

Teorias de Avaliação - CE095

Adilson dos Anjos¹

¹Departamento de Estatística
Universidade Federal do Paraná
aanjos@ufpr.br

Curitiba, PR
15 de março de 2016

Teorias de Avaliação

–Teoria Clássica dos Testes–

Introdução

- Foco no **escore total** do indivíduo: acertos e erros;
- É a metodologia mais utilizada e conhecida; (a TRI ainda é recente e mais complexa);
- Objetivo: medir um **Traço Latente** (variável contínua);
- Existe uma associação entre **Traço latente** *vs* Escore;

O que estamos medindo?

- Traço latente: Característica **não observável**.
- Cumulativo: Quanto maior o traço latente do respondente, maior a probabilidade dele “responder afirmativamente” ao item.
- Não Cumulativo (modelos de desdobramento: *unfolding*): não estudado na disciplina.

Qual a medida?

- Escore ou número de acertos;
- Escala: um número inteiro entre 0 e I ($I =$ número de Itens).

Escala

- É possível construir uma escala.
- Interpretação: especialistas de cada área.

Modelo:

$$X_i = V_i + Erro_i$$

em que, X_i = score observado do i-ésimo indivíduo;

V_i = score verdadeiro do i-ésimo indivíduo

...

... e o "Erro" ?

- De onde surge?
- Como avaliar?
 - ✓ realizar vários testes
- $E(\text{Erro}_i) = 0$
(considerando que os erros são aleatórios).

Os resultados são **dependentes**:

- dos itens ou questões que compõem a prova ou questionário em particular;
- do tamanho do teste (número de itens (I));
- das pessoas que fizeram a prova ou responderam ao questionário.

✓ Todas as interpretações estão associadas à prova e ao grupo de indivíduos;

Comparação entre indivíduos ou grupos de indivíduos:

- somente possível quando submetidos as mesmas provas ou,
- pelo menos, submetidos à testes paralelos (normalmente difíceis de serem construídos).

Testes paralelos:

Dois testes são ditos paralelos (com itens diferentes) se:

- a distribuição dos erros é a mesma em cada um dos testes e;
- os escores verdadeiros são os mesmos em ambos os testes para o mesmo indivíduo.

✓ Difícil!

Análises:

- Estatísticas descritivas dos itens: proporção de acertos ou erros;
- Dificuldade: proporção de respostas corretas em cada item;
- índices (coeficientes) para avaliar a qualidade dos itens;
- Discriminação: baseado em proporções de acertos;

✓ Todas as interpretações estão associadas à prova e ao grupo de indivíduos;

Dificuldade:

$$D_i = \frac{C_i}{N_i}$$

C_i : número de indivíduos que responderam corretamente o item;

N_i : número de indivíduos submetidos ao item.

Dificuldade:

	Dificuldade	Percentual de acerto por alternativa				
ITEM		A	B	C	D	E
1	0,73	0,12	0,05	0,73	0,07	0,03
2	0,15	0,15	0,16	0,30	0,11	0,28
3	0,55	0,02	0,02	0,21	0,20	0,55

Coeficiente de correlação ponto-bisserial (discriminação)

O coeficiente de correlação ponto-bisserial (ρ_{pb}) é a correlação de Pearson entre uma variável dicotômica e o escore do teste e é definido por:

$$\rho_{pb} = \frac{\bar{X}_A - \bar{X}_T}{S_T} \sqrt{\frac{p}{1-p}} \quad (1)$$

em que,

\bar{X}_A é a média dos escores dos respondentes que acertaram o item;

\bar{X}_T é a média global dos escores do teste;

S_T é o desvio padrão dos escores obtidos pelos respondentes no teste;

p é a proporção de respondentes que acertaram o item.

Coeficiente de correlação ponto-bisserial

- valores negativos: indicam problemas;
- avaliação do gabarito.
- Software BILOG: exclui o item por *default*;
- obtido com e sem o item (R faz automaticamente);
- valores acima de 0.3 são considerados bons;

Coeficiente de correlação bisserial

O coeficiente de correlação bisserial (ρ_b) é uma medida de associação entre uma variável dicotomizada e uma variável contínua, e é definido por:

$$\rho_b = \rho_{pb} \frac{\sqrt{p(1-p)}}{h(p)} \quad (2)$$

em que,

ρ_{pb} é a correlação ponto-bisserial;

p é a proporção de respondentes que acertaram o item;

$h(p)$ é o valor da densidade da distribuição normal padrão no ponto em que a área da curva à esquerda deste ponto é igual a p .

Coeficiente alfa de Cronbach

O coeficiente alfa de Cronbach é utilizado para medir a consistência interna do instrumento de medida, e é definido por:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum s_i^2}{s_T^2} \right) \quad (3)$$

em que,

n é o número de itens;

$\sum s_i^2$ é a soma das variâncias dos n itens;

s_T^2 é a variância global dos escores dos testes.

Coeficiente alfa de Cronbach

- Esse coeficiente varia de 0 a 1. Quanto mais próximo de 0 menor a consistência e quanto mais próximo de 1 maior a consistência do teste.
- Em geral, valores acima de 0,7 são considerados satisfatórios

Índice de discriminação

É a diferença entre a proporção de acertos dos participantes com maior Escore daqueles com menor Escore (27% dos respondentes com pontuações mais altas) (27% dos respondentes com pontuações mais baixas);

Espera-se que a proporção de acertos para um item seja maior no grupo superior em relação ao grupo inferior.

Quanto maior a diferença, maior será a discriminação do item.

Índice de discriminação

Em geral,

- Discriminação > 40 : item com boa discriminação;
- Discriminação entre 30 e 40: item que pode melhorar;
- Discriminação entre 20 e 30: item que precisa ser refeito;
- Discriminação < 20 : item ruim, deve ser eliminado.

Recomendações de leitura

- 1 Artigo (pdf): Psicometria: fundamentos matemáticos da Teoria Clássica dos Testes
- 2 Artigo (pdf): Da Teoria Clássica dos Testes para os Modelos de Resposta ao Item

Softwares

- R: pacotes ltm e CTT;
- ITEMAN (comercial);