Métodos de equalização
na teoria clássica e na
teoria da resposta ao item

Rosangela Molini Senno

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO DE MESTRE
EM
ESTATÍSTICA

Área de Concentração: Estatística
Orientador: Prof. Dr. Dalton Francisco de Andrade

São Paulo, fevereiro de 2006.

Métodos de equalização
na teoria clássica e na
teoria da resposta ao item

Este exemplar corresponde à redação final da
dissertação devidamente corrigida e defendida
por Rosangela Molini Senno
e aprovada pela Comissão Julgadora.

São Paulo, 02 de fevereiro de 2006.

Banca Examinadora:

Prof. Dr. Dalton Francisco de Andrade (orientador) – UFSC
Prof. Dr. Luis Carlos Freitas – UNICAMP
Prof. Dr. Heliton Ribeiro Tavares – UFPA
Prof. Dr. Heraldo Vianna – FCC

Suplente:

Prof. Dra. Márcia D'Elia Branco – IME/USP

Aos meus pais, Ana Rosa e João,
à minha irmã Roseli,
ao meu marido Marcio
e ao meu filho Giovanni,
pelo apoio e incentivo,
em todos os momentos.

Agradecimentos

Primeiramente a Deus, sempre presente em minha vida, ajudando-me a superar mais um obstáculo e conquistar mais um objetivo.

Ao meu orientador, Professor Dalton Francisco de Andrade, pela dedicação e seriedade com que assumiu este trabalho, com atenção, conselhos, apoio, confiança e carinho.

À Professora Cláudia Monteiro Peixoto pelo seu apoio e pelos conselhos desde o início desta etapa como orientadora de curso.

Ao Professor Michael J. Kolen por fornecer os programas computacionais necessários para a equalização na Teoria Clássica, e importantes para a realização deste trabalho.

À Raquel da Cunha Valle pelo artigo cedido sob autoria de Petersen, N. S. et al(1989).

A todos os professores do IME-USP pela contribuição à minha formação, em especial à Professora Roseli Fernandez, pela amizade e pela torcida.

Ao grande amigo Conrad pelo imenso apoio e companheirismo, em todos os momentos de alegria, de desespero, de dificuldade e de ansiedade.

A amiga Andréia pela amizade sincera, e pela grande ajuda nas traduções e, logicamente, pela torcida.

A amiga Iracema pela amizade, pela ajuda com o programa Latex e pelos valiosos conselhos.

Agradeço aos amigos do IME-USP Caio, Fernanda, Gisela, Humberto, Jonas,

Regina e Valéria pela amizade, carinho e apoio sempre.

Agradeço a todos os parentes, amigos de trabalho e alunos pela enorme torcida.

Agradeço à Banca Examinadora pelas correções e sugestões apresentadas.

Resumo

Este trabalho tem como objetivo a apresentação dos vários métodos de equalização de testes e de grupos segundo a Teoria Clássica e a TRI - Teoria da Resposta ao Item. Consideramos testes aplicados a grupos equivalentes e não equivalentes. A equalização segundo a Teoria Clássica é descrita por métodos lineares e do equipercentil, e a equalização na TRI é descrita utilizando métodos de transformação e o método dos grupos múltiplos.

Abstract

The purpose of this work is to apresent, according to the Classical Theory and IRT – Item Response Theory, the various methods of tests and groups equating. We consider tests which were applied to equivalent and nonequivalent groups. According to the Classical Theory the equating is described by linear and equipercentile methods, and the equating in IRT is described using transformation methods and multiple groups methods.

Índice

1. Equalização de testes e de grupos – Introdução	1
2. Equalização de testes com grupos equivalentes – Teoria Clássica	5
2.1. Equalização linear	6
2.2. Equalização do equipercentil	9
2.3. Suavização	18
2.3.1. Métodos de suavização a priori	19
2.3.1.1. Modelo Log-linear	20
2.3.1.2. Modelo Beta4	20
2.3.2. Métodos de suavização a posteriori	22
2.4. Três grupos ou mais	25
3. Equalização de grupos não equivalentes – Teoria Clássica	28
3.1. Métodos lineares para grupos não equivalentes	28
3.1.1. Método de Tucker	31
3.1.2. Método do Escore Observado de Levine	34
3.1.3. Método do Escore Verdadeiro de Levine	39
3.1.4. Comparação entre os métodos	41
3.2. Métodos de equipercentil para grupos não equivalentes	44
3.2.1. Método de estimação de frequência	44
3.2.2. Método linear de Braun-Holland	49
3.3. Três grupos ou mais	50
4. Equalização na Teoria da Resposta ao Item	52
4.1. Modelo Logístico Unidimensional de Três Parâmetros	53
4.1.1. Parâmetros dos itens	55
4.1.2. Parâmetros dos respondentes	57
4.2. Equalização na TRI	57
4.2.1. Equalização a posteriori	57
4.2.2. Equalização via modelos de grupos múltiplos	68
4.2.2.1. Parâmetros dos itens	69
4.2.2.2. Parâmetros populacionais	71
5. Recursos computacionais e aplicação dos métodos	73

5.1. Recursos computacionais	73
5.2. Aplicação dos métodos	74
5.2.1. Equalização pela Teoria Clássica	75
5.2.2. Equalização pela TRI	80
5.2.2.1. Equalização a posteriori	81
5.2.2.2. Equalização via grupos múltiplos	86
Conclusão	89
Referências bibliográficas	91
Apêndice	94

Capítulo 1

Equalização de testes e de grupos – Introdução

Testes e/ou provas são formados por um conjunto de itens julgados serem bons, devido a terem passado anteriormente por uma análise de itens (Pinheiro,2006). Os itens podem ser avaliados de várias formas como, por exemplo:

Múltipla escolha Os itens são constituídos de 2, 3 ou mais alternativas, nas quais somente uma é correta. Nos itens que possuem duas alternativas podemos considerar uma prova onde se responda falso ou verdadeiro, ou uma prova já corrigida onde leva-se em consideração se a questão está respondida corretamente ou não.

Notas Os itens são classificados através de notas que podem ser, por exemplo, de 1 a 5, ou de 1 a 10.

Classificados Podemos considerar o caso em que respondemos a uma pesquisa no qual os itens possam estar com as respostas classificadas como 1 para Ótimo, 2 para Bom, 3 para Regular e 4 para Ruim.

Os testes devem ser construídos através de informações detalhadas do que se deseja avaliar, ou seja, precisa ser formulado conforme as características sejam conhecidas e muito bem definidas, fornecendo assim uma linha de desenvolvimento de modo a torná-los indistinguíveis. Existem normas que contribuem para que se tenha um teste adequado como normas nacionais, locais, por idade e série, entre outras discutidas por Angoff(1971).

Devido à grande variedade de grupos, definidos como uma amostra aleatória de uma população mais geral, e variedade de escalas, como, por exemplo, uma escala de

conhecimento em português e uma escala de conhecimento em matemática, existe uma dificuldade na interpretação dos testes. É necessário que se retire o máximo de informação dos testes considerados apropriados.

Na teoria clássica, o conceito estatístico mais utilizado se refere aos escores de cada examinado em um teste, enquanto a Teoria da Resposta ao Item (TRI) trabalha utilizando o traço latente, ou a habilidade de cada examinado.

O conceito inicial de escore define 0 para cada resposta incorreta e 1 para cada resposta correta, onde o escore se processa pela soma de todos os itens respondidos corretamente. Isto é:

$$X_i = \begin{cases} 1 & \text{se correta} \\ 0 & \text{se c.c.} \end{cases}$$

$$X = \sum_{i=1}^K X_i,$$

onde K corresponde ao número de itens do teste.

No entanto, o escore por si só não tem valor algum, a não ser que esteja acompanhado de um suporte de interpretação dessa medida. É natural que os testes sejam diferentes em dificuldade, definindo que um determinado escore em um teste não representa o mesmo nível de habilidade ou terá a mesma interpretação que o mesmo escore em outro teste. O escore é uma das formas utilizadas para a comparação entre indivíduos.

A interpretação dos escores é dada por uma escala de habilidade construída quando o teste tem metas bem especificadas e limites muito bem detalhados de conhecimento. Por exemplo, um examinado que acertou 85% de um teste A_1 pode ter habilidade equivalente a 60% em um teste A_2 que seja mais difícil.

Quando desejamos comparar o desempenho de grupos de indivíduos sobre determinados aspectos em relação à mesma série, ou séries diferentes, entre outros, é necessário adotarmos um sistema confiável que permita a transformação para uma escala comum de habilidade. Para que isso possa acontecer é necessário determinarmos um processo estatístico que ajuste diferenças em dificuldade, e não em conteúdo, denominado *equalização*.

Na equalização, o sistema de unidades de um teste é convertido no sistema de unidades de outro, de modo que os novos escores sejam diretamente equivalentes, porém

determinando uma certa variação nos momentos centrais. Desta forma podemos comparar indivíduos que tenham realizado diferentes testes, os quais não necessariamente necessitam ter o mesmo número de itens.

Por exemplo, os testes devem sempre ser diferentes, pois no caso de uma pessoa fazer duas vezes o mesmo teste, o resultado provavelmente será inflacionado devido ter sido apresentado anteriormente. Ou, no caso de duas pessoas fazerem testes com o mesmo objetivo, porém em datas diferentes, um deles, eventualmente, poderia receber um teste dito mais fácil, tendo assim uma certa vantagem sobre o outro, verificando a existência de uma diferença em dificuldade.

Existem algumas propriedades importantes para os procedimentos de equalização. São elas:

Especificações As características de medida devem ser as mesmas. Por exemplo, um teste que esteja avaliando um certo conhecimento em português não pode ser comparado a um teste que avalie um certo conhecimento em história.

