

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Departamento de Estatística e Investigação Operacional

Modelos Rasch e Escalas de Likert e Thurstone na  
medição de atitudes

Luísa Margarida Antunes da Cunha

Mestrado em Probabilidades e Estatística

2007

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS

Departamento de Estatística e Investigação Operacional

Modelos Rasch e Escalas de Likert e Thurstone na  
medição de atitudes

Luísa Margarida Antunes da Cunha

Dissertação orientada pela  
Professora Doutora Luisa da Conceição S. Canto e Castro Loura

Mestrado em Probabilidades e Estatística

2007

E se um quilo de massambala não for um quilo de massambala?

À memória do meu pai

## Agradecimentos

Agradeço à Professora Doutora Luísa Castro Loura, pela receptividade que manifestou relativamente a um tema que lhe era estranho, pela curiosidade crescente, pelo incentivo e pela enorme disponibilidade.

À minha colega Dulce Gomes agradeço as tardes de trabalho em que construímos um programa que, depois, sozinha, ‘assassinei’.

Agradeço à companheira de percurso, a Regina Bispo, que me fez acreditar que era possível recomeçar, pelo incentivo constante, pela amizade e pela facilidade com que resolve todos os problemas que o computador introduz na minha vida.

Aos colegas de mestrado, em particular ao Rui e ao Zé, agradeço terem tornado esta caminhada conjunta num alegre convívio e desafio.

Agradeço aos colegas do Ispa que contribuíram com documentos que puseram à minha disposição: Glória, Pinto Coelho e Rui Bártolo, e aos outros que simplesmente me ouviram e apoiaram: Ângela, Olívia e Marta.

Obrigada, Zé Manel, pelo apoio nos momentos de desalento e simplesmente por existires.

Ao Gonçalo e ao Miguel, agradeço o Amor que vejo nos vossos olhos.

À João agradeço tudo. Sem o teu apoio, a tua confiança em mim, as nossas gargalhadas e as nossas lágrimas, não teria concluído a caminhada.

À Alexandria, agradeço as sopas quentinhas a meio da manhã e muito mais.

## Resumo

A medida tem um papel importante na sociedade moderna. Na psicologia, em particular, existem testes para medir toda a variedade de capacidades, aptidões, atitudes. Que qualidades métricas possuem esses testes, que capacidade diferenciadora lhes atribuir? A incompreensão das ciências duras relativamente à medida que se utiliza na psicologia reabriu o diálogo inesgotável entre as Ciências Empíricas e a Matemática, daí resultando a taxionomia proposta em 1946 por Stevens estabelecendo vários níveis de medição até se chegar à moderna Teoria Representacional da Medição.

São vários os tipos de testes que se usam em psicologia, tendo sido construídos de acordo com uma ou outra corrente de pensamento. Sendo a construção de cada tipo de teste inerente à natureza do que se está a medir, deu-se ênfase à forma de construção das escalas de atitudes usadas preferencialmente na Psicologia Social desde os anos 20 do século passado. Remonta-se assim, a Thurstone, Likert, Guttman e Osgood, encontrando em Rasch a sùmula de todos eles. Simulam-se, com base nos modelos Rasch, dados de uma escala de Thurstone, e de uma escala de Likert, para se compararem os resultados e avaliar da capacidade de discriminação de tais instrumentos.

## Sumário

Apresentação .....	7
Introdução .....	8
1. Medir em Psicologia .....	9
1.1. A proposta de Stevens.....	9
1.2. A Teoria representacional da Medida.....	13
1.3. Os níveis de medida e a psicologia moderna.....	16
2. Construindo escalas de atitudes .....	18
2.1. Escalas tipo Thurstone .....	20
2.2. Escalas tipo Likert.....	24
2.3. Escalas tipo Guttman.....	26
2.4. Escalas tipo Diferenciador Semântico .....	29
3. Modelação da atitude usando TRI .....	31
3.1. Escalas de atitudes usando o Modelo Rasch.....	31
3.2. Apresentação geral do modelo.....	32
3.3. O modelo Rasch dicotómico.....	36
3.4. Construção de uma escala de atitudes usando o modelo Rasch dicotómico.....	37
3.5. Estimação da atitude .....	43
3.6. O modelo Rasch politómico.....	45
3.7. Thurstone, Likert, Guttman e o modelo Rasch.....	49
3.8. O modelo seleccionado .....	51
<i>O modelo do co-seno hiperbólico .....</i>	<i>51</i>
<i>O modelo GUM.....</i>	<i>54</i>
4. Os resultados simulados .....	55
5. Análise dos resultados .....	57
Conclusão .....	65
Referências.....	67
Apêndice .....	71
Índice de figuras.....	74
Índice de fórmulas.....	75
Índice de tabelas .....	76
Índice geral .....	77

## **Apresentação**

Pareceu-nos oportuno realizar a presente Dissertação de Mestrado numa área em que os saberes da estatística estão ao serviço da psicologia. A medição está na génese da psicologia como ciência autónoma e está presente em várias áreas. A utilização de muitas escalas, umas mais divulgadas do que outras, umas mais cuidadas na construção do que outras, gerou a curiosidade necessária para focalizarmos o nosso trabalho nas escalas de atitudes.

Para se estudarem as propriedades métricas das escalas de atitudes há que explorar a questão dos fundamentos da medição, tentando compreender argumentos de dois lados: os que defendem que a medição é possível em psicologia e os seus detractores.

Uma das opções feitas à partida foi não explorar campos específicos da psicologia, assim, sabemos ser a atitude uma noção complexa com várias abordagens e várias expressões limitando-nos, neste trabalho, ao estudo das capacidades métricas das escalas.

Por várias vezes nos encontramos em encruzilhadas onde os caminhos pareciam igualmente válidos. Havendo vários modelos de construção de escalas de atitudes impôs-se partir à sua descoberta. Thurstone surgiu como o nome que se impõe ao estudar as escalas de atitudes, por ser o pioneiro: acreditando nessa medição, propondo um modelo e uma forma de construção do instrumento de medida. Likert, que simplifica o processo de construção do instrumento proposto por Thurstone, contribuiu para a grande explosão de utilização das referidas escalas. Perseguindo sempre o objectivo do estudo das capacidades métricas das escalas de atitudes compararam-se estes dois tipos de escalas com base em dados simulados.

A simulação dos dados conduziu-nos à descoberta das teorias mais modernas da medição em psicologia que se nos afigurou como indispensável aprofundar. Sendo uma área vastíssima em pleno desenvolvimento constituiu a parte mais estimulante do trabalho, que foi necessário delimitar e restringir ao papel que lhe havíamos atribuído à partida: um modelo para alicerce da simulação de resultados de escalas hipotéticas seja ao estilo de Thurstone, ou de Likert.

## Introdução

As escalas de atitudes, utilizadas em larga escala na psicologia social, conduziram ao estudo da sua construção e das teorias subjacentes à medição. Na construção de uma escala de atitudes estão envolvidos vários saberes que se entrecruzam: a psicologia que define os constructos e analisa a relação entre os processos mentais e as manifestações comportamentais que são susceptíveis de medição; a teoria da medição, dando fundamento à existência de uma escala, isto é, analisando as relações entre os processos mentais, por forma a legitimar as correspondências com sistemas numéricos mais ou menos complexos; a estatística que dá fundamento matemático aos instrumentos de medida e a computação, permitindo a elaboração de cálculos cada vez mais complexos.

Thurstone, em 1929, propõe uma teoria sobre a medição das atitudes baseada nos modelos da psicofísica. Se, por um lado, os instrumentos de medida se desenvolviam e permitiam o nascimento de novas ciências, por outro, debatia-se a qualidade e pertinência de se efectuarem essas medições. Em 1946, Stevens apresenta a proposta de classificação das escalas de medição em vários níveis, de acordo com o tipo de transformações válidas. Só com a teoria representacional da medição, que surge no final do século, se tem um enquadramento teórico para a construção de escalas. Axiomatizar a medição é um complemento aos métodos de avaliação, pois permite perceber quais as variáveis que podem e as que não podem ser medidas e permite fazer escolhas sobre a melhor forma de medir.

Ao modelo proposto por Thurstone outros se sucederam: Guttman e Osgood, são dois nomes indissociáveis das escalas de atitudes, mas é Likert, que apenas propôs uma forma alternativa de construção das escalas, que se associa mais frequentemente à medição na psicologia social. Os diversos tipos de escalas de atitudes são apresentados e comparados, nos seus fundamentos e resultados. Havendo actualmente dois grandes paradigmas na medição em Psicologia, a teoria clássica dos testes (TCT), em que se enquadram os modelos referidos, e a teoria da resposta ao item (TRI), as duas abordagens são comparadas, ainda que superficialmente. É na teoria de resposta ao item que se ancora a simulação de dados, em particular nos modelos Rasch, por serem modelos com plasticidade, que decorrem da proposta de Thurstone e por haver vários estudos que deles se socorrem para analisar a complexidade da medição de atitudes.

Evidenciam-se as capacidades métricas de escalas de atitudes do tipo Thurstone ou Likert, contribuindo para uma utilização melhor e mais consciente da medida em psicologia.



## 1. Medir em Psicologia

Medir: “determinar, avaliar por meio de instrumento ou utensílio de medida, ou algo usado como padrão “...“avaliar, aferir por meio de testes ou de outros procedimentos próprios da psicologia experimental” (Dicionário Houaiss de língua portuguesa)

Ao longo dos séculos foram surgindo problemas quer teóricos, quer práticos ligados à mensuração, isto é, à tentativa de associar números a objectos ou a fenómenos. Os géometras Gregos debateram-se com questões práticas de medida associando esta a contagens, depois à noção de proporcionalidade provando em consequência a incomensurabilidade do lado de um quadrado com a sua diagonal. Desenvolveram a teoria da medida para dar resposta a problemas práticos, mas também com fins teóricos de organização do conhecimento. Durante séculos a discussão da teoria ou das práticas de medida fez-se com base no legado de Euclides.

O desenvolvimento da teoria da medida a partir de um conjunto fundamental de conceitos e axiomas que se conseguiu na viragem do sec XIX não pôs termo às questões que a medição levanta quer pelos novos desenvolvimentos da Física, quer pelo surgimento de novas Ciências, em particular as Ciências do Homem. A Psicologia que se desenvolveu graças aos conhecimentos adquiridos na Estatística, vem a constituir um desafio na Teoria da Medição e o contributo de vários psicofísicos revela-se como fundamental (Luce, 2001).

Luce (1996) define dois períodos no desenvolvimento da relação entre as ciências empíricas e a teoria da medida desde a publicação em 1887 de “Counting and Measuring” de von Helmholtz em que se desenvolve a teoria da medição a partir de conceitos fundamentais e axiomas. A saber, uma primeira fase, baseada em medições da física que conduzem a representações numéricas, em que se tentou perceber a natureza empírica dos números, designando-a o autor por *Representação Clássica da Medida*. A segunda fase, a *Representação Contemporânea da Medida* começa nos anos 70 e continua até à actualidade, mas terá a sua génese na classificação das escalas de medida proposta por Stevens em 1946.

### 1.1. A proposta de Stevens

Em diversos artigos, Stevens retoma a definição de 1946, de quatro níveis de medida que se tornaram familiares a quem trabalha na área da Psicologia: *nominal, ordinal, intervalar e de razão*,

descrevendo uma hierarquia na estrutura matemática e classificando posteriormente os procedimentos estatísticos admissíveis em cada uma delas. As Estatísticas “permitidas” serão as que se mantêm invariantes para as transformações matemáticas que cada um dos níveis de mensuração admite. A classificação é feita com base nas propriedades do grupo de transformações associado a cada escala o que gerou alguns malentendidos por se confundir as transformações “permitidas” com mudanças de escala (Luce *et al.*, 1990)<sup>1</sup>. De acordo com Stevens (1968, p. 699):

é uma convenção útil definir a medição como a atribuição de números a objectos ou acontecimentos segundo uma regra rigorosa. Uma vez que há diferentes regras, há diferentes formas de medida; mas para cada uma delas obtém-se um certo grau de isomorfismo entre as relações empíricas dos objectos e as relações formais dos números.

A medida é tratada como um isomorfismo ( $s$ ) entre dois conjuntos: o dos objectos (no caso: observações psicológicas) que se passa a designar por  $\Theta$ , e um conjunto de categorias simbólicas que se designa por  $\mathcal{C}$ .

A natureza das diferentes escalas reside no conceito de invariância: “como se podem transformar os números sem que haja perda de informação empírica? (...) cada uma das quatro escalas possui um grupo característico de transformações permitidas” (Stevens, 1968, p.700). Será nominal qualquer escala que admita transformações biunívocas, se as transformações permitidas forem estritamente monótonas a escala será ordinal, as transformações afins são as de uma escala de intervalos e as do tipo  $f(x)=cx$  ( $c>0$ ) definem uma escala de razão.

As escalas nominais não requerem uma associação a valores numéricos, são invariantes para qualquer transformação que preserve a relação entre os objectos e as categorias. Neste nível de medida apenas se estabelecem relações de equivalência entre os objectos, a cada classe de equivalência de  $\Theta$  corresponde um símbolo de  $\mathcal{C}$ :

$$\forall i, j \in \Theta, i \sim j \Leftrightarrow s(i) = s(j) . \quad (1)$$

O conjunto  $\mathcal{C}$  pode ser constituído por números, mas apenas enquanto numerais (significantes sem significado matemático). Então uma escala nominal é um isomorfismo de  $(\Theta, \sim)$  em  $(\mathcal{C}, =)$

O exemplo que Stevens apresentou é o da numeração dos jogadores de futebol americano. Entre as escalas de atitude, são escalas nominais as que determinam apenas o grupo a que o sujeito pertence, como por exemplo a escala de atitudes face à polícia (Lima, 2000) que se apresentará mais à frente.

---

<sup>1</sup> “A unidade é uma convenção do cientista à cerca da dimensão e, mudá-la é uma questão meramente numérica; ao passo que uma mudança automórfica é uma substituição sistemática no próprio estímulo, não na sua representação”(p. 94).

A escala  $s$  é ordinal se

$$\forall i, j \in \mathcal{O}, i \succsim j \Leftrightarrow s(i) \geq s(j), \quad (2)$$

sendo o símbolo  $\succsim$  entendido como "é preferido a", "é mais alto que", "é mais difícil do que". Os elementos de  $\mathcal{O}$  são ordenáveis. Estabelece-se um isomorfismo entre  $(\mathcal{O}, \succsim)$  e  $(\mathcal{C}, \geq)$ . Pode-se então efectuar qualquer transformação  $f$  na escala que preserve a ordem dando origem a outra escala que mantém a mesma relação:

$$s(i) \geq s(j) \Rightarrow f(i) \geq f(j). \quad (3)$$

Então,  $f$  pode ser qualquer transformação estritamente monótona.

Um dos exemplos que Stevens apresentou foi a escala de dureza dos minerais. Há diversas escalas de atitudes com este nível de mensuração, em particular as construídas de acordo com os princípios de Likert como a escala de atitudes face ao ambiente desenvolvida por Soczka em 1983 (Lima, 2000) e que se apresentará mais à frente.

Uma escala de intervalos envolve a noção de distância. Os elementos de uma sequência mantêm sempre o mesmo espaçamento (Luce, 2001) mas a origem e o zero são arbitrários. "Os números podem servir para representar as diferenças ou as distâncias entre os elementos" (Stevens, 1968, p.700).

Qualquer transformação em  $\mathcal{C}$  deve preservar tanto a ordem como as diferenças entre os objectos, ou seja, apenas se devem efectuar transformações lineares porque só estas mantêm as distâncias:

$$f(x) = \alpha s(x) + \beta \quad (4)$$

A designação advém de as razões entre intervalos se manterem invariantes:

$$\frac{f(i) - f(j)}{f(k) - f(l)} = \frac{[\alpha s(i) + \beta] - [\alpha s(j) + \beta]}{[\alpha s(k) + \beta] - [\alpha s(l) + \beta]} = \frac{s(i) - s(j)}{s(k) - s(l)}. \quad (5)$$

Uma escala intervalar é um isomorfismo de  $(\mathcal{O}, \succsim, \circ)$  e  $(\mathcal{C}, \geq, +)$  o que levanta a questão da definição de  $\circ$  designada por concatenação. A concatenação entende-se de forma simples nalgumas medições físicas como por exemplo o comprimento: considerando duas varetas ( $a$  e  $b$ ) de comprimentos  $s(a)$  e  $s(b)$  colocadas uma após a outra, isto é, de maneira a que duas das extremidades se colem obtém-se uma vareta  $a \circ b$  cujo comprimento  $s(a \circ b)$  é a soma dos dois comprimentos  $s(a)$  e  $s(b)$ . Esta é uma das questões fundamentais da medida em Ciências Sociais, em particular na Psicologia, em que a concatenação (também referida como aditividade) não é válida ou

conduz até a contradições da própria definição de medida. este obstáculo pode ser contornado usando axiomas próprios definidos na *Representação Contemporânea da Medida* como se poderá ver em Krantz *et al.*, 1971.

Entre os exemplos apresentados por Stevens está a escala Celsius de temperaturas. O próprio Stevens (1946, p. 679) admite que “de facto, a maioria das escalas usadas extensa e efectivamente pelos psicólogos são escalas ordinais”.

São tradicionalmente consideradas escalas de atitude intervalares as construídas de acordo com as teorias de Thurstone, por exemplo, a escala de Thurstone para medir atitudes face à igreja (Lima, 2000), que será apresentada mais à frente.

As escalas de razão preservam os quocientes pois o zero é absoluto. O quociente de quaisquer dois valores é independente da unidade de medida. “Os números associados a valores numa escala de razão são “verdadeiros” números com um verdadeiro zero” (Siegel, 1956/1988, p. 32). Uma escala de razão é um isomorfismo com a estrutura aritmética:  $(\mathbb{R}_0^+, >, +, \times)$ .

Então, as transformações possíveis são as que satisfazem a condição:

$$f(x) = \alpha s(x) \text{ com } \alpha > 0. \quad (6)$$

A designação advém do facto de serem invariantes as razões entre valores da escala:

$$\frac{f(i)}{f(j)} = \frac{\alpha s(i)}{\alpha s(j)} = \frac{s(i)}{s(j)} \quad (\alpha > 0) \quad (7)$$

Constata-se então que à medida que aumenta a complexidade do nível de medida se perde liberdade nas transformações: a transformação logarítmica que se pode efectuar numa escala ordinal, ao deformar as distâncias não é permitida numa escala intervalar ou de razão, a transformação linear sendo permitida numa escala de intervalos, já o não é numa de razão pois não se pode deslocar o zero da escala (Velleman & Wilkinson, 1993).

A escala de temperaturas em graus Kelvin é um dos exemplos apresentados por Stevens. Entre as escalas de atitude nenhuma atingirá tal nível de mensuração, pois não se pode falar em ausência de atitude que se possa associar ao verdadeiro zero da escala.

Vários autores propuseram taxionomias alternativas, entre os quais Mosteller e Tukey (Velleman & Wilkinson, 1993), mostrando que os níveis de mensuração propostos por Stevens não cobrem todas as possibilidades de medições.

Apesar desta taxionomia não ser exaustiva, apesar de deixar, do ponto de vista formal, algumas questões em aberto, dir-se-á com Velleman & Wilkinson (1993) que é errado concluir que não tem valor. Efectivamente era necessário diferenciar as variáveis qualitativas (nominais e ordinais) das quantitativas (intervalares e de razão) e discutir a aplicabilidade das técnicas estatísticas aos diversos níveis de mensuração.

A terminologia de Stevens é tão universalmente difundida que se torna necessária quando se abordam as questões de medição em geral, e em particular no âmbito da psicologia. Contra os seus detractores Stevens argumentava que o facto de as sensações não poderem ser separadas em partes constituintes ou colocadas uma após a outra, obtendo-se ainda uma sensação (alusão à concatenação) não era argumento contra a possibilidade da sua medição pois

a distinção entre os aspectos formais, arbitrários, vazios lúdicos das matemáticas e as investigações das disciplinas “concretas” fez aparecer claramente que o domínio da medição se estendia a tudo onde o nosso engenho podia conceber regras sistemáticas permitindo associar números a coisas.

O sistema numérico tornou-se um modelo que se pode utilizar de forma arbitrária. É um modelo rico, certamente, de que se pode usar um ou outro dos aspectos da sua sintaxe para explicar uma ou outra das propriedades dos objectos ou dos acontecimentos. (Stevens, 1968, p.700)

Stevens tentava dar suporte à escala de sonoridade que acabara de criar e justificar a legitimidade da medição em psicologia com os mesmos critérios das escalas usadas noutras ciências mais desenvolvidas. “O trabalho que Stevens começou desenvolveu-se numa especialização da matemática que produz resultados que confirmam alguns aspectos que Stevens intuiu, e infirmam outros aspectos do seu trabalho” (Teghtsoonian, 2001). Neste esforço de sistematizar e dar forma matemática às diversas formas de medir encontram-se, entre outros, Krantz, Luce, Suppes e Tversky (1971; 1989; 1990) que se propuseram “organizar os resultados centrais de uma forma cumulativa” (1971, p. xviii).

