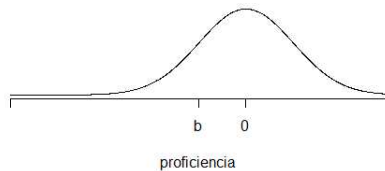
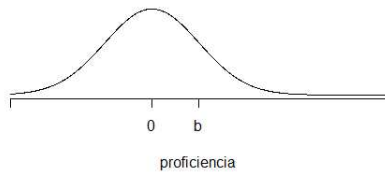


Teoria de Resposta ao Item Equalização

Antonio Eduardo Gomes
aegomes@unb.br

Nestes slides, veremos como expressar, na mesma escala, as estimativas das proficiências de dois grupos de indivíduos que responderam testes distintos, mas com alguns itens em comum. Ao utilizarmos os dados referentes às respostas dos indivíduos para obter estimativas dos parâmetros dos itens e das proficiências separadamente para cada grupo, teremos estimativas diferentes para os parâmetros dos itens em comum nos dois testes. Isso ocorre porque a eventual diferença na distribuição da proficiência nos dois grupos de respondentes fará com que um item que tenha sido respondido pelos dois grupos apresente diferenças nas estimativas de seus parâmetros. Os gráficos a seguir ilustram esse fato para o parâmetro de dificuldade.

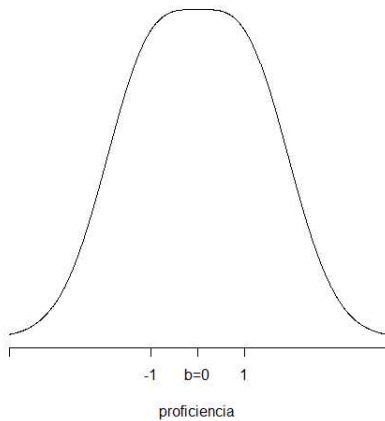


Considere que um item tenha sido aplicado a dois grupos distintos, tendo o primeiro grupo proficiências mais baixas do que o segundo. Suponha que, inicialmente, tenhamos feito a estimação dos parâmetros dos itens e das proficiências utilizando as respostas do primeiro grupo. Lembre-se que assumimos uma distribuição para a proficiência em torno de 0 (em geral, a distribuição Normal padrão) e que o parâmetro de dificuldade é expresso na mesma escala da proficiência. Na figura superior, vemos a distribuição das proficiências. O valor “b” corresponde ao valor do parâmetro de dificuldade do item comum aos dois grupos. Para grupo 1, o item é mais difícil, pois apresenta estimativa do parâmetro de dificuldade em torno de 1.

Ao procedermos à estimação dos parâmetros dos itens e das proficiências utilizando apenas as respostas do segundo grupo, temos as estimativas distribuídas em torno de 0 (figura inferior). Porém, como assumimos que os respondentes do segundo grupo têm proficiência superior aos respondentes do primeiro grupo, o item aplicado aos dois grupos teria grau de dificuldade mais baixo para o segundo grupo do que para o primeiro. No nosso exemplo, o mesmo item seria mais fácil, pois apresenta estimativa igual a -1.

Esta diferença nas estimativas do parâmetro de dificuldade do mesmo item ocorre porque as proficiências e os parâmetros de dificuldade estão expressos na mesma escala. Se as estimativas dos parâmetros dos itens e das proficiências tivessem sido feitas conjuntamente para os dois grupos, teríamos uma única estimativa para o parâmetro de dificuldade do item comum aos dois grupos. Esta estimativa ficaria em torno de um valor intermediário entre as estimativas obtidas separadamente para cada grupo (entre -1 e 1, próximo de 0, no nosso exemplo). Também teríamos estimativas das proficiências para os respondentes dos dois grupos expressas na mesma escala. Para o primeiro grupo, as estimativas das proficiências ficariam concentradas em torno de um valor negativo, enquanto que, para o segundo grupo, as estimativas das proficiências ficariam concentradas em torno de um valor positivo. A imagem a seguir ilustra isso.