Simetria A equação utilizada para converter escores da escala do teste A_1 para a escala do teste A_2 é a mesma utilizada para o caso contrário, devendo assim ser única. Por exemplo, um escore 12 na escala do teste A_1 pode corresponder a um escore 15 na escala do teste A_2 e vice-versa. O mesmo não ocorre na regressão linear utilizando o método dos mínimos quadrados ordinários, a menos que o coeficiente de correlação seja 1.

Invariância A equalização deve ser a mesma indiferente do grupo utilizado para conduzi-la, isto é, deve ser válida para quaisquer grupos de qualquer população.

Equidade Garante que a qualquer examinado possa ser aplicado qualquer teste sem prejudicá-lo na sua competência. Kolen & Brennan(1995) e Lord(1980) descrevem em detalhes essa propriedade, a qual determina que para todo grupo de idêntica habilidade (capacidade) τ , a distribuição de frequência acumulada condicional dos escores sobre o teste A_2 seja a mesma distribuição de frequência acumulada condicional após a transformação da escala. Ou seja:

$$G^*[eq_Y(x)|\tau] = G(y|\tau), \text{ para todo } \tau \quad (1.1)$$

onde x representa o escore observado no teste A_1 , y representa o escore observado no teste A_2 , $eq_Y(x)$ representa a função de equalização utilizada na transformação da escala, G a distribuição acumulada sobre o teste A_2 e G^* a distribuição acumulada sobre a transformada.

Para o nosso estudo consideraremos dois tipos de grupos mais comuns: equivalentes e não equivalentes.

Grupos Equivalentes Aos examinados, os testes são distribuídos de forma aleatória em um processo em espiral. Para o grupo de examinados os testes são distribuídos alternadamente, ou seja, o primeiro examinado recebe o teste A_1 , o segundo recebe o teste A_2 , e assim por diante. Porém uma restrição ocorre no caso dos examinados já estarem sentados seguindo uma ordem, por exemplo, menino, menina, e assim por diante, pois os grupos não serão considerados aleatórios.

Grupos Não Equivalentes Ao trabalharmos com populações não equivalentes, os examinados são analisados através de itens comuns existentes nas provas, os quais devem ser exatamente iguais em ambos testes podendo ser diretamente compável o rendimento de um grupo em relação ao outro grupo. Por exemplo, provas realizadas em dias diferentes, ou em anos diferentes, ou em séries diferentes.

Existem ainda outros tipos de grupos (Andrade et al, 2000), os quais não estarão em estudo nesse trabalho: um único grupo realizando um único teste, no qual não existe a necessidade da equalização; dois grupos realizando um único teste, onde o teste é considerado comum para ambos grupos, e dois grupos realizando dois testes totalmente distintos, no qual não se permite a equalização.

Este trabalho tem como objetivo descrever o tratamento dado para a equalização na teoria clássica e na teoria da resposta ao item, mostrando métodos e aplicações. O segundo capítulo tratará somente de dois tipos introdutórios de equalização na teoria clássica. O capítulo 3 descreverá métodos de equalização na teoria clássica quando os testes possuem itens comuns. O quarto capítulo tratará dos métodos de equalização na teoria da resposta ao item e, no capítulo 5 serão apresentadas algumas aplicações. O capítulo 6 apresentará a conclusão do trabalho.

Capítulo 2

Equalização de testes com grupos equivalentes – Teoria Clássica

Na teoria clássica, a análise é realizada em função do escore obtido no teste. Ou seja, se tivermos um teste A_1 com K_{A_1} itens, o escore x pode variar de modo que $0 \leq x \leq K_{A_1}$.

É necessário a introdução de dois tipos de equalização como uma indicação da diferença em dificuldade dos testes. Esses métodos são denominados ***Equalização Linear*** e ***Equalização do Equipercantil***. A apresentação de ambos métodos será realizada considerando somente o caso de termos grupos equivalentes cujos testes não possuem itens comuns.

O objetivo da equalização na teoria clássica é que os escores transformados tenham algumas características comuns com o grupo a qual está sendo equalizado. Consideraremos, no momento, somente a existência de dois grupos de indivíduos.

É necessário que a propriedade das especificações esteja muito clara. A propriedade simétrica também é essencial para qualquer forma de equalização assim como as outras propriedades. Porém, na prática, infelizmente nem todas as condições são totalmente satisfeitas.

2.1 Equalização linear

Permite a variação ao longo da escala do escore para as diferenças de dificuldade entre os testes. Nesse caso, os escores que se encontram em igual distância dessas médias em desvio padrão são conjuntos iguais ($z_x = z_y$), de tal forma que:

$$\frac{x - \mu_x}{\sigma_x} = \frac{y - \mu_y}{\sigma_y}, \quad (2.1)$$

onde x representa o escore observado no teste A_1 , y representa o escore observado no teste A_2 , μ_x corresponde ao escore médio obtido no teste A_1 , σ_x o desvio padrão dos escores no teste A_1 , μ_y corresponde ao escore médio obtido no teste A_2 e σ_y o desvio padrão dos escores no teste A_2 . Assim, reescrevendo a expressão (2.1), obtemos:

$$\sigma_y[x - \mu_x] = \sigma_x[y - \mu_y]. \quad (2.2)$$

Portanto, a expressão da equalização linear, rearranjando os termos em (2.2), é dada por:

$$l_y(x) = y = \frac{\sigma_y}{\sigma_x}x + \left[\mu_y - \mu_x \frac{\sigma_y}{\sigma_x} \right] \quad (2.3)$$

onde $l_y(x)$ é denominada a função de conversão linear para transformar os escores observados na escala do teste A_1 para a escala do teste A_2 . Temos $\beta = \frac{\sigma_y}{\sigma_x}$ como a inclinação e $\alpha = \mu_y - \beta\mu_x$ como o intercepto dessa função.

Vamos considerar que tenham sido aplicados dois testes hipotéticos A_1 e A_2 com as mesmas especificações e estruturas, ambos com 15 itens, e realizados por dois grupos equivalentes. As distribuições de frequência dos grupos são dadas na Tabela 2.1, onde $f(x)$ corresponde à frequência relativa do teste A_1 e $g(y)$ corresponde à frequência relativa do teste A_2 .

Utilizando os dados da Tabela 2.1 temos que $\mu_x = 10,0392$, $\mu_y = 8,0489$, $\sigma_x = 2,7446$ e $\sigma_y = 2,8725$ determinam o intercepto e a inclinação da equalização linear, e dados por:

$$\beta = \frac{\sigma_y}{\sigma_x} = \frac{2,8725}{2,7446} = 1,0466, \text{ e}$$

Tabela 2.1: Frequência dos testes A_1 e A_2 para um exemplo hipotético de 15 itens

Escore	$g(y)$	$f(x)$
0	0,0007	0,0007
1	0,0013	0,0007
2	0,0117	0,0053
3	0,0293	0,0100
4	0,0535	0,0233
5	0,0965	0,0266
6	0,1330	0,0339
7	0,1402	0,0631
8	0,1284	0,1090
9	0,1063	0,1276
10	0,0847	0,1462
11	0,0652	0,1329
12	0,0639	0,1216
13	0,0561	0,0997
14	0,0228	0,0664
15	0,0065	0,0332

$$\alpha = \mu_Y - \frac{\sigma_Y}{\sigma_X} \cdot \mu_X = 8,0489 - 1,0466 \cdot 10,0392 = -2,4581.$$

Portanto, a equalização linear resultante é $l_Y(x) = 1,0466x - 2,4581$. Assim:

$$l_Y(6) = 1,0466(6) - 2,4581 = 3,8215,$$

$$l_Y(9) = 1,0466(9) - 2,4581 = 6,9613.$$

Esses valores equiparados mostram que a diferença em dificuldade no teste varia de acordo com o escore. A diferença em dificuldade entre os testes A_1 e A_2 para um escore 6 no teste A_1 é -2,1875 (3,8215 - 6), enquanto que para um escore 9 no teste A_1 será -2,0387 (6,9613 - 9), indicando que o teste A_2 parece ser mais difícil que o teste A_1 .

O processo para determinar a equalização linear do teste A_1 em função do teste A_2 é análogo.

Podemos notar que a função de equalização linear se parece com uma regressão linear onde o termo $\beta = \frac{\sigma_Y}{\sigma_X}$ é multiplicado pela correlação entre os testes. Porém, a regressão linear com o método dos mínimos quadrados ordinários se comporta diferente-

mente da equalização linear pelo fato de não ser simétrica, isto é, a regressão de X em Y é diferente da regressão de Y em X, a não ser que o coeficiente de correlação seja 1.

Na equalização linear a separação entre sucessivas unidades de escore bruto, ou entre unidades de escore escalonado correspondentes a sucessivos escores brutos são considerados iguais somente no senso operacional de que cada escore representa mais um item respondido corretamente em relação ao escore anterior.

Podemos fazer uma verificação da equalização linear com respeito aos dois primeiros momentos. Essa verificação pode nos dar uma idéia dos resultados que devem ser encontrados após a equalização.

$$\begin{aligned} E[l_Y(X)] &= E\left[\frac{\sigma_Y}{\sigma_X}X + \mu_Y - \mu_X \frac{\sigma_Y}{\sigma_X}\right] \\ &= \frac{\sigma_Y}{\sigma_X}E(X) + \mu_Y - \mu_X \frac{\sigma_Y}{\sigma_X} \\ &= \frac{\sigma_Y}{\sigma_X}[E(X) - \mu_X] + \mu_Y = \mu_Y, \end{aligned}$$

onde $E(X) = \mu_X$.