## **1.2. A Teoria representacional da Medida**

Nos três volumes de *Foundations of Measurement* estão organizados os principais resultados matemáticos usados para construir representações numéricas de estruturas qualitativas, cobrem exemplos da física, da geometria, da utilidade e também da psicologia, e axiomatizam a teoria da medição. Dão portanto sentido às diversas formas de medir que se desenvolveram nas ciências empíricas. Retomam a taxionomia de Stevens dando uma definição matemática a cada um dos níveis de medida e discutindo algumas lacunas. Criam assim uniformidade nas várias formas de

medir permitindo que se tenha um corpo teórico onde as diversas ciências empíricas podem fundamentar as, também diversas, formas de medir.

A axiomatização da medição baseia-se em dois resultados:

- A representação: se uma estrutura de relação empírica satisfaz um determinado conjunto de axiomas, então pode ser construído um homomorfismo dessa estrutura para uma estrutura relacional numérica escolhida.

Sendo  $R_i$  ( $i = 1, \dots, m$ ) uma relação definida num conjunto  $A$  de objectos, designa-se  $\langle A, R_1, R_2, \dots, R_m \rangle$  por estrutura de relação empírica. Se  $S_i$  ( $i = 1, \dots, m$ ) é uma relação numérica em  $\mathbb{R}$ ,  $\langle \mathbb{R}, S_1, S_2, \dots, S_m \rangle$  será uma estrutura de relação numérica. Uma função  $s$  de  $A$  em  $\mathbb{R}$  é um homomorfismo se transforma cada  $R_i$  em  $S_i$  ( $i = 1, \dots, m$ ). Os autores justificam a escolha da designação de homomorfismo em lugar de isomorfismo porque  $s$  não é habitualmente uma relação de um para um: duas varetas terem o mesmo comprimento não quer dizer que as duas varetas sejam iguais.

“Um homomorfismo no conjunto dos números reais é frequentemente referido como uma escala na literatura da medição psicológica” (p. 9).

O conceito de transformações permitidas que Stevens usou, torna-se muito mais claro quando se faz a análise dos homomorfismos do que das escolhas arbitrárias no procedimento da medição. “Uma transformação  $f \rightarrow f'$  é permitida se e só se  $f$  e  $f'$  são ambos homomorfismos de  $\langle A, R_1, R_2, \dots, R_m \rangle$  na mesma estrutura numérica  $\langle \mathbb{R}, S_1, S_2, \dots, S_m \rangle$ ” (p. 12).

- A unicidade que determina o tipo de escala de medida resultante do processo de medição. A escolha do termo unicidade é justificada por Suppes & Zinnes (1963, p.10): “do ponto de vista da matemática a determinação do tipo de escala de medida resultante de um dado sistema de relações empíricas consiste em determinar a maneira como quaisquer dois sistemas numéricos estão relacionados quando usam a mesma relação numérica e são homomorfos de dado sistema empírico”.

Esta noção percebe-se melhor quando ilustrada com um exemplo. No caso do peso, as quatro afirmações seguintes são equivalentes (Suppes & Zinnes, 1963):

- (1) O peso mede-se numa escala de razão.

(2) A medição do peso é única para a mudança de unidade

$$\left( \frac{s_1(i)}{s_1(e_1)} = \frac{s_2(i)}{s_2(e_2)} \right), \text{ sendo } e_1 \text{ e } e_2 \text{ as unidades de medida}.$$

(3) A medição do peso é única para a transformação de semelhança ( $s_2(i) = \alpha s_1(i)$ ).

(4) Considerando um qualquer sistema empírico de medição do peso, então quaisquer dois sistemas numéricos que usem as mesmas relações numéricas e sejam homomorfos com o sistema empírico considerado estão relacionados por uma transformação de semelhança.

Um atributo é medido extensivamente<sup>2</sup> se verifica os axiomas de unicidade:

(1) Os números obtidos respeitam a relação de ordem:

$$\forall i, j \in A, i \succsim j \Leftrightarrow s(i) \geq s(j) \quad (8)$$

(2) Há aditividade em relação à concatenação:

$$\forall i, j \in A, s(i \circ j) = s(i) + s(j) \quad (9)$$

(3) Independentemente da escolha da unidade, as razões dos números são únicas:

Se  $n$  cópias de  $a$  podem ser concatenadas para se obter  $i$  e  $n'$  para se obter  $j$ , então

$$\frac{s(i)}{s(j)} = \frac{n}{n'} \quad (10)$$

Para que se tenha uma escala ordinal é preciso que se verifique o axioma (1), para uma escala intervalar o (1) e o (2), e para uma escala de razão, os três.

Já se viu que a grande controvérsia nas ciências sociais e do comportamento é se (e quando) se atinge o nível de medida intervalar. Krantz *et al.* (1971) demonstram (e exemplificam) que há, para além da medição extensiva (com os três axiomas referidos), a medição intensiva e a medição conjunta que poderão ser de nível quer de razão, quer intervalar, mas em que as relações de estrutura empírica não se estabelecem de forma evidente. “Noutros contextos empíricos, os axiomas podem ser bastante interessantes e não óbvios. Em tais casos, o desenvolvimento de escalas de medição está fortemente ligado à formulação e testagem de leis qualitativas apropriadas” (Krantz, 1971, p.13).

---

<sup>2</sup> A medida definida extensivamente é referida como sendo a medida fundamental. Na física existem muitos atributos que são derivados, pois são definidos, e medidos, à custa de outros atributos.

Axiomatizar a medição é um complemento aos métodos de avaliação, pois permite perceber quais as variáveis que podem e as que não podem ser medidas e permite fazer escolhas sobre a melhor forma de medir. Mais do que identificar o nível de medida de cada escala importa conhecer intrinsecamente o atributo e as relações que aí se podem estabelecer para se construir o instrumento de medida que fará corresponder a medição a números interpretáveis. “Por exemplo, na aptidão, inteligência ou atitudes sociais, os resultados de testes ou classificações numéricas são usualmente interpretados como medidas do atributo em questão. Mas na ausência de um homomorfismo bem definido está longe de ser claro como interpretar tais números” (Krantz, 1971, p. 33).

A identificação do nível de medida de uma escala faz-se portanto a montante, isto é, antes mesmo da sua construção. Os utilizadores de uma escala, que não a construíram, deveriam reportar-se ao nível de medida que o autor lhe atribui. Não havendo essa indicação, é tarefa árdua identificar o nível de medida de uma escala, sendo que alguns atributos são medidos sem que tenha havido o cuidado de atribuir propriedades métricas à escala.

### 1.3. Os níveis de medida e a psicologia moderna

Michell (2002) analisa alguns dos autores mais citados na área de tratamento de dados na psicologia para dar a conhecer o papel das escalas de medida definidas por Stevens na psicologia moderna, chamando a atenção para o facto de a taxionomia de Stevens ser largamente difundida enquanto que os recentes trabalhos de Luce, Krantz, Suppes, Tversky, Zinnes, ou Narens, entre outros, são completamente ignorados. Dá conta de que modernamente se continua a afirmar com Stevens que

de facto, a maioria das escalas usadas extensa e efectivamente pelos psicólogos são escalas ordinais. Para ser estritamente correcto, as estatísticas usuais envolvendo médias e desvios padrão não deveriam ser usadas nestas escalas (...). Por outro lado, (...) pode-se evocar uma espécie de confirmação pragmática: em inúmeras situações o seu uso conduziu a resultados frutuozos. (Stevens, 1946, p. 679)

Ora, no volume III dos *Foundations of Measurement* (Luce, 1990), partindo da ideia de que nem todo o enunciado quantitativo é forçosamente admissível enquanto descrição de um fenómeno ou de uma situação empírica, isto é, baseando-se no conceito de significado<sup>3</sup>, os autores reafirmam e provam que algumas estatísticas (ou tratamentos estatísticos) são desapropriadas para alguns níveis de mensuração e recomendam que se prefiram os tratamentos estatísticos adequados ao nível de medição.

---

<sup>3</sup> Em inglês *meaningfulness*: “requer que a forma de uma lei científica não seja alterada por uma mudança de unidades da escala de medição”(Falmagne, 2004, p. 1341)



A teoria de Stevens apenas levantava o problema, mas não fornecia ao investigador um guia explícito de como identificar o tipo de escala e foi (é) usada sem que se investigue cientificamente a questão do tipo de escala, dando origem a que se considere que a atribuição das propriedades de uma escala de intervalos ou de razão é sempre uma convenção (Michell, 2002).

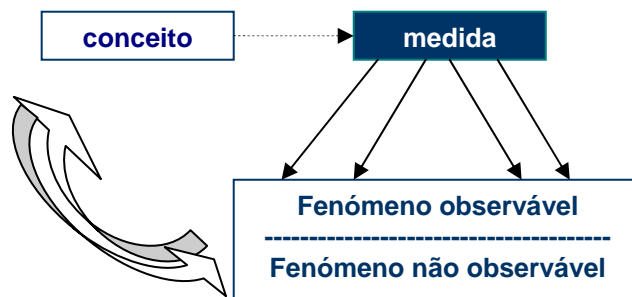
A partir dos anos 50, vários autores adoptaram a taxionomia de Stevens “talvez porque apareceu para constituir um guia simples e proteger analistas ingénuos de dados, de erros ao aplicar a estatística” (Velleman & Wilkinson, 1993, p. 5) e grande parte dos manuais de estatística para as Ciências Humanas referem os procedimentos adequados a cada um dos tipos de escalas de medida (cf. Siegel, 1956/1988). Recentemente, alguns dos packages de estatística exigem que o utilizador identifique o nível de medida de cada uma das variáveis antes de se proceder a qualquer tratamento estatístico estando vedadas algumas abordagens em concordância com as limitações preconizadas na taxionomia de Stevens.

O que é importante salientar é que o nível de medida não é inerente aos dados, está antes ligado à estrutura do atributo que se mede e também à experiência que se pretende efectuar (Velleman & Wilkinson, 1993).

O desafio para as novas gerações de psicólogos é deixarem de pretender que podem medir sem se preocuparem com a questão fundamental dos fundamentos da medição (Michell, 2002).

## 2. Construindo escalas de atitudes

Há três níveis a reter quando se pensa em medir em psicologia: o conceito teórico, o observável e o instrumento de medida, definindo assim três campos de acção: o campo teórico, o empírico e o metodológico.



A definição do conceito teórico é fundamental, alicerçando-se numa ou noutra corrente de pensamento da Psicologia aparecendo até definições operacionais, ou seja, que se apoiam no instrumento de medida<sup>4</sup>.

A necessidade de medir pelo menos o observável conduz à construção de instrumentos de medida que se usam em psicologia, sejam os testes de aptidões, de personalidade, de conhecimentos ou as escalas de atitude. O processo de medida é um sistema de correspondência entre o campo teórico e o campo empírico que privilegia, de entre os fenómenos observáveis, apenas os que são mais relevantes para o conceito.

Se por um lado se estuda a natureza do que se quer medir, por outro desenvolvem-se técnicas para a medir. Aparecem claramente três campos de investigação: um de cariz psicológico procurando definir o conceito e os fenómenos observáveis que o revelam, outro de cariz matemático, procurando definir a medida de uma forma alargada e precisa, permitindo a construção de escalas (campo metodológico) que traduzam os fenómenos observáveis em números. As três frentes estão em constante diálogo, influenciando-se e promovendo os avanços e recuos próprios do conhecimento. A Teoria Representacional da Medição desenvolve novas generalizações matemáticas aperfeiçoando a compreensão matemática do que está subjacente à medida, dando informação às ciências das diversas formas de medição, contribuindo assim para o seu desenvolvimento.

---

<sup>4</sup> "L'intelligence, c'est ce que mesure mon test" atribuído a Binet segundo Pierre Pichot (1981, p. 14)

A construção de um instrumento de medida exige um profundo conhecimento do atributo e das relações que se podem estabelecer, como se viu anteriormente.

Na construção de uma escala de atitudes também os três campos estão envolvidos. A noção de atitude tem evoluído ao longo do tempo, mas tem à partida um conteúdo que lhe é atribuído espontaneamente que decorre da palavra existir na linguagem corrente.

Uma atitude é muitas vezes definida como uma tendência para reagir relativamente a uma designada classe de estímulos tais como um grupo étnico ou nacional, um hábito ou uma instituição. (...) Na prática actual, o termo 'atitude' é frequentemente associado ao estímulo social e às respostas matizadas emocionalmente. Também envolve frequentemente juízos de valor. (Anastasi, 1990, p. 405)

A atitude é um constructo que se tenta avaliar através da expressão favorável ou desfavorável, o que indica que as atitudes não são directamente observáveis. (Lima, 2000).

Por trás do observável escondem-se por vezes as verdadeiras razões e atitudes dos sujeitos e por isso se desenvolveram várias técnicas (umas mais deontológicas do que outras) para se ter acesso a esse mundo que insistimos em preservar do olhar dos outros (Garcia-Marques, 2000).

A atitude intervém em todas as relações do sujeito com o exterior: a nível físico, social, abstracto... e estas relações são observáveis. São estas relações que se tornam o objecto de medição quando se constrói uma escala de atitudes.

Apesar de não haver uma distinção clara entre opinião e atitude, em termos metodológicos há diferenças. Os inquéritos de opinião são constituídos por diversas questões que podem estar ou não relacionadas e destinam-se preferencialmente a identificar as causas de determinadas escolhas. As escalas de atitudes geram um score total que indica a intensidade do sentir do sujeito relativamente a um determinado objecto.

A maior parte das escalas de atitudes foram desenvolvidas para serem usadas em projectos de investigação particulares. Algumas foram desenhadas para a investigação de atitudes e moral de funcionários. Outras foram usadas para aceder aos resultados de programas educacionais ou de treino. (...) Uma das aplicações mais importantes da medida das atitudes encontra-se na investigação em psicologia social. Praticamente todos os textos de psicologia social contêm secções referentes às atitudes e à sua medição. (...) Foram publicadas relativamente poucas escalas de atitudes, se bem que a maioria esteja descrita de forma exaustiva na literatura de investigação. (Anastasi, 1990 p. 407)

É longo o caminho percorrido na construção de tais escalas. Referindo apenas as questões metodológicas e as limitações que a matemática impõe, apresentam-se várias abordagens dessa construção, desde a preconizada por Thurstone em 1929, passando por Likert, Guttman e Osgood.

## 2.1. Escalas tipo Thurstone

As escalas tipo *Thurstone* são constituídas por um conjunto de frases (itens) em relação às quais o sujeito avaliado deve manifestar o seu acordo ou desacordo.

Mede-se a atitude do sujeito fazendo a média ponderada dos itens em que houve acordo. Os factores de ponderação são calculados na fase de construção da escala.

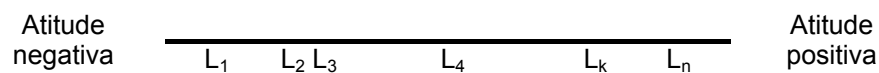
Exemplo: Escala de Thurstone para medir atitudes face à igreja (Lima, 2000).

<b>ESCALA DE THURSTONE PARA MEDIR ATITUDES FACE À IGREJA (THURSTONE E CHAVE, 1929)</b>	
<p>Este estudo pretende analisar as atitudes face à igreja. Irá encontrar 24 frases que expressam atitudes diferentes face à igreja. Faça um visto (v) se concordar com a frase. Faça uma cruz (x) se discordar da frase. Se não se conseguir decidir pode marcar a frase com um ponto de interrogação (?). Isto não é um exame. Nesta questão as pessoas têm posições diferentes acerca do que está certo ou errado. Por favor, indique a sua atitude fazendo um visto quando concordar e uma cruz quando discordar.</p>	
Valor da Escala*	Item
3.3	<input type="checkbox"/> Gosto da minha igreja porque há lá um espírito de amizade.
5.1	<input type="checkbox"/> Gosto das cerimónias que se realizam na minha igreja, mas não sinto a falta delas quando estou fora.
8.8	<input type="checkbox"/> Respeito as crenças dos membros de qualquer igreja, mas acho que é tudo «treta».
6.1	<input type="checkbox"/> Sinto necessidade da religião, mas não encontro o que quero em nenhuma igreja.
8.3	<input type="checkbox"/> Acho que os ensinamentos da igreja são demasiado superficiais para terem algum significado social.
11.0	<input type="checkbox"/> Acho que a igreja é um parasita da sociedade.
6.7	<input type="checkbox"/> Acredito na sinceridade e na bondade sem nenhuma cerimónia religiosas.
3.1	<input type="checkbox"/> Não compreendo os dogmas ou credos da igreja, mas acho que a igreja me ajuda a ser mais honesto e respeitável.
9.6	<input type="checkbox"/> Acho que a igreja é um obstáculo à religião, porque ainda depende da magia, da superstição e de mitos.
9.2	<input type="checkbox"/> Acho que a igreja tenta impor uma série de dogmas ultrapassados e de superstições medievais.
4.0	<input type="checkbox"/> Quando vou à igreja gosto de assistir a uma bela cerimónia ritual acompanhada de boa música.
0.8	<input type="checkbox"/> Acho que a igreja perpetua os valores que são mais importantes para a sua filosofia de vida.
5.6	<input type="checkbox"/> Às vezes acho que a igreja e a religião são necessárias, mas outras vezes duvido disso.
7.5	<input type="checkbox"/> Acho que se gasta demasiado dinheiro com a igreja, em comparação com os benefícios que daí derivam.
10.7	<input type="checkbox"/> Acho que a igreja organizada é um inimigo da verdade e da ciência.
2.2	<input type="checkbox"/> Gosto de ir à igreja, porque fico com ideias importantes para reflectir e cheio de bons pensamentos.
1.2	<input type="checkbox"/> Acho que a igreja é um importante agente de promoção tanto da rectidão individual como social.
7.2	<input type="checkbox"/> Acho que as igrejas estão demasiadamente divididas em facções para serem uma força de rectidão.
4.5	<input type="checkbox"/> Acredito no que a igreja diz, mas com reservas mentais.
0.2	<input type="checkbox"/> Acho que a igreja é a maior instituição da América actual.
4.7	<input type="checkbox"/> Eu tenho descuidado as minhas relações com a religião e a igreja, mas não gostava que a minha posição fosse geral.
10.4	<input type="checkbox"/> A igreja representa superficialidade, hipocrisia e preconceito.
1.7	<input type="checkbox"/> As cerimónias religiosas inspiram-me e ajudam-me a dar o meu melhor durante o resto da semana.
2.6	<input type="checkbox"/> A igreja faz com que os negócios e a política tenham uma posição importante, que de outra forma não teriam.
<p>* O valor da escala encontra-se aqui com fins meramente ilustrativos. Na versão a apresentar aos respondentes estes valores não estão impressos.</p>	

Em 1929, Thurstone e Chave elencaram os princípios de construção de escalas de atitudes baseados na psicofísica demonstrando assim a possibilidade de medir atitudes (Andrich, 1998). O enfoque na construção da escala é dado à selecção objectiva de frases em relação às quais os sujeitos vão manifestar o seu acordo ou desacordo. À semelhança da psicofísica, mede-se a sensação que um determinado estímulo produz. O método foi sendo alvo de sucessivos melhoramentos, quer pelos próprios Thurstone e Chave, quer por outros autores. Ainda hoje se constroem escalas de atitude fiéis a esses princípios e os mais recentes trabalhos em psicometria vão de encontro a muitas das conjecturas elaboradas por Thurstone (cf.3.7.).

Para se acompanhar a metodologia de construção de tal tipo de escalas, bem como os pressupostos teóricos dessa construção teve-se em conta os textos de Ghiglione e Matalon (1985/1993) e Alexandre (1971).

Admite-se que há um continuum psicológico na atitude, desde uma atitude negativa até uma atitude positiva e que os vários estímulos (itens) capazes de medir a atitude não estão todos à mesma distância quer entre si, quer em relação aos extremos. Assim, considerando que os estímulos são os itens  $i$  ( $i = 1, \dots, n$ ) e  $L_i$  representa a localização do item no continuum da atitude, podemos admitir o esquema:



Decorre deste postulado que se devem atribuir pesos diferentes a cada um dos itens.

A construção de uma escala deste tipo é feita em duas fases: a primeira em que se escolhem os estímulos (itens, frases) que poderão medir a atitude; a segunda em que se atribuem pesos a cada um dos estímulos.

As frases escolhidas devem cobrir a maior diversidade possível de posições (estímulos) que se relacionem com a atitude a medir.

Ao atribuir o grau de proximidade entre os diferentes itens constrói-se uma escala intervalar.

*Forma de construção de uma escala de Thurstone (Alexandre, 1971):*

1) Elaboração de uma lista de frases que manifestem opinião relativamente a assuntos directamente relacionados com a atitude que se pretende medir.

Estas frases devem ser curtas, claras e cobrir todas as posições possíveis face à atitude que se está a medir. Devem ser frases que manifestam posições desde muito favoráveis até às muito desfavoráveis.

2) Selecciona-se de forma criteriosa um grupo de sujeitos (vários autores sugerem a dimensão do grupo, indo essa sugestão dos 50 aos 300) que se designarão por juízes e a quem se pede para se abstrair de manifestar a sua opinião de acordo ou desacordo com a afirmação feita, mas apenas que julgue a frase num continuum de opinião do muito desfavorável ao objecto relacionado com atitude até ao muito favorável.

