

## MODELO RASCH: INTRODUÇÃO, SIMULAÇÃO E ESTIMAÇÃO

Pedro Florentino 1; Gilberto Matos 2

1 Universidade Federal de Campina Grande, pedroitalloep@gmail.com

2 Universidade Federal de Campina Grande, gsmatos@gmail.com

### Introdução

O modelo Rasch (Georg Rasch, 1960) é um modelo probabilístico que propõe explicar a probabilidade de uma certa resposta a um determinado item dicotômico,  $X = 0$  ou  $1$ , em função de um traço latente associado ao respondente ( $\theta$ : habilidade ou conhecimento, ansiedade, depressão, qualidade percebida, etc.), e uma característica do item mais conhecida por dificuldade do item,  $b$ . Este modelo, bem como extensões do mesmo vêm sendo utilizados em diversas áreas do conhecimento, dentre elas: psicomетria, avaliação educacional, marketing, ciências da área da saúde, dentre outras.

Ainda que o modelo Rasch possa ser considerado como um modelo particular da Teoria da Resposta ao Item, o que simplifica consideravelmente sua interpretação bem como o processo de estimação e inferência; seu uso pode ser relativamente ampliado e divulgado no Brasil. Deste modo, torna-se de extrema importância um estudo teórico e aplicado dos modelos Rasch para a análise de dados que surgem frequentemente na prática.

O presente trabalho tem como objetivo fazer um levantamento teórico para compreender o modelo Rasch, bem como os seus métodos de estimação dos parâmetros para que, posteriormente, possam ser analisados dados reais e simulados que envolvam características de respondentes (habilidades) e itens (dificuldades) que não podem ser medidas de forma direta, mas apenas de forma indireta, ou seja, através de indicadores que constituem conjuntos de itens (testes, questionários, etc.) aplicados aos respondentes. Um objetivo adicional é, através da análise dos dados simulados, verificar a eficácia da estimação dos parâmetros dos itens.

### Metodologia

Para facilitar o entendimento dos conceitos e métodos utilizados na construção e estimação dos parâmetros do modelo Rasch, estamos realizando uma simulação no software R através do ambiente de desenvolvimento integrado RStudio. Nela, estamos adotando uma população de 250 habilidades (respondentes aos itens) seguindo distribuição normal padrão; isto é  $N(0,1)$  e, num primeiro momento, estamos determinando dificuldades dos itens fixados em: -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5, ou seja, duzentas e cinquenta pessoas com habilidades individuais respondendo a sete itens ou questões.

De posse desses valores, aplicamos na fórmula do modelo Rasch e obtemos a matriz com as probabilidades de acerto de cada sujeito com sua habilidade individual respondendo a cada item. Logo após, calculamos as probabilidades de respostas corretas aleatórias através da distribuição uniforme. Com esses valores obtidos, comparamos cada valor e geramos uma nova matriz de zeros e uns com as respostas dicotômicas aleatórias simuladas; em que se o valor da probabilidade de resposta correta aleatória for menor ou igual que o valor da probabilidade de resposta

correta, assume-se valor/resposta 1, que representa que o sujeito respondeu corretamente ao item em questão e, caso contrário, irá aparecer 0 significando que o sujeito respondeu incorretamente ao item em questão. A partir daí utilizamos o *pacote ltm* para ajustar o modelo de Rasch e estimar os parâmetros de dificuldade dos itens que inicialmente foram pré-determinados. O primeiro passo foi criar uma restrição que o quarto elemento da matriz será zero, pois ele significa a dificuldade média dos itens. Tal procedimento faz com que todos os demais parâmetros de dificuldade sejam estimados em comparação ao quarto item fixado em 0.

## Resultados e discussão

A principal contribuição do Modelo de Rasch diz respeito ao entendimento de um sujeito A, com habilidade maior do que um sujeito B, deve ter uma probabilidade maior de responder corretamente a um determinado item. Paralelamente, um item com grau de dificuldade C deve ser mais frequentemente acertado por um sujeito com habilidade D, quando comparado com um item E de maior dificuldade (Rasch, 1960). Esse modelo pode ser representado pela função logística de um parâmetro, a qual considera que as respostas de um sujeito a um conjunto de itens dependem apenas de sua habilidade e da dificuldade do respectivo item (Baker, 2001; Linacre e Wright, 2002).

Sua expressão matemática é dada pela probabilidade de se responder corretamente ao item  $i$  dado/conhecido um determinado valor de  $\theta$  (valor do traço latente do respondente ou variável que se tem interesse medir) e o valor de  $b_i$  (índice de dificuldade do item  $i$ ).

No modelo Rasch temos que estimar dois tipos de parâmetros, os parâmetros dos itens e as habilidades dos respondentes. Então, do ponto de vista teórico, temos três cenários: quando já conhecemos os parâmetros dos itens e temos que estimar as habilidades; quando temos as habilidades dos indivíduos e iremos estimar os parâmetros dos itens; ou, a situação mais comum, em que precisamos estimar as habilidades dos respondentes e os parâmetros dos itens simultaneamente.

O processo de estimação é, em geral, feito pelo método de máxima verossimilhança através da aplicação de algum processo iterativo, como o algoritmo Newton-Raphson ou “Scoring” de Fisher.

Para podermos fazer a estimação dos parâmetros no modelo, precisamos admitir duas suposições, a saber:

- As respostas oriundas de indivíduos diferentes são independentes;
- Os itens são respondidos de forma independente por cada indivíduo, fixada sua habilidade (independência local).

Diante do modelo descrito acima e dos dados obtidos por simulação conforme descrito na sessão anterior, nós estudamos o processo de estimação dos parâmetros dos itens pelo método da Máxima Verossimilhança Marginal (MVM). Posteriormente, fizemos uma comparação com os valores pré-estabelecidos da simulação para saber se foi feita uma boa estimação e observamos que os valores dos parâmetros dos itens foi próxima dos valores pré-fixados, mas com o sinal inverso. Isso acontece porque no pacote *ltm* do R, o que se estima é o parâmetro de facilidade do item, ao invés da dificuldade.

## Conclusões

Ao fazermos o processo de estimação dos parâmetros, percebemos que o *pacote ltm* utilizado do R estima a facilidade dos itens ao invés de estimar a dificuldade como é proposto pelo modelo Rasch. Tal estimação nesse pacote é realizada pelo método da Máxima Verossimilhança Marginal (MVM), a qual foi estudada em detalhes por nós. Em geral, foi possível observar que a estimação dos parâmetros resultou em valores condizentes com os valores previamente fixados na simulação.

O desenvolvimento deste projeto requer muitos conhecimentos em nível básico e avançado não somente de probabilidade e estatística, mas também da Teoria de Resposta ao Item. Tendo isto em vista, até o momento nós estudamos o modelo Rasch bem como poucos métodos de estimação dos parâmetros de tal modelo, restando ainda avançarmos para outros métodos de estimação; bem como o estudo de Extensões do Modelo Rasch juntamente com seus métodos de estimação e fazermos as simulações e análises de dados reais.

**Palavras-Chave:** Teoria de Resposta ao item; Modelo Rasch; Estimação dos parâmetros; Software R.

## Fomento

Ao CNPq, a UFCG e ao LANEST/UAESt/UFCG.

## Referências

Baker, F.B. (2001). The basics of item response theory. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation.

Linacre, J.M. & Wright, B.D. (2002). Understanding Rasch measurement: Construction of measures from many-facet data. *Journal of Applied Measurement*, 3 (4), 486–512.

Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests.*, Copenhagen: Danish Institute for Educational Research.