E o desvio padrão da equalização linear é dado por:

$$\sigma[l_Y(X)] = \sigma\left[\frac{\sigma_Y}{\sigma_X}X + \mu_Y - \mu_X \frac{\sigma_Y}{\sigma_X}\right] = \sigma\left[\frac{\sigma_Y}{\sigma_X}X\right] = \sigma_Y.$$

Como todo procedimento estatístico, a equalização considera a existência de um erro de medida (Angoff, 1971), que na equalização linear é dado por:

$$SE^2[l_Y(x)] \cong \frac{2\sigma_Y^2}{N}(z_x^2 + 2),$$

onde N corresponde a amostra total ($n_X + n_Y$), σ_Y^2 a variância amostral de Y e z_x o escore padronizado de X, sendo X a variável aleatória dos escores do teste A_1 e Y a variável aleatória dos escores do teste A_2 .

2.2 Equalização do equipercenil

Nesse método de equalização, escores nos testes são equivalentes se eles correspondem às mesmas ordens percentis. A ordem percentil se refere a uma escala que contribui na distribuição dos escores de grupos que possuam diferentes médias e desvios padrões. Ela informa a porcentagem de indivíduos num grupo que se encontra abaixo do ponto médio do intervalo do escore, isto é, a porcentagem utilizada para verificar a posição de um certo escore no grupo em relação aos demais daquele grupo. Sua escala é dividida em 99 grupos de mesma amplitude.

Seja X a variável aleatória dos escores do teste A_1 e x uma realização de X , Y a variável aleatória dos escores do teste A_2 e y uma realização de Y , F a função densidade acumulada de X na população, G a função densidade acumulada de Y na mesma população, e_Y a função de equalização simétrica para converter escores do teste A_1 para a escala do teste A_2 , e G^* a função densidade acumulada de e_Y na mesma população. Por consequência, a função de equalização do equipercenil é dada por:

$$e_Y(x) = G^{-1}[F(x)], \quad (2.4)$$

onde G^{-1} é a função inversa da G .

O processo é análogo para encontrarmos $e_X(y) = F^{-1}[G(y)]$. Pela propriedade simétrica temos que $e_X^{-1}(x) = e_Y(x)$ e $e_Y^{-1}(y) = e_X(y)$.

Ao trabalharmos com variáveis aleatórias discretas, o que ocorre na maioria dos casos, utilizaremos o ponto médio do intervalo que contém o escore equivalente, de modo a torná-lo contínuo, cujos escores se encontram distribuídos uniformemente num intervalo. Nesse caso, considerando o escore 17, este será representado no intervalo 16,5 - 17,5. Assim, examinados com escore 17 estão uniformemente distribuídos nesse intervalo.

Seja $f(x)$ a função densidade correspondente à proporção de examinados que obtiveram um escore x , $F(x)$ a função de distribuição acumulada e $P(x)$ como a graduação dos percentis dada por:

$$P(x) = 100[F(x-1) + f(x)/2]. \quad (2.5)$$

Suponhamos que em um teste A_1 , 25% dos examinados receberam escore menor ou igual a 15. Em um teste A_2 , tivemos 21% dos examinados com escore menor ou igual a 16 e 28% com um escore menor ou igual a 17. Pela equação (2.5), a graduação de percentil de um escore inteiro de 15 no teste A_1 é a graduação de percentil da porcentagem mais próxima no teste A_2 acrescido do ponto médio do intervalo que contém esse escore. Portanto, temos $21\% + 1/2(28\%-21\%) = 21\% + 1/2 (7\%) = 24,5\%$.

A Tabela 2.2 se refere aos testes A_1 e A_2 acrescidos da frequência acumulada e da graduação dos percentis. Por exemplo, queremos encontrar a ordem percentil no teste A_1 do escore equivalente a 8. Então $P(x) = 100[F(7)+f(8)/2] = 100[0,1635 + 0,1090/2] = 100[0,1635 + 0,0545] = 21,7940$. O processo é análogo para encontrarmos a ordem percentil equivalente ao teste A_2 , onde $Q(y) = 100 [G(y-1)+g(y)/2]$.

Tabela 2.2: Distribuição do escore dos testes A_1 e A_2

Escore	$f(x)$	$F(x)$	$P(x)$	$g(y)$	$G(y)$	$Q(y)$
0	0,0007	0,0007	0,0332	0,0007	0,0007	0,0326
1	0,0007	0,0013	0,0997	0,0013	0,0020	0,1304
2	0,0053	0,0066	0,3987	0,0117	0,0137	0,7823
3	0,0100	0,0166	1,1628	0,0293	0,0430	2,8357
4	0,0233	0,0399	2,8239	0,0535	0,0965	6,9752
5	0,0266	0,0664	5,3156	0,0965	0,1930	14,4720
6	0,0339	0,1003	8,3389	0,1330	0,3259	25,9452
7	0,0631	0,1635	13,1894	0,1402	0,4661	39,6023
8	0,1090	0,2724	21,7940	0,1284	0,5945	53,0313
9	0,1276	0,4000	33,6213	0,1063	0,7008	64,7653
10	0,1462	0,5462	47,3090	0,0847	0,7855	74,3155
11	0,1329	0,6791	61,2625	0,0652	0,8507	81,8123
12	0,1216	0,8007	73,9867	0,0639	0,9146	88,2660
13	0,0997	0,9003	85,0498	0,0561	0,9707	94,2634
14	0,0664	0,9668	93,3555	0,0228	0,9935	98,2073
15	0,0332	1,0000	98,3389	0,0065	1,0000	99,6741

Existem procedimentos gráficos que utilizam os limites superior e inferior de cada escore para cada teste, os quais são comparados para obtermos os escores equivalentes. Através da frequência relativa dos testes dada pela Figura 2.1 percebemos que ambas distribuições são positivamente inclinadas e, da mesma forma, o teste A_1 parece ser mais fácil que o teste A_2 . Isso indica que o desempenho do grupo de examinados que fizer o

teste A_1 será um pouco melhor em relação ao grupo que fizer o teste A_2 . Observando a Figura 2.2 verificamos que o teste A_1 se encontra abaixo do teste A_2 , cujos escores se encontram distribuídos uniformemente no intervalo entre -0,5 e 15,5. Como o teste A_1 aparece um pouco deslocado para a direita, este parece ser um pouco mais fácil que o teste A_2 .

Figura 2.1: Frequência relativa dos testes A_1 e A_2

Se acontecer de nenhum examinado obter um escore particular, teremos que a graduação do percentil pode não ser única, de modo que o escore médio seja condizente com os procedimentos. Por exemplo, se nenhum examinado recebeu um escore de 3 e 4, a sua ordem percentil será equivalente ao intervalo de 2,5 a 4,5, cujo ponto médio correspondente será 3,5.

Assim, sejam K_{A_1} o número de itens do teste A_1 , X a variável aleatória dos escores do mesmo teste variando de $0, 1, \dots, K_{A_1}$, $f(x)$ a função densidade discreta para X e $F(x)$ a função densidade acumulada discreta do teste, onde temos as seguintes definições:

Figura 2.2: Ordem percentil dos testes A_1 e A_2

$$f(x) \geq 0 \text{ para } x = 0, 1, \dots, K_{A_1};$$

$$f(x) = 0, \text{ se caso contrário, e}$$

$$\sum f(x) = 1,$$

e

$$0 \leq F(x) \leq 1 \text{ para } x = 0, 1, \dots, K_{A_1};$$

$$F(x) = 0 \text{ para } x < 0, \text{ e}$$

$$F(x) = 1 \text{ para } x > K_{A_1}.$$

Dessa forma, considerando que o escore x não seja um valor inteiro, devido ao fato de estar distribuído uniformemente no intervalo, definiremos x^* como aquele obtido

por arredondamento tal que $x^* - 0,5 \leq x < x^* + 0,5$, e, generalizando a função de graduação do percentil, temos:

$$\begin{aligned} P(x) &= 100F(x^* - 1) + [x - (x^* - 0,5)][F(x^*) - F(x^* - 1)], \text{ se } -0,5 \leq x < K_{A_1} + 0,5 \\ &= 0, \text{ se } x < -0,5 \\ &= 100, \text{ se } x \geq K_{A_1} + 0,5. \end{aligned}$$

Por exemplo, suponha que tenhamos $x = 2,4$. Utilizando a equação acima, e baseado na Tabela 2.2 em relação ao teste A_1 , sabendo que $x^* = 2$ temos:

$$\begin{aligned} P(2,4) &= 100\{F(1) + [2,4 - (2 - 0,5)][F(2) - F(1)]\} \\ &= 100\{0,0013 + [2,4 - 1,5][0,0066 - 0,0013]\} \\ &= 100\{0,0013 + 0,0048\} = 0,61. \end{aligned}$$

Dessa forma, a ordem percentil de um escore de 2,4 na escala do teste A_1 equivale a 0,6100 na escala do teste A_2 .