Essa classificação pode ser feita considerando cada par de itens (comparações emparelhadas ou julgamento comparativo) ou pedindo que coloquem as frases numa escala de 11 pontos que varia da categoria “completamente desfavorável” à “completamente favorável” (intervalos sucessivos ou julgamento categorial).

Com base nesta ordenação são retirados alguns itens de acordo com critérios que a seguir se explicitam.

3) A cada item seleccionado é atribuído um peso, esse peso é calculado a partir dos resultados obtidos anteriormente e considerando que a atitude se distribui normalmente.

Começando pelas comparações emparelhadas apresentam-se com pormenor os cálculos que se efectuam para se seleccionarem os itens e para se chegar ao peso a atribuir a cada um deles.

Este procedimento traduz a ideia da existência de um continuum na atitude que corresponde ao continuum psicológico que se admite ter distribuição normal na população. Pretende-se saber a distância entre  $L_i$  e  $L_j$  ( $i, j = 1, \dots, n$ ), sendo o item  $i$  mais próximo da atitude negativa do que o item  $j$ . Estima-se essa distância postulando que os juízes estabelecerão uma ordem para  $i$  e  $j$  de tal forma que a distribuição da selecção da localização de qualquer item  $i$  ( $A_i$ ) será uma distribuição normal, de média  $L_i$  e desvio padrão 1, sendo  $L_i$  a mediana da distribuição empírica. Então, por consulta da tabela de distribuição normal estima-se a distância entre  $L_i$  e  $L_j$  ( $d(L_j, L_i)$ ) com base na percentagem de juízes que ordenou os itens de forma contrária.

Exemplo: Seja  $p$  a proporção de juízes que colocou  $L_i$  à direita de  $L_j$ . Atendendo a que  $A_i \sim N(L_i, 1)$ ,  $A_j - A_i \sim N(L_j - L_i, \sqrt{2 - 2\text{cov}(A_i, A_j)})$ ,  $\hat{P}[A_j - A_i < 0] = p \Leftrightarrow \hat{P}[Z < L_i - L_j] = p$  e então

$\phi(L_i - L_j) = p$ , donde  $\hat{d}(L_i, L_j) = \phi^{-1}(p)$ , supondo adicionalmente (Ghiglione, 1993) que o desvio padrão da distribuição da diferença é unitário ( $\sqrt{2 - 2\text{cov}(A_i, A_j)} = 1$ ).

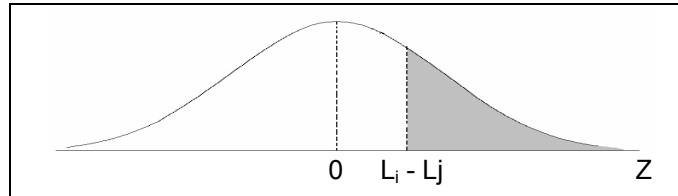


Figura 1: Localização das medianas

Repete-se o procedimento para o par  $i$  e  $k$ , e assim por diante. Fixando a localização de um dos itens no ponto de atitude negativa e atribuindo-lhe o valor zero, tem-se a escala construída e a distância de cada um dos itens à origem é a ponderação que se lhe atribui.

Este método é muito fastidioso para os juízes que teriam de ordenar muitos pares de itens ( $C_2^n = \frac{n(n-1)}{2}$ ). Assim, Thurstone prevê a possibilidade de pedir aos juízes a atribuição de proximidade do muito desfavorável ao muito favorável numa escala de 11 pontos tornando-se o cálculo dos pesos a atribuir a cada item mais complexo.

Começa por se construir uma tabela de frequências relativas (percentagens), ou seja, contabiliza-se o número de juízes que colocou o item em cada uma das classes de ordem. Eliminam-se os juízes que colocam muitos itens na mesma posição e há que notar que as posições 2, 6 e 10 são sempre as que apresentam maior frequência.

Na fase seguinte da construção, eliminam-se os itens que aparecem nos extremos opostos da escala de opinião dos juízes, isto é, que apresentam grande variabilidade, pois serão frases que encerram ambiguidade. Eliminam-se também os que têm pouca variabilidade, pois não discriminarão opiniões. Para isso usa-se como coeficiente de ambiguidade a amplitude semi-interquartílica.

A distribuição empírica, considerando apenas dois juízes, será do tipo:

	1	2	3	4	5	6	7	8	9	10	11	
Atitude negativa	$L_1$	$L_2$	$L_4$	$L_6$			$L_i$			$L_j$		Atitude positiva
	$L_3$		$L_5$									
			$L_7$									
	$L_2$	$L_4$	$L_1$	$L_7$		$L_i$	$L_j$					
	$L_3$		$L_5$									
			$L_6$									

Figura 2: Exemplo de localizações de itens atribuídas por dois juízes

Para cada item  $k$  ( $k = 1, \dots, n$ ) constrói-se a tabela de frequências referente à classificação atribuída pelo conjunto dos juízes a esse item e faz-se um ajustamento à distribuição normal com desvio padrão unitário e média  $L_k$ , em que  $L_k$  é a mediana da distribuição. Obtêm-se, assim, os limites de cada uma das 11 categorias de resposta:  $Z_{ki}$ , ( $i = 0, \dots, 11$ ), para o item  $k$ . Repete-se o procedimento para os  $n$  itens da escala. Considerar-se-á o limite de cada uma das categorias a média destes

valores, por exemplo:  $Z_4 = \frac{\sum_{k=1}^n Z_{k4}}{n}$ . Está desta forma dividido o continuum da atitude em 11 intervalos.

É preciso localizar o item  $k$  nesses intervalos para se lhe atribuir a sua localização final  $L_{fk}$ . Na tabela de frequências do item  $k$  substituem-se as categorias de resposta pelos intervalos  $[Z_{i-1}, Z_i[$  ( $i = 1, \dots, 11$ ) e  $L_{fk}$  será a mediana dessa distribuição. Assim, identificado o intervalo em que se situa a mediana,  $[Z_{i-1}, Z_i[$ ,

$$L_{fk} = Z_{i-1} + \frac{0.5 + F_{i-1}}{f_i} (Z_i - Z_{i-1}), \quad (11)$$

em que  $F_i$  se refere às frequências relativas acumuladas e  $f_i$  às frequências relativas desse intervalo.

## 2.2. Escalas tipo Likert

Uma escala tipo *Likert* é composta por um conjunto de frases (itens) em relação a cada uma das quais se pede ao sujeito que está a ser avaliado para manifestar o grau de concordância desde o *discordo totalmente* (nível 1), até ao *concordo totalmente* (nível 5, 7 ou 11).

Mede-se a atitude do sujeito somando, ou calculando a média, do nível seleccionado para cada item<sup>5</sup>.

Exemplo: Escala de atitudes face ao ambiente desenvolvida por Soczka em 1983 (Lima, 2000).

---

<sup>5</sup> Inicialmente Likert propôs um método de cálculo do resultado final através de uma média ponderada das respostas dadas, atribuindo em cada item um peso a cada nível de concordância expresso pelo sujeito. Esses pesos são calculados de acordo com a assumpção de que está subjacente à atitude uma distribuição Normal e tomando como referência as frequências de resposta aquando da aferição da escala (Ghiglione, 1993, p. 292).



**EXEMPLO DE ESCALA DE LIKERT:  
ESCALA DE ATITUDES FACE AO AMBIENTE (SOCZKA, 1983)**

Em seguida encontra algumas frases relativas a questões ambientais. Para cada uma delas indique, por favor, o seu grau de concordância, utilizando a seguinte escala:

<i>Concordo totalmente</i> .....	<b>5</b>
<i>Concordo</i> .....	<b>4</b>
<i>Nem concordo nem discordo</i> .....	<b>3</b>
<i>Discordo</i> .....	<b>2</b>
<i>Discordo totalmente</i> .....	<b>1</b>

- 1** Grande parte do que se diz acerca da crise ecológica que o mundo atravessa é exagerado e alarmista.
- 2** Os males provocados pela construção de uma central nuclear no nosso país não compensam os seus eventuais benefícios.
- 3** A comida que se consome nos países industrializados já não oferece segurança para a saúde por estar contaminada por pesticidas e doenças.
- 4** O mundo acabará no prazo de algumas décadas se não se travar imediatamente o crescimento industrial e tecnológico de todos os países industrializados.
- 5** Todas as modernas sociedades urbanas e industriais são antinaturais: só pelo regresso à terra e à natureza é que a humanidade poderá sobreviver e evitar o caos ecológico.
- 6** As pessoas que protestam contra o crescimento económico e industrial em nome da ecologia são idealistas que não sabem o que dizem e não têm o sentido das realidades.

**Nota:** A cotação da escala é dada pela média das respostas dadas, invertendo os itens 1 e 6. Valores elevados (maiores que 4) indicam atitudes muito ambientalistas e valores baixos (inferiores a 2) indicam atitudes pouco ambientalistas.

No resumo do artigo de 1934, Likert, afirma: “É apresentado um método simples de atribuir scores em Escalas de Atitude de Thurstone, que não envolve o uso de um grupo de juizes e ainda se baseia em várias amostras para ser consistentemente mais fidedigno do que o método original.” Não se muda de paradigma mas apenas de forma de construção do instrumento de medida.

*Forma de construção de uma escala de Likert (Lima, 2000):*

- 1) Elaboração de uma lista de frases que manifestem opiniões radicais (claramente positivas ou negativas) em relação à atitude que se está a estudar tendo o cuidado de cobrir as diferentes vertentes que se relacionam com o assunto.

2) Para validar a escala usa-se uma amostra representativa da população em que se pretende aplicar a escala de atitude. É pedido aos sujeitos que manifestem o seu grau de concordância com cada uma das frases numa escala de 3, 4, 5, 7 ou 11 pontos<sup>6</sup>.

O estudo que se fará destes resultados conduzirá à eliminação de itens. Começam por se eliminar os itens que não apresentem distribuição aproximadamente Normal o que conduz também à eliminação dos itens que não discriminam a opinião. Posteriormente eliminam-se os itens que não estão relacionados com a atitude, o que na prática consiste em eliminar os itens que não garantem boa consistência interna da escala.

O processo de escolha dos itens envolve o cálculo de um coeficiente de correlação item - teste. Obter-se-ão tantos coeficientes de correlação quantos os itens e sempre que o seu valor for próximo de zero elimina-se o item.

O conjunto final de itens deve ser sempre analisado para se perceber se as diferentes vertentes da atitude que se queriam avaliar continuam presentes. A escala só está pronta para ser aplicada quando está garantida a objectividade, a fidelidade, a validade e a consistência da escala.

### **2.3. Escalas tipo Guttman**

Uma escala tipo *Guttman* ou Cumulativa é composta por um conjunto de frases (itens) organizadas de forma hierárquica e em relação aos quais se pede ao sujeito avaliado que manifeste a sua concordância ou não.

O resultado de cada sujeito é dado pelo nº de respostas positivas, de acordo com os padrões de resposta preestabelecidos (Anderson, 1990<sup>b</sup>).

Exemplo: Escala de Bogardus para estudar as atitudes face ao racismo (Lima, 2000).

---

<sup>6</sup> Likert preconizava uma escala de 5 pontos, mas actualmente, na mais diversa bibliografia preconiza-se a utilização de escalas de 3, 4, 7 ou 11 pontos, alegando a falta de poder discriminatório de cada sujeito quando a escala tem muitas possibilidades de resposta, ou inversamente, alegando que só com muitos pontos a escala se assemelha ao continuum da nossa opinião, ou traduzindo os resultados de diversas experiências que demonstram que há tendência para se responder na classe central devendo por isso evitá-la, etc.



conjunto de 5 itens, uma escala de Guttman perfeita forneceria a seguinte matriz de possíveis

resultados: 
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$
, isto é, quem concorda do item 3 concorda com todos os anteriores;

discordando do item 3, discorda-se do 4 e do 5. Assim teria pontuação 1, quem discordasse de todos os itens, a pontuação 2 atribuir-se-ia a quem concordasse apenas com o item 1, a pontuação 3 seria atribuída a quem concordasse com os itens 1 e 2, e assim por diante. Como não há escalas perfeitas e o ser humano tem características de unicidade, o que se define então são padrões de resposta, passando a considerar que todos os itens são equivalentes. Assim, por exemplo, terá score 2, o sujeito com o padrão 10001, tal como o sujeito que responda 10100, ou ainda aquele que, como se esperava, responda 11000.

*Forma de construção de uma escala de Guttman (Matalon, 1978):*

- 1) Elaboração de uma lista de frases relativas a um único tema, muito restritas e com um carácter muito repetitivo.
- 2) A partir das respostas dadas numa amostra e utilizando o escalograma procede-se à eliminação de itens que não são ordenáveis (apresentam “1” à direita de “0”) ou que se revelam dependentes uns dos outros e estabelece-se a ordem de apresentação dos itens seleccionados. As frases serão então ordenadas da mais negativa até à mais positiva, de tal forma que, a concordância com uma delas, conduz tendencialmente à concordância com todas as anteriores.

Depois de escolhida a ordem de apresentação dos itens é ainda conveniente avaliar a reprodutibilidade, isto é, se o número de respostas desajustadas ao modelo hierárquico (erros), é aceitável. O índice de reprodutibilidade deve ser superior a 0.9 e é de cálculo fácil ( $1 - (\text{n}^\circ \text{ de erros} / \text{n}^\circ \text{ de padrões de resposta possíveis})$ ), o índice de escalabilidade, de cálculo mais elaborado, e que avalia se a escala tem as propriedades de uma escala de Guttman, deve ser superior a 0.60 (Anderson, 1990). Por vezes as frases parecem encadear-se de forma cumulativa, mas só os valores destes dois índices permitirão dar suporte, ou não, a essa aparência.

- 3) A partir da matriz das respostas possíveis estabelece-se o valor da atitude correspondente. Obtém-se uma escala ordinal com resultados de 1 a  $n+1$ , se  $n$  for o número de itens que integram a escala.

Apesar das dificuldades para desenvolver uma escala Guttman, tais escalas têm muitas vantagens em relação às Thurstone ou Likert. Primeiro (...) é possível determinar todo o padrão de respostas aos itens a partir de um só score total. Assim, é possível ter mais informação sobre a natureza da atitude do inquirido. Segundo, a natureza cumulativa das escalas de Guttman torna possível avaliar a mudança de atitude. Um aumento de um ponto no score total significa que o inquirido subiu no continuum da atitude um item. A mudança torna-se aditiva.

Em resumo, então, as escalas de Guttman são difíceis de construir mas extremamente úteis como medidas de atitude. (Anderson, 1990<sup>b</sup>, p. 334)

## 2.4. Escalas tipo Diferenciador Semântico

Uma escala tipo *Diferenciador Semântico* é composta por um conjunto de pares de adjectivos com significados o mais opostos possível, entre os quais se estabelece uma escala de sete pontos. É pedido ao sujeito avaliado que assinale na referida escala o que sente relativamente ao objecto de estudo para cada par de adjectivos.

A atitude do sujeito é dada pela soma dos postos atribuídos em cada par de adjectivos.

Exemplo: Escala de atitudes face à polícia (Lima, 2000).

<b>EXEMPLO DE ESCALA TIPO DIFERENCIADOR SEMÂNTICO: ESCALA DE ATITUDES FACE À POLÍCIA</b>		
<p>Pretendemos saber a sua opinião sobre a polícia. Encontra a seguir uma série de adjectivos opostos, e pedimos-lhe para assinalar a sua posição nos espaços que estão entre os dois. Use a casa do meio quando achar que nenhum dos adjectivos se aplica, ou se a sua posição for média:</p>		
<b>POLÍCIA</b>		
Boa	_ : _ : _ : _ : _ : _ : _	Má
Simpática	_ : _ : _ : _ : _ : _ : _	Antipática
Honesta	_ : _ : _ : _ : _ : _ : _	Desonesta
Justa	_ : _ : _ : _ : _ : _ : _	Injusta
Agradável	_ : _ : _ : _ : _ : _ : _	Desagradável
Prestável	_ : _ : _ : _ : _ : _ : _	Cruel
+3 : +2 : +1 : 0 : -1 : -2 : -3		
<p><b>Nota:</b> A cotação da escala é pela soma das respostas dadas, de acordo com a pontuação indicada na linha de baixo. Valores positivos indicam atitudes favoráveis face à polícia, enquanto que valores negativos indicam o contrário.</p>		

As escalas de Diferenciador Semântico foram introduzidas por Osgood em 1952 na sequência de estudos desenvolvidos no quadro de teorias da aprendizagem. Ao estudar vários pares de adjectivos opostos, identificaram-se através da análise factorial, três factores: avaliação, potência e acti-

vidade. Em que, BOM – MAU e AGRADÁVEL – DESAGRADÁVEL constituem exemplos de pares de adjectivos antagónicos que integram a dimensão avaliação. A dimensão actividade pode ser ilustrada com os adjectivos do tipo RÁPIDO – LENTO e ACTIVO – PASSIVO. Sendo da esfera da potência os pares do tipo RIJO – MACIO, GRANDE – PEQUENO (Lima, 2000).

Considerando que a atitude de cada sujeito se pode localizar num continuum, pode-se medir a atitude face a qualquer objecto tomando sempre os mesmos adjectivos seleccionados entre os da dimensão avaliativa.

*Forma de construção de uma escala de tipo Diferenciador Semântico (Lima, 2000):*

- 1) Elaboração de uma lista de pares de adjectivos opostos, pertencentes à dimensão avaliativa e que se apliquem ao objecto em estudo.
- 2) Procede-se à aferição da escala usando uma amostra representativa da população. A partir das respostas dadas eliminam-se os pares que não estão significativamente correlacionados com o resultado total, bem como aqueles que não discriminam os sujeitos.

### **3. Modelação da atitude usando TRI**

Para avaliar as capacidades métricas das escalas de atitudes procedeu-se à simulação de resultados, para se poder averiguar quer a consistência das medições, quer da importância da análise, na prática, do nível de medição. Para obter os resultados numa escala de Thurstone e numa escala de Likert havia que encontrar um modelo onde ancorar a simulação. Os modelos propostos na Teoria da Resposta ao Item (TRI) ou *Item Response Theory* (IRT) pareceram responder a essa necessidade.

TRI, mais do que uma técnica ao serviço da medida, é uma interpretação da realidade psicológica, são propostas de modelação de vários constructos, sejam eles capacidades, aptidões, atitudes, interpretando-os e integrando-os na sua multidimensionalidade. A complexidade dos modelos aumenta com a proximidade da realidade. Vários modelos foram já estudados e só van der Linden (1997) reporta 27 deles. Foram modelos desenvolvidos a partir de 1960, com um grande boom nos anos 80, mas “algumas das famílias de modelos (...) não têm qualquer interesse para a estimação de parâmetros de capacidade, são apresentados essencialmente como veículos para especificar hipóteses concorrentes relativamente, por exemplo à estratégia de *Problem-Solving*, composição de competências, ou crescimento cognitivo” (van der Linden, 1997, p.22). Os modelos TRI, estendem-se ao conhecimento das capacidades e das estratégias que cada sujeito disponibiliza para produzir determinado resultado.

Alguns dos modelos de TRI foram desenvolvidos exactamente para as escalas de atitudes, sendo portanto modelos que dão conta da forma como se pensa que se estruturam as atitudes. Assim, começa-se por fazer uma apresentação do modelo geral de TRI passando a seguir para as particularidades do modelo Rasch dicotómico e do politómico. Embora não se utilize neste trabalho dados fornecidos por uma escala construída de acordo com o modelo, apresenta-se uma escala calibrada no modelo Rasch e a forma de a utilizar. Também se acompanha passo a passo a construção de uma tal escala. De entre os vários modelos foi escolhido o *Graded Unfolding Model* (GUM) (Roberts, 1995), como modelo subjacente aos dados gerados em simulação, que é apresentado e analisado.

#### **3.1. Escalas de atitudes usando o Modelo Rasch**

Uma escala de atitudes construída de acordo com o modelo Rasch é constituída por um conjunto de frases (itens) relativamente às quais é pedido, ao sujeito avaliado, que manifeste o seu acordo

ou desacordo (escala dicotômica), ou que manifeste o grau de concordância numa escala politômica de  $m$  pontos (1- “discordo totalmente”;  $m$  - “concordo totalmente”).

Estima-se o nível da atitude do sujeito a partir dos itens utilizados, que terão sido calibrados na fase da construção do banco de itens disponíveis para a constituição da escala. A atitude de um sujeito é medida através do padrão de resposta. Veja-se o exemplo: num teste composto por 10 itens o padrão 0011011111, que corresponde a um score 7, é diferente do padrão 1111111000, com o mesmo score. Neste modelo todos os itens têm o mesmo poder de discriminação (ver p. 34) e assim, o padrão deixa de ter importância para apenas se considerar o nº de respostas positivas<sup>7</sup> (Baker, 2001). O score do sujeito corresponderá a um valor  $\hat{\theta}$  da atitude, que é estimado (ver p. 43) em função dos parâmetros de localização dos itens (ver p. 34) que foram estimados na fase de construção da escala (ver p. 37).

