dos desenvolvedores de testes.
	Os direitos e responsabilidades dos usuários de testes.
Aplicações dos Testes	Avaliação e Testes psicológicos
	Testes do mundo laboral e credenciamento.
	Avaliação e testes educacionais.

- As considerações de validade e confiabilidade dos instrumentos psicométricos pelo geral são vistos como elementos essenciais para determinar a qualidade de qualquer teste.

- A *confiabilidade* ou fidedignidade é uma característica da medida que faz referência ao grau de consistência ou reprodutibilidade das medidas quando os procedimentos das avaliações são replicados sob as mesmas condições.

- A *validez* da medida faz referência ao grau pelo qual a evidência e a teoria suportam as interpretações a partir dos valores das medidas

- Quais normas são utilizadas no pais?

- Considerando as normas vigentes, existe uma necessidade de reformulação de disciplinas em Estatística e Psicologia, por exemplo de Teoria da Resposta ao Item, Variáveis latentes, Estatística em Psicologia, Medição Psicológica, Construção de Testes, Psicometria, etc

- Existem normas mais especializadas para outros ambitos
  http://www.fsassessments.org/
  https://www2.ed.gov/admins/lead/account/saa.html

3. MODELOS DE MEDIÇÃO

3.1 Quais são os princípios de medição?

- Se 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 prática 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.

- Assume-se 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 que toma o lugar da sua posição na escala de habilidade.

- Para elaborar tais testes para medir uma determinada característica de interesse nós podemos considerar como marco metodológico uma versão adaptada da proposta de Duckor, Draney e Wilson (2009) também discutido em Wilson (2005).

Estes autores apresentam uma proposta para a construção de medidas com base em quatro etapas e princípios do sistema de avaliação, os quais são apresentados na Figura 5.

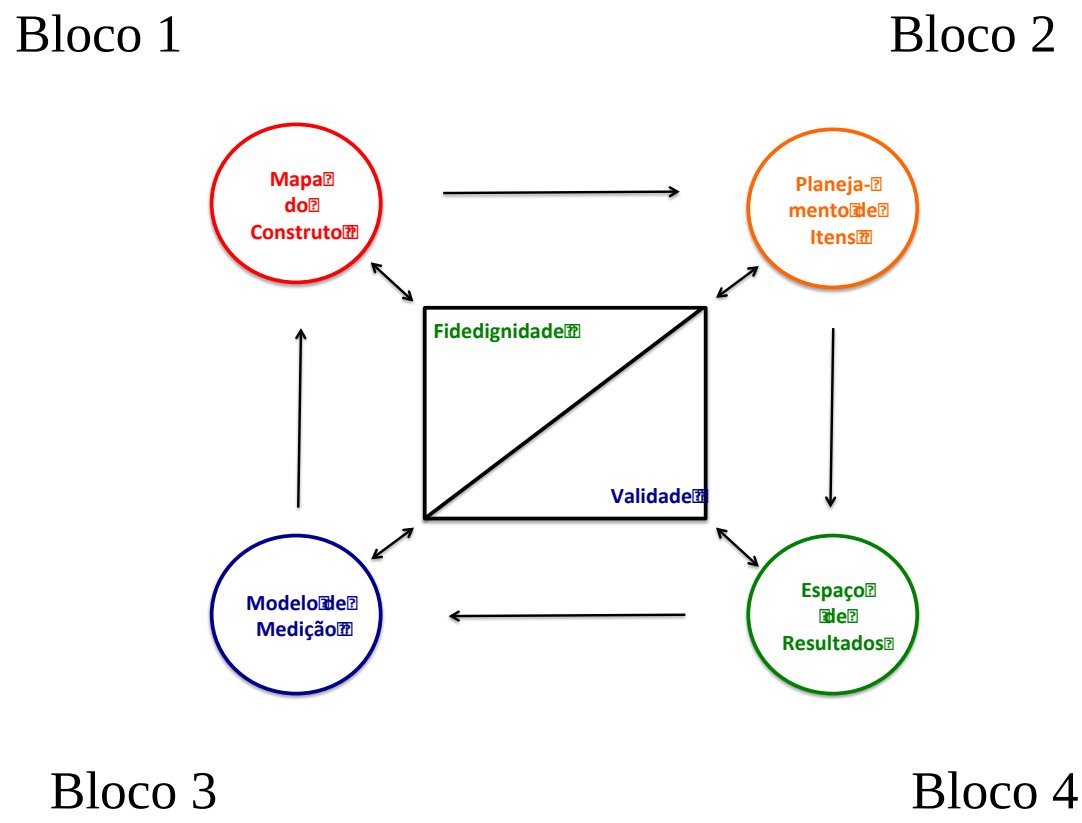


Figura 5. Relações entre os quatro blocos para a construção de medidas (tomado de Duckor, Draney e Wilson, 2009).