Dada a função de graduação do percentil, sua função inversa é utilizada para encontrar o escore correspondente a ele. Essa função inversa será denominada por P^{-1} , onde $P(x)$ será o percentil. Então:

$$\begin{aligned} P(x) &= 100F(x^* - 1) + [x - (x^* - 0,5)][F(x^*) - F(x^* - 1)], \\ F(x^* - 1) + [x - (x^* - 0,5)][F(x^*) - F(x^* - 1)] &= P(x)/100, \\ [x - (x^* - 0,5)][F(x^*) - F(x^* - 1)] &= P(x)/100 - F(x^* - 1), \\ x[P(x)] = P^{-1}[P(x)] &= \frac{P(x)/100 - F(x^* - 1)}{F(x^*) - F(x^* - 1)} + (x^* - 0,5). \end{aligned}$$

Mas temos duas funções alternativas que produzem os mesmos resultados. São elas:

$$\begin{aligned} x_u[P(x)] &= P^{-1}[P(x)] = \frac{P(x)/100 - F(x_u^* - 1)}{F(x_u^*) - F(x_u^* - 1)} + (x_u^* - 0,5), \text{ se } 0 \leq P(x) < 100 \\ &= K_x + 0,5, \text{ se } P(x) = 100, \end{aligned}$$

onde x_u^* é o menor escore inteiro com a porcentagem acumulada maior que $P(x)$, e

$$\begin{aligned} x_L[P(x)] &= P^{-1}[P(x)] = \frac{P(x)/100 - F(x_L^*)}{F(x_L^* + 1) - F(x_L^*)} + (x_L^* + 0,5), \text{ se } 0 < P(x) \leq 100 \\ &= -0,5, \text{ se } P(x) = 0, \end{aligned}$$

onde x_L^* é o maior escore inteiro com a porcentagem acumulada menor que $P(x)$.

Se $f(x) \neq 0$ para todo escore $0, 1, \dots, K_{A_1}$ então $x = x_L = x_U$. Se $f(x) = 0$ para algum escore então $x_L \neq x_U$ e, nesse caso, convencionamos o uso de $x = \frac{x_L + x_U}{2}$. O processo é análogo para determinar $Q^{-1}[Q(y)]$.

Utilizando o equipercntil equivalente para encontrar no teste A_2 o escore correspondente ao do teste A_1 temos a seguinte equalização:

$$e_Y = y = Q^{-1}[P(x)], \text{ se } -0,5 \leq x \leq K_{A_1} + 0,5. \quad (2.6)$$

Como é válida a propriedade simétrica, temos que $e_X = x = P^{-1}[Q(y)]$, $-0,5 \leq y \leq K_{A_2} + 0,5$ onde K_{A_2} se refere ao número de itens do teste A_2 , $g(y)$ a sua densidade discreta, $G(y)$ sua função densidade acumulada, $Q(y)$ o percentil graduado e Q^{-1} a inversa da função do percentil.

Utilizando o exemplo citado anteriormente, através dos resultados da Tabela 2.2, queremos determinar o equipercntil equivalente ao escore 8 no teste A_1 , no qual $P(8) = 21,7940$, o correspondente no teste A_2 , onde $y_u^* = 6$, pois esse é o menor $G(y)$ que é maior que 0,2179. Portanto,

$$e_Y = Q^{-1}[P(x)] = \frac{P(x)/100 - G(y_u^* - 1)}{G(y_u^*) - G(y_u^* - 1)} + (y_u^* - 0,5),$$

e, para $x = 8$,

$$\begin{aligned} e_Y(8) &= \frac{21,7940/100 - G(5)}{G(6) - G(5)} + (6 - 0,5) \\ &= \frac{0,2179 - 0,1930}{0,3259 - 0,1930} + 5,5 = 5,6874. \end{aligned}$$

Uma vantagem da equalização do equipercentil é que os escores equiparados sempre se encontram no intervalo $-0,5 \leq e_Y(x) \leq K_Y + 0,5$, o que não ocorre com a equalização linear. Podemos observar esse fato através das equações (2.3) e (2.6) comparando os resultados obtidos através da Tabela 2.3 referente ao nosso exemplo.

Tabela 2.3: Conversão do escore bruto do Teste A_1 para o Teste A_2

Escore	Linear	Equipercentil
0	-2,4581	-0,4949
1	-1,4115	0,7645
2	-0,3649	1,6731
3	0,6817	2,3243
4	1,7283	2,9960
5	2,7749	3,6895
6	3,8215	4,2551
7	4,8681	4,8671
8	5,9147	5,6878
9	6,9613	6,5733
10	8,0079	7,5544
11	9,0545	8,6703
12	10,1011	9,9612
13	11,1477	11,4966
14	12,1943	12,8381
15	13,2409	14,0577

Somente através da Figura 2.3 não conseguimos determinar a diferença entre ambos tipos de equalização. No entanto, observando a Figura 2.4, a equalização do equipercentil se mostra não linear, o que é bastante plausível, porém está mais sujeita a erros e, portanto, existe a necessidade do uso de métodos de suavização. Na Figura 2.4 comparamos ambas equalizações com a equalização identidade, a qual um escore no teste A_1 corresponde ao mesmo escore no teste A_2 , sugerindo assim que ambos testes possuem a mesma distribuição e não diferem em dificuldade, o que não ocorreu nesse exemplo. Se tivermos um examinado com escore 8 no teste A_1 , pela Figura 2.4, acompanhando a equalização linear temos que $8 - 2,1 = 5,9$ que é aproximadamente o valor apresentado na Tabela 2.3. O mesmo escore 8, observando a equalização do equipercentil na Figura 2.4, temos que $8 - 2,3 = 5,7$, e cujo valor é aproximadamente o resultado apresentado na Tabela 2.3 para esse escore.

Figura 2.3: Escore do teste A_1 equalizado para a escala do teste A_2

Para termos uma idéia mais abrangente do que acontece em cada teste e podermos compará-los, verificaremos os quatro primeiros momentos centrais: média, desvio padrão, assimetria e curtose que se encontram na Tabela 2.4, onde a assimetria e a curtose são dadas, respectivamente, pelas equações a seguir:

$$ass_x = \frac{E(X - \mu_x)^3}{[\sigma_x]^3}, \text{ e}$$

$$curt_x = \frac{E(X - \mu_x)^4}{[\sigma_x]^4}.$$

Da Tabela 2.4 podemos ver que, ambos testes possuem média μ maior que 7,5, equivalente a 50% dos 15 itens do teste, indicando que os testes são razoavelmente fáceis para estes examinados. O teste A_1 é, em média, aproximadamente 2 pontos mais fácil que o teste A_2 . Observando o desvio padrão σ percebemos que o teste A_2 tem maior variação em relação ao teste A_1 . A assimetria ass indica que o teste A_2 tem distribuição positi-

Figura 2.4: Diferença entre as equalizações realizadas e o escore do teste A_1

Tabela 2.4: Conversão do escore bruto no Teste A_1 para o Teste A_2

	μ	σ	ass	$curt$
Teste A_2	8,0489	2,8725	0,2271	2,4279
Teste A_1	10,0392	2,7446	-0,4446	2,9681
$l_Y(x)$	8,0489	2,8725	-0,4446	2,9681
$e_Y(x)$	8,0434	2,8596	0,2028	2,4113

vamente assimétrica enquanto o teste A_1 possui distribuição negativamente assimétrica. Com base na curtose estimada $curt$ notamos que o teste A_1 possui distribuição aproximada da normal, cuja curtose vale 3, porém o teste A_2 possui curtose abaixo da distribuição normal (platicúrtica).

Analisando os momentos centrais para os métodos de equalização na Tabela 2.4 observamos que a equalização linear obteve resultados semelhantes ao teste A_2 nos dois primeiros momentos, no entanto, a assimetria e curtose são equivalentes ao teste A_1 . Já na equalização do equipercantil os resultados se encontram mais aproximados em todos

os momentos daqueles encontrados para o teste A_2 , indicando que este método é o que melhor se ajusta a nível de equiparação entre testes neste exemplo.

Logo, o desempenho médio de um grupo de examinados que fizer o teste A_1 tende a ser melhor do que ao fazer o teste A_2 , apesar do desempenho em ambos testes se encontrar acima da média.

O erro de medida associado a esse método de equalização (Angoff, 1971) é dado por:

$$\hat{se}[e_Y] \cong \sigma_Y^2 \frac{[\frac{P(x_i)}{100}][1 - \frac{P(x_i)}{100}]}{\phi^2} \left[\frac{1}{n_X} + \frac{1}{n_Y} \right]^{1/2},$$

onde x_i corresponde ao escore do teste A_1 com $0 \leq x_i \leq K_{A_1}$ e ϕ se refere à ordenada da densidade normal padrão.

2.3 Suavização

Quando se utiliza percentis amostrais ou ordem percentil na estimação das relações de equipercartil, estas não são suficientemente precisas na prática devido ao erro amostral, e as distribuições do escore e as relações do equipercartil são irregulares na estimação do percentil equivalente.

Existe então a necessidade de uma estrutura conceitual que trabalhe com os erros envolvidos na suavização. Segundo Kolen & Brennan (1995), temos esse erro dado pela relação a seguir:

$$\underbrace{\hat{t}_Y(x_i) - e_Y(x_i)}_{\text{Erro Total}} = \underbrace{\hat{t}_Y(x_i) - t_Y(x_i)}_{\text{Erro Aleatório}} + \underbrace{t_Y(x_i) - e_Y(x_i)}_{\text{Erro Sistemático}},$$

onde $e_Y(x_i)$ corresponde ao equipercartil equivalente a um determinado escore, $\hat{t}_Y(x_i)$ um estimador alternativo de $e_Y(x_i)$ e resultante do método de suavização utilizado e $E[\hat{t}_Y(x_i)] = t_Y(x_i)$.

