3
Traços Latentes e a *Teoria de Resposta ao Item* - TRI

Em muitas situações de medidas sociológicas, psicológicas ou educacionais a variável de interesse é de entendimento intuitivo para todos. Porém, na maioria das vezes, não é observável diretamente. É isto que a psicometria chama de variáveis não observáveis ou habilidades ou traços latentes. Embora essas variáveis possam ser facilmente descritas e listadas, como por exemplo, a inteligência, a habilidade em executar uma tarefa, ansiedade, o nível de entendimento de texto, e etc, elas não podem ser medidas diretamente como o peso ou altura de uma pessoa. Apesar de todas serem características implícitas a cada ser humano. A meta das medidas educacionais e psicológicas é determinar como os traços latentes se processam na pessoa.

Se for de interesse medir tal traço latente é necessário então criar uma escala de medida segundo a qual essa variável assumirá seus valores. Por inúmeras razões técnicas a definição da escala de medidas, o número na escala e a interpretabilidade da mesma em relação ao traço latente é muito difícil, necessitando de um formalismo maior.

Do ponto de vista prático, questões de resposta livre são de difícil uso na TRI (exceto se a resposta for caracterizada como certo ou errada ou, ainda, com algum tipo de graduação). Como resultado, a maioria dos testes usados na TRI são de múltipla escolha e os itens podem ser dicotômicos (certo ou errado) ou politômicos (incorporam variáveis categóricas em suas respostas).

È razoável admitir a hipótese de que cada examinando responda a um item de acordo com habilidades implícitas. Por motivos que depois será explicado, admitiremos que se deseje medir apenas uma habilidade que representaremos pela letra grega θ.

Por exemplo, no caso dicotômico a cada nível de habilidade existirá uma certa probabilidade que o respondente j com esta habilidade dará uma resposta correta ao item i. Esta probabilidade será denotada por $P_i(\theta_j)$. No caso típico de testes com itens dicotômicos a probabilidade será pequena se a habilidade do respondente for pequena ou será grande se o mesmo também o for.

A curva que caracteriza essa $P_i(\theta_j)$ tem, **em geral**, uma forma de "S amortecida" (figura 1). Isso não é arbitrário, mas advém de muitos estudos empíricos e pioneiros. Ela é conhecida como *curva característica do item* (CCI) e é a base da construção da TRI, todas as outras construções dependem dela.

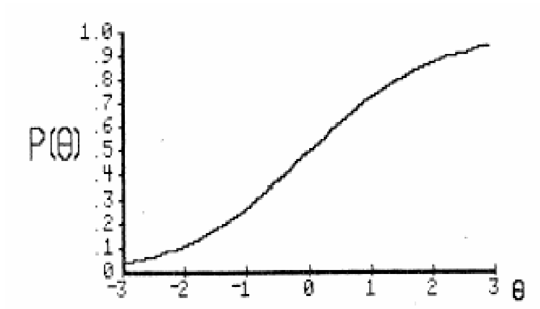

Figura 1 - Curva Característica do Item

Existem duas propriedades técnicas da CCI que são usadas para descrevê-las: dificuldade do item e seu poder de discriminação.

Na figura 2 abaixo estão representadas 3 curvas características do item no mesmo gráfico. Todas têm o mesmo nível de discriminação, porém diferentes com respeito à dificuldade.

Num contexto em que as respostas podem ser certas ou erradas, a curva da esquerda representa um item fácil porque a probabilidade de resposta correta é alta para habilidade baixa.

A curva do meio representa um item de dificuldade médio e, a curva da direita representa um item difícil porque a probabilidade de resposta correta é baixa para boa parte da escala exceto para os níveis mais altos de habilidade.

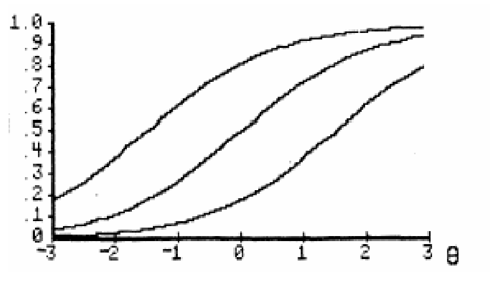

Figura 2 - CCI com mesma discriminação e diferentes dificuldades

O conceito de discriminação é mostrado na figura 3 abaixo. Estas três curvas contêm três CCI tendo o mesmo nível de dificuldade, mas diferem com respeito à discriminação.

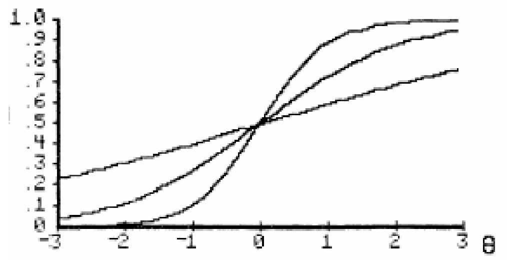

Figura 3 - CCI com dificuldades iguais, porem com diferentes discriminações

A curva com maior formato de S tem um alto nível de discriminação, enquanto para a curva mais amortecida tem um poder de discriminação médio e por fim a última possui um poder de discriminação baixo.