Modelagem do constructo é uma estratégia para o desenvolvimento de um instrumento usando cada um dos quatro tijolos da construção:

- Mapa do constructo
- Plano para o desenvolvimento dos itens
- Espaço dos resultados
- Modelo de medição

3.2 O que é o modelo de medição?

- Nós destacamos neste trabalho o modelo de medição, o qual é utilizado para relacionar as variáveis observadas, registradas e medidas (respostas aos itens) com as variáveis latentes ou não observadas (habilidade) o construto a serem medido.

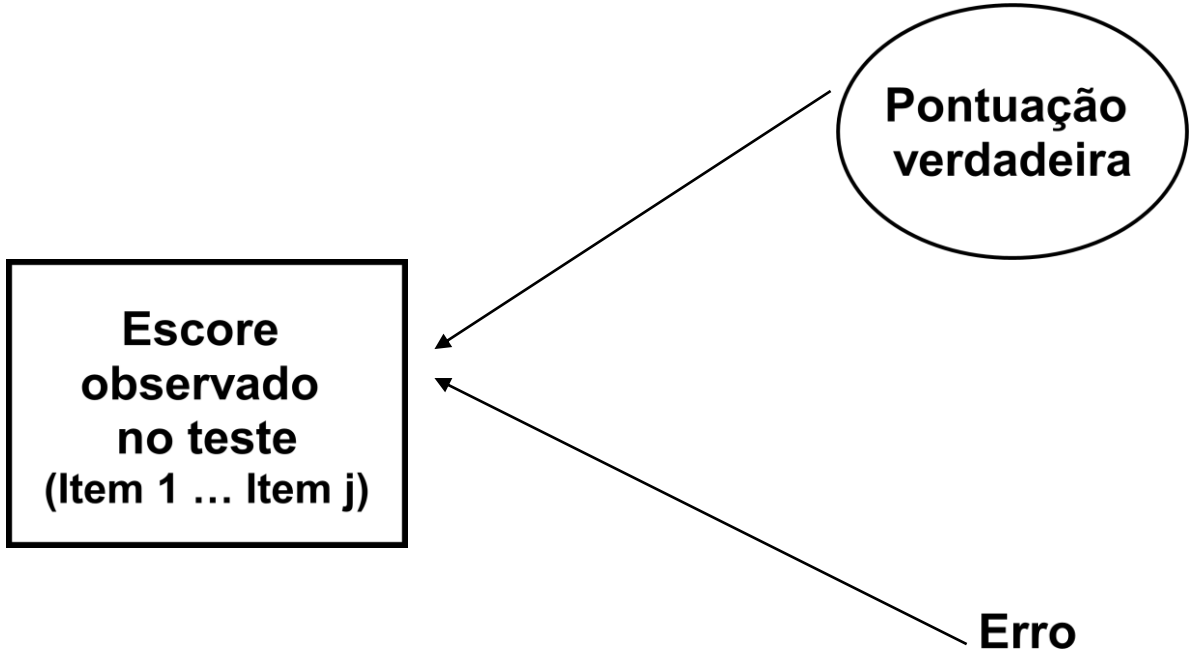
- Existem dois modelos de medição que se destacam: O “Modelo Clássico dos Testes” e o “Modelo de Resposta ao Item”.

- Embora estes modelos não sejam os únicos, eles são as mais consolidadas.
- Uma recente abordagem são os chamados Modelos de Diagnostico Cognitivo. Eles dominaram o último Conbratri http://conbratri.org/

4. MODELOS DE TESTES CLASICOS OU TEORIA CLASSICA DOS TESTES

- O Modelo clássico dos testes, chamada Teoria Clássica dos testes (TCT) é um enfoque da psicometria que prediz as respostas dos testes tais como a dificuldade dos itens ou a habilidade dos respondentes sendo o principal propósito a compreensão e melhora da confiabilidade dos testes.
- Ela também é considerada como sinônimo da teoria do escore verdadeiro como foi formulada por Spearman em 1904 sendo posteriormente sistematizada em Novick (1966) é descrita no clássico livro de Lord e Novick (1968).
- A TCT se baseia em três ideias principais: reconhecimento da presença de erros de medida, a concepção de que o erro é uma variável aleatória e por último a concepção de que através de uma determinada medida de correlação, isto é através de uma medida da associação entre o valor verdadeiro e o valor observado, é possível estimar a pontuação verdadeira.

- Especificamente se postula que existe uma relação linear entre o verdadeiro valor de habilidade e o escore de habilidade observado.



Escore observado = pontuação verdadeira + erro

Figura 6. Decomposição do escore observado na TCT

- O resultado do teste ou escore de linha é a soma das pontuações recebidas sobre os itens do teste.
- Tradicionalmente, a teoria da medição foi estabelecida baseado num analise da escala ou do 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.

- Uma ferramenta estatística usada é o ANOVA dos efeitos aleatórios, ou análise de componentes de variância, cujo principal objetivo é medir a quantidade de erro na medida associado com determinadas fontes.

- A principal ferramenta usada é um conjunto de índices que fazem parte dos *analises de itens*. Estes são: proporção de acerto, porcentagem de omissão, discriminação, correlação pergunta-prova, alfa de cronbanch se o item é desconsiderado, média e variância entre outros.