Em termos do erro médio quadrático temos:

$$\begin{aligned} mse[\hat{t}_Y(x_i)] &= var[\hat{t}_Y(x_i)] + \{viés[t_Y(x_i)]\}^2 \\ &= E[\hat{t}_Y(x_i) - t_Y(x_i)]^2 + [t_Y(x_i) - e_Y(x_i)]^2. \end{aligned}$$

Os métodos de suavização devem reduzir a variabilidade do erro de forma a diminuir o erro aleatório compensado pela introdução do erro sistemático, diminuindo assim, o erro total. Dessa forma, a suavização tem o objetivo de melhorar a precisão dos resultados da equalização do equipercentil se comparado àquelas não suavizadas. Esses métodos devem ser flexíveis quanto à variedade de distribuições e às relações de equipercentil encontradas na prática, levando em consideração uma estrutura estatística de ajuste considerável.

Para grandes amostras, em geral, as distribuições do escore e as relações do equipercentil deverão ficar razoavelmente suavizadas. Porém, no caso de termos amostras muito pequenas, a relação pode se apresentar irregular mesmo após a suavização devido à descontinuidade.

Existem dois métodos de suavização : suavização a priori e suavização a posteriori. O método de suavização a priori envolve a suavização das distribuições do escore para equiparar ambos testes A_1 e A_2 antes que a equalização do equipercentil seja efetuada. O método de suavização a posteriori envolve a suavização da função de equalização do equipercentil produzida da distribuição do escore não suavizado.

Os métodos de suavização a priori possuem uma propriedade importante quanto à exatidão, e denominada preservação dos momentos, cuja distribuição suavizada tem pelo menos um dos mesmos momentos centrais que a distribuição observada. São mais comumente usados dois métodos de suavização a priori: o modelo log-linear e o modelo beta4 (Lord, 1965). No caso de suavização a posteriori, é mais utilizado o modelo de "cubic splines" por possuir grande flexibilidade em relação ao uso de polinômios (Kolen, 1995).

2.3.1 Métodos de suavização a priori

Se baseiam na premissa de que as distribuições de X e Y têm o potencial de melhorar sua estimação e, provavelmente, suas funções de equalização do equipercentil.

Possuem uma propriedade importante quanto a exatidão denominada preservação dos momentos, cuja distribuição suavizada tem pelo menos algum dos mesmos momentos centrais que a distribuição observada. Por exemplo, se o método preservar os três primeiros momentos centrais, significa que a média, o desvio padrão e a assimetria são os mesmos para as distribuições suavizadas e não suavizadas.

Discutiremos dois métodos de pré suavização: o modelo log-linear e o modelo beta4.

2.3.1.1 Modelo Log-linear

O modelo log-linear, para a distribuição de X , tem a seguinte forma:

$$\begin{aligned}\log[f(x)] &= \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_C x^C \\ &= \alpha_0 + \sum_{K=1}^C \alpha_K x^K.\end{aligned}$$

Os parâmetros α são estimados através do método de máxima verossimilhança cujos procedimentos para modelos log-lineares são dados por Haberman(1974).

Neste modelo, a aplicação do logaritmo permite a sua utilização de forma aditiva. O modelo garante que os primeiros C momentos da distribuição ajustada sejam iguais aos obtidos pela distribuição amostral.

Na escolha do melhor C é utilizado o teste da razão de verossimilhança, no qual a qualidade do ajuste do modelo é verificada comparando-se o valor observado da estatística χ^2 com o valor crítico da distribuição para 1 grau de liberdade. São preferíveis modelos mais simples que ajustem adequadamente a distribuição, considerando que modelos muito complicados definem maior erro de equalização. Haberman(1974) também descreve a escolha do melhor modelo. Neste método procedimentos gráficos são utilizados para verificar qual das distribuições ajustadas são suavizadas e não divergem muito da distribuição empírica.

2.3.1.1 Modelo Beta4

Lord(1965) generaliza o modelo beta binomial utilizando modelos paramétricos para escores verdadeiros. O modelo beta4 é também denominado modelo robusto do escore verdadeiro pois estima a sua distribuição. A distribuição dos escores observados é dada da seguinte forma:

$$f(x) = \int_i^s f(x|\eta)\psi(\eta)d\eta,$$

onde i corresponde ao limite inferior, s se refere ao limite superior, $\psi(\eta)$ é a proporção

correta de escores ajustados e $f(x|\eta)$ é a condicional da distribuição dos escores observados dados os escores verdadeiros.

Seja a distribuição beta de 4 parâmetros dada por:

$$\psi(\eta) = \begin{cases} \frac{(-a+\eta)^{\alpha-1} (b-\eta)^{\beta-1}}{A^{\alpha+\beta-1} B(\alpha, \beta)} & , \text{ se } 0 \leq a \leq \eta \leq b \leq 1 \\ 0 & , \text{ se } \eta < a \text{ ou } \eta > b, \end{cases}$$

onde

$$A = b - a, \quad e$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Assumindo que $f(x|\eta)$ seja uma binomial composta, temos:

$$f(x|\eta) = p_n(x) + k\eta(1-\eta)C_r(x) \quad \text{para } x = 0, 1, \dots, n,$$

onde

$$p_n(x) = \begin{cases} \binom{n}{x} \eta^x (1-\eta)^{n-x} & , \text{ para } x = 0, 1, \dots, n \\ 0 & , \text{ se } x < 0 \text{ ou } x > n, \end{cases}$$

$$C_r(x) = \sum_{u=0}^r (-1)^{u+1} \binom{r}{u} p_{n-r}(x-u), \quad r = 2, 3, \dots, n,$$

no qual k é um parâmetro da distribuição estimada a partir de um coeficiente de fidedignidade (Vianna, 1987), ou seja, define a precisão de um teste, o qual é consistente com a distribuição dos escores e estável quanto a característica no tempo.

Neste modelo, os 4 primeiros momentos centrais da distribuição ajustada concordam com os da distribuição amostral, contanto que não existam estimativas de parâmetros inválidas. Caso contrário, teremos um número menor de momentos centrais concordantes.

O ajuste do modelo é feito graficamente e pela análise dos momentos centrais. A boa qualidade do ajuste é feita através da estatística χ^2 padrão com $k - 4 = (k + 1) - 1 - 4$ graus de liberdade ($k+1$ escores de $0, \dots, k$ menos 1, e menos o número de parâmetros ajustados). Lord (1965) apresenta todos os passos deste método, e também apresenta variações deste modelo. Esse método, porém, não terá sua aplicação realizada neste trabalho.

2.3.2 Métodos de suavização a posteriori

Nestes métodos o procedimento de suavização é aplicado na função de equalização do equipercentil, ajustando uma curva para essa relação, que deve ser suavizada sem divergir muito da observada.

Neste trabalho será apresentado somente o modelo de "cubic splines" (Kolen, 1995 e Hanson et al, 1994), que possui grande flexibilidade em relação ao uso de polinômios.

Neste modelo, para escores inteiros x_i temos a seguinte função "spline":

$$\hat{d}_Y(x) = v_{0i} + v_{1i}(x - x_i) + v_{2i}(x - x_i)^2 + v_{3i}(x - x_i)^3, \quad x_i \leq x \leq x_{i+1}.$$

Os pesos ($v_{0i}, v_{1i}, v_{2i}, v_{3i}$) são diferentes para cada escore pois existe uma equação cúbica para cada um. A função "spline" é ajustada sobre o intervalo dos escores entre x_{min} e x_{max} tal que $0 \leq x_{min} \leq x \leq x_{max} \leq K_X$ onde x_{min} é o menor escore inteiro e x_{max} é o maior escore inteiro do teste A_1 .

O grau de suavização é controlado por um parâmetro S tal que:

$$\frac{\sum_{i=min}^{max} \frac{\hat{d}_Y(x_i) - \hat{e}_Y(x_i)}{\hat{s}_e[\hat{e}_Y(x_i)]}}{x_{max} - x_{min} + 1} \leq S,$$

onde $\hat{s}_e$ é o erro padrão estimado da equalização do equipercentil. (Kolen & Brennan, 1995)

Para $S = 0$, a função de equalização do equipercentil permanece a mesma. Para S muito grande temos uma função linear. Porém, na prática, quando $0 < S < 1$, os resultados obtidos se mostram mais adequados.

No caso do maior e menor escore, utiliza-se interpolação linear para obtermos a suavização destes através deste modelo (Kolen & Brennan, 1995).

Kolen & Brennan (1995) recomendam que escores com ordem percentil abaixo de 0,5 e acima de 99,5 sejam excluídos pois não influenciarão na função do modelo.

A estimativa final da função de equalização do equipercentil suavizado é dada por:

$$\hat{d}_Y^*(x) = \frac{\hat{d}_Y(x) + \hat{d}_X^{-1}(x)}{2}, \quad -0,5 \leq x \leq K_X + 0,5,$$

onde $\hat{d}_Y(x)$ é a função "cubic spline" suavizada de X em Y , e $\hat{d}_X^{-1}(x)$ é a função "cubic spline" inversa suavizada de Y em X .

Como exemplo ilustrativo podemos considerar os mesmos grupos que realizaram os testes A_1 e A_2 para os métodos de equalização citados anteriormente. Para termos uma idéia de como é realizada a suavização nos testes, faremos para o modelo log-linear na suavização a priori e para o modelo de "cubic splines" no caso da suavização a posteriori, utilizando o pacote computacional RAGE (Kolen, 1995).

No método log-linear da suavização a priori podemos verificar a propriedade de preservação dos momentos através da Tabela 2.5. Para $C=1$ somente o primeiro momento (média) é preservado; para $C=2$ temos a preservação da média e desvio; para $C=3$ são preservados a média, o desvio e a assimetria; e para $C=4$ temos preservados os quatro primeiros momentos da distribuição, e os modelos seguintes ($C \geq 4$) também preservam, no mínimo, os quatro primeiros momentos. No entanto, devemos escolher o modelo mais simples que ajuste adequadamente os dados e que direcione a um menor erro de estimação. Para a escolha do melhor modelo utilizaremos a estatística χ^2 obtida da diferença entre um modelo e seu anterior. O modelo anterior é rejeitado quando o valor da estatística for maior do que o valor crítico 3,84.