3.1
Modelos da *Curva Característica do Item* – CCI

Segundo Backer (1995), muitos autores ajudaram na construção da teoria de resposta ao item (TRI). Em especial 3 desses indivíduos, para os quais é importante citar seus nomes e respectivos trabalhos. D.N. Lawley (1943) publicou um trabalho mostrando que muitos dos *constructos* produzidos por testes baseados na Teoria Clássica de Medidas (TCM) poderiam ser expressos em termos de parâmetros da curva característica do item nos moldes dos modelos apresentados acima. Este trabalho marcou o início da teoria de medida da TRI. O trabalho de F. M. Lord ganhou força nos anos 50 com a aplicação em testes educacionais. Lord iniciou o desenvolvimento formal da teoria, e também contribuiu para o desenvolvimento de programas para computadores necessários para por a teoria em prática. O que levou posteriormente a livros clássicos, redigidos junto com M. Novick em 1968 e 1980. Em 1960 o matemático dinamarquês Georg Rasch reforçou os estudos da TRI criando o modelo de 1 parâmetro e, para muitos seu trabalho é tão pioneiro quando o de Lord.

3.2
Comentário

Na seção anterior, as propriedades da CCI foram abordadas apenas como uma descrição verbal. Agora podemos introduzir de maneira formal os modelos matemáticos para CCI.

Estes modelos são definidos a partir de uma equação matemática para a relação entre a probabilidade do respondente ao teste responder corretamente ao item e a habilidade do respondente. Cada modelo emprega um ou mais parâmetros, cujos valores definem a forma da CCI.

Esses modelos e seus parâmetros fornecem um veículo para a comunicação de informações sobre as propriedades técnicas do item.

A função logística é muito utilizada nas ciências biológicas e nos modelos de crescimento de plantas e animais. Seu uso se dá pelo fato de sua simplicidade e também por ser uma função explícita dos parâmetros dos itens e da habilidade do respondente. Anteriormente se utilizava a distribuição normal acumulada, que é uma função implícita dos parâmetros.

3.3
Teoria de Resposta ao Item - TRI

A TRI constitui-se a partir de um conjunto de modelos matemáticos que buscam representar a relação entre a probabilidade de um indivíduo dar uma determinada resposta a um item como função dos parâmetros do item e da habilidade do respondente, caso paramétrico. No caso dicotômico esta relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acertos ao item. O termo acerto é utilizado apenas devido à origem da TRI estar associada a testes educacionais. **No presente contexto, em que a variável latente que é objeto do estudo representa uma condição sócio-econômica familiar**, a resposta dicotômica pode representar, por exemplo, o fato de se possuir ou não determinado bem.

Os primeiros modelos matemáticos da TRI foram os dicotômicos, cujos resultados eram certo ou errado. Seja a resposta a um item $U_{ij} = 1$ para certo ou 0 para item errado.

Existia a necessidade de encontrar uma função não linear que expressasse a probabilidade do respondente em função de sua habilidade a dar uma certa resposta em função dos parâmetros do item. A própria necessidade desta função já impunha a restrição da CCI ser monótona crescente.

Birnbaum foi motivado pelo trabalho de LORD (1952). Uma contribuição importante de Birnbaum foi sugerir a troca da função ogiva normal pelo modelo logístico de dois parâmetros, por questões de conveniência. Além disso, foi Birnbaum, também, quem introduziu o terceiro parâmetro (vulgarmente conhecido como parâmetro de acerto casual, que modela um acerto em um teste educacional devido a um *chute* na questão). Uma outra contribuição de Birnbaum para a teoria psicométrica foi a introdução da medida de Fisher para descrever a estrutura de informação do teste. O conteúdo de informação de um teste para uma habilidade do respondente desconhecida é a soma de todas as informações individuais de cada item.

Paralelamente ao trabalho de Birnbaum, RASH que nos anos 40 já trabalhava em medidas psicométricas (Linden e Hambleton (1997)), pelos anos 50 começou a desenvolver seu trabalho para modelos dicotômicos que por razões históricas ficaram conhecidos como modelo de RASH. Vale lembrar que, formalmente, os modelos de RASH são um caso particular dos de Birnbaum, mas que conduzem a uma teoria da análise do teste completamente diferente.

A própria necessidade dos testes psicométricos de introduzir respostas que não fossem consideradas somente dicotômicas encarregou-se de motivar a ciência a desenvolver modelos para TRI que tratassem de tais necessidades como: Modelo de Resposta graduada, Modelo de Crédito Parcial, Modelo de Escala Gradual, entre outros. Para maior detalhe destes modelos recomenda-se (Linden e Hambleton (1997)).

