Disciplina:
Teoria da Resposta ao Item

O objetivo da disciplina é estudar técnicas para o desenvolvimento de escalas apropriadas para a medir características de indivíduos que não podem ser medidas diretamente, as quais são comumente denominadas de *traços latentes*.

Portanto, um *traço latente* é uma característica do indivíduo que não pode ser medida diretamente. Por exemplo: *conhecimento sobre um assunto*, *nível de depressão*, *capacidade de concentração*, *grau de dinamismo da pessoa, propensão ao fanatismo, vaidade*.

Por outro lado, variáveis como: *altura*, *peso*, *nível de colesterol no sangue*, não são traços latentes pois temos como medir estas variáveis diretamente.

- O traço latente é medido indiretamente através de questionários, com itens (questões) que podem ser:
- dicotômicos (respostas binárias, i.e., *sim* ou *não*, *certo* ou *errado*).
- politômicos (questões de múltipla escolha em que não há uma única resposta correta, mas a alternativa escolhida pelo respondente influencia na estimativa do traço latente). Por exemplo, no caso de um questionário para avaliação do nível de depressão, considere o item "Você se sente triste: ( ) *nunca*; ( ) *de vez em quando*; ( ) *com frequência*; ( ) *sempre*". As alternativas estão em escala ordinal, indicando um grau crescente de depressão.

- Duas são as teorias utilizadas para este fim. A Teoria Clássica dos Testes – TCT, que utiliza o escore no teste como sua referência de medida, e a Teoria da Resposta ao Item – TRI, cujo foco principal, como bem diz o seu nome, é o item e não o teste como um todo.
- Ambas contemplam a análise de itens através das estimativas de seus parâmetros, e a análise do instrumento de medida como um todo. A TRI foi desenvolvida com o propósito de resolver um problema da TCT que é a dependência da medida de proficiência (traço latente) em relação ao teste aplicado e dos parâmetros dos itens em relação ao conjunto dos respondentes, i.e., para um mesmo indivíduo, obtemos estimativas diferentes de sua proficiência se ele responde questionários distintos, e as estimativas de parâmetros de um item, como seu grau de dificuldade, variam se o item é respondido por grupos diferentes de indivíduos.
- Já dentro do contexto da TRI, a medida de proficiência de um aluno não depende dos itens apresentados a ele, e os parâmetros de discriminação e de dificuldade do item não dependem do grupo de respondentes. Em outras palavras, um item mede determinado conhecimento, independentemente de quem o está respondendo, e a proficiência de um aluno não depende dos itens que estão sendo apresentados a ele.

- Notação (para itens dicotômicos):

$U_{ij} = 1$ se o j-ésimo indivíduo fornece resposta positiva (ou correta) para o i-ésimo item,
e $U_{ij} = 0$ se o j-ésimo indivíduo fornece resposta negativa (ou incorreta) para o i-ésimo item,
$i = 1, \ldots, I$; $j = 1, \ldots, n$ onde
I é o número de itens no questionário, e
n é o número de respondentes (indivíduos).

Teoria Clássica dos Testes - TCT

Def.: o *escore* de um indivíduo é o total de itens para os quais ele(a) forneceu resposta positiva, i.e.,

$$T_j = \sum_{i=1}^{I} U_{ij}$$

Na teoria clássica dos testes, além do número total de acertos, podem ser utilizadas algumas medidas para se avaliar a qualidade do instrumento de medida. Algumas dessas medidas, que podem ser obtidas com o uso do R, são: o *coeficiente de correlação ponto-bisserial*, o *coeficiente de correlação bisserial* e o *coeficiente alfa de Cronbach*.

Coeficiente de correlação ponto-bisserial

É calculado para cada item.