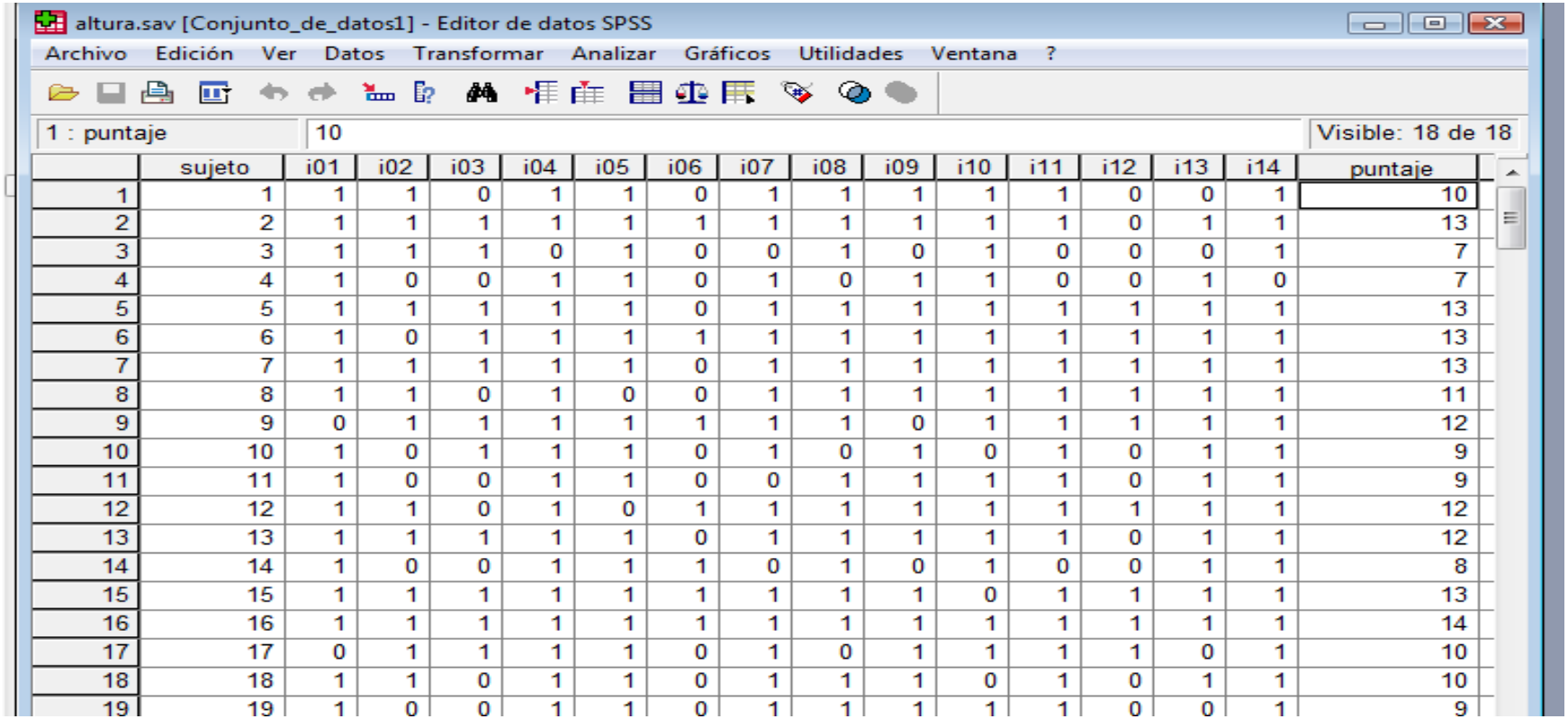
- Em geral procuramos testes com o menor número de itens que conservem as melhores propriedades estatísticas logo da analise de itens.

- Uma medida general de consistência interna da prova se baseia no índice alfa de Cronbach o qual é visto como una medida apropriada da confiabilidade do teste.

- Esta teoria é valida para qualquer formato de pontuação ou escore dos itens. É aplicado tanto para itens dicotômicos quanto para itens politômicos o qualquer subtipo destes ou para itens misturados.

- A TCT em principio não envolve um modelagem probabilístico pois não é assumida nenhuma distribuição para os dados observados logo da aplicação de um determinado teste numa amostra de pessoas.

- A TCT é formada por um conjunto de técnicas de classificação e é susceptível de ser enfocada desde a perspectiva de aprendizado de maquinas.



	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

Figura 7. Exemplo de uma matriz de respostas de um teste de 14 itens aplicado a 131 examinados

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 eleimina 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

Figura 8. Exemplo de analise de itens do teste de 14 itens.

O índice alfa de cronbach é definido por

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum_{i=1}^{k} \sigma_i^2}{\left(\sum_{i=1}^{k} \sigma_i \rho_{iX}\right)^2}\right]$$

onde k é o número de itens, σ_i^2 é a variância de cada item e ρ_{iX} é a correlação de cada item com o escore total do teste.