3.4
O Contexto das Avaliações Educacionais

Nas últimas décadas a TRI vem tornando-se a técnica predominante no contexto de avaliações educacionais em vários países. No Brasil a primeira experiência ocorreu em 1995, na análise de dados do SAEB (Sistema Nacional de Avaliação Básica).

A TRI permite uma melhor análise de cada item que constitui o instrumento de avaliação e/ou medida, levando em consideração suas características na produção das habilidades, o que facilita, também, a interpretação da escala produzida.

Além disso, a TRI é extremamente vantajosa, pois permite a comparabilidade dos resultados produzidos para grupos de indivíduos diferentes, mesmo quando são aplicados testes parcialmente diferentes.

Essa propriedade é útil, particularmente, quando se deseja avaliar a evolução da medida de uma mesma escala ao longo do tempo, o que se denomina de estudos longitudinais.

É freqüente autores afirmarem que a medida de habilidade produzida através da TRI é independente do teste aplicado e independente do grupo de alunos a que é aplicado. Claro que cuidados e técnicas específicas devem ser empregadas para garantir essa propriedade. A TRI tem como base de todo seu desenvolvimento dois postulados: i) a *performance* de um respondente em um teste pode ser prevista por um conjunto de fatores inerentes ao indivíduo, chamados de habilidade ou traço latente. Ressalta-se aqui que a habilidade é uma variável não observável diretamente pelo teste empregado; ii) a relação entre a habilidade do indivíduo e a probabilidade de escolha no item pode ser descrita por uma função característica ou curva característica item (CCI), parametrizada por características do item de um teste (1 – poder de discriminação; 2 – poder de dificuldade)

Antes de apresentar os modelos faz-se necessário explicação sobre duas hipóteses básicas na TRI: unidimensionalidade e invariância.

3.5
Hipóteses Básicas

Os modelos matemáticos empregados na TRI especificam que a probabilidade de resposta certa de um respondente depende de sua habilidade ou habilidades e as características do item. Modelos da TRI aqui tratados incluem um conjunto de hipóteses sobre os dados em que o modelo é aplicado.

A primeira hipótese é chamada de unidimensionalidade. Esta hipótese admite que somente uma habilidade é medida pelo modelo. Isto é, o conjunto de itens deve estar medindo um único traço latente. Parece claro que dentro das instâncias do ser humano exista uma variedade de habilidades responsáveis por um processo de execução de uma tarefa. Porém, para satisfazer tal postulado, é suficiente admitir que haja uma habilidade dominante, responsável pelo conjunto de itens. Este fator é que se supõe estar sendo medido pelo teste (Hambleton, 1991). Tipicamente a análise de unidimensionalidade é obtida através de métodos de análise fatorial.

Somente uma habilidade é medida pelo conjunto de itens que compõe o teste. Fatores como motivação, ansiedade, habilidades cognitivas, dentre outros, compõem o processo de execução de um teste. Como foi dito antes, basta aceitar a presença de um fator dominante, isto é, aquele que influencia a performance do teste. Modelos que incorporam mais de uma habilidade para performance de um teste são chamados de multidimensionais (Linden, 1997).

Uma outra suposição é a chamada independência local ou invariância, a qual postula que, para uma dada habilidade, as respostas atribuídas aos diferentes itens são independentes entre si. Tal pressuposto será importante para a estimação dos parâmetros nos modelos.

Teoricamente, a unidimensionalidade implica na independência local (Hambleton, 1991) e assim apenas uma e não duas hipóteses devem ser verificadas. De fato, se houver dependência local entre os itens essa produzirá falsas dimensões na análise fatorial.

3.6
Modelos

Teoricamente pode existir uma infinidade de modelos da TRI. Porém, poucos modelos são usados na prática. Entre os modelos propostos na literatura dependem fundamentalmente de três fatores:

(i) da natureza do item – dicotômicos ou politômicos;

(ii) do número de populações envolvidas;

(iii) da quantidade de traços latentes que está sendo medida.

Na seção 3.6.1 apresentaremos os modelos unidimensionais para itens dicotômicos mais utilizados e na seção 3.7 apresentaremos um dos modelos para itens politômicos utilizado no trabalho proposto.

3.6.1
Modelos unidimensionais para itens dicotômicos

1. ***ML1** - Modelo logístico (unidimensional) de um parâmetro ou modelo de Rasch.*

Rasch começou seu trabalho em medidas educacionais e psicométricas por volta de 1940. Por volta da década de 50, Rasch usando a função de Poisson desenvolveu dois modelos, um para leitura de testes e um modelo para aproveitamento e inteligência de testes, com a finalidade de produção de *scores.* Para maiores detalhes (Linden, 1991 p: 6-7).