### 3.2. Apresentação geral do modelo

As grandes ideias da psicometria foram desenvolvidas no princípio do século XX e com as críticas quer internas (perda de informação do sujeito, massificação de resultados, scores não comparáveis inter-culturas), quer externas (atente-se à discussão já referida sobre os níveis de mensuração das escalas) conheceu alguma estagnação, para retomar fôlego com os trabalhos de Rasch nos anos 60. Nas medições das atitudes, o mesmo se verificou, não tendo havido progressos relevantes, quer a nível prático, quer teórico, até 1978, quando Andrich aplicou o modelo Rasch à análise das atitudes medidas à Likert ou à Thurstone (Keats, 1990).

Paralelamente à teoria de Rasch desenvolveu-se a Teoria da Resposta ao Item (TRI) ou *Item Response Theory* que é hoje apresentada como uma ferramenta indispensável aos psicometristas. Vários autores se referem à TRI como uma teoria fundamental para a avaliação de vários constructos na área da psicologia (Nunnally, 1994, Kline, 1986), mas em todos aparece a associação privilegiada destes processos à avaliação de conhecimentos. Com efeito são métodos de construção de escalas em que o conhecimento de cada sujeito vai ser avaliado independentemente do conjunto de itens que as integram. Por outro lado, ao reconhecer que as escalas de medição podem ser desenvolvidas a partir das respostas individuais a cada um dos itens, Rasch introduz a grande diferença relativamente aos modelos anteriores da psicometria que se baseavam numa outra conjectura: a distribuição da variável latente de acordo com a lei normal, na população. Por

---

<sup>7</sup> Note-se que o modelo de cotação vai de encontro ao proposto por Guttman.



esta razão, a não necessidade de pressupor a forma da distribuição da variável latente na população, os modelos TRI são considerados independentes da amostra. Os scores obtidos dizem-se *sample-free*. Na verdade o valor da capacidade de cada sujeito é estimado em função das suas respostas. Os itens são calibrados de acordo com uma matriz de respostas obtida numa amostra suficientemente grande. Há estudos que comprovam que a calibração de uma escala é invariante para diversas culturas (Kline, 1986).

O primeiro pressuposto da TRI é a unidimensionalidade, há apenas uma dimensão latente subjacente aos itens, isto é, o conjunto de itens mede apenas uma dimensão, uma variável latente, e por isso se eliminam os itens que violam este princípio (ver p. 37). Na fase de construção da escala, feita uma análise factorial devem reter-se os itens que permitem identificar apenas um factor. O segundo pressuposto é a independência local dos itens, ou seja, nenhum item deve conter informação que possa ser usada para responder a um outro: a correlação entre os vários itens deve ser nula. Deste pressuposto decorre que a probabilidade de um sujeito acertar k itens é igual ao produto das probabilidades de acertar cada um deles.

A TRI baseia-se na análise de cada item que integra uma escala. Qualquer dos modelos de TRI assume que é possível descrever matematicamente a relação entre a característica latente de um sujeito e a sua performance num item. Essa relação é designada por função de resposta do item ou curva característica do item (CCI). A forma desta função nasce nas leis da psicofísica<sup>8</sup> que Fechner, Stevens e Thurstone, entre outros, propuseram e que relacionam a intensidade de um estímulo com a aparente magnitude da sensação que aquele causa (Engelhard, 1992), sendo que a dissemelhança relativamente à psicofísica é a variável independente ser usualmente não observável. Como o constructo não é observável, conjectura-se que há potência do contínuo, à semelhança das magnitudes físicas que a psicofísica manipula.

A performance no item i do sujeito v com capacidade latente  $\theta_v$ , é modelada por uma função logística<sup>9</sup>:

$$P[X_{vi} = 1] = c_i + \frac{(1 - c_i) \exp(\alpha_i(\theta_v - \sigma_i))}{1 + \exp(\alpha_i(\theta_v - \sigma_i))} \quad (12)$$

<sup>8</sup> Veja-se a propósito das leis da psicofísica as recentes publicações de Duncan Luce

<sup>9</sup> Uma função logística é definida por  $f(x) = \frac{\exp(Dx)}{1 + \exp(Dx)}$ ,  $x \in \mathbb{R}$ . Quando  $D = 1,7$  a distribuição logística não dista mais de 0,01 da distribuição normal, qualquer que seja o valor de  $x$ . Assim, por exemplo: Nos modelos de TRI aparece por vezes a logística com o parâmetro  $D = 1,7$  para que a probabilidade de resposta acertada se aproxime da normal. Os investigadores que trabalham com Rasch nunca referem este factor (Andrich, 2004).

A Figura 3 auxilia a compreensão das curvas características dos itens. Considere-se o item  $i$ , no eixo dos XX está a aptidão real do sujeito  $v$ , ou seja a real capacidade ( $\theta_v$ ) que tem para responder correctamente ao item  $i$ . No eixo dos YY está a probabilidade do sujeito  $v$  responder correctamente ao item  $i$ .

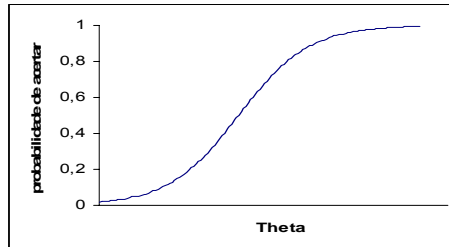


Figura 3: Curva Característica do item

Os parâmetros do modelo da equação (12) têm interpretação geométrica:

- A abcissa do ponto de inflexão da curva (com coordenadas  $(\sigma_i, 0.5)$ ) é o parâmetro de dificuldade do item (Figura 4) e corresponde à mediana da distribuição logística subjacente ao modelo, sendo portanto um parâmetro de localização do item. O item 2 é mais difícil do que o item 1, pois a mediana do item 2 é superior à do item 1: para o mesmo valor de theta há maior probabilidade de acertar no item 1 do que no 2.

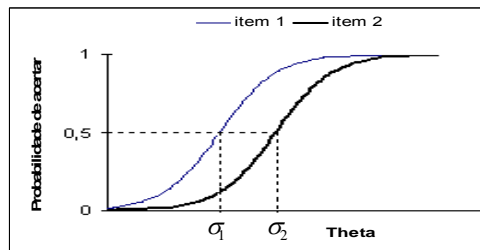


Figura 4: Dois itens com localização diferente

- O declive da parte central da curva é o grau de discriminação ( $\alpha_i$ ) do item (Figura 5). O poder discriminatório dos 3 itens ordena-se do item 1 (menor declive) para o item 3 (maior declive).

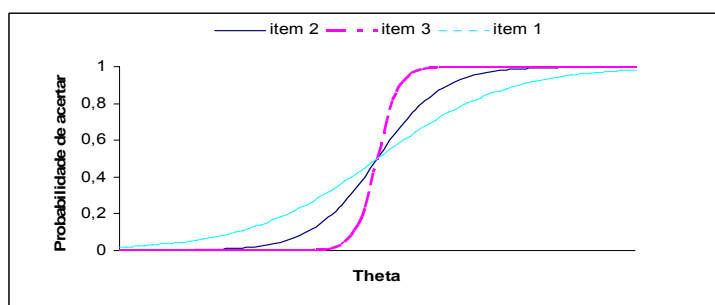
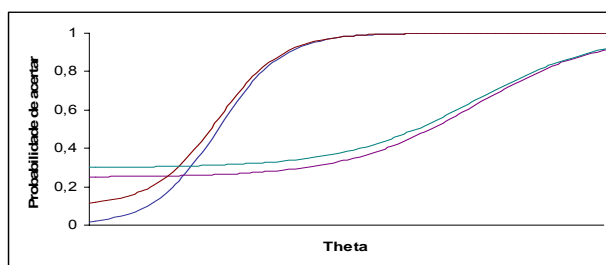


Figura 5: Três itens com discriminação diferente e mesma localização

- O parâmetro de aleatoriedade ( $c_i$ ) corresponde à ordenada na origem (Figura 6). É portanto a probabilidade de qualquer sujeito, independentemente da sua capacidade, acertar a resposta no item  $i$  (em respostas de escolha múltipla, com 4 alternativas de resposta, este parâmetro será 0,25).



**Figura 6:** Quatro itens com aleatoriedade diferente

Há vários modelos baseados nas curvas características dos itens que, conforme consideram 1, 2 ou os 3 parâmetros referidos, assim se denominam modelos 1LP (one parameter logistic), 2LP ou 3LP. Os mais simples são os modelos 1LP, entre os quais está o modelo proposto por Rasch<sup>10</sup>.

A proximidade da realidade aumenta a complexidade do modelo e por conseguinte, o número de parâmetros a estimar, o que exige, no presente, grandes amostras, nalguns casos tão grandes que se tornam projectos inviáveis, comprometendo a estabilidade das estimações (van der Linden, 1997, p.23).

“Uma pergunta importante, nesta altura, é até onde pode ir a modelação TRI sem perder a possibilidade de tratamento estatístico dos seus modelos” (van der Linden, 1997, p.23). O futuro está no desenvolvimento de modelos TRI que se basearão em idealizações e que portanto não se ajustarão completamente à realidade, mas também no desenvolvimento da Estatística e da Computação que permitirão ultrapassar as restrições técnicas que se anunciam.

---

<sup>10</sup> Dizer que o modelo Rasch é um modelo TRI apenas com um parâmetro é uma forma simplista de abordar a questão. Com efeito, Andrich (2004) dá de diferenças nos paradigmas dos modelos Rasch e dos modelos TRI, por exemplo, nos modelos de TRI aparece por vezes a logística com o parâmetro  $D= 1.7$  e os investigadores que trabalham com Rasch nunca referem este factor (Andrich, 2004).

### 3.3. O modelo Rasch dicotómico

No modelo Rasch assume-se que a probabilidade de resposta correcta, do sujeito  $v$  no item  $i$ ,  $P[X_{vi}=1]$ , depende apenas da capacidade do sujeito,  $\theta_v$ , e da dificuldade do item,  $\sigma_i$ . Todos os itens têm o mesmo poder de discriminação ( $\alpha_i=1, \forall i$ ), bem como o mesmo parâmetro de aleatoriedade ( $c_i=0, \forall i$ ) e portanto,

$$P[X_{vi} = 1] = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)} \quad (13)$$

ou ainda

$$P[X_{vi} = x] = \frac{\exp(x(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)}, \quad x \in \{0, 1\} \quad (14)$$

Então, a cada item estão associadas duas estimações: o parâmetro de localização ( $\sigma_i$ ) e a capacidade do sujeito ( $\theta_v$ ), que constitui a medição propriamente dita. Um conceito interessante dos TRI é o uso da mesma escala para ambos os parâmetros, tanto os itens como os sujeitos ocupam posições no mesmo continuum da variável latente.

Considerando o vector  $u_v = [u_{v1}, u_{v2}, \dots, u_{vn}]$  constituído por respostas a  $n$  itens independentes,

$r_v = \sum_{i=1}^n u_{vi}$  é o score do sujeito  $v$ , obtendo-se

$$P[u_v] = \prod_{i=1}^n \frac{\exp(u_{vi}(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)} = \frac{\exp\left[\sum_{i=1}^n u_{vi}(\theta_v - \sigma_i)\right]}{\prod_{i=1}^n [1 + \exp(\theta_v - \sigma_i)]} \quad (15)$$

Há vários procedimentos para se estimarem quer os parâmetros dos itens, quer a atitude de cada sujeito, sendo para isso utilizado software muito diverso. Muitos dos procedimentos baseiam-se na abordagem bayesiana, outros na estimação de máxima verosimilhança, que será ilustrada com um exemplo (ver p.43).

Habitualmente, é seleccionada uma amostra aleatória (ou amostra de calibragem) de uma população alvo, obtendo as respostas a um conjunto de itens. Dadas as respostas aos itens, a capacidade e os parâmetros dos itens são estimados. Os parâmetros estimados a partir da amostra podem ser tratados como conhecidos, e nesta suposição podem construir-se bancos de itens. Nas aplicações sucessivas, estes itens, dos quais se conhecem

os valores dos parâmetros, são administrados aos examinandos e as suas capacidades estimadas. (Hambleton, 1990, p. 75)

Numa perspectiva de compreensão da utilização deste paradigma nas escalas de atitudes, ilustrar-se-ão os procedimentos de estimação (calibragem) de parâmetros de dificuldade (localização) na fase de construção de um banco de itens (ver 3.4), bem como a estimação da atitude de um sujeito a partir das respostas a um conjunto de itens previamente calibrados (ver 3.5).

### **3.4. Construção de uma escala de atitudes usando o modelo Rasch dicotómico**

Para construir, com base no modelo Rasch, um banco de itens utilizáveis numa escala de atitudes, começa por se elaborar um vasto conjunto de frases (itens) que manifestem opiniões diversas, desde muito negativas até muito positivas relativamente à atitude que se quer estudar.

Constituída uma amostra acima de 500 sujeitos (Muñiz, 1997, p. 173), que servirá para calibrar a escala, pede-se a cada sujeito que manifeste a sua opinião sob a forma de “concordo” (1), “não concordo” (0). Resulta uma matriz de dados que se usará para eliminar itens, bem como para estimar, quer a dificuldade de cada item ( $\sigma_i$ ), quer a atitude de cada sujeito ( $\theta_v$ ). O parâmetro dificuldade desempenha um papel de regulador da distância: introduz uma métrica, é por isso designado por parâmetro de localização do item.

Antes de proceder à estimação dos parâmetros, há que não esquecer a questão da unidimensionalidade, que é um dos pressupostos dos modelos TRI e que se avalia através de uma análise factorial, mantendo-se apenas os itens que integram um dos factores<sup>11</sup> (Pasquali, 2003).

Nos diversos modelos existentes há diversas formas de calibrar os itens, todas elas fazendo recurso de diversos softwares (Pasquali (2003) apresenta em anexo uma vasta lista), pois a estimação dos parâmetros é feita por processos iterativos estando fora de questão executá-la sem apoio informático específico. Para ilustrar um dos procedimentos (Henson, 1997; Cantrell, 1997; Henard, 1998), tome-se por exemplo a matriz de resultados da (Tabela 1) onde se exibem, para além das respostas dadas pelos 36 sujeitos (nas linhas) aos 18 itens (nas colunas), as proporções de respostas 1, quer por linha ( $p_{si}$ ), quer por coluna ( $p_{pj}$ ).

---

<sup>11</sup> Bock desenvolveu um método de análise de itens baseado no modelo TRI: a Full Information Factor Analysis (Pasquali, 2003).

Tabela 1: Matriz de resultados da aplicação de 18 itens a 36 sujeitos

sujei- to	item																		score por suei- to	P <sub>si</sub>
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	6	0,33
2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	9	0,50
3	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	9	0,50
4	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	5	0,28
5	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	9	0,50
6	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	9	0,50
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	15	0,83
8	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	9	0,50
9	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	9	0,50
10	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0,56
11	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	8	0,44
12	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	8	0,44
13	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0,56
14	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	10	0,56
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	14	0,78
16	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0,56
17	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	8	0,44
18	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	10	0,56
19	1	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0	8	0,44
20	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	12	0,67
21	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	12	0,67
22	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	13	0,72
23	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	13	0,72
24	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	15	0,83
25	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	5	0,28
26	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0,56
27	1	1	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	7	0,39
28	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	10	0,56
29	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	10	0,56
30	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	9	0,50
31	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	10	0,56
32	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	11	0,61
33	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	6	0,33
34	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0	0	13	0,72
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	17	0,94
score por item	35	35	35	33	33	27	31	26	27	24	14	7	10	5	3	2	2	0		
P <sub>oi</sub>	0,97	0,97	0,97	0,92	0,92	0,75	0,86	0,72	0,75	0,67	0,39	0,19	0,28	0,14	0,08	0,06	0,06	0,00		

Na (Tabela 2) ordenaram-se tanto os sujeitos quanto os itens por ordem crescente das percentagens. Devem eliminar-se os itens com que todos os sujeitos concordam (coluna só com 1) e também aqueles com que todos discordam (coluna só com 0), pois não discriminam as verdadeiras atitudes dos sujeitos: o modelo pressupõe que todos os itens têm o mesmo grau de discriminação, que é 1. Também os sujeitos que concordam com todos os itens (linha só com 1) ou que discor-

dam de todos (linha só com 0), são eliminados: não permitem a introdução da métrica na escala pois as suas atitudes estão muito abaixo da localização dos itens que foram elaborados ou muito acima.

**Tabela 2:** Matriz de resultados da aplicação de 20 itens a 36 sujeitos ordenados por  $p_{si}$  e  $p_{pj}$ .

sujei- to	item																		score por sujei- to	$p_{si}$		
	1	2	3	4	5	7	6	9	8	10	11	13	12	14	15	16	17	18				
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00	
4	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0,28
25	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0,28
1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0,33
33	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	6	0,33
27	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	7	0,39
11	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	8	0,44
12	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	8	0,44
17	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	8	0,44
19	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	8	0,44
2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	9	0,50
3	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	9	0,50
5	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	9	0,50
6	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	9	0,50
8	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	9	0,50
9	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	9	0,50
30	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	9	0,50
10	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	10	0,56
13	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	10	0,56
14	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	10	0,56
16	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	10	0,56
18	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	10	0,56
26	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	10	0,56
28	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	10	0,56
29	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	10	0,56
31	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	10	0,56
32	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	11	0,61
20	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	12	0,67
21	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	12	0,67
22	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	13	0,72
23	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	13	0,72
34	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	13	0,72
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	14	0,78
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	15	0,83
24	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	15	0,83
36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	17	0,94
score por item	35	35	35	33	33	31	27	27	26	24	14	10	7	5	3	2	2	0				
$p_{pj}$	0,97	0,97	0,97	0,92	0,92	0,86	0,75	0,75	0,72	0,67	0,39	0,28	0,19	0,14	0,08	0,06	0,06	0,00				
$1-p_{pj}$	0,03	0,03	0,03	0,08	0,08	0,14	0,25	0,25	0,28	0,33	0,61	0,72	0,81	0,86	0,92	0,94	0,94	1,00				

Na Tabela 3 pode-se verificar que foram eliminados os itens 1, 2, 3, 18 e os sujeitos 35 e 36. A eliminação de itens pode dar origem à eliminação de sujeitos e vice-versa. O processo termina quando a matriz não exibe qualquer linha ou coluna só com zeros ou uns.

**Tabela 3:** Matriz de resultados da aplicação de 20 itens a 36 sujeitos ordenados por  $p_{si}$  e  $p_{pj}$ .

sujeito	item															score por sujeito	$p_{si}$
	4	5	7	6	9	8	10	11	13	12	14	15	16	17			
4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2	0,143	
25	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2	0,143	
1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	3	0,214	
33	1	0	0	0	1	1	0	0	0	0	0	0	0	0	3	0,214	
27	1	0	0	1	1	1	0	0	0	0	0	0	0	0	4	0,286	
11	1	1	1	1	1	0	0	0	0	0	0	0	0	0	5	0,357	
12	1	1	1	0	1	0	1	0	0	0	0	0	0	0	5	0,357	
17	1	1	1	0	0	1	1	0	0	0	0	0	0	0	5	0,357	
19	1	1	1	0	0	1	0	0	1	0	0	0	0	0	5	0,357	
2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0,429	
3	1	1	1	1	0	1	1	0	0	0	0	0	0	0	6	0,429	
5	1	1	1	0	1	1	1	0	0	0	0	0	0	0	6	0,429	
6	1	1	1	0	1	1	1	0	0	0	0	0	0	0	6	0,429	
8	1	1	1	1	0	1	1	0	0	0	0	0	0	0	6	0,429	
9	1	1	1	1	0	0	1	1	0	0	0	0	0	0	6	0,429	
30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0,429	
10	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0,500	
13	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0,500	
14	1	1	1	1	1	1	0	0	1	0	0	0	0	0	7	0,500	
16	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0,500	
18	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0,500	
26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0,500	
28	1	1	1	0	1	1	1	1	0	0	0	0	0	0	7	0,500	
29	1	1	1	1	1	0	1	1	0	0	0	0	0	0	7	0,500	
31	1	1	1	1	1	0	1	1	0	0	0	0	0	0	7	0,500	
32	1	1	1	1	1	1	1	0	0	1	0	0	0	0	8	0,571	
20	1	1	1	1	1	1	1	1	1	0	0	0	0	0	9	0,643	
21	1	1	1	1	1	1	1	1	0	1	0	0	0	0	9	0,643	
22	1	1	1	1	1	1	1	1	1	1	0	0	0	0	10	0,714	
23	1	1	1	1	1	1	1	1	1	0	1	0	0	0	10	0,714	
34	1	1	1	1	1	0	1	1	1	1	1	0	0	0	10	0,714	
15	1	1	1	1	1	1	1	1	1	1	0	1	0	0	11	0,786	
7	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12	0,857	
24	1	1	1	1	1	1	1	1	1	0	1	0	1	1	12	0,857	
score por item	32	32	30	26	26	25	23	13	9	6	4	2	1	1			
$p_{pj}$	0,941	0,941	0,882	0,765	0,765	0,735	0,676	0,382	0,265	0,176	0,118	0,059	0,029	0,029			
$1-p_{pj}$	0,06	0,06	0,12	0,24	0,24	0,26	0,32	0,62	0,74	0,82	0,88	0,94	0,97	0,97			

Limpa a matriz de dados, calculam-se de novo as proporções de respostas 1, quer por linha ( $p_{si}$ ),

quer por coluna ( $p_{pj}$ ) e os respectivos logits  $\left( \ln \left( \frac{p_{si}}{1-p_{si}} \right) ; \ln \left( \frac{1-p_{pj}}{p_{pj}} \right) \right)$ . Os resultados aparecem



nas tabelas 5 e 4, respectivamente.