É escolhido o modelo mais simples que satisfaça os critérios de um teste de significância χ^2 para ambos testes, isto é, o que corresponde a um menor valor de C no modelo log-linear. Observando a Tabela 2.5 verificamos que para o teste A_2 é aceito o modelo para o qual $C \geq 6$, e para o teste A_1 temos aceito os modelos para $C \geq 8$. Como devemos decidir pelo modelo mais simples que contemple os quatro primeiros momentos em ambos testes, utilizaremos o modelo para $C = 8$. Na Tabela 2.6 apresentamos os resultados dos quatro primeiros momentos para ambos testes e para a equalização do equipercenil não suavizada e suavizado pelo modelo para $C = 8$.

Para o modelo de "cubic splines" no método de suavização a posteriori, observando a Tabela 2.7 verificamos os quatro primeiros momentos para ambos testes e o teste equalizado para a escala do teste A_2 em valores para S entre 0 e 1. Valores de S mais próximos de 0 parecem fornecer resultados mais adequados se comparado a valores mais próximos de 1. É necessário observarmos para vários valores de S onde compararmos os resultados. Obsevando os resultados obtidos percebemos que para S entre 0,01 e 0,10 obtivemos melhores estimativas, pois os valores dos quatro primeiros momentos se

Tabela 2.5: Momentos dos grupos comparado ao método log-linear de suavização a priori utilizando a estatística χ^2

Teste	Método	μ	σ	<i>ass</i>	<i>curt</i>	χ^2	<i>gl</i>	$\chi_c^2 - \chi_{c+1}^2$
A_2		8,0489	2,8725	0,2271	2,4279			
	$C=10$	8,0489	2,8725	0,2271	2,4279	0,328	5	
	$C=9$	8,0489	2,8725	0,2271	2,4279	3,879	6	3,551
	$C=8$	8,0489	2,8725	0,2271	2,4279	3,922	7	0,043
	$C=7$	8,0489	2,8725	0,2271	2,4279	5,648	8	1,725
	$C=6$	8,0489	2,8725	0,2271	2,4279	6,018	9	0,370
	$C=5$	8,0489	2,8725	0,2271	2,4279	10,195	10	4,177
	$C=4$	8,0489	2,8725	0,2271	2,4279	25,631	11	15,437
	$C=3$	8,0489	2,8725	0,2271	2,6828	45,817	12	20,186
	$C=2$	8,0489	2,8725	-0,0546	2,7138	83,017	13	37,200
	$C=1$	8,0489	4,5900	-0,1443	1,8192	997,732	14	914,714
A_1		10,0392	2,7446	-0,4446	2,9681			
	$C=10$	10,0392	2,7446	-0,4446	2,9681	3,910	5	
	$C=9$	10,0392	2,7446	-0,4446	2,9681	4,393	6	0,483
	$C=8$	10,0392	2,7446	-0,4446	2,9681	4,393	7	0,000
	$C=7$	10,0392	2,7446	-0,4446	2,9681	9,234	8	4,841
	$C=6$	10,0392	2,7446	-0,4446	2,9681	11,422	9	2,188
	$C=5$	10,0392	2,7446	-0,4446	2,9681	16,294	10	4,872
	$C=4$	10,0392	2,7446	-0,4446	2,9681	17,701	11	1,407
	$C=3$	10,0392	2,7446	-0,4446	3,0052	17,948	12	0,247
	$C=2$	10,0392	2,7446	-0,3081	2,7238	26,047	13	8,100
	$C=1$	10,0392	4,1768	-0,6984	2,4722	682,140	14	656,093

Tabela 2.6: Momentos dos grupos comparado ao método de suavização a priori log-linear equalizado

	μ	σ	<i>ass</i>	<i>curt</i>
Teste A_2	8,0489	2,8725	0,2271	2,4279
Teste A_1	10,0392	2,7446	-0,4446	2,9681
Não suavizado	8,0434	2,8596	0,2028	2,4113
$C=8$	8,0436	2,8550	0,2102	2,4017

encontram mais próximos aos mesmos momentos no teste A_2 .

Podemos observar na Figura 2.5, utilizando o modelo log-linear para a suavização a priori, o qual confirma que modelos com poucos parâmetros ($C=2$) não são convenientes pois se encontram fora das bandas de confiança. Já o modelo para $C=8$ se encontra

Tabela 2.7: Momentos dos grupos comparado ao método de suavização a posteriori "cubic splines" equalizado

	μ	σ	<i>ass</i>	<i>curt</i>
Teste A_2	8,0489	2,8725	0,2271	2,4279
Teste A_1	10,0392	2,7446	-0,4446	2,9681
Não suavizado	8,0437	2,8586	0,2056	2,4031
S=0,01	8,0430	2,8675	0,1984	2,4114
S=0,05	8,0450	2,8652	0,1939	2,4186
S=0,10	8,0479	2,8621	0,1866	2,4317
S=0,20	8,0535	2,8562	0,1671	2,4522
S=0,30	8,0577	2,8527	0,1509	2,4656
S=0,40	8,0609	2,8502	0,1376	2,4755
S=0,50	8,0636	2,8483	0,1258	2,4835
S=0,75	8,0691	2,8451	0,1000	2,4986
S=1,00	8,0737	2,8430	0,0775	2,5093

completamente dentro das bandas de confiança, reforçando que esse modelo deve ser bastante razoável. Já utilizando o modelo de suavização a posteriori "cubic splines", observando a Figura 2.6, podemos verificar que para valores de S mais próximos de zero, a suavização se apresenta mais plausível.

2.4 Três grupos ou mais

Quando trabalhamos somente com dois grupos equivalentes de indivíduos realizando, cada um, testes distintos, a equalização sempre está sujeita a um erro de medida, como apresentado anteriormente, pois a equalização não se trata de um processo exato. A teoria clássica permite a equalização somente de dois em dois grupos, indicando que, ao trabalharmos com mais de dois grupos, tenhamos um acúmulo quanto aos erros de medida.

A princípio, para podermos comparar o desempenho de vários grupos equivalentes de indivíduos, e portanto, com a mesma habilidade, é necessário que todos os testes aplicados tenham sido construídos levando-se em consideração a propriedade das mesmas especificações e a propriedade da invariância.

Podemos considerar a situação em que são aplicados quatro testes (A_1, A_2, A_3, A_4) distribuídos espiralmente para uma população de respondentes, obtendo assim quatro

Figura 2.5: Diferença entre as equalizações realizadas e o escore do teste A_1 utilizando a suavização a priori.

grupos equivalentes de indivíduos dessa população. Consideraremos X como a variável aleatória dos escores do teste A_1 , x uma realização de X , Y como a variável aleatória dos escores do teste A_2 , y uma realização de Y , Z como a variável aleatória dos escores do teste A_3 , z uma realização de Z , e W como a variável aleatória do teste A_4 , w uma realização de W . Se desejamos comparar os testes devemos realizar as equalizações aos pares.

No exemplo, podemos considerar que desejamos transformar a escala dos testes A_2 , A_3 e A_4 na escala do teste A_1 . Assim, devemos realizar três equalizações, em que, ao trabalharmos com equalização linear, teremos os seguintes processos:

- $l_x(y)$ correspondente à equalização linear que transforma os escores do teste A_2 para a escala do teste A_1 ,
- $l_x(z)$ correspondente à equalização linear que transforma os escores do teste A_3 para

Figura 2.6: Diferença entre as equalizações realizadas e o escore do teste A_1 utilizando a suavização a posteriori.

a escala do teste A_1 ,

- $l_x(w)$ correspondente à equalização linear que transforma os escores do teste A_4 para a escala do teste A_1 .

Se trabalharmos com a equalização do equipercentil teremos um processo análogo. Essas equalizações são realizadas no caso de fazermos a transformação direta da escala de cada um dos testes A_2 , A_3 e A_4 na escala do teste A_1 .

Pelo fato de deverem ser realizadas três equalizações, duas a duas, esse processo torna-se trabalhoso no caso de termos vários grupos equivalentes para serem analisados. Porém, a teoria clássica não possui embasamento teórico que permita trabalharmos com vários grupos de uma só vez, ou seja, todas as equalizações devem ser realizadas aos pares.

Capítulo 3

Equalização de grupos não equivalentes - Teoria Clássica

3.1 Métodos lineares para grupos não equivalentes

Os grupos não equivalentes são considerados na situação em que tenhamos que garantir a segurança do teste, tornando-se inconveniente a introdução de novos testes a priori. Como os testes são realizados em épocas diferentes não existe garantia de que os grupos que tenham feito ambos testes sejam da mesma população. Portanto, com o intuito de minimizar essa diferença entre os grupos, deve-se tomar algumas medidas para realizar tais ajustes.

Quando trabalhamos com populações não equivalentes devemos analisá-las através de testes que contenham itens comuns, os quais devem ser o mais equivalentes possível em conteúdo e características estatísticas. Mais que isso, os itens comuns devem ser idênticos em ambos testes. Para entendermos melhor podemos analisar a seguinte situação:

- Seja X uma variável aleatória dos escores sobre o teste A_1 dado para a população 1.
- Seja Y uma variável aleatória dos escores sobre o teste A_2 dado para a população 2.
- Seja V uma variável aleatória dos escores correspondente aos itens comuns nos testes A_1 e A_2 .