Estimação conjunta para os dois grupos



Para sanarmos o problema de termos duas estimativas distintas para os parâmetros de itens comuns aos dois grupos e estimativas para as proficiências com distribuição similar para os dois grupos quando as proficiências são, na verdade, diferentes, adotamos o procedimento de “equalização”.

“Equalizar” significa colocar parâmetros de itens vindos de provas distintas, ou habilidades de respondentes de diferentes grupos, na mesma métrica, isto é, numa escala comum, tornando as habilidades dos dois grupos comparáveis.

Este procedimento é que torna a aplicação da TRI útil para avaliação do desempenho de estudantes ao longo do tempo, como ocorre, por exemplo, com o SAEB (Sistema de Avaliação da Educação Básica), ENADE (exame Nacional de Desempenho dos Estudantes), ENEM (Exame Nacional do Ensino Médio).

Aqui, veremos a equalização “a posteriori”, feita após o fim do processo de calibração dos itens e estimação das proficiências feito separadamente para cada grupo de respondentes.

Adotamos a escala com média 0 e desvio-padrão 1 para as proficiências, mas podemos expressar a proficiência (e consequentemente os parâmetros de dificuldade e de discriminação dos itens) em outra escala através de uma transformação linear. Por exemplo, um indivíduo com habilidade 1 está 1 desvio-padrão acima da média 0. Este mesmo indivíduo teria habilidade 240 e estaria 1 desvio-padrão acima da média se a escala utilizada fosse com média 200 e desvio-padrão 40.

Isto porque, através da propriedade de invariância, temos

$$\begin{aligned}P(U = 1 \mid \theta) &= c + \frac{1 - c}{1 + \exp[-a(\theta - b)]} \\&= c + \frac{1 - c}{1 + \exp\{-\frac{a}{\alpha}[\alpha\theta + \beta - (\alpha b + \beta)]\}} \\&= P(U = 1 \mid \theta^*)\end{aligned}$$

sendo $\theta^* = \alpha\theta + \beta$.

Na escala (200,40), o parâmetro de dificuldade do item passaria a ser dado por $b^* = \alpha b + \beta = 40b + 200$ e o parâmetro de discriminação passaria a ser dado por $a^* = a/\alpha = a/40$.

Sejam b_{G_1} e b_{G_2} os parâmetros de dificuldade para um mesmo item presente nos testes aplicados aos grupos 1 e 2, respectivamente. Analogamente, sejam a_{G_1} e a_{G_2} os parâmetros de discriminação para um mesmo item presente nos testes aplicados aos grupos 1 e 2, respectivamente.

Então, devemos ter $b_{G_1} = \alpha b_{G_2} + \beta$ e $a_{G_1} = (1/\alpha)a_{G_2}$. Uma vez determinados os valores de α e β , as estimativas dos parâmetros dos itens do grupo 2 podem facilmente ser colocadas na mesma escala das estimativas do grupo 1.

O processo de equalização poderia ser feito através de uma transformação linear obtida através da regressão linear aplicada às estimativas do parâmetro de dificuldade obtidas para os itens comuns nos dois grupos.

Suponha, sem perda de generalidade, que os primeiros k itens dos testes aplicados aos dois grupos de respondentes sejam os itens em comum. Sejam, então, $\hat{b}_1^1, \dots, \hat{b}_k^1$ as estimativas dos parâmetros b_1^1, \dots, b_k^1 para o grupo 1 e $\hat{b}_1^2, \dots, \hat{b}_k^2$ as estimativas dos parâmetros b_1^2, \dots, b_k^2 para o grupo 2. Podemos obter estimativas de α e β através do ajuste de uma regressão linear

$$\hat{b}_i^1 = \alpha \hat{b}_i^2 + \beta + \varepsilon_i, i = 1, \dots, k.$$

sendo

$$\varepsilon_i \sim N(0, 1), i = 1, \dots, k,$$

o erro aleatório.