É dado pelo coeficiente de correlação de Pearson entre a variável indicadora de respostas positivas para o item e os escores, i.e., entre

$$U_{ij} \text{ e } T_j, \; j = 1, \dots, n$$

$$\begin{aligned}
\rho_i^{PB} &= \frac{\sum_{j=1}^{n} U_{ij} T_j - (\sum_{j=1}^{n} U_{ij})(\sum_{j=1}^{n} T_j)/n}{\sqrt{[\sum_{j=1}^{n}(U_{ij} - \bar{U}_i)^2][\sum_{j=1}^{n}(T_j - \bar{T})^2]}} \\
&= \frac{n_i \bar{T}_A - n_i \bar{T}}{n\sqrt{p_i(1-p_i)}S_T} = \left(\frac{\bar{T}_A - \bar{T}}{S_T}\right)\sqrt{\frac{p_i}{1-p_i}}
\end{aligned}$$

onde n_i é o número de respondentes com resposta positiva para o i-ésimo item,
$\bar{T}_A$ é o escore médio dos respondentes com resposta positiva para o i-ésimo item,
$\bar{T}$ é o escore médio dos respondentes,
S_T é o desvio-padrão dos escores, e
p_i é a proporção de indivíduos que deram resposta positiva para o i-ésimo item.

Espera-se que o coeficiente de correlação ponto-bisserial apresente valores positivos para todos os itens.

Se algum item apresentar valor negativo para o coeficiente, isto indica incoerência entre os escores dos indivíduos e suas respostas para o item, já que o coeficiente indicaria uma tendência a observar respostas negativas para o item conforme o escore dos indivíduos cresce, i.e., teríamos uma tendência a observar mais respostas positivas para indivíduos com escores baixos do que para indivíduos com escores altos, o que é incoerente.

Isto indicaria que o item é incoerente com o teste (questionário) e deve ser removido (ou reformulado).

Coeficiente de correlação bisserial

O coeficiente de correlação bisserial é uma medida de associação entre uma variável dicotomizada e uma variável contínua (não observada) associada ao *construto* (traço latente). Tem a mesma interpretação da correlação ponto-bisserial.

$\theta \sim N(0,1)$ é a variável (não observada) que representa o traço latente.

Assumimos que o escore é função linear do traço latente mais um erro aleatório ε:

$T = A\theta + B + \varepsilon$, com $E(\varepsilon) = 0$, ε independente de θ.

$\rho^B = corr(T,\theta) = \frac{E[\theta(A\theta+B+\varepsilon)]-E(T)E(\theta)}{\sigma_\theta \sigma_T} = \frac{A\sigma_\theta}{\sigma_T} = \frac{A}{\sigma_T}$ onde σ_T é o desvio-padrão de T e σ_θ é o desvio-padrão de θ.

Seja U a v.a. indicadora de resposta positiva para o item.
$E(T \mid U = 0) = AE(\theta \mid U = 0) + B$
$E(T \mid U = 1) = AE(\theta \mid U = 1) + B$
$\Rightarrow E(T \mid U = 0) - AE(\theta \mid U = 0) = E(T \mid U = 1) - AE(\theta \mid U = 1)$
$\Rightarrow A = \frac{E(T|U=0) - E(T|U=1)}{E(\theta|U=0) - E(\theta|U=1)}$.
Podemos estimar $E(T \mid U = 0)$ por $\bar{T}_E$ (escore médio dos respondentes que forneceram resposta negativa para o item) e $E(T \mid U = 1)$ por $\bar{T}_A$ (escore médio dos respondentes que forneceram resposta positiva para o item).

Vamos admitir que os respondentes para os quais $U = 1$ são aqueles com $\theta > z_p$, onde p é a proporção de respondentes com resposta positiva para o item.
Como z_p é tal que $\int_{z_p}^{\infty} \phi(x)dx = p$, temos que

$$
\begin{aligned}
E(\theta \mid U = 1) &= \frac{1}{p}\int_{z_p}^{\infty} x\phi(x)dx = \frac{1}{p}\int_{z_p}^{\infty} x\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}dx \\
&= \frac{1}{p}\int_{z_p^2/2}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-u}du = -\frac{1}{p\sqrt{2\pi}}\mathrm{e}^{-u}\bigg|_{z_p^2/2}^{\infty} \\
&= \frac{1}{p\sqrt{2\pi}}\mathrm{e}^{-z_p^2/2} = \frac{\phi(z_p)}{p}
\end{aligned}
$$