Existem dois casos para grupos não equivalentes de itens comuns:

Interno Ocorre quando os escores sobre os itens comuns contribuem para o escore total em ambos testes.

Externo Ocorre quando os escores sobre os itens comuns não contribuem para o escore total dos testes.

Por exemplo, duas populações 1 e 2 fazem os testes A_1 e A_2 , respectivamente. Se os testes possuem 80 itens cada, sendo 60 não comuns e 20 comuns, para o caso Interno o escore total é calculado sobre os 80 itens e, para o caso Externo, o escore total é dado pelos 60 itens não comuns.

Esses itens comuns, em geral, são utilizados para ajustar as diferenças entre as populações. Para isso é necessária uma estatística robusta, pois cada população faz somente um tipo de prova.

Como estamos trabalhando com duas populações, a relação de equalização deve ser uma combinação das mesmas, onde introduzimos o conceito de população sintética cujas populações são ponderadas por pesos relativos aos grupos 1 e 2 e denotados por w_1 e w_2 , respectivamente, de tal forma que $w_1 + w_2 = 1$ e $w_1, w_2 \geq 0$.

Da mesma forma que a equalização linear utilizada anteriormente, temos a equalização dos escores observados sobre X transformado para a escala dos escores observados sobre Y pela seguinte equação:

$$l_{Ys} = \frac{\sigma_{Ys}}{\sigma_{Xs}}[x - \mu_{Xs}] + \mu_{Ys}, \quad (3.1)$$

sendo s referente à população sintética, e onde os parâmetros são expressos em função dos pesos ponderados das populações tal que:

$$X = w_1 X_1 + w_2 X_2. \quad (3.2)$$

O processo é análogo para o teste A_2 . Assim, para determinarmos a média e variância da população sintética utilizamos a equação (3.2) cujas demonstrações são dadas a seguir:

$$\begin{aligned} \mu_{Xs} &= E[w_1 X_1 + w_2 X_2] \\ &= w_1 \mu_1 + w_2 \mu_2, \end{aligned}$$

onde $E(X_1) = \mu_1$ e $E(X_2) = \mu_2$.

Para demonstrarmos a variância precisamos ainda desenvolver um resultado que será útil nessa prova.

$$\begin{aligned}
w_1 E(X_1 - \mu)^2 &= w_1 E(X_1 - w_1 \mu_1 - w_2 \mu_2)^2 \\
&= w_1 E[X_1 - (1 - w_2) \mu_1 - w_2 \mu_2]^2 \\
&= w_1 E[X_1 - \mu_1 + w_2 \mu_1 - w_2 \mu_2]^2 \\
&= w_1 E[(X_1 - \mu_1) + w_2 (\mu_1 - \mu_2)]^2 \\
&= w_1 E[(X_1 - \mu_1)^2 + 2(X_1 - \mu_1)w_2(\mu_1 - \mu_2) + w_2^2(\mu_1 - \mu_2)^2] \\
&= w_1 \sigma_1^2 + w_1 w_2^2 (\mu_1 - \mu_2)^2.
\end{aligned}$$

Por consequência, para a variância da população sintética, temos:

$$\begin{aligned}
\sigma_{X_s} &= w_1 E(X_1 - \mu_{X_s})^2 + w_2 E(X_2 - \mu_{X_s})^2 \\
&= w_1 E(X_1 - w_1 \mu_1 - w_2 \mu_2)^2 + w_2 E(X_2 - w_1 \mu_1 - w_2 \mu_2)^2 \\
&= w_1 \sigma_1^2 + w_1 w_2^2 (\mu_1 - \mu_2)^2 + w_2 \sigma_2^2 + w_1^2 w_2 (\mu_1 - \mu_2)^2 \\
&= w_1 \sigma_1^2 + w_2 \sigma_2^2 + w_1 w_2 (w_1 + w_2) (\mu_1 - \mu_2)^2 \\
&= w_1 \sigma_1^2 + w_2 \sigma_2^2 + w_1 w_2 (\mu_1 - \mu_2)^2.
\end{aligned}$$

Assim, obtemos os seguintes resultados para ambos testes A_1 e A_2 :

$$\mu_{X_s} = w_1 \mu_{X_1} + w_2 \mu_{X_2}, \quad (3.3)$$

$$\mu_{Y_s} = w_1 \mu_{Y_1} + w_2 \mu_{Y_2}, \quad (3.4)$$

$$\sigma_{X_s}^2 = w_1 \sigma_{X_1}^2 + w_2 \sigma_{X_2}^2 + w_1 w_2 [\mu_{X_1} - \mu_{X_2}]^2, \quad (3.5)$$

$$\sigma_{Y_s}^2 = w_1 \sigma_{Y_1}^2 + w_2 \sigma_{Y_2}^2 + w_1 w_2 [\mu_{Y_1} - \mu_{Y_2}]^2. \quad (3.6)$$

Porém os parâmetros μ_{X_2} , μ_{Y_1} , $\sigma_{X_2}^2$ e $\sigma_{Y_1}^2$ não podem ser determinados diretamente. Assim, para podermos determiná-los apresentaremos os métodos de Tucker e Levine para os escores observados, que os expressam como função dos parâmetros obtidos diretamente.

3.1.1 Método de Tucker

Para utilizarmos esse método é necessário que os testes aplicados para grupos não equivalentes sejam compostos por um conjunto de itens comuns, e denominado teste âncora. Os testes devem ser similares em fidedignidade (Thorndike,1971 e Vianna,1987), cuja razão entre a variância do escore verdadeiro e a variância do escore obtido indica de que forma as diferenças nos escores são provenientes de variações na característica examinada e não de erros casuais. Os itens âncora são utilizados com o objetivo de ajustar as diferenças existentes entre os grupos 1 e 2.

Esse método requer três suposições com relação à estimação dos parâmetros. Para isso consideramos a regressão representada por $y = \alpha + \beta x$, com α como o intercepto e β como a inclinação da reta de regressão.

Inicialmente assumimos uma regressão linear de X sobre V e Y sobre V para ambas populações. Dessa forma, temos a primeira suposição considerando a igualdade do intercepto para ambos grupos, e a segunda suposição levando em consideração a igualdade da inclinação da reta de regressão. Assim:

$$\alpha_{(X|V)1} = \alpha_{(X|V)2}, \quad (3.7)$$

$$\mu_{X1} - \beta_{(X|V)1}\mu_{V1} = \mu_{X2} - \beta_{(X|V)2}\mu_{V2}, \quad e \quad (3.8)$$

$$\beta_{(X|V)1} = \beta_{(X|V)2}, \quad (3.9)$$

onde $\alpha_{(X|V)1}$ define o intercepto da regressão do escore do teste A_1 dado o escore dos itens comuns no grupo 1, $\alpha_{(X|V)2}$ define o intercepto da regressão do escore do teste A_1 dado o escore dos itens comuns no grupo 2, $\beta_{(X|V)1} = \frac{\sigma_{XV1}}{\sigma_{V1}^2}$ corresponde à razão da covariância dos escores do teste A_1 e os escores dos itens comuns no grupo 1 e a variância dos escores dos itens comuns no grupo 1, $\beta_{(X|V)2} = \frac{\sigma_{XV2}}{\sigma_{V2}^2}$ corresponde à razão da covariância dos escores do teste A_1 e os escores dos itens comuns no grupo 2 e a variância dos escores dos itens comuns no grupo 2 (Bussab,1988,p.69), μ_{X1} é a média dos escores do teste A_1 no grupo 1, μ_{X2} é a média dos escores do teste A_1 no grupo 2, μ_{V1} é o escore médio do teste âncora no grupo 1, e μ_{V2} é o escore médio do teste âncora no grupo 2.

Com relação à terceira suposição, também assumimos que a variância condicional (Gulliksen, 1950) para ambos grupos são similares, tanto para X sobre V como para Y sobre V, isto é, $\sigma_{(X|V)1}^2 = \sigma_{x2}^2 [1 - \rho_{xv2}^2]$ e $\sigma_{(X|V)2}^2 = \sigma_{x1}^2 [1 - \rho_{xv1}^2]$. Assim,

$$\sigma_{x2}^2 [1 - \rho_{xv2}^2] = \sigma_{x1}^2 [1 - \rho_{xv1}^2], \quad (3.10)$$

onde $\rho_{xv1} = \frac{\sigma_{xv1}}{\sigma_{x1}\sigma_{v1}}$ é a correlação dos escores do teste A_1 e os escores sobre os itens comuns no grupo 1, $\rho_{xv2} = \frac{\sigma_{xv2}}{\sigma_{x2}\sigma_{v2}}$ é a correlação dos escores do teste A_1 e os escores sobre os itens comuns no grupo 2, σ_{x1}^2 determina a variância dos escores do teste A_1 no grupo 1, σ_{x2}^2 determina a variância dos escores do teste A_1 no grupo 2, σ_{x1} define o desvio médio dos escores do teste A_1 no grupo 1, σ_{x2} define o desvio médio dos escores do teste A_1 no grupo 2, σ_{v1} define o desvio padrão dos escores sobre os itens comuns no grupo 1, e σ_{v2} define o desvio padrão dos escores sobre os itens comuns no grupo 2.