Com base em seus trabalhos a motivação de Rasch foi representar a probabilidade de resposta como função da habilidade do respondente e a característica do item. Seja θ o parâmetro de habilidade do respondente j e b a dificuldade do item i .

O sucesso do respondente é a razão entre sua habilidade e a soma, da habilidade com a dificuldade do item. Dessa forma, Rasch construiu:

$$P_i(U_{ij}=1/\theta)=\frac{\frac{\theta_j}{\delta_i}}{1+\frac{\theta_j}{\delta_i}}=\frac{\theta_j}{\theta_j+\delta_i}$$ por questões de nomenclatura usaremos $\theta_j=\theta$

e $\delta_i=b_i$

Tomando em uma escala logarítmica para os parâmetros será adotada por convenção daqui por diante. Tem-se o modelo unidimensional de um parâmetro:

$$P(U_{ij}=1/\theta_j)=\frac{1}{1+e^{-(\theta_j-b_i)}}$$ onde:

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i. $P(U_{ij}=1/\theta_j)$ é a probabilidade de um indivíduo *j* com habilidade θ_j responder corretamente o item *i*. O parâmetro de dificuldade (ou de posição) do item *i*, medido na mesma escala da habilidade, é denotado por b_i.

2. **ML2 -** Modelo logístico de 2 parâmetros

Esse modelo, proposto por Birnbaum,1968, a partir da substituição da função de distribuição normal, proposta no modelo de Lord, 1952, pela função logística, pressupõe a relação monótona entre o valor da variável latente (a ser estimada) do indivíduo e a sua probabilidade de escolha por uma das duas alternativas segundo uma função de distribuição logística parametrizada por coeficientes que representam determinadas características do item. É muito natural no contexto de avaliação educacional, onde a variável latente é identificada com a habilidade cognitiva do aluno, e as possibilidades de escolha são acertar ou não o item.

Assim, admita que U_{ij} seja uma variável aleatória dicotômica assumindo os valores 0, ou, 1. No caso específico de um teste educacional o valor 0 está associado a uma resposta errada e, o valor 1 a uma resposta certa por parte do aluno. O modelo de dois parâmetros expressa a relação entre a variável latente θ e a resposta dada ao item da seguinte forma:

$$P(U_{ij} = 1/\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

a_i é o parâmetro de discriminação (ou de inclinação) do item *i*, com valor proporcional à inclinação da Curva Característica do Item – CCI no ponto b_i.

D é um fator de escala constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultados semelhantes ao da função ogiva normal, diz-se, então, que o modelo está na métrica normal. O índice *i* representa o número do item e j o respondente.

Como se pode notar, o parâmetro b_i representa o ponto na escala de habilidade em que um examinando tem 50% de probabilidade de responder ao item *i* corretamente. Num contexto mais geral, b_i representa o valor da variável latente θ, para o qual há 50% de chance de escolha da resposta representada por U_{ij} =1 pelo indivíduo. É fácil observar que se (6) for derivada em relação à θ, a função resultante atinge seu máximo em θ = b_i com um valor diretamente proporcional a a_i (0.425 a_i), indicando que a inclinação da curva do modelo atinge seu maior valor onde a probabilidade de ocorrer uma resposta representada por U_{ij} =1 (isto é, a resposta correta, no caso de modelos para avaliação educacional) é 0.5. Portanto, quanto maior for o valor do parâmetro a_i, mais sensível torna-se o modelo a variações na habilidade em torno de seu ponto de dificuldade. Isto é, maiores valores para o parâmetro a_i produzirão maior capacidade de distinção entre dois indivíduos com habilidades diferentes no nível da escala em torno do nível de dificuldade do item. Por isso, ele é conhecido como parâmetro de discriminação do item. A seguir, apresenta-se a denominada curva característica de um item, isto é, a representação dos valores, sob forma de gráfico, de um particular modelo θ, enfatizando as propriedades de seus parâmetros:

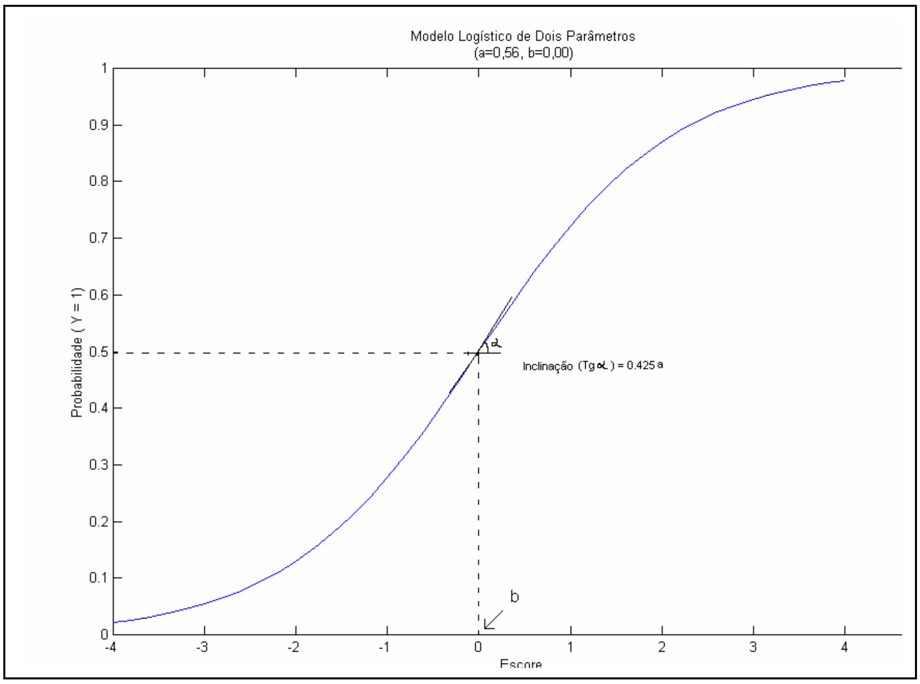

Figura 4 – Curva Característica do Item com valores especificados

Muito embora, a motivação principal na proposição desse modelo tenha sido o de sua utilização em avaliação educacional, pode-se empregá-lo com finalidade diversa. È o caso de, por exemplo, utilizá-lo para a construção de um índice que mede a condição sócio-econômica de indivíduos de uma população. Podem ser consideradas como variáveis indicadoras da condição sócio-econômica, a posse de determinados bens como, por exemplo, eletrodomésticos, automóvel, etc.

3. **ML3 –** Modelo Logístico de 3 parâmetros

Dos modelos propostos pela TRI, o modelo logístico unidimensional de 3 parâmetros é dado por:

$$P(U_{ij} = 1/\theta_j) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

Onde c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item *i* (muitas vezes referido como a probabilidade de acerto casual). E os outros parâmetros já foram citados acima.

3.7
Modelos para Itens com Formato de Resposta Politômica

Será apresentado agora o modelo de respostas graduadas pelo fato de este ter sido o modelo usado no desenvolvimento deste trabalho para a produção do *score* indicador NSE da condição sócio-econômica. Recomenda-se fortemente a leitura do livro (Linden, 1991) para aqueles que desejam se aprofundar no conhecimento dos inúmeros modelos existentes hoje. Itens politômicos são itens que não se caracterizam apenas pela presença do fator certo ou errado, mas por todo um conjunto ordenado de respostas. Em particular, o modelo de resposta graduada foi desenvolvido por Samejima (1968).

Modelo de respostas graduadas (MRG).

O MRG de Samejima (1962) assume que as categorias de respostas de um item podem ser ordenadas entre si.

Suponha que os *scores* das categorias de um item i estão dispostos em ordem crescente denotamos por $k = 0,1,\ldots,m_i$ onde $(m_i + 1)$ é o número de categorias do i-ésimo item. A probabilidade de um indivíduo j escolher uma particular categoria ou outra mais alta do item i é representada por:

$$P_{i,k}^{*}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}}$$

um modelo logístico de dois parâmetros, com $i = 1,\ldots,I$; $j = 1,\ldots,n$; $b_{i,k}$ é o parâmetro de dificuldade da k-ésima categoria do item i.

Por definição do modelo temos: $b_{i,1} \leq b_{i,2} \leq \ldots \leq b_{i,m_i}$ ordenação entre os níveis de dificuldade das categorias de um dado item.

A probabilidade de um indivíduo j receber um *score* k no item i é dada pela expressão:

$$P_{i,k}(\theta_j) = P^*_{i,k}(\theta_j) - P^*_{i,k+1}(\theta_j)$$

E Samejima também admite que $P^*_{i,0}(\theta_j) = 1$ e, $P^*_{i,m_i+1}(\theta_j) = 0$. Tem-se então:

$$P_{i,0} = P^*_{i,0}(\theta_j) - P^*_{i,1}(\theta_j) = 1 - P^*_{i,1}(\theta_j)$$

$$P_{i,1} = P^*_{i,1}(\theta_j) - P^*_{i,2}(\theta_j)$$

$$P_{i,m} = P^*_{i,m}(\theta_j) - P^*_{i,m+1}(\theta_j) = P^*_{i,m}(\theta_j),$$

em geral tem-se: $P_{i,k} = P^*_{i,k}(\theta_j) - P^*_{i,k+1}(\theta_j) \geq 0$.