Analogamente, $E(\theta \mid U = 0) = -\frac{\phi(z_p)}{1-p}$ e

$$\rho^B = \left[\frac{\bar{T}_A - \bar{T}_E}{\frac{\phi(z_p)}{1-p} + \frac{\phi(z_p)}{p}}\right]\frac{1}{\sigma_T}$$

sendo ϕ a f.d.p. da distribuição $N(0,1)$, e σ_T estimado por S_T.
Como este coeficiente é calculado para cada item, temos

$$\rho_i^B = \left[\frac{\bar{T}_A - \bar{T}_E}{\frac{\phi(z_{p_i})}{1-p_i} + \frac{\phi(z_{p_i})}{p_i}}\right]\frac{1}{S_T}, \ \ i = 1, \ldots, I.$$

Note que podemos estabelecer uma relação entre ρ_i^B e ρ_i^{PB}.

$$\begin{aligned}
\rho_i^B &= \left[\frac{\bar{T}_A - \bar{T}_E}{\frac{\phi(z_{p_i})}{1-p_i} + \frac{\phi(z_{p_i})}{p_i}}\right]\frac{1}{S_T} = \frac{p_i(1-p_i)}{S_T\phi(z_{p_i})}(\bar{T}_A - \bar{T}_E) \\
&= \frac{p_i(1-p_i)}{S_T\phi(z_{p_i})}\left[\frac{\sum_{j=1}^{n} U_{ij}T_j}{n_i} - \frac{\sum_{j=1}^{n}(1-U_{ij})T_j}{n-n_i}\right] \\
&= \frac{p_i(1-p_i)}{S_T\phi(z_{p_i})}\left[\frac{n\bar{T}_A}{n-n_i} - \frac{n\bar{T}}{n-n_i}\right] \\
&= \frac{p_i}{\phi(z_{p_i})}\left(\frac{\bar{T}_A - \bar{T}}{S_T}\right) = \frac{\sqrt{p_i(1-p_i)}}{\phi(z_{p_i})}\left(\frac{\bar{T}_A - \bar{T}}{S_T}\right)\sqrt{\frac{p_i}{1-p_i}} \\
&= \frac{\sqrt{p_i(1-p_i)}}{\phi(z_{p_i})}\rho_i^{PB}
\end{aligned}$$

Coeficiente α de Cronbach

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

$$\alpha = \frac{I}{I-1}\left(1 - \frac{\sum_{i=1}^{I} S_i^2}{S_T^2}\right)$$

sendo S_i^2 a variância das respostas para o i-ésimo item,i.e.,

$$
\begin{aligned}
S_i^2 &= \frac{1}{n}\sum_{j=1}^{n}(U_{ij}-\bar{U}_i)^2 = \frac{1}{n}\sum_{j=1}^{n}(U_{ij}^2-2\bar{U}_iU_{ij}+\bar{U}_i^2) \\
&= \frac{1}{n}(\sum_{j=1}^{n}U_{ij}-2\bar{U}_i\sum_{j=1}^{n}U_{ij}+n\bar{U}_i^2) = \bar{U}_i-2\bar{U}_i^2+\bar{U}_i^2 \\
&= \bar{U}_i-\bar{U}_i^2 = \bar{U}_i(1-\bar{U}_i) = p_i(1-p_i)
\end{aligned}
$$

pois $U_{ij}^2 = U_{ij}$, já que U_{ij} só pode assumir os valores 0 e 1.

Sendo assim, $\alpha = \frac{I}{I-1}\left(1-\frac{\sum_{i=1}^{I}p_i(1-p_i)}{S_T^2}\right)$.

A confiabilidade do Coeficiente α de Cronbach normalmente varia entre 0 e 1.

O valor mínimo aceitável para α é 0,70.

A consistência interna dos itens é considerada baixa para valores abaixo desse limite.

O valor 1-p_i pode ser utilizado também como um *Índice de dificuldade do item*.

Temos, ainda, na TCT, o *Índice de discriminação do item*, que é dado pela diferença entre a proporção de resposta positiva para o item no grupo superior e a proporção de resposta positiva para o item no grupo inferior (varia de -1 a 1).

Grupo superior: 27% dos respondentes com os escores mais altos

Grupo inferior: 27% dos respondentes com os escores mais baixos

Temos, ainda, o *erro padrão de medida* (do teste): medida de precisão do teste

$$EPM = S_T\sqrt{1-\alpha} \approx S_T\sqrt{\sum_{i=1}^{i} p_i(1-p_i)}$$