Procede-se então à primeira estimação dos índices de localização dos itens (Tabela 4), que correspondem aos logits centrados (distância dos logits em relação à média destes).

**Tabela 4:** Primeira calibragem dos itens

item	score do item	$p_{si}$	$1 - p_{si}$	$1 - p_{si} / p_{si}$	logit	parâmetro de localização inicial
4	32	0,941	0,059	0,063	-2,772588722	-2,997
5	32	0,941	0,059	0,063	-2,772588722	-2,997
7	30	0,882	0,118	0,133	-2,014903021	-2,240
6	26	0,765	0,235	0,308	-1,178654996	-1,403
9	26	0,765	0,235	0,308	-1,178654996	-1,403
8	25	0,735	0,265	0,360	-1,021651248	-1,246
10	23	0,676	0,324	0,478	-0,737598943	-0,962
11	13	0,382	0,618	1,615	0,47957308	0,255
13	9	0,265	0,735	2,778	1,021651248	0,797
12	6	0,176	0,824	4,667	1,540445041	1,316
14	4	0,118	0,882	7,500	2,014903021	1,790
15	2	0,059	0,941	16,000	2,772588722	2,548
16	1	0,029	0,971	33,000	3,496507561	3,272
17	1	0,029	0,971	33,000	3,496507561	3,272

média dos logit =0,225

$VAR_{item} =$  variância dos logit =4,841

A capacidade de cada sujeito é estimada através dos logits calculados para cada sujeito (Tabela 5).

**Tabela 5:** Primeira calibragem dos sujeitos

Número de respostas correctas	FR	$p_{pj}$	$1 - p_{pj}$	$p_{pj} / 1 - p_{pj}$	logit	Atitude
1	0	0,071	0,929	0,077	-2,5649	-2,5649
2	2	0,143	0,857	0,167	-1,7918	-1,7918
3	2	0,214	0,786	0,273	-1,2993	-1,2993
4	1	0,286	0,714	0,400	-0,9163	-0,9163
5	4	0,357	0,643	0,556	-0,5878	-0,5878
6	7	0,429	0,571	0,750	-0,2877	-0,2877
7	9	0,500	0,500	1,000	0,0000	0,0000
8	1	0,571	0,429	1,333	0,2877	0,2877
9	2	0,643	0,357	1,800	0,5878	0,5878
10	3	0,714	0,286	2,500	0,9163	0,9163
11	1	0,786	0,214	3,667	1,2993	1,2993
12	2	0,857	0,143	6,000	1,7918	1,7918
13	0	0,929	0,071	13,000	2,5649	2,5649

média dos logit =-0,0697

$VAR_{sujeito} =$  variância dos logit =0,722

Aplicam-se então, factores de correcção para retirar o efeito da amostra e o dos itens, obtendo-se a primeira estimação dos parâmetros do modelo.

Estando os parâmetros estimados há que verificar se o modelo se ajusta aos dados: a matriz esperada de respostas de acordo com o modelo será uma matriz em que figurarão 1 na parte inferior esquerda, que corresponde a itens fáceis relativamente à capacidade elevada dos sujeitos, e 0 na parte superior direita, correspondendo a itens com localização à direita da verdadeira atitude dos sujeitos (itens mais difíceis do que a capacidade dos sujeitos).

Identificadas as respostas aberrantes, procede-se a um estudo sujeito a sujeito e item a item, para eliminar tanto os sujeitos quanto os itens que não se ajustam ao modelo.

Este processo exige algumas precauções, pois a remoção de alguns itens pode resolver a falta de ajustamento de algum sujeito, ou vice-versa, por isso se deve verificar caso a caso.

Retirados os sujeitos e os itens que não se ajustam ao modelo, o processo recomeça, novos logits são calculados, os parâmetros estimados e todos os itens e sujeitos são de novo avaliados. Fazem-se sucessivas estimações tanto da localização dos itens como da capacidade dos sujeitos e eliminam-se sucessivamente sujeitos e itens. O processo iterativo termina quando tanto os sujeitos quanto os itens se ajustam ao modelo.

Cantrell (1997) sugere que se avalie se as localizações dos itens são independentes da amostra. Para isso, dividem-se os sujeitos que ficaram na estimação final em seis grupos. Se o modelo Rasch estiver a funcionar bem, a localização estimada para cada item em cada um dos seis grupos deve ser semelhante à estimada quando se considera o grupo total.

Ao eliminar sujeitos e itens que se desviam das expectativas, pode-se criar um banco de itens independentes e que permitem a medição da capacidade dos sujeitos, pois foi criada uma escala que põe em paralelo a localização dos itens e essa capacidade.

Ilustrou-se assim o procedimento para a construção de um banco de itens.

Um banco de itens é uma colecção de itens de teste, organizados, classificados e catalogados, como os livros de uma biblioteca, no sentido de facilitar a construção de uma variedade de testes de capacidades e outros tipos de testes mentais. (Choppin, 1990, p. 351)

A escolha dos itens que integrarão a escala propriamente dita é feita em função dos objectivos da escala escolhendo seguidamente os itens com maior quantidade de informação que, no caso do moledo apresentado, Rasch (TRI – 1LP), é dada por  $p_{pj} (1-p_{pj})$  (Baker, 2001). Por exemplo, estando a estudar a atitude face a determinado assunto, querendo identificar os sujeitos que têm uma

atitude claramente favorável devem escolher-se os itens com localização (parâmetro de dificuldade) mais elevada e, dentre estes, os que apresentam maior quantidade de informação (Muñiz, 2005).

A estimação da atitude de um sujeito é feita com base no score obtido e na dificuldade dos itens seleccionados para integrarem a escala de atitudes.

### 3.5. Estimação da atitude

Construída uma escala de atitudes, a partir de um banco de itens de que se «conhecem» os parâmetros, isto é, estando estimados os  $\sigma_i$ , há que estimar a atitude de um sujeito com determinado padrão de resposta. Como se deduz da equação (15), o score  $r_v$  ( $r_v = \sum_{i=1}^n u_{vi}$ ) do sujeito  $v$  é

uma estimativa suficiente de  $\theta_v$ , pois pertencendo a função à família exponencial ( $f(x | \theta_v) = h(x)c(\theta_v) \exp\left[\sum_{i=1}^n w_i(\theta_v)t_i(x)\right]$  em que  $h(x) = 1$ ,  $c(\theta) = \frac{1}{\prod_{i=1}^n (1 + \exp(\theta_v - \hat{\sigma}_i))}$ ,

$w_i(\theta_v) = (\theta_v - \sigma_i)$  e  $t_i(x) = u_{vi}$ ),  $\sum_{i=1}^n u_{vi}$  é uma estatística suficiente de  $\theta_v$ .

De entre os vários procedimentos para estimar  $\theta_v$ , tome-se, como exemplo, o procedimento da máxima verosimilhança (Hambleton, 1990).

Seja  $u_v = [u_{v1}, u_{v2}, \dots, u_{vm}]$  o vector resposta do sujeito  $v$ , sendo  $u_{vi}$  a resposta (0 ou 1) no item  $i$ , tem-se a função de verosimilhança:

$$\mathcal{L}[\theta_v | u_v] = \prod_{i=1}^n \frac{\exp(u_{vi}(\theta_v - \hat{\sigma}_i))}{1 + \exp(\theta_v - \hat{\sigma}_i)} = \frac{\exp\left[\sum_{i=1}^n u_{vi}(\theta_v - \hat{\sigma}_i)\right]}{1 + \exp(\theta_v - \hat{\sigma}_i)} \quad (16)$$

O estimador de máxima verosimilhança da atitude do sujeito  $v$ , será  $\hat{\theta}_v$  tal que

$$\frac{d}{d\theta_v} \ln \mathcal{L}(\theta_v - \hat{\sigma}_i) = 0 \quad (17)$$

Ou seja,

$$\sum_{i=1}^n \frac{\exp(\hat{\theta}_v - \hat{\sigma}_i)}{1 + \exp(\hat{\theta}_v - \hat{\sigma}_i)} = r_v. \quad (18)$$

Recorrendo ao método de Newton-Raphson:

$$\theta_{m+1} = \theta_m - h_m \quad \text{em que } h_m = \left[ \frac{\frac{d}{d\theta} \ln \mathcal{L}(\theta_v | u_v)}{\frac{d^2}{d\theta^2} \ln \mathcal{L}(\theta_v | u_v)} \right] \quad (19)$$

O valor de  $\hat{\theta}_v$  vai ser calculado numericamente, por aproximações sucessivas, tomando o valor de  $\theta_{m+1}$  quando o processo convergir ( $|\theta_{m+1} - \theta_m| < \varepsilon$ ).

Exemplifica-se na Tabela 6 com o cálculo de  $\hat{\theta}_v$ , numa escala de 5 itens com localização: -1; 0; 1; 1.5 e 2, para um sujeito v, que respondeu positivamente a três dos itens, ou seja, teve  $r_v = 3$ .

$P_{mj} = \frac{\exp(\theta_{m-1} - \hat{\sigma}_i)}{1 + \exp(\theta_{m-1} - \hat{\sigma}_i)}$ , e a primeira estimativa de  $\theta$ , será  $\hat{\theta}_0 = \ln\left(\frac{3}{5-3}\right) = 0.4054$ , atendendo a que  $\theta = \ln\left(\frac{P}{1-p}\right)$ .

**Tabela 6:** Estimação de Theta

item	sigma	(m=1) P <sub>mi</sub>	(m=2) P <sub>mi</sub>	(m=3) P <sub>mi</sub>
1	-1	0,803039	0,90599	0,904402309
2	0	0,599984	0,779993	0,776801897
3	1	0,35558	0,566018	0,561468894
4	1,5	0,250753	0,441673	0,437116145
5	2	0,168738	0,324235	0,320195312
soma P		2,178095	3,017909	2,999984557
soma PQ		0,955454	0,968123	0,969776952
h <sub>m</sub>		-0,86022	0,018499	-1,5924E-05
θ <sub>m</sub>		1,265625	1,247126	1,247141552

Então  $\hat{\theta}_v = 1.247$ .

### 3.6. O modelo Rasch politômico

Até aqui pensou-se na possibilidade da resposta ao item ser 1 ou 0, seja certo ou errado, concordo/ não concordo. Idealmente, todos os sujeitos com capacidade ( $\theta_v$ ) superior à localização do item ( $\sigma_i$ ) deixarão de dar resposta 0 para passar a dar resposta 1. O parâmetro dificuldade ou de localização ( $\sigma_i$ ) pode então ser interpretado como uma barreira ou parâmetro de transição que separa o 0 do 1. Quanto maior for  $\sigma_i$ , maior a dificuldade<sup>12</sup> do item e portanto menor a probabilidade de acertar (dar resposta 1).

Quando há mais do que duas possibilidades de resposta, por exemplo "0" - discordo totalmente até "m" – concordo totalmente, o número de parâmetros de transição, em cada item  $i$ , passa a ser  $m$  e serão designados por  $\tau_{ix}$ , com  $x = 0, \dots, m$ .

O estudo dos parâmetros de transição é essencial pois  $\tau_{ix}$  é um ponto de corte, isto é, para o item  $i$ , a partir deste ponto a resposta deixa de ser na categoria  $x-1$ , para passar a ser na categoria  $x$ .

A curva característica do item sofre algumas alterações. Fixando o item  $i$ , a curva de resposta resulta sempre da conjugação do valor da variável latente com o valor da localização do item: os sujeitos localizados à esquerda ( $\theta_v$  baixo) terão tendência para dar resposta na categoria 0, diminuindo a probabilidade desta resposta com o aumento da variável latente. Haverá um ponto  $\tau_{i0}$ , em que a resposta deixa de ser 0 para ser 1 (Figura 7).

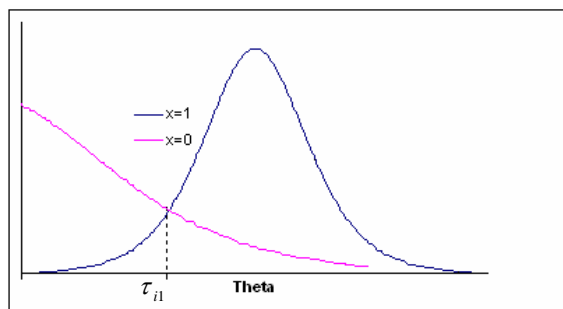


Figura 7: Primeira barreira

A resposta na categoria  $m$  terá distribuição do tipo da que se apresenta na Figura 8, pois serão os sujeitos com nível elevado na variável latente que tenderão a responder na última categoria.  $\tau_{im}$

---

<sup>12</sup> No continuum da atitude, a dificuldade traduz-se na barreira entre concordo/não concordo, ou seja localiza o item mais para o lado favorável ou desfavorável.

será a última barreira a considerar correspondendo ao valor da variável latente em que a resposta deixa de estar na categoria  $m-1$  para estar na categoria  $m$  (acordo mais extremo). Terão tendência para responder maioritariamente na categoria  $x$  os sujeitos cuja variável latente se situa entre  $\tau_{ix}$  e  $\tau_{i(x+1)}$ .

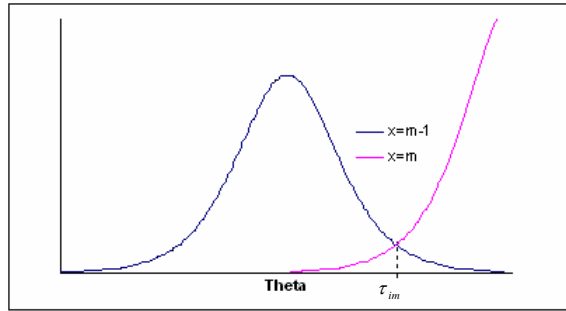


Figura 8: Última barreira

Como o modelo politómico é uma generalização do modelo dicotómico, começa-se por reescrever a expressão (2), do modelo dicotómico, como uma probabilidade condicionada:

$$\frac{P[X_{vi} = x]}{P[X_{vi} = 1] + P[X_{vi} = 0]} = \frac{\exp(x(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)}, \quad x \in \{0, 1\} \quad (20)$$

A barreira  $\tau_{i1}$  está localmente condicionada à resposta na categoria 0 ou 1, portanto ter-se-á:

$$\frac{P[X_{vi} = 0]}{P[X_{vi} = 1] + P[X_{vi} = 0]} = \frac{1}{1 + \exp(\theta_v - \tau_{i1})} \quad (21)$$

$$\frac{P[X_{vi} = 1]}{P[X_{vi} = 1] + P[X_{vi} = 0]} = \frac{\exp(\theta_v - \tau_{i1})}{1 + \exp(\theta_v - \tau_{i1})} \quad (22)$$

Havendo que considerar que  $\sum_{x=0}^m P[X_{vi} = x] = 1$  e sendo  $\eta_{vi}$  o denominador que o garante, ter-se-á:

$$P[X_{vi} = 0] = \frac{1}{\eta_{vi}} \quad (23)$$

$$P[X_{vi} = 1] = \frac{\exp(\theta_v - \tau_{i1})}{\eta_{vi}} \quad (24)$$

Pensando na barreira  $\tau_{i2}$ , ter-se-á:

$$\frac{P[X_{vi} = 2]}{P[X_{vi} = 1] + P[X_{vi} = 2]} = \frac{\exp(\theta_v - \tau_{i2})}{1 + \exp(\theta_v - \tau_{i2})} \quad (25)$$

Substituindo  $P[X_{vi} = 1]$  pelo valor de (17), aparece

$$P[X_{vi} = 2] = \frac{\exp(2\theta_v - (\tau_{i1} + \tau_{i2}))}{\eta_{vi}} \quad (26)$$

Pode-se então generalizar, para qualquer  $x = 0, \dots, m$ , (Figura 9) aparecendo

$$P[X_{vi} = x] = \frac{\exp(x\theta_v - \sum_{s=1}^x \tau_{is})}{\eta_{vi}}, \quad x = 0, \dots, m \quad (27)$$

convencionando que  $\sum_{s=1}^0 \tau_{is} = 0$ , pois não há barreira para a categoria 0.

Em que,

$$\eta_{vi} = \sum_{y=0}^m \exp(y\theta_v - \sum_{s=1}^y \tau_{is}) \quad (28)$$

Assim, neste modelo há  $nm+1$  parâmetros a estimar:  $\theta_v$ , e os  $nm$   $\tau_{ix}$  ( $i = 1, \dots, n$  e  $x = 1, \dots, m$ ).

As barreiras  $\tau_{ix}$  ( $x = 1, \dots, m$ ) são os parâmetros do modelo que o munem de uma partição em  $m+1$  secções onde cada uma das categorias de resposta tem maior probabilidade, por exemplo:

$P[X_{vi} = 2]$  é máxima quando  $\theta_v \in ]\tau_{i1}, \tau_{i2}]$ .

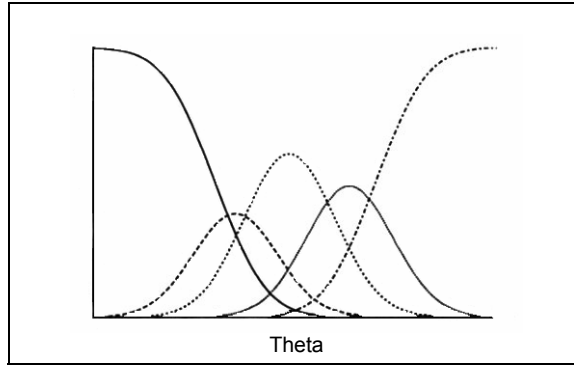


Figura 9: Exemplo de funções de resposta em 5 categorias

Pode-se restringir o modelo (27) à condição da distância entre as barreiras ser constante para todos os itens, isto é, qualquer que seja o item  $\tau_{ix} - \tau_{i(x-1)} = \varphi_x$ , ou ainda  $\sigma_i - \tau_{ix} = \psi_x$ , qualquer que seja o item  $i$ , com  $\sum_{x=1}^m \psi_x = 0$ , o que é razoável admitir quando há um formato de resposta constante em todos os itens (Rost & Langeheine, 1997). Tendo-se então,

$$P[X_{vi} = x] = \frac{\exp(x(\theta_v - \sigma_i) - \sum_{s=1}^x \psi_s)}{\sum_{y=0}^m \exp(y(\theta_v - \sigma_i) - \sum_{s=1}^y \psi_s)}, \quad x = 0, \dots, m \quad (29)$$

e passando o modelo a ter menos parâmetros a estimar:  $\theta_v$ ,  $\sigma_i (i = 1, \dots, n)$  e  $\psi_x (x = 1, \dots, m)$ , ou seja,  $n+m+1$ .

O modelo (27) é o mais geral e designa-se por PCM (Parcial Credit Model). Designa-se por RSM (Rating Scale Model) o modelo (29), com a restrição das barreiras estarem em todos os itens à mesma distância.

A Figura 10 é um exemplo de dois itens  $i$  e  $j$ , de acordo com o modelo RSM, com cinco categorias de resposta.



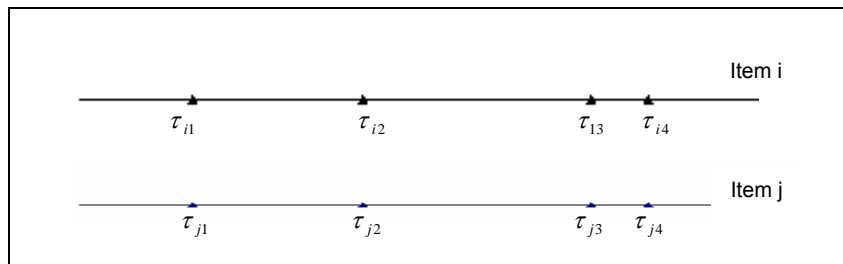


Figura 10: Localização das barreiras em dois itens no modelo RSM

### 3.7. Thurstone, Likert, Guttman e o modelo Rasch

O modelo apresentado por Thurstone em 1929 para medir as atitudes é ainda o modelo vigente. “Essencialmente, Thurstone propôs o que seria hoje reconhecido como curvas características do sujeito.” (Engelhard, 1992, p. 283).

Nos modelos Rasch, mantém-se a ideia de um continuum na atitude (e de resto em todas as aplicações de medição de características psicológicas, avaliação de conhecimentos ou de aptidões) e, quando se assume que um sujeito concordará apenas com os itens que reflectem a sua própria atitude e estará em desacordo com todos os que reflectam uma atitude menos ou mais favorável está-se a encontrar uma localização do sujeito nesse continuum. Esta localização é por vezes designada por *ideal point* do sujeito (Andrich, 1998). Em relação à metodologia proposta por Thurstone ganhou-se imenso, mantendo as preocupações de rigor na medição, dispensando o trabalho fastidioso do grupo de juizes (que alguns estudos provaram não serem imparciais!), constrói-se uma escala calibrada que permite a localização quer do sujeito, quer do item no mesmo continuum de atitude (na mesma escala).