Na forma logística o modelo de respostas graduadas é dado por:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}}$$

O modelo na figura 5 abaixo foi o encontrado para a questão que perguntava sobre o número de banheiros que a família possuía (as possibilidades de respostas foram as seguintes: nenhum, tem 1, tem 2, tem 3, tem 4 ou mais). Para esse tipo de item, a probabilidade de escolha de cada resposta pode ser modelada através do modelo de respostas graduadas. Foi obtido o seguinte modelo: (a_i = 1.511, b_i 0.592, $b_{i,0}$ = ∞, $b_{i,1}$ = 3.81, $b_{i,2}$ = 0.191, $b_{i,3}$ = 1.183, $b_{i,4}$=1.807), a probabilidade de cada resposta está representada no gráfico abaixo:

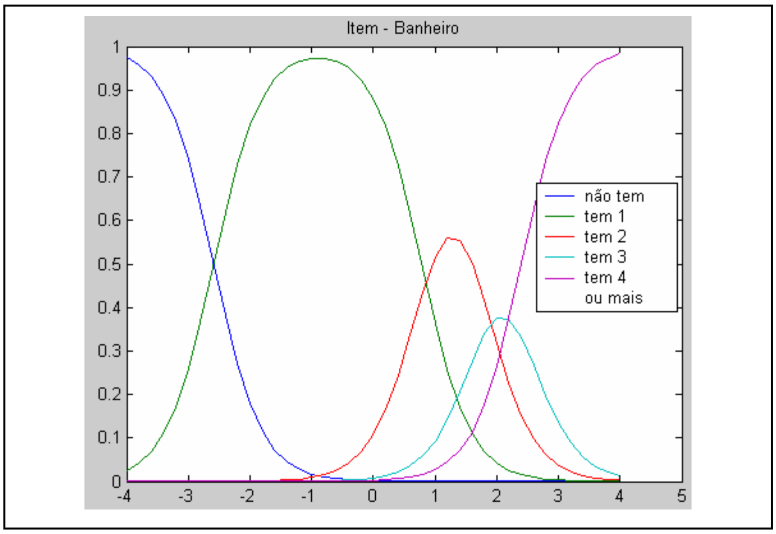

Figura 5 - Modelo de Respostas Graduadas para o item banheiro

Podemos extrair, com base na figura acima, informações valiosas sobre o item do questionário possuir (um ou mais banheiros) ou não, como:

i) Pessoas com *score* estimado entre -4 e -3, a probabilidade de não possuir banheiro é superior a 60 %; A medida que aumenta o valor do *score*, também aumenta a chance do respondente possuir pelo menos um banheiro.

ii) Pessoas com *score* estimado entre -3 e -2, começam a apresentar um aumento na probabilidade de possuir um banheiro;

iii) Já entre o *score* -2 e 0 podemos afirmar sem perda de generalidade que tal *score* é caracterizado por àqueles que possuem pelo menos um banheiro.

A construção da interpretabilidade para leitura dos *scores* estimados, para cada item, segue como descrito acima. Faz-se necessário chamar a atenção que o valor da probabilidade de uma curva qualquer (para um *score* dado) é 1 menos o valor da probabilidade de outra curva abaixo da primeira, Por exemplo, exatamente no *score* 3, temos 80 % de possuir quatro banheiros mais, aproximadamente 15 % de possuir 3 banheiros mais, 5 % (aproximadamente) de possuir 2 banheiros.

3.8
Estimação dos parâmetros

O primeiro e mais importante passo na aplicação da TRI a dados de testes é a estimação dos parâmetros que caracterizam o modelo de resposta ao item.

Nos modelos da TRI, a probabilidade de uma resposta depende, da habilidade do examinando, θ_j, e os parâmetros que caracterizam o item. Ambos, habilidade e parâmetros do item, na maioria das vezes, são desconhecidos; o que é conhecido são as respostas dos examinados aos itens do teste. O problema da estimação é determinar o valor de θ_j para cada examinado e os parâmetros que compõe cada item do teste. Fazendo um paralelo com modelos clássicos de regressão, onde os parâmetros que caracterizam o modelo de regressão (os coeficientes de regressão) devem ser estimados, nota-se, no entanto, duas grandes diferenças entre os modelos de regressão e os modelos da TRI. Primeiro, o modelo de regressão é usualmente linear, enquanto os modelos de resposta ao item são não lineares. Segundo, e mais importante, o regressor (variável independente) na análise de regressão é observável. Na TRI a "variável regressora" θ é não observável. Esse aspecto dificulta substancialmente o problema de estimação dos parâmetros do modelo.

Na regressão linear o melhor ajuste do modelo é definido pelo critério dos mínimos quadrados. Nos modelos da TRI tal critério não é usado porque seria difícil determinar as propriedades requeridas para seu uso em modelos não lineares. Alternativamente, os parâmetros podem ser estimados usando o método da máxima verossimilhança através da aplicação de algum processo iterativo, como o algoritmo de *Newton-Raphson* ou "*Scoring*" de Fisher. Alguns procedimentos Bayesianos também são aplicados com alguma freqüência (ver Mislev (1986[a])).