Nós desejamos testes com α perto de 1 e descartamos itens que não contribuam ou contribuam muito pouco neste valor.

Recursos - TCT

Software

- Pacotes estatísticos (Excel, SPSS, SAS, STATA, JMP, R)
- ITEMAN (disponível a partir de http://www.assess.com/iteman/?nabe=656643662571110 4:1)

Leitura

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

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Harcourt Brace.

5. MODELOS DE RESPUESTA AL ITEM

5.1. Os modelos TRI

Os modelos de Teoria de Resposta ao Item (TRI) desde uma perspectiva mais psicológica é um modelo de medição que defende que o interesse primário nos testes é saber se o examinando responde um determinado item corretamente ou não, ao invés de saber a pontuação ou escore total.

A ênfases agora é nas características dos itens para posteriormente obter a medida de interesse que é assumida latente isto é, não observável.

- A TRI especifica como o traço latente e as características do item estão relacionados com as respostas das pessoas aos itens a través de um determinado modelo probabilístico.

- Isto é, as respostas são assumidas aleatórias e então deve ser especificada uma determinada distribuição de probabilidades.
- As probabilidades ficam especificadas considerando parâmetros associados que precisam ser estimados considerando um processo de inferência.
- Os parâmetros são de dois tipos: associados com os indivíduos avaliados, e associados com os itens do teste.

Desde uma perspectiva estatística nos chamaremos os correspondentes modelos estatísticos para diferentes situações de resposta como Modelos de Resposta ao item. Este será a ênfase neste minicurso.

5.2. Modelo mas simples: modelo Rasch

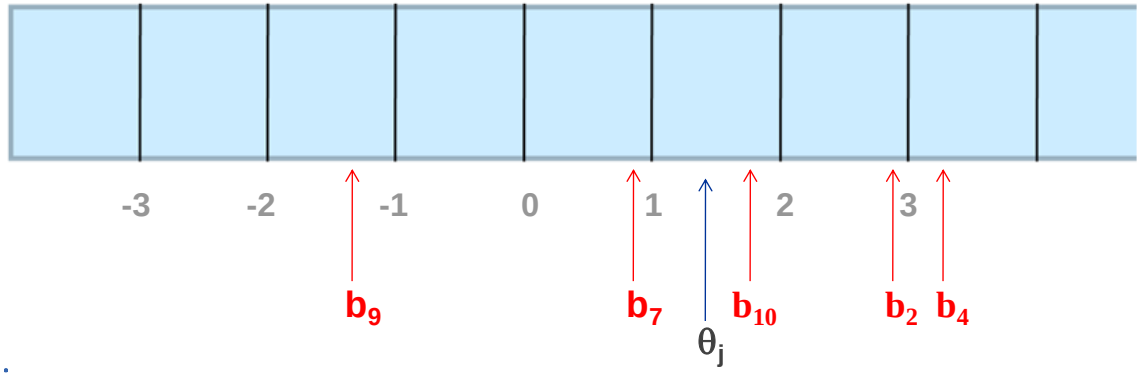
Resposta correta $\rightarrow$ Y=1
Resposta errada $\rightarrow$ Y=0

Y é aleatório! $\rightarrow$ Y $\sim$ Bernoulli (P)

P (resposta correta item) = F(parâmetro dos examinados, parâmetro dos itens)

- A probabilidade de uma resposta "correta" para um item é modelada como função dos parâmetros do examinando e dos itens.

Precisamos especificar esta função para definir os parâmetros de interesse



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

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

$(\theta_j - b_i) > 0$ examinado está "acima" do item $\rightarrow$ $\uparrow$ probabilidade de acertar

$(\theta_j - b_i) \approx 0$ é considerado" próximo "do item $\rightarrow$ 0.5 probabilidade de acertar

$(\theta_j - b_i) < 0$ é considerado "abaixo" do item $\rightarrow$ $\downarrow$ probabilidade de acertar