Como as estimativas das proficiências são expressas na mesma métrica dos parâmetros de dificuldade, podemos utilizar os coeficientes α e β obtidos através do ajuste da reta de regressão para expressar as proficiências do grupo 2 na escala do grupo 1 através da transformação

$$\hat{\theta}_{G_2}^1 = \alpha \hat{\theta}_{G_2} + \beta,$$

sendo $\hat{\theta}_{G_2}$ a proficiência de um indivíduo do grupo 2 expressa na métrica do grupo 2, e $\hat{\theta}_{G_2}^1$ a proficiência do mesmo indivíduo do grupo 2 expressa na métrica do grupo 1.

No entanto, este método não é o mais utilizado. O método mais comumente utilizado é denominado “Média-Desvio”.

Sejam M_1 a média, e S_1 o desvio-padrão de $\hat{b}_1^1, \dots, \hat{b}_k^1$, e M_2 a média, e S_2 o desvio-padrão de $\hat{b}_1^2, \dots, \hat{b}_k^2$.

Padronizando as estimativas, devemos ter

$$(\hat{b}_i^1 - M_1)/S_1 \approx (\hat{b}_i^2 - M_2)/S_2, \quad i = 1, \dots, k.$$

Portanto,

$$\hat{b}_i^1 \approx (S_1/S_2)\hat{b}_i^2 + M_1 - S_1M_2/S_2, \quad i = 1, \dots, k.$$

Sendo assim, podemos adotar $\alpha = S_1/S_2$ e $\beta = M_1 - \alpha M_2$.

Obtidos os coeficientes α e β , podemos utilizar o fato de que as proficiências estão expressas na mesma escala do parâmetro de dificuldade em cada grupo para expressar as estimativas das proficiências do grupo 2 na escala do grupo 1.

Sejam θ_{G_2} a proficiência na escala do grupo 2 e $\theta_{G_2}^1$ o valor da proficiência θ_{G_2} na escala do grupo 1. Obtidas estimativas de θ_{G_2} , obtemos estimativas de $\theta_{G_2}^1$ através da relação $\theta_{G_2}^1 = \alpha\theta_{G_2} + \beta$.

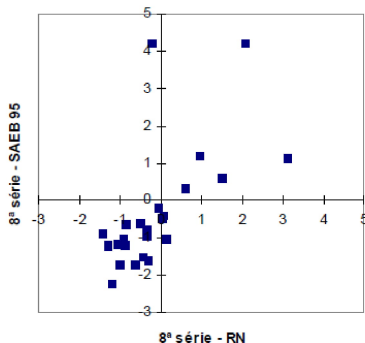
Exemplo: Avaliação de Língua Portuguesa para alunos da 8ª série do Ensino Fundamental feita no Rio Grande do Norte (RN) pela Fundação Carlos Chagas utilizou alguns itens do SAEB 95 com o intuito de colocar os resultados obtidos na mesma métrica do SAEB.

Utilizando o método “Média-Desvio”, obtemos

$$\alpha = \frac{S_{SAEB}}{S_{RN}} = \frac{1,614}{1,104} = 1,462$$

$$\text{e } \beta = M_{SAEB} - \alpha M_{RN} = -0,363 - 1,462 \times (-0,162) = -0,126.$$

Temos, então $\theta_{RN}^{NOVO} = \alpha\theta_{RN} + \beta$, sendo θ_{RN} a proficiência de um indivíduo na escala do teste aplicado no RN, e θ_{RN}^{NOVO} a proficiência do mesmo indivíduo expressa na métrica do teste do SAEB 95. O gráfico a seguir mostra o diagrama de dispersão das estimativas do parâmetro de dificuldade dos itens comuns ao SAEB e à avaliação no RN.



Mais detalhes podem ser obtidos nos livros de Andrade, Tavares e Valle (2000), e de dos Anjos e Andrade (2012).