Na situação em que se deseja estimar tanto os parâmetros dos itens, quanto as habilidades, há duas formas de se abordar o problema de estimação: estimação conjunta ou em duas partes (máxima verossimilhança marginal) primeiro a estimação dos itens e, posteriormente, das habilidades. Utilizou-se no presente trabalho a estimação em duas partes.

Não é a finalidade do presente trabalho prolongar-se sobre o assunto de estimação, haja visto a existência de inúmeros trabalhos expostos na literatura corrente. Entendemos que, se decidíssemos apresentar de maneira formal o assunto de estimação, nos desviaríamos de nosso real objetivo. Recomenda-se para aprofundamento em estimação de parâmetros dos modelos de resposta ao item: Hambleton [3] e/ou Linden [4], Andrade et al, 2001.

3.9 Métodos Clássicos para Discriminação dos Itens e Análise da Dimensionalidade

3.9.1 Correlação Bisserial e Correlação Ponto Bisserial

Considere o caso de testes constituídos por itens binários ou *dicotômicos,* isto é, itens para os quais se admite duas respostas possíveis. A correlação bisserial e a correlação ponto bisserial são medidas estatísticas que medem a correlação do resultado de um item em particular do teste com o resultado do teste (isto é, o escore bruto total), sendo, portanto, uma medida da capacidade de discriminação do item em relação ao resultado do teste. Elas são muito usadas dentro da teoria clássica de testes psicométricos. A correlação ponto bisserial pode ser derivada diretamente da correlação de Pearson. Para tanto, admita que S represente o escore bruto obtido no teste. Admita que Y represente o resultado da resposta atribuída a um item, uma variável dicotômica (no caso de testes educacionais, por exemplo, atribui-se o valor $Y = 0$, se a resposta for errada, e $Y = 1$, para uma resposta correta; e, no caso de um item que avalia a condição sócio-econômica $Y = 1$ representa a posse de um bem, por exemplo). O índice de correlação de Pearson é definido por:

$$\rho_{SY} = \frac{E(SY) - E(S)E(Y)}{\sigma_Y \sigma_S}$$

Se p é a probabilidade de se acertar o item , então $E(Y) = p$ e $\sigma_Y = \sqrt{pq}$, onde q = 1 – p. Assim, tendo em vista que Y é uma variável discreta, tem-se:

$$\rho_{SY} = \frac{E(S\,Y \setminus Y = 0)\;P(Y = 0) + E(S\;Y \setminus Y = 1)\;P(Y = 1) - E(S)\,p}{\sigma_S\;\sqrt{\rho\, q}}$$

tal que

$$\rho_{SY} = 0 + \frac{E(S \setminus Y = 1)\;p - E(S)\;p}{\sigma_S\;\sqrt{p\,q}}$$

Uma estimativa natural obtida sobre o resultado do teste é a seguinte:

$$\overline{\rho}_{\rho b} = \frac{\overline{S}_p - \overline{S}}{\overline{\sigma}_S} \qquad \sqrt{\frac{\overline{p}}{\overline{q}}}$$

onde $\overline{S}_p$ é o escore médio no teste para os que acertaram o item e $\overline{S}$ é o escore médio no teste para todos. Aqui $\overline{\sigma}_S$ é o desvio padrão dos escores obtidos nos testes pelos respondentes e, a estimativa $\overline{\rho}_{pb}$ é o que se freqüentemente se denomina na literatura de correlação ponto bisserial. $\overline{p}$ é a proporção dos que acertaram o item no teste. Um desenvolvimento para a correlação bisserial pode ser o seguinte.

Seja Z uma variável aleatória artificial (e, portanto, não observada), associada ao *constructo* latente do respondente, tal que $Z \sim$ N(0,1). Admita ainda que o escore bruto do respondente no teste se associa linearmente a essa variável da seguinte forma: $S = AZ + B + \varepsilon$, onde $E(\varepsilon) = 0$ e $E(\varepsilon Z) = 0$. Note-se que $E(S) = A\;E(Z) + B$ e, então a correlação de Pearson para S e Z é dada por:

$$\rho_{SZ} = \frac{E(AZ^2 + BZ) - E(S)E(Z)}{\sigma_Z\;\;\sigma_S} = \frac{A}{\sigma_S}\quad \sigma_Z$$

Sejam dois conjuntos de possíveis respondentes, os que acertam o item e os que erram o item. Assim:

$$E(S \setminus Y = 0) = A\;E(Z \setminus Y = 0) + B$$

e , $$E(S \backslash Y = 1) = A\, E(Z \backslash Y = 1) + B$$

tal que:

$$A = \frac{E(S \backslash Y = 0) - E(S \backslash Y = 1)}{E(Z \backslash Y = 0) - E(Z \backslash Y = 1)} \tag{1}$$