Figura 9. Interpretação da posição dos examinados e itens numa escala

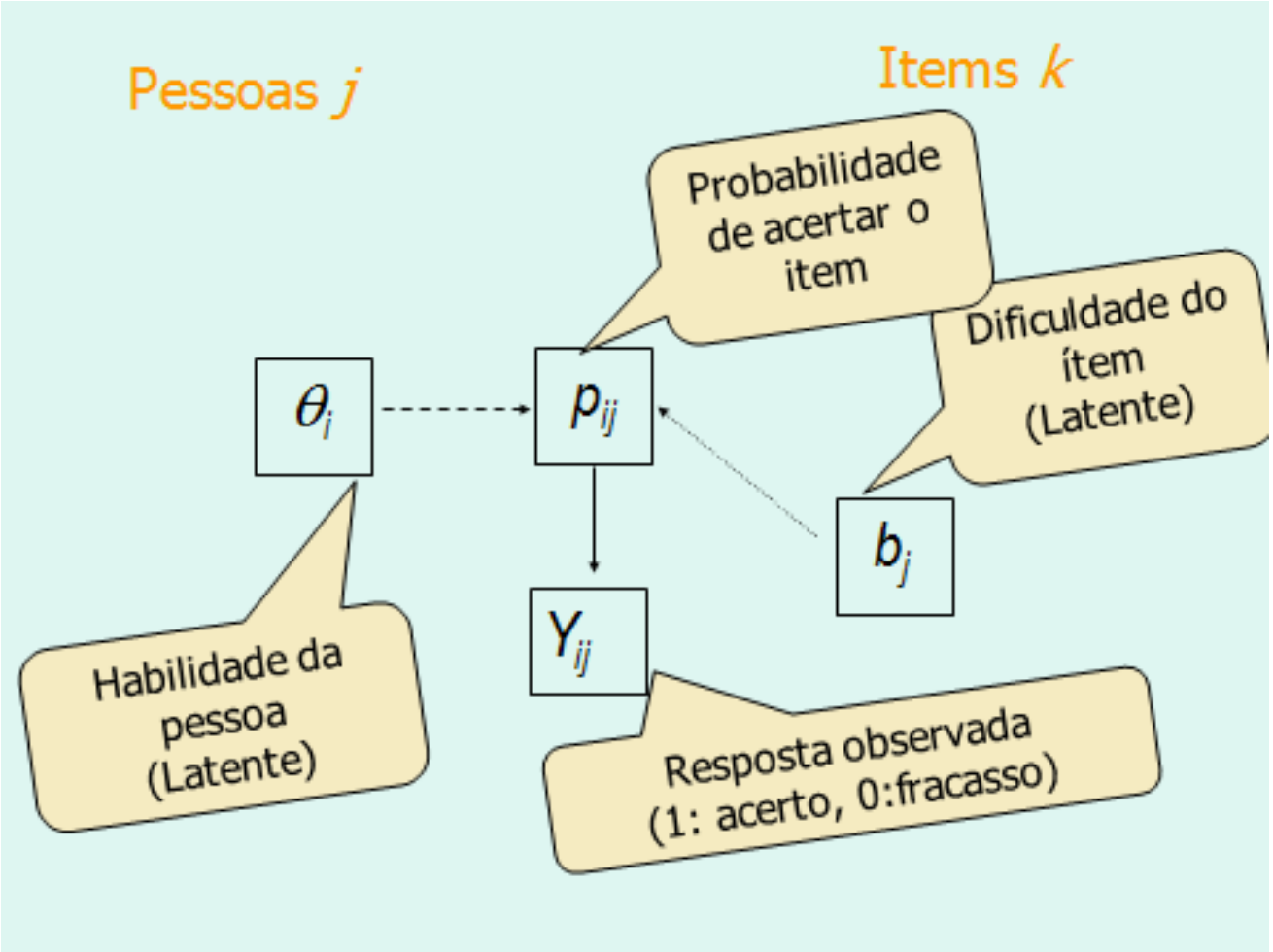


Figura 10. Os componentes do modelo de Rasch

- Baseado nas respostas dos itens de um teste a uma determinada amostra desejasse estimar:
  - parâmetros dos Itens ou dificuldades (chamada Etapa de calibração)
  - parâmetros dos indivíduos ou traços latentes dos examinados (chamada Etapa de estimação)
  - Parâmetros da população (distribuição dos traços latentes): média, desvio padrão, etc
    (chamada Etapa de caracterização)

    Na linguagem da área, precisamos calibrar os itens para estimar os traços latentes das pessoas e posteriormente caracterizar o grupo avaliado.

As curvas das probabilidades são chamadas funções de resposta ou curvas características dos itens. No caso Rasch tomam a seguinte forma:

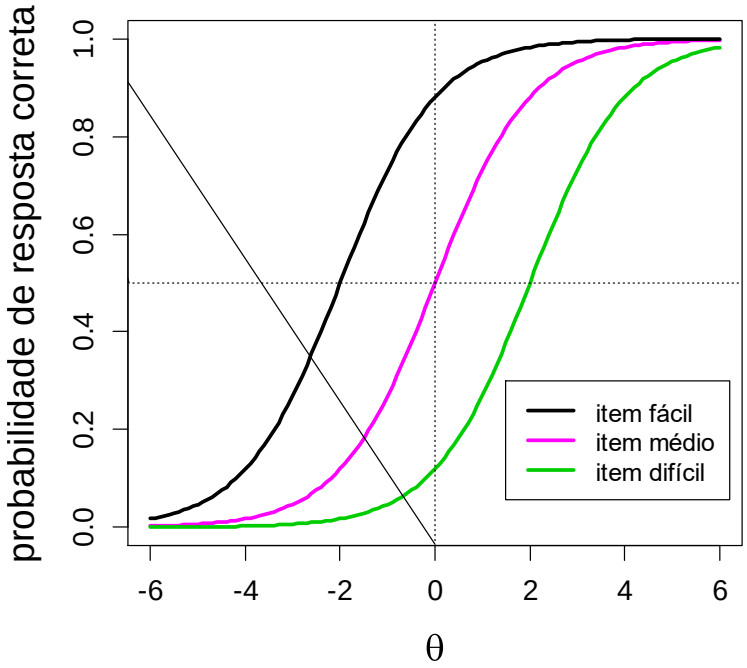


Figura 11. Curvas características de três itens com diferentes níveis de dificuldade: fácil (b=-2), médio (b=0) e difícil (b=2) no modelo de Rasch