Nos modelos TRI, a distribuição normal inerente ao modelo de Thurstone foi em larga escala substituída pela distribuição logística<sup>13</sup> dada a simplicidade desta.

Likert não sugere outro modelo, o que propõe é outra forma de construção de escalas de atitude por forma a simplificar a questão controversa da intervenção de juizes para se poderem atribuir pesos aos diferentes itens.

---

<sup>13</sup> Nos modelos de TRI aparece por vezes a logística com o parâmetro  $D=1.7$  para que a probabilidade de resposta acertada se aproxime da normal. Os investigadores que trabalham com Rasch nunca referem este factor (Andrich, 2004).

Guttman propõe outra forma engenhosa de construir escalas de atitudes retomando o que Bogardus havia feito de forma intuitiva com a escala de proximidade social. A base é também a do continuum da atitude e ao analisar as curvas características dos itens de uma escala Guttman encontram-se funções em escada, isto é, os itens têm o poder de discriminação máximo: todos os índices de discriminação ( $\alpha_i$ ) são infinito. Outro ponto em comum com o modelo Rasch é a forma de determinar a atitude do sujeito, uma vez que em ambos os casos não se utiliza o padrão de resposta, mas apenas o número de respostas positivas<sup>14</sup>.

Na Tabela 7, adaptada de Anderson (1990<sup>a</sup>), apresentam-se as semelhanças e dissemelhanças destes 3 tipos de escalas.

**Tabela 7:** Comparação das diversas escalas de atitudes

	<b>Thurstone</b>	<b>Likert</b>	<b>Guttman</b>
<b>Formato</b>	frases	frases	frases
<b>Posição dos itens no continuum da atitude</b>	Pontos ao longo do continuum	Pontos nos extremos	Pontos ao longo do continuum
<b>Natureza do continuum</b>	Não cumulativo	Não cumulativo	cumulativo

Qualquer destes tipos de escalas de atitudes se constrói e utiliza até hoje, mas sem dúvida que a proposta de Likert é a mais divulgada. Os modelos TRI, parecendo ser o futuro da medição em psicologia, ainda não estão suficientemente divulgados sendo por isso pouco utilizados na área da medida de atitudes.

Há que evidenciar que no meio se considera uma escala de Thurstone quando as respostas sejam apenas sim e não e uma escala de Likert quando há várias modalidades de resposta sendo preferencialmente utilizada a escala de 5 pontos. Não há propriamente escalas TRI ou escalas de Rasch. Estes modelos utilizam-se para estimar parâmetros de escalas tipo Likert, Thurstone ou mesmo Guttman.

Nos modelos clássicos a atitude de cada sujeito é sempre resultado da comparação de um sujeito com um grupo de referência. Nos modelos Rasch pretende-se avaliar o sujeito tendo em conta as

<sup>14</sup>Consequência de todos os itens terem o mesmo índice de discriminação  $\alpha_i = 1, \forall i, i = 1, \dots, n$ .

suas respostas e a probabilidade da ocorrência dessa resposta: são técnicas em que é estimada a atitude de cada sujeito.

Uma escala de atitudes é, no modelo clássico, um conjunto estanque de questões que funcionam num todo. Nos modelos Rasch, o que se avalia são os itens de *per si*, constituindo-se um banco de itens, e cada escala é construída com um conjunto de itens adequado à avaliação que se pretende (Pasquali, 2007).

### 3.8. O modelo seleccionado

Andrich e Luo, em 1993, adaptaram o modelo Rasch às escalas de atitudes construindo o modelo do co-seno hiperbólico (*Hyperbolic Cosine Model – HCM*), quando se consideram itens dicotómicos (sim/não) e em 1996, Andrich e Roberts separadamente generalizaram-no para itens politómicos (respostas com vários níveis: discordo totalmente até concordo totalmente).

#### *O modelo do co-seno hiperbólico*

O modelo que a seguir se apresenta é o proposto por Andrich e Luo, em 1993, tomando como referência os textos de Andrich de 1997 e 1998.

O ponto de partida é a curva característica do item, no modelo Rasch, isto é, com um único parâmetro: o de dificuldade, que assume o papel de índice de distância em relação à atitude mais negativa e que se designará por  $\sigma_i$ .

Andrich faz notar que numa escala de atitudes não se trata de responder correctamente ao item, mas de concordar (ou não) com a afirmação que o item contém. E concordar ou não com a afirmação contida num item depende da localização do item no continuum da atitude e também do valor da atitude do sujeito (variável latente). Assim, se um item  $i$  está posicionado ( $\sigma_i$ ) à esquerda da verdadeira atitude do sujeito ( $\theta_v$ ), a curva característica do item será consistente com a ideia que a probabilidade de estar de acordo decresce à medida que vai aumentando  $\theta_v$  (Figura 11). De forma simétrica se comportará a curva característica de um item que se situe à direita da verdadeira atitude do sujeito: a probabilidade de estar de acordo cresce quando  $\theta_v$  aumenta (Figura 13). Num item localizado ao centro, serão os sujeitos indiferentes que terão maior tendência para estar de acordo com a afirmação (Figura 12). Para um qualquer item ter-se-á então uma curva característica que toma em consideração o que se acabou de referir, tomando o aspecto da Figura 14.

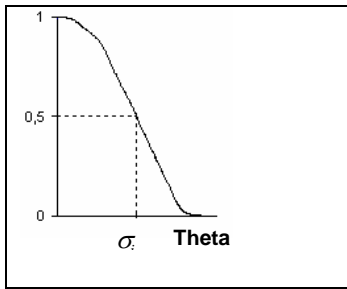


Figura 11: Item localizado à esquerda da atitude ( $\sigma_i < \theta_v$ )

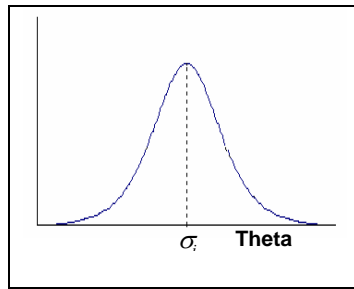


Figura 12: Item com  $\sigma_i = \theta_v$

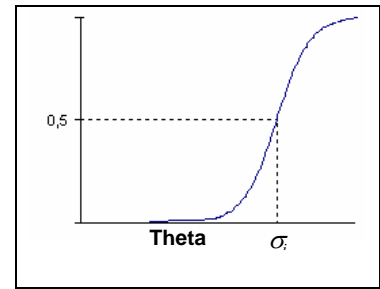


Figura 13: Item localizado à direita da atitude ( $\sigma_i > \theta_v$ )

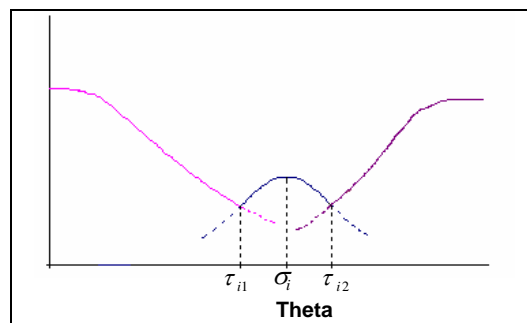


Figura 14: Curva característica de um item geral

Uma resposta A (acordo) é tanto mais provável quanto a verdadeira atitude do sujeito ( $\theta_v$ ) se encontre perto da localização do item ( $\sigma_i$ ).

Com efeito, o sujeito  $v$  pode discordar do item  $i$  por duas razões: o item está localizado ( $\sigma_i$ ) à esquerda da posição real do sujeito ( $\theta_v$ ), que se designa por desacordo à esquerda; ou o item está localizado à direita da posição real do sujeito e, nesse caso, o desacordo designa-se por desacordo à direita. O desacordo passa a ser um acontecimento composto. A variável resposta ( $Y$ ) deixa de ser dicotômica para passar a ser ordenada, tomando os valores: 0 (desacordo à esquerda); 1 (acordo) e 2 (desacordo à direita). Tendo em conta o modelo da equação (1), em que  $\sigma_i - \tau_{ix} = \psi_x$ ,

qualquer que seja o item  $i$ , atendendo a que  $\sum_{x=1}^m \psi_x = 0$ , pode considerar-se  $\psi_2 = -\psi_1 = \lambda$ , e

portanto:

$$P[Y_{vi} = 0] = \frac{1}{1 + \exp(\theta_v - \sigma_i + \lambda) + \exp(2(\theta_v - \sigma_i))} \quad (30)$$

$$P[Y_{vi} = 1] = \frac{\exp(\lambda + \theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i + \lambda) + \exp(2(\theta_v - \sigma_i))} \quad (31)$$

$$P[Y_{vi} = 1] = \frac{\exp(2(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i + \lambda) + \exp(2(\theta_v - \sigma_i))} \quad (32)$$

Considerando agora,  $X = 0$  quando  $Y = 0$  ou  $Y = 2$  e  $X = 1$  quando  $Y = 1$ , a variável  $X$  é modelada por

$$P[Y_{vi} = 1] = \frac{1 + \exp(2(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i + \lambda) + \exp(2(\theta_v - \sigma_i))} \quad (33)$$

$$P[Y_{vi} = 1] = \frac{\exp(\lambda + \theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i + \lambda) + \exp(2(\theta_v - \sigma_i))} \quad (34)$$

Multiplicando tanto o numerador como o denominador por  $\exp(-(\theta_v - \sigma_i))$  e simplificando obtém-se

$$P[Y_{vi} = 1] = \frac{\exp \lambda}{\exp(-(\theta_v - \sigma_i)) + \exp \lambda + \exp(\theta_v - \sigma_i)} \quad (35)$$

E, atendendo a que  $\cosh(x) = \frac{\exp(x) + \exp(-x)}{2}$ ,

$$P[Y_{vi} = 1] = \frac{\exp \lambda}{\exp \lambda + 2 \cosh(\theta_v - \sigma_i)} \quad (36)$$

Como

$$P[X_{vi} = 0] = 1 - P[Y_{vi} = 1] = \frac{2 \cosh(\theta_v - \sigma_i)}{\exp \lambda + 2 \cosh(\theta_v - \sigma_i)} \quad (37)$$

pode-se escrever a equação geral do modelo do co-seno hiperbólico (HCM):

$$P[X_{vi} = x] = \frac{\exp(\lambda x) 2 \cosh^{1-x}(\theta_v - \sigma_i)}{\exp \lambda + 2 \cosh(\theta_v - \sigma_i)}, \quad x = 0, 1 \quad (38)$$

Em que o parâmetro  $\lambda$  pode ser interpretado como a unidade de medida da escala, pois a maior probabilidade de acordo ocorre quando a verdadeira atitude de um sujeito  $v$  ( $\theta_v$ ) se localiza no intervalo  $\sigma_i \pm \lambda$ , o que reflecte que o sujeito se localiza próximo do item  $i$ . Quanto maior é o valor de  $\lambda$ , maior é a região na qual é verosímil que um sujeito manifeste acordo.

#### O modelo GUM

Baseado nas questões enunciadas por Andrich relativamente às atitudes, ou seja no desacordo por se estar seja à esquerda do item, seja à direita deste, Roberts (1995) desenvolveu um modelo adequado a itens de resposta politômica: o *Graded Unfolding Model* (GUM), que tem a seguinte função massa de probabilidade:

$$P[X_{vi} = x] = \frac{\exp[x(2m+1-x)\lambda + x(\theta_v - \sigma_i)] + \exp[(2m+1-x)(x\lambda + (\theta_v - \sigma_i))]}{\sum_{y=0}^{2m+1} \exp[y(2m+1-y)\lambda + y(\theta_v - \sigma_i)]}, \quad (39)$$

$$x = 0, \dots, m$$

Este modelo deriva directamente de (29) considerando que em vez das  $m+1$  categorias objectivas ( $X = 0, \dots, m$ ) de resposta, haverá  $2(m+1)$  categorias subjectivas ( $Y = 0, \dots, 2m+1$ ), que se agruparão  $Y = i$  e  $Y = 2m+1-i$ ,  $i = 0, \dots, 2m+1$ , para se obter as probabilidades  $P(X = i)$ ,  $i = 0, \dots, m$  respectivamente. O parâmetro  $\lambda$  traduz a distância entre duas barreiras sucessivas das categorias subjectivas e que se considera fixa para todos os itens e todas as categorias:

$$\lambda = \tau_{i(x+1)} - \tau_{ix}, \quad \forall i = 1, \dots, n, \quad \forall x = 1, \dots, m.$$

Este modelo não é uma generalização do HCM, pois no caso de respostas dicotômicas considera também que o acordo se pode considerar dividido em acordo à esquerda e acordo à direita o que dará origem a um modelo ) com 4 categorias subjectivas de resposta.

Para uniformizar resultados na simulação, foi com base neste modelo que se simularam respostas de 500 sujeitos, a escalas do tipo Thurstone ( $m=1$ ) e Likert de 5 ( $m=4$ ) pontos, compostas por 10, 20 e 30 itens.

#### 4. Os resultados simulados

Na simulação utilizou-se o programa Visual Fortran 6.0. As verdadeiras atitudes foram geradas de acordo com uma lei normal de parâmetros 0 e 2.

Nas escalas tipo Thurstone, os índices de localização foram estabelecidos no intervalo de -3.8 a 3.8, a iguais distâncias. Nas escalas tipo Likert, foram considerados apenas itens com índices de localização à esquerda (de -3 a -1) ou à direita (de 1 a 3), tendo-se, portanto, eliminado os itens moderados (Roberts, 1997). As barreiras de cada uma das categorias de resposta foram estabelecidas de forma a que fossem simétricas em relação à localização e iguais para os diferentes itens que integram a escala.

Calculadas, para cada item e cada sujeito, de acordo com o modelo GUM, as probabilidades de resposta em cada uma das categorias, atribuiu-se de forma aleatória uma resposta do sujeito  $v$  ao item  $i$  por comparação dessas probabilidades com números de uma uniforme (0,1) gerados aleatoriamente.

Os resultados em cada uma das supostas escalas de Likert foram calculados por soma das respostas aos diversos itens, tendo-se feito a conversão das respostas aos itens negativos. Nas escalas de Thurstone foi calculada a média ponderada das respostas positivas, sendo os factores de ponderação os índices de localização dos itens.

Replicou-se o procedimento 50 vezes para se ter como resultado de “um sujeito” a média dos scores obtidos em cada uma das escalas.

Os parâmetros dos sujeitos ( $\theta_v$ ) e de localização ( $\sigma_i$ ) foram seleccionados de acordo com os valores utilizados por Roberts em 1997, para simulação do mesmo tipo de dados, após se terem feito algumas experiências. Como se pode ver na Figura 15, escolhendo parâmetros de localização mais extremos as escalas de atitudes tendem a sobrevalorizar as atitudes dos sujeitos localizados na parte central da distribuição e a desvalorizar atitudes mais extremas. Optou-se pelos parâmetros de localização negativa situados entre -3 e -1 e os positivos entre 1 e 3, pois na parte central são os que manifestamente se comportam melhor: para  $\theta=0$ , o resultado da aplicação da escala de Likert com 20 itens está na vizinhança de 50.

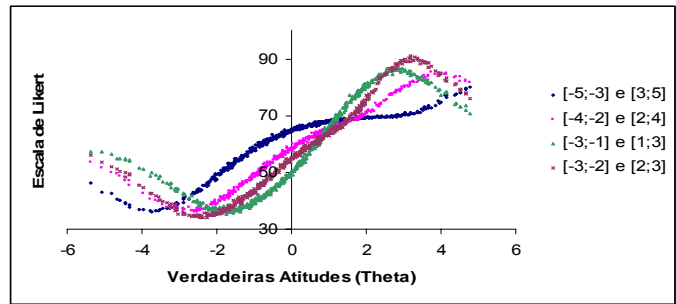


Figura 15: Comparação de vários parâmetros de localização

Obteve-se uma base de dados referentes a 499 “sujeitos” onde se registaram os scores obtidos nos diversos tipos de escalas.



## 5. Análise dos resultados

A distribuição das verdadeiras atitudes (Figura 16) apresenta uma ligeira assimetria negativa (Skewness= - 0.096), registrando-se dois outliers à esquerda (Figura 17). O teste de Kolmogorov-Smirnov indicou tratar-se de valores provenientes de uma distribuição normal ( $p > 0.200$ ).

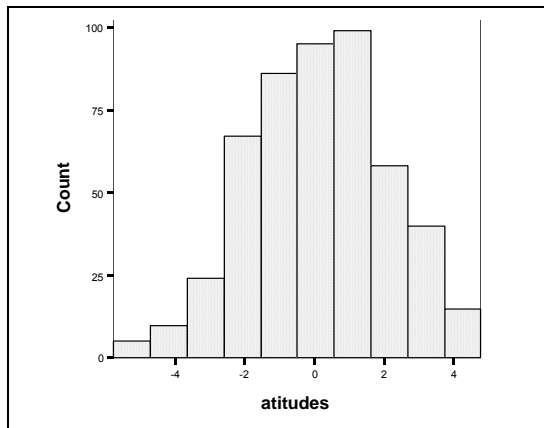


Figura 16: Histograma das verdadeiras atitudes

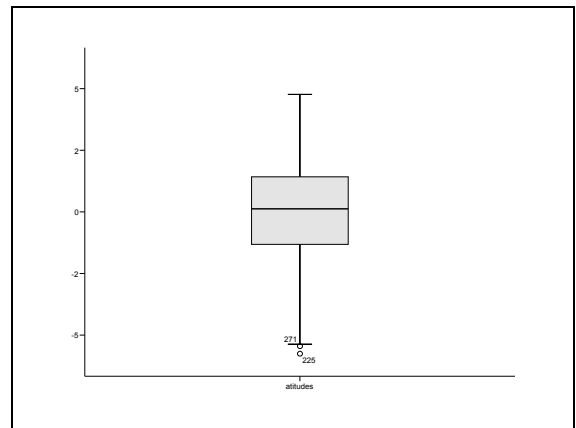


Figura 17: Caixa de bigodes das verdadeiras atitudes

Nenhuma das escalas tem distribuição normal ( $p < 0.027$ ) sendo os resultados nas escalas de Likert mais assimétricos do que os das escalas de Thurstone (Figura 18 e Figura 19).

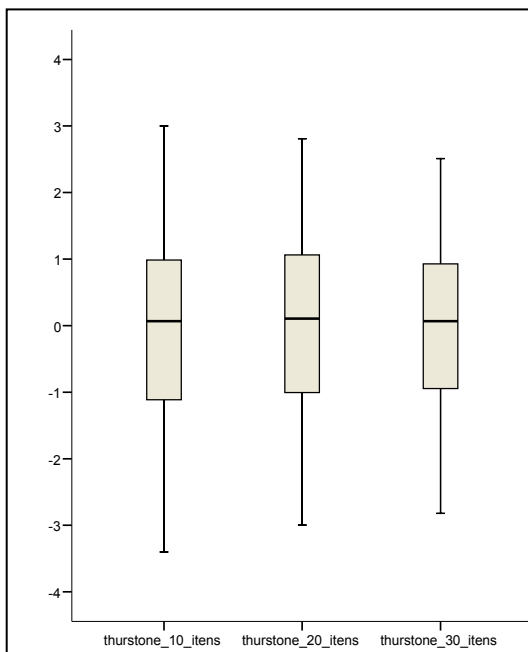


Figura 18: Caixas de bigodes das três escalas tipo Thurstone

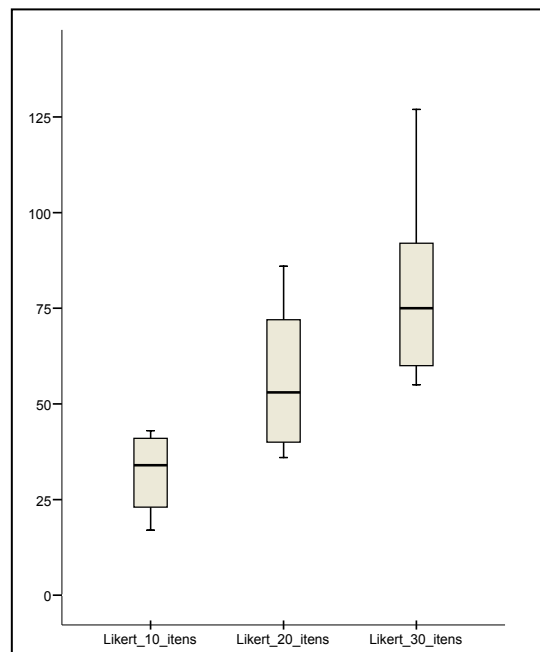
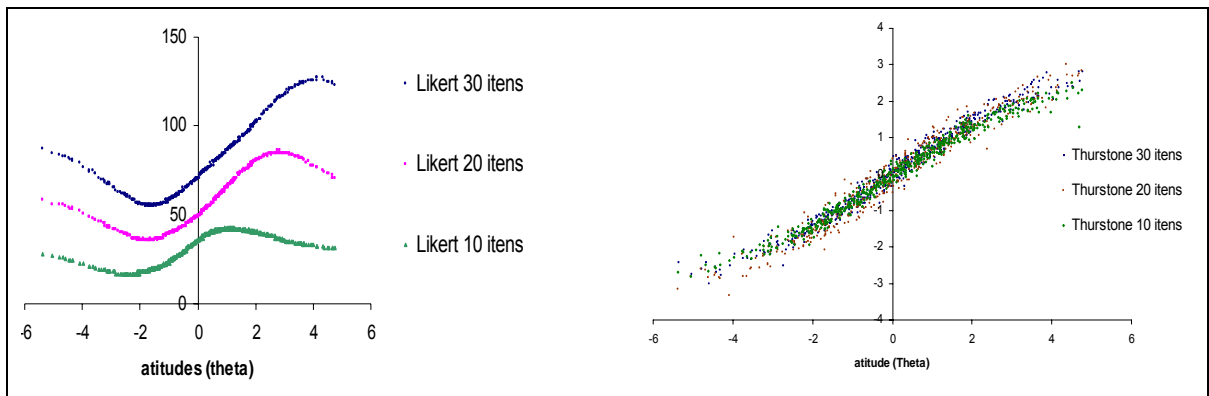


Figura 19: Caixas de bigodes das três escalas tipo Likert

Os diagramas de dispersão obtidos (Figura 20) confirmam as análises de Roberts (1997) que sugerem medições desajustadas ao utilizar escalas de Likert nos sujeitos com atitudes mais extremas, tendendo a apresentar valores mais elevados do que a verdadeira atitude quando esta é muito negativa e valores abaixo da verdadeira atitude quando esta é muito positiva. As escalas de Thurstone apresentam um comportamento mais adequado ao longo de todo o continuum das atitudes.