É fácil obter estimativas para os termos no numerador da equação (1). Basta tomar a média dos escores em todo o teste dos que acertam e dos que erram o item. O mesmo não ocorre em relação ao denominador por se tratarem de variáveis latentes. Admite-se, então, que os respondentes que acertam o item são os que apresentam valores para Z superiores à Z_p, onde Z_p é tal que $\int_{Z_P}^{\infty} \frac{e^{\frac{-z2}{2}}}{\sqrt{2\pi}}\, dz = p$. Logo, sob essa hipótese:

$$E(Z \backslash Y = 1) = \frac{h(Z_p) = \int_{Z_p}^{\infty} \frac{e^{\frac{-Z_p^2}{2}}}{\sqrt{2\pi}}}{q} \quad \text{e,} \quad E(Z \backslash Y = 0) = \frac{-h(z_p)}{q}$$

Assim, de (1):

$$A = \frac{E(S \backslash Y = 0) - E(S \backslash Y = 1)}{-\frac{h(z_p)}{q} - \frac{h(z_p)}{p}}$$

Uma estimativa para a correlação de Pearson é dada, então, por:

$$\overline{\rho}_{bis} = \overline{\rho}_{sz} = \frac{\overline{S}_q - \overline{S}_p}{-\frac{h(Z_{\overline{p}})}{\overline{q}} - \frac{h(Z_{\overline{p}})}{\overline{p}}} \qquad \frac{1}{\overline{\sigma}_S}$$

onde $\overline{S}_q$ é o escore bruto médio para os que erram o item, $\overline{S}_p$ é o escore bruto médio para os que acertam o item, $h(z_{\overline{p}})$ é o valor da função de densidade normal padrão em $z_{\overline{p}}$, $\overline{p}$ é a proporção dos que acertaram o item no teste.

Finalmente, $\overline{\sigma}_S$ é o desvio padrão dos escores brutos obtidos no teste. Note-se que:

$$\overline{\rho}_{bis} = \frac{\overline{S}_q - \overline{S}_p}{\frac{-h(z_{\overline{p}})}{\overline{q}} - \frac{h(z_{\overline{p}})}{\overline{p}}} \quad \frac{1}{\overline{\sigma}_S} = \frac{1}{\overline{\sigma}_S \quad h(z_{\overline{p}})} \left(\frac{\overline{S}_p - \overline{S}_q}{\overline{p} + \overline{q}} \right) \overline{q} \quad \overline{p} =$$

$$= \frac{1}{\overline{\sigma}_S \quad h(z_{\overline{p}})} \left(\frac{\overline{S}_p - \overline{S}_q}{1} \right) \overline{q} \quad \overline{p} \ .$$

mas, $\overline{S} = \overline{p} \ \overline{S}_p + \overline{q} \ \overline{S}_q$, de tal forma que $\overline{S}_q = \dfrac{\overline{S} - \overline{S}_p \ \overline{p}}{\overline{q}}$ e, portanto:

$$\begin{aligned} \overline{\rho}_{bis} &= \frac{1}{\overline{\sigma}_S \quad h(z_{\overline{p}})} \left(\frac{\overline{S}_p - \overline{S}}{\overline{q}} \right) \overline{q} \quad \overline{p} = \left(\frac{\overline{S}_p - \overline{S}}{\overline{\sigma}_S} \right) \sqrt{\frac{\overline{p}}{\overline{q}}} \quad \frac{\sqrt{\overline{q} \quad \overline{p}}}{h(z_{\overline{p}})} \\ &= \overline{\rho}_{pb} \frac{\sqrt{\overline{p} \quad \overline{q}}}{h(z_{\overline{p}})}. \end{aligned} \tag{2}$$

A fórmula (2) expressa a relação entre a correlação bisserial e correlação ponto bisserial. Maiores detalhes em Machado [20].

3.9.2
Correlação Polisserial e Ponto Polisserial

Os conceitos de correlação ponto bisserial e bisserial podem ser estendidos para o caso de itens politômicos, os quais apresentam mais de duas categorias ordenadas de respostas $(T_0, T_1, ..., T_m), T_{K+1} \geq T_K$. A correlação ponto polisserial (ρ_{ppol}) é definida, simplesmente, como sendo a correlação de Pearson entre o escore bruto do teste (S) e o escore do item, medido segundo uma escala ordenada de inteiros cujas diferenças entre dois valores sucessivos seja sempre a mesma (por exemplo, (0, 1, 2, ..., m)). A correlação polisserial é definida com base na relação (2) da seguinte forma:

$$\bar{\rho}_{pol} = \bar{\rho}_{ppol} \frac{\bar{\sigma}}{\sum_{k=0}^{m-1} h(z_{\bar{p}_{k+1}})(T_{k+1} - T_k)} \tag{3}$$

onde $\bar{p}_{k+1}$ é a proporção dos que alcançaram o k-ésimo escore e $\bar{\sigma}$ é o desvio padrão dos escores alcançados no item.