- Como consequência dos resultados do processo de calibração e estimação nós conseguimos estimar os parâmetros dos itens (dificuldades) e os parâmetros dos examinados (habilidades). Os resultados são mostrados nas figuras 12 e 13.

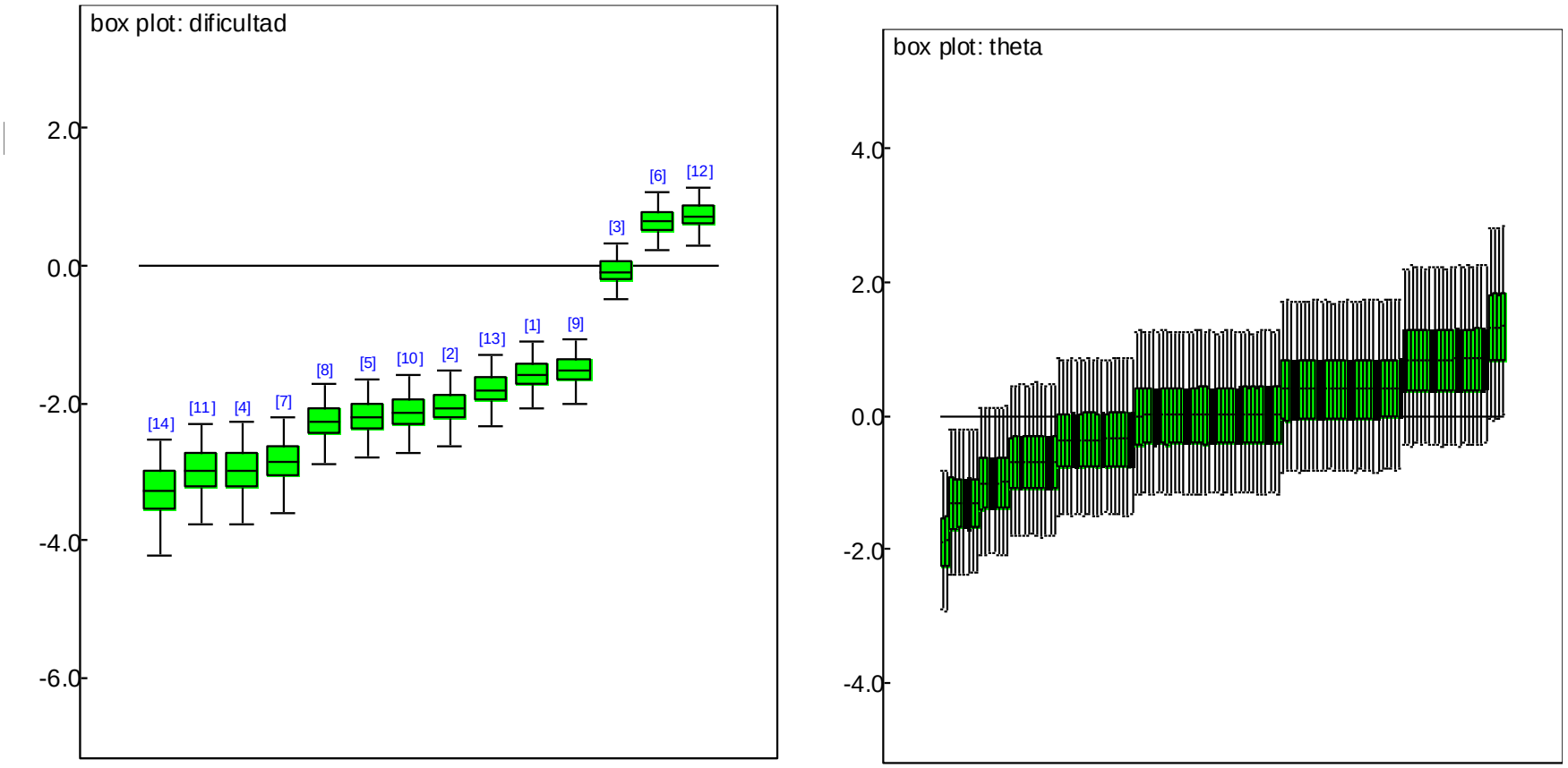


Figura 12. Estimativas dos parâmetros dos itens e dos examinados

5.3 Software Psicométrico

Embora as bases teóricas da TRI aconteceram entre 1950 e 1960 os métodos não foram amplamente utilizados até os anos de 1970, devido à complexidade dos cálculos que foi solucionada usando os computadores. Há uma extensa bibliografia cada vez maior sobre a TRI[1], assim como diferentes softwares livres e pagos.