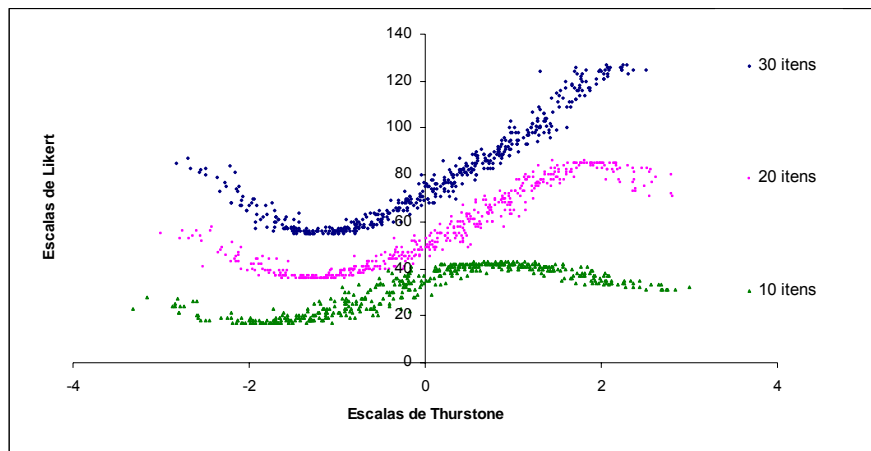
**Figura 20:** Diagramas de dispersão

Note-se que os scores obtidos numa escala de Likert não devem ser considerados com nível de medida intervalar, não se devendo portanto utilizar nesse caso o coeficiente de correlação de Bravais-Pearson. Optou-se então pelo coeficiente de correlação de Spearman em relação a todas as escalas. No estudo da correlação entre as verdadeiras atitudes e as diversas escalas encontraram-se coeficientes de correlação de Spearman significativamente positivos (Tabela 8), sendo os mais elevados nas diversas escalas de Thurstone seguido da escala de Likert com 20 itens o que vem confirmar ser este o número ideal de itens numa escala deste tipo (Roberts, 1997) e que as escalas de Thurstone avaliam melhor as atitudes do que as de Likert.

**Tabela 8:** Coeficientes de correlação não linear

	Spearman's rho		
	atitudes		
	Correlation Coefficient	Sig. (1-tailed)	N
Likert_20_itens	,907	,000	497
Likert_10_itens	,766	,000	497
Likert_30_itens	,886	,000	497
thurstone_20_itens	,992	,000	497
thurstone_10_itens	,982	,000	497
thurstone_30_itens	,993	,000	497

Os diagramas de dispersão da Figura 21 exibem as relações entre os resultados nas escalas de Likert e de Thurstone com o mesmo número de itens.



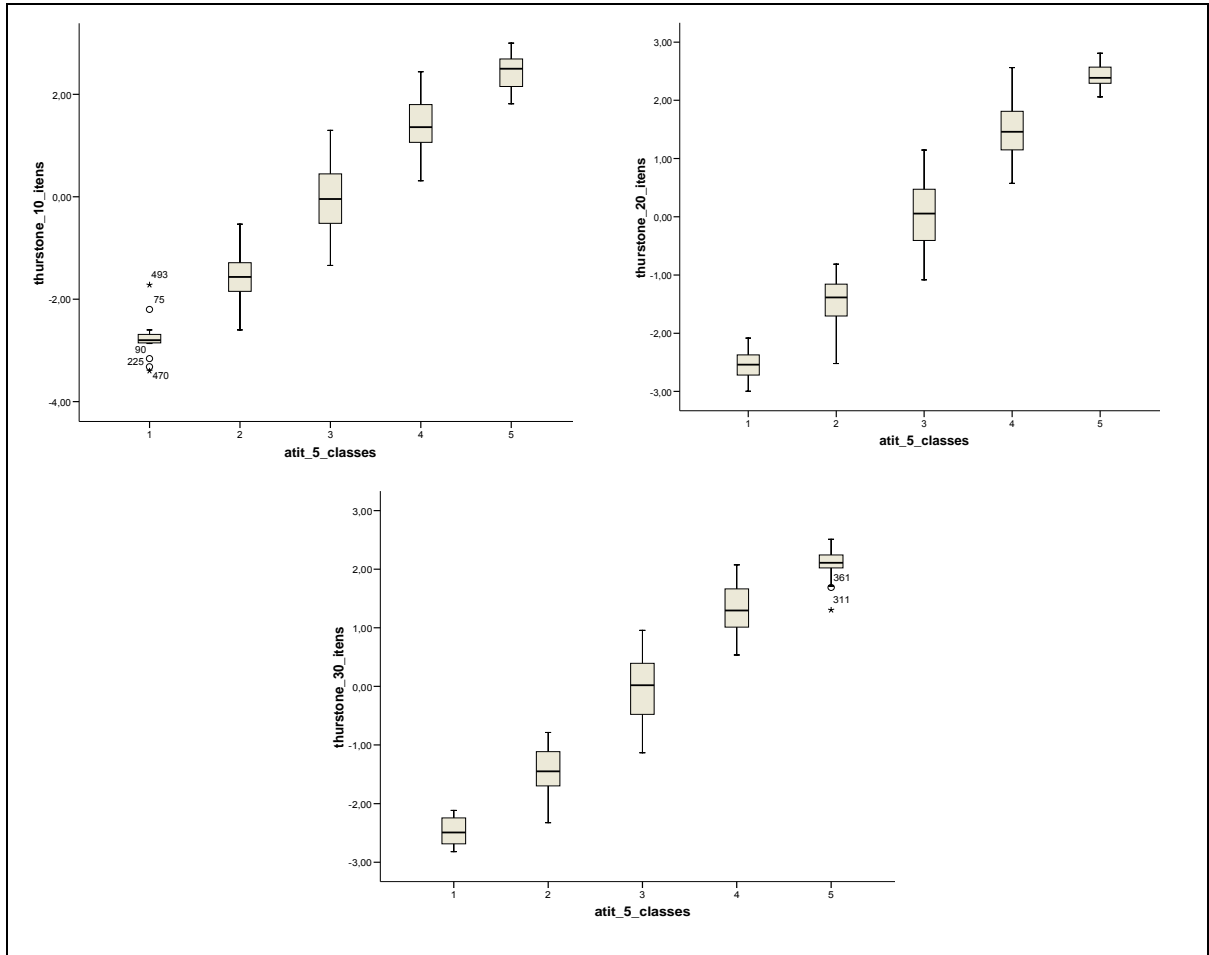
**Figura 21:** Relação entre os resultados em escalas de Likert e de Thurstone com o mesmo número de itens

Embora muitas vezes se utilizem as somas como resultados da aplicação de uma escala de atitudes, também é prática corrente introduzirem-se pontos de corte, dando origem a resultados do tipo: atitude muito negativa (1), atitude negativa (2), atitude moderada (3), atitude positiva (4) e atitude muito positiva (5). Os pontos de corte foram localizados nos percentis 3,6, 27,4, 72,6, 96,4 que corresponderiam à divisão da amplitude de variação de uma distribuição normal de média 0 e desvio padrão 2 (entre  $-3\sigma$  e  $3\sigma$ ) em intervalos de igual amplitude ( $1,2\sigma = 2,4$ ).

**Tabela 9:** Pontos de corte nas diversas escalas

	atitudes	Likert_10_itens	Likert_20_itens	Likert_30_itens	thurstone_10_itens	thurstone_20_itens	thurstone_30_itens
Percentiles 3,6	-3,53627600	17,00	36,00	55,00	-2,5467	-2,1994	-2,1387
27,4	-1,12018800	23,00	41,00	61,00	-,9547	-,9098	-,8735
72,6	1,27621200	40,00	70,00	90,00	,8933	,9661	,8613
96,4	3,65569400	42,00	85,00	124,00	2,2667	2,3589	2,0348

Transformando apenas as verdadeiras atitudes numa variável com 5 classes, os resultados nas escalas de Thurstone apresentam comportamento diferenciado (Figura 22): na escala de 10 itens são poucos os sujeitos de verdadeira atitude muito negativa a terem a mesma classificação que os sujeitos de atitude apenas negativa, mas nos outros níveis há sujeitos com o mesmo resultado que pertencem efectivamente a níveis diferentes. Esta situação verifica-se nas outras duas escalas em qualquer dos níveis. Ao comparar as distribuições dos scores nos cinco escalões verifica-se que são todas significativamente diferentes ( $p < 0,05$ , no teste de Kruskal-Wallis, seguido das comparações múltiplas).



**Figura 22:** Caixas de bigodes das distribuições das classificações de acordo com as três escalas de Thurstone, nos cinco níveis das verdadeiras atitudes

Antevê-se que introduzindo pontos de corte nas três escalas os sujeitos nem sempre sejam bem avaliados ou seja, haverá escalas que colocam sujeitos em escalões a que efectivamente não pertencem. Procurou-se identificar as situações em que cada uma das escalas classifica mal os sujeitos.

As escalas de Thurstone (Tabela 10) classificam razoavelmente os sujeitos nas cinco categorias, apesar de a tendência para os colocar nas categorias adjacentes aumentar nos extremos. A classificação na categoria três que se obtém com a escala de 30 itens parece ser a mais consistente, mas é esta escala que funciona pior nas categorias extremas.

Das três escalas a que funciona melhor é a de 20 itens, apesar de haver metade dos sujeitos com verdadeira atitude muito negativa classificados apenas com atitude negativa.

**Tabela 10:** Classificação dos sujeitos nas diversas escalas de Thurstone

		Thurstone 10 itens				
		1	2	3	4	5
atitudes	1	87,5%	12,5%			
	2	3,5%	90,4%	6,1%		
	3		5,8%	89,3%	4,9%	
	4			16,0%	79,2%	4,8%
	5				31,6%	68,4%
Total		3,6%	23,8%	45,5%	23,2%	3,8%

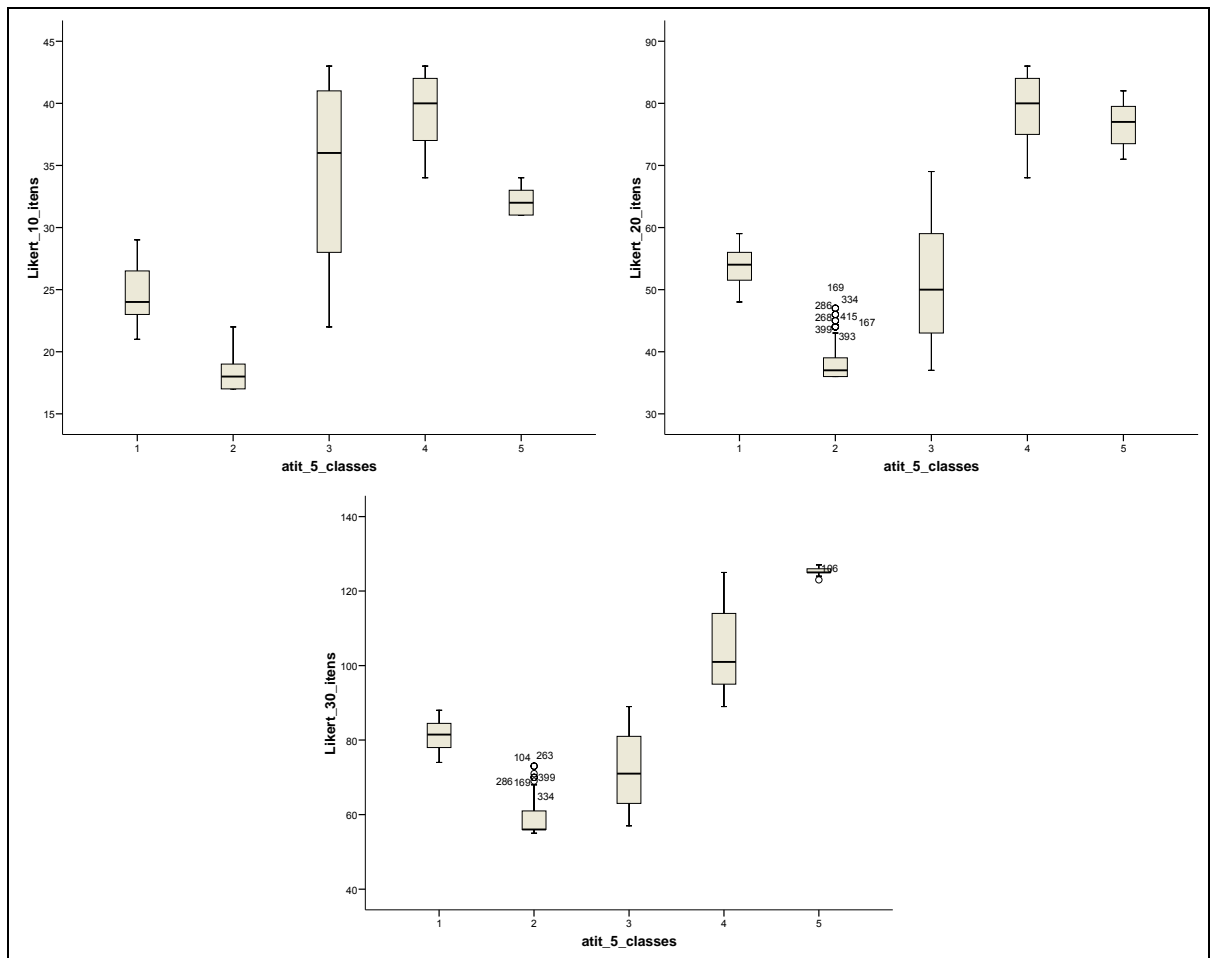
  

		Thurstone 20 itens				
		1	2	3	4	5
atitudes	1	56,3%	43,8%			
	2	,9%	93,9%	5,2%		
	3		3,1%	90,6%	6,3%	
	4			8,8%	86,4%	4,8%
	5				15,8%	84,2%
Total		2,0%	24,4%	44,1%	25,1%	4,4%

		Thurstone 30 itens				
		1	2	3	4	5
atitudes	1	43,8%	56,3%			
	2		90,4%	9,6%		
	3		1,8%	97,3%	,9%	
	4			14,4%	85,6%	
	5				52,6%	47,4%
Total		1,4%	23,4%	49,5%	23,8%	1,8%

A consistência das escalas de Likert é muito pior, transformando apenas as verdadeiras atitudes numa variável com cinco classes, os resultados distribuem-se nessas classes de forma muito heterogénea (Figura 23). Com efeito, para os sujeitos de atitudes muito negativas, as três escalas de Likert dão resultados superiores aos dos sujeitos com atitudes negativas. Na escala de 10 itens os sujeitos de nível 3 e 4 não são diferenciados e muitos deles obtêm resultados superiores aos de verdadeira atitude muito positiva. O resultado no teste de Kruskal-Wallis confirma-o, havendo diferenças estatisticamente significativas nos cinco níveis ( $p < 0.001$ ), ao efectuar as comparações múltiplas os scores na escala de 10 itens, nos níveis 3 e 5 ( $p = 0.315$ ), não são significativamente diferentes o que se traduz por uma desvalorização das atitudes muito positivas. Na escala de 20 itens também não diferem estatisticamente os resultados nos escalões 1 e 3 ( $p = 0,717$ ) e nos escalões 4 e 5 ( $p = 0.758$ ). Então a escala de Likert com 20 itens não classifica bem: atribui “iguais” scores a sujeitos com atitudes muito negativas e sujeitos com atitude moderada.



**Figura 23:** Caixas de bigodes das distribuições das classificações de acordo com as três escalas de Likert, nos cinco níveis das verdadeiras atitudes

Antevê-se que introduzindo pontos de corte nas três escalas os sujeitos nem sempre sejam bem avaliados, ou seja, haverá escalas que colocam sujeitos em escalões a que efectivamente não pertencem. Procurou-se identificar as situações em que cada uma das escalas classifica mal os sujeitos (Tabela 11). Com efeito, as classificações têm um desajuste grande que se manifesta não apenas na fuga de sujeitos para classes adjacentes, em todas estas escalas os sujeitos com atitudes muito negativas são classificados maioritariamente (nas de 20 e 30 itens, é na totalidade) como moderados. Também se verifica o mesmo nos sujeitos de atitude muito positiva, sendo neste caso a escala com 10 itens que funciona pior.

**Tabela 11:** Classificação dos sujeitos nas diversas escalas de Likert

		Likert 10 itens				
		1	2	3	4	5
atitudes	1		31,3%	68,8%		
	2	29,6%	70,4%			
	3		8,9%	59,4%	27,2%	4,5%
	4			51,2%	46,4%	2,4%
	5			100,0%		
Total		6,8%	21,2%	45,5%	23,8%	2,6%

		Likert 20 itens				
		1	2	3	4	5
atitudes	1			100,0%		
	2	37,4%	46,1%	16,5%		
	3		21,4%	78,6%		
	4			9,6%	88,8%	1,6%
	5				100,0%	
Total		8,6%	20,2%	44,7%	26,1%	,4%

		Likert 30 itens				
		1	2	3	4	5
atitudes	1			100,0%		
	2	19,1%	59,1%	21,7%		
	3		21,9%	78,1%		
	4			8,8%	90,4%	,8%
	5				15,8%	84,2%
Total		4,4%	23,4%	45,5%	23,2%	3,4%

Na escala de 10 itens há dez sujeitos de atitude moderada que estão classificados como muito positivos.

Estas escalas discriminam razoavelmente os sujeitos localizados na parte central, mas distorcem os resultados da cauda esquerda e os da cauda direita. Tanto na escala de 20 itens quanto na de 30 itens todos os sujeitos de verdadeira atitude muito negativa aparecem classificados como moderados.

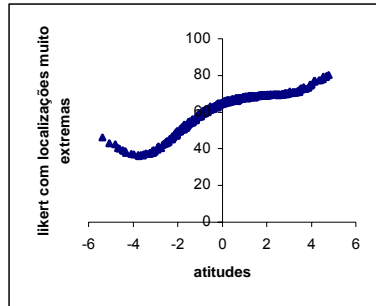
Com a classificação em cinco níveis mantêm-se níveis de correlação elevados em qualquer uma das escalas (Tabela 12).

**Tabela 12:** Coeficientes de correlação não linear para as classificações em cinco níveis.

			Likert 10 itens_5 classes	Likert 20 itens_5 classes	Likert 30 itens_5 classes	Thurstone 10 itens_5 classes	Thurstone 20 itens_5 classes	Thurstone 30 itens_5 classes
Spearman's rho	atit_5_ classes	Correlation Coefficient	,628	,822	,809	,913	,934	,936
		Sig. (1-tailed)	,000	,000	,000	,000	,000	,000
		N	499	499	499	499	499	499

Nos estudos preliminares à simulação utilizaram-se outros índices de localização, como foi referido (p. 55), daí resultando scores finais, numa escala de Likert de 20 itens, mantendo uma forte correlação ( $r_{\text{Spearman}} = 0.996$ ), com sobrevalorização ao longo de toda a escala (Figura 24) mas em que a distribuição pelas cinco classes constituídas é mais consentânea com a que resulta das verdadeiras atitudes dos sujeitos.

Apesar disso esta escala discrimina melhor os sujeitos do que qualquer outra do tipo Likert utilizada neste estudo quer pela distribuição que se pode ver na Tabela 13 quer pelas comparações



**Figura 24:** Diagrama de dispersão com escala de itens muito extremos

múltiplas após o teste de Kruskal-Wallis que apresenta diferenças estatisticamente significativas para todos os pares ( $p < 0.001$ ).