[1] Algumas publicações relevantes na área são Baker e Kim. (2004). Bond, e Fox, C.M (2001), De Boeck, P., e Wilson, M. (Eds.) (2004). Embretson e Reise. (2000). Fox (2010). Hambleton, Swaminathan e Rogers (1991), Lord (1980) Van der Linden e Hambleton. (Eds.) (1997).

Pacotes específicos

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

- T-Rasch
- ICL-WIN
- LEM
- Multilog
- Xcalibret

Pacotes estatísticos

- SAS
- R
- Stata
- Systat
- OpenStat

Pacotes com programação intermédia

- WinBUGS
- JAGS
- OpenBUGS
- STAN
- INLA
- JULIA
- MATLAB

Pacotes no R

http://cran.r-project.org/web/views/Psychometrics.html

Classical Test Theory (CTT):

- CTT
- psychometric.
- cocron.
- CMC
- psy
- psych,
- psychometric
- ICC.
- QME

Item Response Theory (IRT):

- eRm
- ltm
- TAM
- mirt.
- MLCIRTwithin.
- IRTShiny.
- mcIRT
- sirt.
- pcIRT.
- kcirt.
- MultiLCIRT
- mRm
- psychomix
- mixRasch
- PP
- equateIRT

- kequate
- SNSequate
- EstCRM
- difR
- lordif
- DIFlasso.
- DFIT.
- difNLR
- catR
- mirtCAT
- plRasch
- pairwise
- lme4, nlme, MCMCglmm, ordinal
- irtrees.
- mokken
- fwdmsa
- KernSmoothIRT
- RaschSampler

- pwrRasch.
- irtProb
- cacIRT
- irtoys.
- VGAM.
- mlirt
- latdiag
- rpf
- classify
- WrightMap

Cognitive diagnostic Model (CDM):

- CDM
- GDINA

5.4 Livros, sites recomendados, compra de software especializado

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

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

- Thissen, D., & Wainer, H. (Eds.). (2001). Test Scoring. Mahwah, NJ: Lawrence Erlbaum.
- Van der Linden, W.J. & Hambleton, R.K. (Eds.) (1997). Handbook of modern item response theory. New York: Springer.

de Ayala, R. J. (2009). Theory and practice of item response theory. Guilford Press, New York, NY.

Sites

- 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

6. Modelos TCT vs Modelos de TRI

Quadro 3. Comparação entre os modelos de medição

Modelo dos testes clássico	**Modelo de resposta ao item**
O modelo é expresso ao 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 de Item)
Estimados da habilidade dependem dos itens	Estimativas da habilidade independente dos itens (Invariância de 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

Na prática, TCT é usado ainda, mas a aproximação TRI está cada vez mas predominante pois a TRI trata novos problemas como a multidimensionalidade das medidas, os possíveis viés das medidas usando o conceito de diferenciabilidade, os diferentes testes (testes de adaptação, teste de velocidades (speedeness), testlet, testes longitudinais), o problema de tornar diferentes medidas comparáveis usando processos de equalização ou equiparação, assim como uso de testes adaptativos computorizadas entre outros tópicos associados com a elaboração das medidas.

7. Exemplos de Avaliações e dados no minicurso

7.1 Exemplos

1. Escala Global de Atitudes frente a Estatística obtida de
Aparicio, A. (2015). AVALIAÇÃO DAS ATITUDES NO CURSO DE ESTATÍSTICA: CONTEXTOS UNIVERSITÁRIOS LATINO-AMERICANOS. Teses de doutorado. FEA USP.

Veja também

http://www.ime.unicamp.br/sinape/sites/default/files/sumissao%20de%20trabalho%20AparicioEstradaBazan%2019Sinape.pdf

2. Prova de conhecimentos em Matemática 6ta serie

Bazán, J. L, Branco, M. D. , Bolfarine, H. (2006). A skew item response model. Bayesian Analysis, 1 (2006), pp. 861–892. https://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-4/A-skew-item-response-model/10.1214/06-BA128.full

Prova de conhecimentos para 6ta Série

Leia com atenção cada questão e responda marcando com uma X sua resposta.