**Tabela 13:** Classificação dos sujeitos numa escala de Likert de 20 itens com localização muito extremada

		Likert 20 itens muito extremos				
		1,00	2,00	3,00	4,00	5,00
atitudes	1	57,1%	42,9%			
	2	12,2%	87,8%			
	3		6,7%	93,3%		
	4			32,8%	66,4%	,8%
	5				21,1%	78,9%
Total		4,4%	24,5%	50,3%	17,5%	3,2%

Há razões para se pensar que as escalas de Thurstone têm qualidades métricas superiores às escalas de Likert. E isto por duas razões que explorámos com a simulação de dados:

- Ao escolher a localização dos itens constroem-se escalas de Likert com propriedades métricas diferentes. Na fase de construção de uma escala de Likert a escolha dos itens que integram a escala final é portanto um passo que se reveste da maior importância. Evidenciou-se que devendo os itens ser claramente positivos ou negativos ao serem muito extremados empolam os scores obtidos com a escala mas discriminam melhor os sujeitos ao estabelecer cinco níveis de classificação.
- As escalas de Likert devem ter 20 itens (Roberts, 1997), mas mesmo essas classificam por vezes os sujeitos com atitudes mais extremas de forma desadequada.



## Conclusão

O estudo das escalas de atitudes presta serviço à psicologia no sentido de se melhorarem os instrumentos de medição. Vários são os saberes que se conjugam quando se pretende medir uma 'qualidade' psicológica. A teoria da medição chamada a intervir desde o início do processo de autonomia da Psicologia, é uma área vastíssima com um desenvolvimento recente, que descobrimos com entusiasmo e que não se esgota na proposta de Stevens.

A teoria de Stevens apenas levantava o problema, mas não fornecia ao investigador um guia explícito de como identificar o tipo de escala e foi (é) usada sem que se investigue cientificamente a questão do tipo de escala, dando origem a que se considere que a atribuição das propriedades de uma escala de intervalos ou de razão é sempre uma convenção (Michell, 2002). A axiomatização da medição proposta por Krantz *et al.* (1971) é um complemento aos métodos de avaliação, pois permite perceber quais as variáveis que podem e as que não podem ser medidas e permite fazer escolhas sobre a melhor forma de medir. Mais do que identificar o nível de medida de cada escala importa conhecer intrinsecamente o atributo e as relações que aí se podem estabelecer para se construir o instrumento de medida que fará corresponder a medição a números interpretáveis.

As escalas de atitudes construídas de acordo com os modelos clássicos propostos por Thurstone, Likert, Guttman e Osgood, que são apresentados de maneira detalhada neste trabalho, garantem a medição ao nível ordinal, mas nem todas podem ser interpretadas como medições que ultrapassem esse nível. Os modelos TRI são o futuro da avaliação em psicologia e muitos são os trabalhos recentes nesta área, havendo a destacar aqueles que, como Keats (1990), não descoram os aspectos formais da teoria representacional da medição para justificarem que as escalas baseadas no modelo Rasch são intervalares.

Este estudo, como outros (Roberts, 1997) levam a considerar que as escalas de Thurstone têm qualidades métricas superiores às escalas de Likert. A construção das primeiras, exigindo a colaboração de um grupo de juízes, tornavam-se mais difíceis de elaborar, o que permite compreender a grande popularidade que as escalas de tipo Likert atingiram.

Actualmente, as atitudes podem ser estimadas através dos modelos TRI, em particular com o modelo GUM que, sendo apresentado nesta dissertação, não é explorado nas suas potencialidades. É nesse sentido que o nosso trabalho poderá evoluir: comparar resultados obtidos em escalas de Likert e de Thurstone numa abordagem clássica (TCT) com estimações resultantes da aplicação do modelo GUM, com base em dados simulados. Dados simulados que podem também

servir para otimizar a dimensão da amostra que se utiliza para estimar os parâmetros do modelo GUM.

Estimando as atitudes com base nos modelos TRI, a partir de um banco de itens, serão ainda as escalas dicotômicas (tipo Thurstone) as que melhor diferenciam as verdadeiras atitudes? A má discriminação das escalas politômicas (tipo Likert) nos extremos, pode ser ultrapassada considerando que 'prevendo' que há atitudes muito positivas (negativas) se use uma escala constituída por itens com localização muito à direita (à esquerda), correndo nesse caso o risco de sobrevalorizar as atitudes moderadas.

## Referências

- Alexandre, V. (1971). *Les échelles d'attitude*. Paris: Éditions Universitaires.
- Anastasi, A. (1990). *Psychological testing*. New-York: MacMillan.
- Anderson, L. W. (1990<sup>a</sup>). Attitudes and their Measurement. In: H. J. Walberg & G. D. Haertel (Eds.), *The International Encyclopedia of Educational Evaluation* (pp 368-374). Oxford: Pergamon.
- Anderson, L. W. (1990<sup>b</sup>). Guttman scales. In: H. J. Walberg & G. D. Haertel (Eds.), *The International Encyclopedia of Educational Evaluation* (pp 333-334). Oxford: Pergamon.
- Andrich, D. (1990). Thurstone scales. In: H. J. Walberg & G. D. Haertel (Eds.), *The International Encyclopedia of Educational Evaluation* (pp 329-332). Oxford: Pergamon.
- Andrich, D. (1997). An hyperbolic cosine IRT model for unfolding direct responses of persons to items. In: W. van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory*. (pp 1-28). New-York: Springer.
- Andrich, D. & Styles, I. (1998). The structural relationship between attitude and behavior statements from the unfolding perspective. *Psychological Methods*, 3, (4), 454-469.
- Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care*, 42 (1-suppl), 7-16.
- Baker, F. (2001). *The basic of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Cantrell C. (1997). *Item response theory: Understanding the one-parameter Rasch model*. (Paper presented at the annual meeting of the Southwest Educational Research Association). ERIC.
- Choppin, B. H. (1990). Item bank. In: H. J. Walberg & G. D. Haertel (Eds.), *The International Encyclopedia of Educational Evaluation* (pp 359-362). Oxford: Pergamon.
- Engelhard, G. (1992). Historical views of invariance: evidence from the measurement theories of Thorndike, Thurstone, and Rasch. *Educational and Psychological measurement*, 52, 275-271.
- Garcia-Marques, L. (2000). O inferno são os outros: o estudo da influência social. In: J. Vala & M. B. Monteiro (Eds.), *Psicologia social*. Lisboa: Fundação Calouste Gulbenkian.
- Ghiglione, R. & Matalon, B. (1993). *O inquirito*. Oeiras: Celta. (Original publicado em 1985).
- Hambleton, R. & Swaminathan, H. (1990). *Item response Theory*. Boston: Kluwer-Nijhoff.

Henard, D. (1998). *Using spreadsheets to implement the one-parameter item response theory (IRT) model*. (Paper presented at the annual meeting of the Southwest Educational Research Association). ERIC.

Henson, R. (1999). *Understanding the one-parameter Rasch model of item response theory*. (Paper presented at the annual meeting of the Southwest Educational Research Association). ERIC.

Keats, J. A. (1990). Measurement in educational research. In: H. J. Walberg & G. D. Haertel (Eds.), *The International Encyclopedia of Educational Evaluation* (pp 237-244). Oxford: Pergamon.

Kline P. (1986). *A handbook of test construction*. London: Methuen.

Krantz, D., Suppes, P., Luce, R.D. & Tversky, A. (1971). *Foundations of measurement* (Vol. I). New-York: Academic Press.

Likert, R., Roslow, S. & Murphy, G. (1993). A simple and reliable method of scoring the Thurstone attitude scales. *Personnel Psychology*, 46, 689-690. (Original publicado em 1934).

Lima, L. (2000). Atitudes: Estrutura e mudança. In: J. Vala & M. B. Monteiro (Eds.), *Psicologia social*. Lisboa: Fundação Calouste Gulbenkian.

Luce, R. D. (1994). Thurstone and Sensory Scaling: Then and Now. *Psychological Review*, 101, (2), 271-277.

Luce, R. D. (1996). The Ongoing Dialog between Empirical Science and Measurement Theory *Journal of Mathematical Psychology*, 40, 78-96.

Luce, R. D. (2005). Measurement Analogies: Comparisons of Behavioral and Physical Measures. *Psychometrika*, 70, (2), 227-251.

Luce, R.D. & Suppes, P. (2001). Representational measurement theory. In: H. Pashler & J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology (3<sup>rd</sup> Edition, Vol.4, pp 1-41)*. New-York: Wiley.

Luce, R.D., Suppes, P., Krantz, D. & Tversky, A. (1990). *Foundations of measurement* (Vol. III). New-York: Academic Press.

MacRae, A. W. (1988). Measurement scales and statistics: What can significance tests tell us about the world? *British Journal of Psychology*, 79, 161-171.

Matalon, B. (1978). La mesure des attitudes. *Bulletin de Psychologie*, 269, 846-876.

Matalon, B. (1996). Les échelles d'attitudes. In: J. C. Deschamps & J. L. Beauvois (Eds.), *Des attitudes aux attributions*, Grenoble: PUG.

McCamey, R. (2002). *A primer on the one-parameter Rasch model*. (Paper presented at the annual meeting of the Southwest Educational Research Association). ERIC.

Michell, J. (2002). Stevens's Theory of Scales of Measurement and Its Place in Modern Psychology. *Australian Journal of Psychology*, 54, (2), 99-104.

- Muñiz J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide
- Muñiz J. (2005). *Análisis de los ítems*. Madrid: La Muralla.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. New-York: McGraw-Hill.
- Pasquali, L. (2003). *Psicometria*. Petrópolis: Vozes.
- Pasquali, L. (2007). *Teoria da resposta ao item*. Brasília: LabPAM.
- Pichot P. (1981). *Les tests mentaux*. Paris. PUF. 8original publicado em 1954).
- Roberts, J. S. (1996). *Item response theory approaches to attitude measurement* (Doctoral dissertation, University of South Carolina, Columbia, 1995). Dissertation Abstracts International: Section B: The Sciences and Engineering, Vol 56, 7089B. US: ProQuest Information & Learning.
- Roberts, J. S. (1997). *Comparative validity of the Likert and Thurstone approaches to attitudes measurement*. (Paper presented at the annual meeting of the American Educational Research Association). ERIC.
- Rost, J. & Langeheine, R. (1997). A Guide through Latent Structure Models for Categorical Data. In: J. Rost & R. Langeheine (Eds.) *Applications of Latent Trait and Latent Class Models in the Social Sciences*. Münster: Waxmann.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Stevens, S. S. (1968). Le quantitatif et la perception. *Bulletin de Psychologie*, 276, 696-715.
- Suppes, P. & Zinnes, J. (1963). Basic measurement theory. In: Luce, Bush & Galanter (Eds.), *Handbook of Mathematical Psychology*. (pp 4-76). New-York: Wiley.
- Suppes, P., Krantz, D., Luce, R.D. & Tversky, A. (1989). *Foundations of measurement* (Vol. II). New-York: Academic Press.
- Teghtsoonian (2001). *International Encyclopedia of the Social & Behavioral Sciences* (15104-15108). Elsevier.
- Thurstone, L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554.
- Thurstone, L. & Chave, E. (1929). *The measurement of attitude: A psycho-physical method and some experiments with ba scale for measuring attitude toward the Church*. Chicago: University of Chicago Press.
- Van der Linden, W. & Hambleton, R. (1997). Item response theory: Brief history, common models, and extensions. In: W. van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory*. (pp 1-28). New-York: Springer.

Velleman P. & Wilkinson (1993). Nominal, ordinal, interval and ratio typologies are misleading. *The American Statistician*, 47, (1), 65-72.

Wright, B. D. (1990). Rasch measurement models. In: H. J. Walberg & G. D. Haertel (Eds.), *The International Encyclopedia of Educational Evaluation* (pp 289-295). Oxford: Pergamon.

## Apêndice

Programa para simulação das atitudes e respectiva medição com uma escala de Thurstone com 30 itens, utilizando o modelo GUM.

```
program thurstone final 20 itens
c   SIMULA ATITUDES....
    USE NUMERICAL_LIBRARIES
    parameter(nm=3000,m=10,mm=3000)
    real seedval, x1, x2, x3, nor, media, des, desvio, a_unif, b_unif, unifm, Theta(nm), Sigma(0:nm), parcela(m), deno, P(nm,m), tau(m), Sigman (0:nm), Sigmap(0:nm), Pfinal(nm,nm), resultado(nm,nm), resposta(nm,nm)
    integer dim, errval, item, replica

c   SIMULAÇÃO DA ATITUDE: NORMAL(0, 2) Geram-se as dim atitudes ou lêem-se de um ficheiro
    write(*,*)'Simulacao de uma normal'
    write(*,*)'Qual a dimensao? (Corresponde ao numero de sujeitos)'
    read(*,*) dim
    write(*,*)'Quantas replicas?'
    read(*,*) replica
    item=30

    call gettim(x1,x2,x3,seedval)
    call seed(seedval)
    OPEN(unit=3,file='t30.dad',iostat=errval)
    OPEN(unit=4,file='Theta Likert dois.dad',iostat=errval)

c   Rotina RNNOF(): "Generate a pseudorandom number from a standard normal distribution."
    Do j=1,dim
        nor=RNNOF()
c       des=2*nor
c       Theta(j)=des
c       WRITE(4,*)Theta (j)
        read(4,*)theta(j)
    ENDDO
    Close(4)
```

```

c      Definição dos sigmas (para os 'item' itens)
      Sigma(1)=-3.8
      Do k=2,item
      Sigma(k)=Sigma(k-1)+0.25
      enddo

c      Geram-se resultados da escala 'replica' vezes
      Do r=1, replica
      Do j=1,dim
      Do k=1,item
c      Para cada sujeito, calculam-se as probabilidades de cada categoria, em cada item, de acordo com
o modelo

      parcela(1)=1+exp(3*Theta (j)-3*Sigma(k))
      parcela(2)=exp(2+Theta(j)-Sigma(k))+exp(2+2*Theta(j)-2*Sigma(k))
c      parcela(3)=exp(14*0.4+2*Theta(j)-2*Sigma(k))+exp(14*0.4+7*Theta(j)-7*Sigma(k))
c      parcela(4)=exp(18*0.4+3*Theta(j)-3*Sigma(k))+exp(18*0.4+6*Theta(j)-6*Sigma(k))
c      parcela(5)=exp(20*0.4+4*Theta(j)-4*Sigma(k))+exp(20*0.4+5*Theta(j)-5*Sigma(k))

      deno=parcela(1)+parcela(2)

      Do i=1,2
      P(j,i)=parcela(i)/deno
      Enddo

c      Simulação de U(0,1) para a regra de escolha da resposta
      Call RNUN(1,unifm)
      soma=0.0
      contador=1.
      Do While (soma.le.unifm)
      soma=soma+P(j,contador)
      contador=contador+1
      Enddo

c      Resposta final e ponderação
      resposta(k,j)=(contador-2)
      Pfinal(k,j)=resposta(k,j)*Sigma(k)

```



```

c    nesta altura tem-se a resposta do sujeito 1 ao item 1, na 1ª aplicação da escala
      Enddo
c    Somam-se as respostas nos item itens para o sujeito 1
      Ssoma=0.0
          uns=0.0
          Do l=1,item
            uns=uns+resposta(l,j)
            Ssoma=Ssoma+Pfinal(l,j)
          Enddo
c    resultado de 1 sujeito numa aplicação da escala
      resultado(r,j)= Ssoma/uns
      Enddo
c    já se têm resultados para uma escala de 'item' itens, para cada um dos sujeitos
      Enddo
c    já se têm resultados na 'replica' réplicas
c    Do ih=1,dim
c    write(3,*)Theta(ih)
c    Enddo
c    write(3,*)
c    write(3,*)
c    Do ih=1,20
c    write(3,*)Sigma(ih)
c    Enddo
c    write(3,*)
c    write(3,*)
      Do ih=1,dim
        write(3,*)(resultado(ik,ih), ik=1,replica)
      Enddo
      Close(3)
100  format(4x,'p0',15x,'p1',12x,'p2',12x,'p3',12x,'p4')
200  format(100(3x,I3))
      END

```

## Índice de figuras

Figura 1: Localização das medianas.....	23
Figura 2: Exemplo de localizações de itens atribuídas por dois juízes .....	23
Figura 3: Curva Característica do item .....	34
Figura 4: Dois itens com localização diferente .....	34
Figura 5: Três itens com discriminação diferente e mesma localização .....	34
Figura 6: Quatro itens com aleatoriedade diferente.....	35
Figura 7: Primeira barreira .....	45
Figura 8: Última barreira .....	46
Figura 9: Exemplo de funções de resposta em 5 categorias.....	48
Figura 10: Localização das barreiras em dois itens no modelo RSM.....	49
Figura 11: Item localizado à esquerda da atitude ( $\sigma_i < \theta_v$ ).....	52
Figura 12: Item com $\sigma_i = \theta_v$ .....	52
Figura 13: Item localizado à direita da atitude ( $\sigma_i > \theta_v$ ) .....	52
Figura 14: Curva característica de um item geral .....	52
Figura 15: Comparação de vários parâmetros de localização .....	56
Figura 16: Histograma das verdadeiras atitudes .....	57
Figura 17: Caixa de bigodes das verdadeiras atitudes.....	57
Figura 18: Caixas de bigodes das três escalas tipo Thurstone .....	57
Figura 19: Caixas de bigodes das três escalas tipo Likert.....	57
Figura 20: Diagramas de dispersão .....	58
Figura 21: Relação entre os resultados em escalas de Likert e de Thurstone com o mesmo número de itens .....	59
Figura 22: Caixas de bigodes das distribuições das classificações de acordo com as três escalas de Thurstone, nos cinco níveis das verdadeiras atitudes .....	60
Figura 23: Caixas de bigodes das distribuições das classificações de acordo com as três escalas de Likert, nos cinco níveis das verdadeiras atitudes.....	62
Figura 24: Diagrama de dispersão com escala de itens muito extremos.....	64

## Índice de fórmulas

(1).....	10	(21).....	46
(2).....	11	(22).....	46
(3).....	11	(23).....	47
(4).....	11	(24).....	47
(5).....	11	(25).....	47
(6).....	12	(26).....	47
(7).....	12	(27).....	47
(8).....	15	(28).....	47
(9).....	15	(29).....	48
(10).....	15	(30).....	52
(11).....	24	(31).....	53
(12).....	33	(32).....	53
(13).....	36	(33).....	53
(14).....	36	(34).....	53
(15).....	36	(35).....	53
(16).....	43	(36).....	53
(17).....	43	(37).....	53
(18).....	44	(38).....	54
(19).....	44	(39).....	54
(20).....	46		

## Índice de tabelas

Tabela 1: Matriz de resultados da aplicação de 18 itens a 36 sujeitos .....	38
Tabela 2: Matriz de resultados da aplicação de 20 itens a 36 sujeitos ordenados por $p_{si}$ e $p_{pj}$ .....	39
Tabela 3: Matriz de resultados da aplicação de 20 itens a 36 sujeitos ordenados por $p_{si}$ e $p_{pj}$ .....	40
Tabela 4: Primeira calibragem dos itens .....	41
Tabela 5: Primeira calibragem dos sujeitos .....	41
Tabela 6: Estimação de Theta .....	44
Tabela 7: Comparação das diversas escalas de atitudes .....	50
Tabela 8: Coeficientes de correlação não linear .....	58
Tabela 9: Pontos de corte nas diversas escalas .....	59
Tabela 10: Classificação dos sujeitos nas diversas escalas de Thurstone .....	61
Tabela 11: Classificação dos sujeitos nas diversas escalas de Likert .....	63
Tabela 12: Coeficientes de correlação não linear para as classificações em cinco níveis. ....	63
Tabela 13: Classificação dos sujeitos numa escala de Likert de 20 itens com localização muito extremada .....	64

## Índice geral

Resumo.....	5
Sumário.....	6
Apresentação.....	7
Introdução.....	8
1. Medir em Psicologia.....	9
1.1. A proposta de Stevens.....	9
1.2. A Teoria representacional da Medida.....	13
1.3. Os níveis de medida e a psicologia moderna.....	16
2. Construindo escalas de atitudes.....	18
2.1. Escalas tipo Thurstone.....	20
2.2. Escalas tipo Likert.....	24
2.3. Escalas tipo Guttman.....	26
2.4. Escalas tipo Diferenciador Semântico.....	29
3. Modelação da atitude usando TRI.....	31
3.1. Escalas de atitudes usando o Modelo Rasch.....	31
3.2. Apresentação geral do modelo.....	32
3.3. O modelo Rasch dicotómico.....	36
3.4. Construção de uma escala de atitudes usando o modelo Rasch dicotómico.....	37
3.5. Estimação da atitude.....	43
3.6. O modelo Rasch politómico.....	45
3.7. Thurstone, Likert, Guttman e o modelo Rasch.....	49
3.8. O modelo seleccionado.....	51
<i>O modelo do co-seno hiperbólico.....</i>	<i>51</i>
<i>O modelo GUM.....</i>	<i>54</i>
4. Os resultados simulados.....	55

5. Análise dos resultados .....	57
Conclusão .....	65
Referências .....	67
Apêndice .....	71
Índice de figuras .....	74
Índice de fórmulas .....	75
Índice de tabelas .....	76
Índice geral .....	77