1. Em que alternativa os seguintes números estão ordenados do maior ao menor?

A) 567; 756; 765	B) 756; 765; 567	C) 765; 567; 756	D) 765, 756; 567

2. Indica a desigualdade correta.

A) 1/2 > 3/4	B) 7/6 <1/2	C) 3/4 > 7/6	D) 1/2<11/4

3. Um metro de pano custa S.I. 65. Quanto será pago por 0,5 metro?

A) S/. 302,50 B) S/. 121,00 C) S/. 30,25 D) S/. 30,05

4. Pepe dividiu um número entre 17, obtendo-se um quociente de 9 e um residuo de 2. Qual é o número?

A) 155 B) 171 C) 187 D) 306

5. Ao fazer a divisão: 960 ÷ 87 o quociente e o residuo obtido é?

A) quociente: 12; residuo: 16
B) quociente: 11; residuo: 13
C) quociente: 11; residuo: 3
D) quociente: 3; residuo: 11

6. O preço de uma blusa é S /. 30. Se Ana comprou com 20% de desconto, quanto pagou pela blusa?

A) S/. 50 B) S/. 24 C) S/. 20 D) S/. 6

7. Faça a seguinte operação de frações: 2/3 + 3/4

A) 5/7 B) 17/7 C) 6/12 D) 17/12

8. Pela compra de 100 litros de vinho paga-se S /. 1200. Quanto será pago por 200 litros?

A) S/.1 200 B) S/.1 400 C) S/. 1 500 D) S/. 2 400

9. Resolva as seguintes operações com decimais: **0,75 - 0,2 + 1,2 - 0,30**

A) 2,45 B) 2,05 C) 1,45 D) 0,45

10. Se o lado de um quadrado é de 3 cm, qual é seu perímetro?

A) 3 cm B) 6 cm C) 9 cm D) 12 cm

11. Luisa, Dora e Maria compraram pano. Luisa comprou médio metro, Dora comprou 75 cm y Maria comprou 50 cm. Quais delas compraram a mesma quantidade de pano?

A) Luisa e Dora	B) Dora e María	C) Luisa e María	D) Ninguem

12. Um tanque recebe 4,5 litros de água por minuto. Quantos litros de água vai ter o tanque em uma hora e meia?

A) 4050 litros	B) 405 litros	C) 7,2 litros	D) 6,75 litros

13. Qual das seguintes figuras têm retas paralelas?

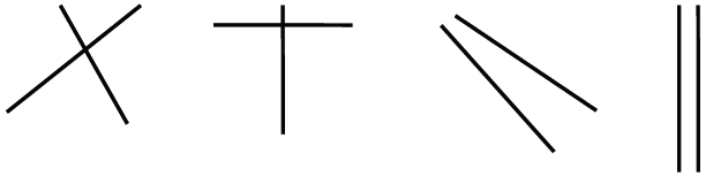
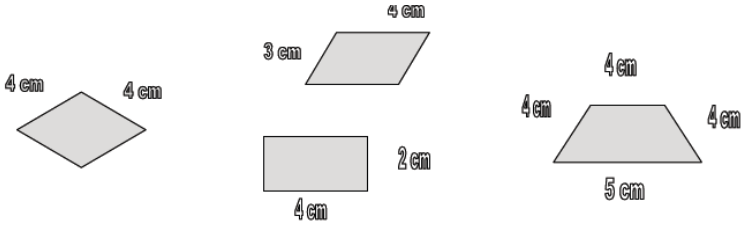


Figura 1 Figura 2 Figura 3 Figura 4

A) A figura	B) A figura	C) A figura	D) A figura

4 3 2 1

14. Observe as seguintes figuras.



Qual é a soma de todos os lados do losango?

A) 17 cm B) 16 cm C) 14 cm D) 11 cm

3. Escala para medir o amor

Sternberg, R. J. (1997). Construct validation of a triangular love scale. *European Journal of Social Psychology, 27*(3), 313-335.
http://vivanautics.com/pdf/Sternberg1997.pdf

Ver também a versão em português
http://www.redalyc.org/pdf/261/26118733005.pdf

7.3 Dados usados no Minicurso

https://jorgeluisbazan.weebly.com/presentations.html

Dichotomous response: Proficiency in Math of Peruvian students Example

MathB.csv

Dataset: Response pattern obtained by the application of a mathematical test to sixt-grade students of the rural Peruvian elementary schools. 131 students to 14 items qualified as binary responses (correct or incorrect). Data are described in Bazán, J. L, Branco, M. D. , Bolfarine, H. (2006). A skew item response model. Bayesian Analysis, 1 (2006), pp. 861–892.

Download

Polytomous Response: latent trait in the Attitudes toward Statistics of

scholar-teachers Example
No usuarios - Atitude2.csv
Dataset: Response pattern obtained by the application of the Scale of Attitudes toward Statistics to 146 scholar-teacher from Peru and Spain. Data contain answers of 146 respondent to 25 items using a Lickert scale which are described in Estrada, A., Bazán, J. and Aparicio, A. (2010). *Un estudio comparativo de las actitudes hacia* la *estadística en profesores españoles y peruanos*. Unión, 24, 45-56.

Download

Multidimensional Dichotomous response: latent traits in the Beck Depression Inventory of College students Example
bdi Univ - Tengdico.csv
Dataset: Response pattern obtained by the application of the Portuguese version of the Beck Depression Inventory (BDI) to 1111 college students gently given by Dr. Teng Chei-Tung from the Hospital
das Clínicas, Faculdade de Medicina, Universidade de Sao Paulo. Data

contain answers of 1111 respondent to 21 items of BDI which are described in Fragoso, T. M. and Cúri, M. (2013).

Download'