Da suposição de igualdade do intercepto da regressão dada pela equação (3.8) podemos obter μ_{x2} .

$$\mu_{x2} = \mu_{x1} - \beta_{(X|V)1}\mu_{v1} + \beta_{(X|V)2}\mu_{v2}. \quad (3.11)$$

Porém, pela suposição (3.9) temos:

$$\mu_{x2} = \mu_{x1} - \beta_{(X|V)1}\mu_{v1} + \beta_{(X|V)2}\mu_{v2} \quad (3.12)$$

$$= \mu_{x1} - \beta_{(X|V)1}[\mu_{v1} - \mu_{v2}]. \quad (3.13)$$

O resultado obtido para Y sobre V é análogo, de tal forma que:

$$\mu_{y1} = \mu_{y2} + \beta_{(Y|V)2}[\mu_{v1} - \mu_{v2}]. \quad (3.14)$$

Do mesmo modo, utilizando a suposição (3.10), calculamos as variâncias desconhecidas em função de resultados conhecidos.

Porém, temos $\rho_{xv2} = \frac{\beta_{(X|V)2}\sigma_{v2}}{\sigma_{x2}}$ e, portanto, $\beta_{(X|V)2}\sigma_{v2} = \rho_{xv2}\sigma_{x2}$. O raciocínio é análogo para o grupo 1.

A obtenção do parâmetro desconhecido, utilizando o resultado acima e a suposição (3.9), então, será dada por:

$$\sigma_{x_2}^2 - \beta_{(x|v)_2}^2 \sigma_{v_2}^2 = \sigma_{x_1}^2 - \beta_{(x|v)_1}^2 \sigma_{v_1}^2, \quad (3.15)$$

$$\sigma_{x_2}^2 - \beta_{(x|v)_1}^2 \sigma_{v_2}^2 = \sigma_{x_1}^2 - \beta_{(x|v)_1}^2 \sigma_{v_1}^2, \quad (3.16)$$

$$\sigma_{x_2}^2 = \sigma_{x_1}^2 - \beta_{(x|v)_1}^2 [\sigma_{v_1}^2 - \sigma_{v_2}^2]. \quad (3.17)$$

O resultado do parâmetro desconhecido para Y sobre V é análogo e dado por:

$$\sigma_{y_1}^2 = \sigma_{y_2}^2 + \beta_{(y|v)_2}^2 [\sigma_{v_1}^2 - \sigma_{v_2}^2]. \quad (3.18)$$

Portanto, substituindo os parâmetros, pelo Método de Tucker, na obtenção dos dois primeiros momentos da população sintética, determinamos as expressões dos ajustes em função das quantidades observadas:

$$\mu_{x_s} = \mu_{x_1} - w_2 \gamma_1 [\mu_{v_1} - \mu_{v_2}], \quad (3.19)$$

$$\mu_{y_s} = \mu_{y_2} + w_1 \gamma_2 [\mu_{v_1} - \mu_{v_2}], \quad (3.20)$$

$$\sigma_{x_s}^2 = \sigma_{x_1}^2 - w_2 \gamma_1^2 [\sigma_{v_1}^2 - \sigma_{v_2}^2] + w_1 w_2 \gamma_1^2 [\mu_{v_1} - \mu_{v_2}]^2, \quad (3.21)$$

$$\sigma_{y_s}^2 = \sigma_{y_2}^2 + w_1 \gamma_2^2 [\sigma_{v_1}^2 - \sigma_{v_2}^2] + w_1 w_2 \gamma_2^2 [\mu_{v_1} - \mu_{v_2}]^2, \quad (3.22)$$

com $\gamma_1 = \beta_{(x|v)_1}$ e $\gamma_2 = \beta_{(y|v)_2}$ e cujas quantidades podemos substituir na equalização linear (3.1).

Se $\mu_{v_1} = \mu_{v_2}$ e $\sigma_{v_1}^2 = \sigma_{v_2}^2$ teremos os parâmetros iguais às médias e variâncias observadas. Esse processo é válido para ambos casos em que V seja interno ou externo.

Existem alguns casos especiais desse método (Kolen, 1995) em relação aos pesos tal que:

- Se $w_1 = 1$ e $w_2 = 0$ é considerada como população sintética somente o grupo 1.
- Se $w_1 = w_2 = 0,5$ considera-se que ambos grupos tem a mesma importância no estudo da população sintética.

- Se $w_1 = \frac{n_1}{N}$ e $w_2 = \frac{n_2}{N}$ onde n_1 e n_2 correspondem ao tamanho da amostra para cada grupo, e $N = n_1 + n_2$.

O primeiro caso parece ter mérito no sentido da simplificação dos resultados, porém os resultados obtidos para os vários pesos não definem mudanças significativas.

3.1.2 Método do Escore Observado de Levine

Esse método é mais aplicado no caso em que os testes são similares em fidedignidade, porém os grupos diferem em habilidade. Além disso não considera explicitamente o conceito de população sintética.

Utiliza-se a equalização linear para equiparar os escores observados do teste A_1 na escala dos escores observados do teste A_2 , porém, as suposições são feitas em relação aos escores verdadeiros, denominados T_x , T_y e T_v , de acordo com o modelo:

$$X = T_x + E_x, \quad (3.23)$$

$$Y = T_y + E_y, \quad (3.24)$$

$$V = T_v + E_v, \quad (3.25)$$

onde $\mu_{E_x} = \mu_{E_y} = \mu_{E_v} = 0$ e não são correlacionados aos escores verdadeiros. Assim, temos que a média dos escores observados é igual a média dos escores verdadeiros, ou seja, $\mu_x = \mu_{T_x}$, $\mu_y = \mu_{T_y}$ e $\mu_v = \mu_{T_v}$. Precisamos de algumas suposições que tornam válido esse método e são análogas àquelas utilizadas no método de Tucker.

Pela suposição de correlação o método assume que T_x e T_v são perfeitamente correlacionadas para ambos grupos 1 e 2. Portanto:

$$\rho_{T_x T_{v1}} = \rho_{T_x T_{v2}} = 1, \quad (3.26)$$

onde $\rho_{T_x T_{v1}}$ é a correlação dos escores verdadeiros sobre o teste A_1 e os escores verdadeiros sobre os itens comuns do grupo 1, e $\rho_{T_x T_{v2}}$ é a correlação dos escores verdadeiros sobre o teste A_1 e os escores verdadeiros sobre os itens comuns do grupo 2.

Pela suposição de regressão linear temos a regressão de T_X sobre T_V como sendo a mesma para ambos grupos. Assim, a igualdade da inclinação de T_X sobre T_V será:

$$\beta_{(T_X|T_V)1} = \beta_{(T_X|T_V)2}, \quad (3.27)$$

$$\frac{\sigma_{T_X1}}{\sigma_{T_V1}} = \frac{\sigma_{T_X2}}{\sigma_{T_V2}}, \quad (3.28)$$

onde σ_{T_X1} é o desvio padrão dos escores verdadeiros sobre o teste A_1 no grupo 1, σ_{T_X2} é o desvio padrão dos escores verdadeiros sobre o teste A_1 no grupo 2, σ_{T_V1} é o desvio padrão dos escores verdadeiros sobre os itens comuns no grupo 1 e σ_{T_V2} é o desvio padrão dos escores verdadeiros sobre os itens comuns no grupo 2.

Da mesma forma temos a igualdade dos interceptos da regressão dos escores verdadeiros sob ambos grupos dada por:

$$\mu_{X1} - \frac{\sigma_{T_X1}}{\sigma_{T_V1}}\mu_{V1} = \mu_{X2} - \frac{\sigma_{T_X2}}{\sigma_{T_V2}}\mu_{V2}. \quad (3.29)$$

Da mesma forma que no Método de Tucker, as quantidades μ_{X2} e μ_{Y1} não podem ser observadas diretamente, o que ocorrerá através das igualdades acima.

Pela terceira suposição também assumimos que a variância do erro de medida para X é a mesma para ambos grupos.

Como os escores verdadeiros e os erros não são correlacionados, pela Teoria Clássica, temos:

$$\begin{cases} \sigma_{X1}^2 - \sigma_{T_X1}^2 = \sigma_{X2}^2 - \sigma_{T_X2}^2, \\ \sigma_{Y1}^2 - \sigma_{T_Y1}^2 = \sigma_{Y2}^2 - \sigma_{T_Y2}^2, \\ \sigma_{V1}^2 - \sigma_{T_V1}^2 = \sigma_{V2}^2 - \sigma_{T_V2}^2. \end{cases} \quad (3.30)$$

As quantidades σ_{X2}^2 e σ_{Y1}^2 são obtidas dessas relações acima.

Calculamos as médias desconhecidas utilizando as suposições (3.28) e (3.29), obtendo:

$$\begin{aligned}
\mu_{x_2} &= \mu_{x_1} - \frac{\sigma_{TX^1}}{\sigma_{TV^1}}\mu_{v_1} + \frac{\sigma_{TX^2}}{\sigma_{TV^2}}\mu_{v_2} \\
&= \mu_{x_1} - \frac{\sigma_{TX^1}}{\sigma_{TV^1}}\mu_{v_1} + \frac{\sigma_{TX^1}}{\sigma_{TV^1}}\mu_{v_2}.
\end{aligned}$$

Assim:

$$\mu_{x_2} = \mu_{x_1} - \frac{\sigma_{TX^1}}{\sigma_{TV^1}}[\mu_{v_1} - \mu_{v_2}]. \quad (3.31)$$

Todo o processo descrito acima é análogo para o teste A_2 e, portanto, os resultados das estimativas são obtidos da mesma forma. Portanto, temos a média estimada de Y para o grupo 1 dado por:

$$\mu_{y_1} = \mu_{y_2} + \frac{\sigma_{TY^2}}{\sigma_{TV^2}}[\mu_{v_1} - \mu_{v_2}]. \quad (3.32)$$

Utilizando (3.26), (3.28) e (3.30) determinamos os resultados para as variâncias desconhecidas. Por (3.28) temos que $\sigma_{TX^2}^2 = \frac{\sigma_{TX^1}^2 \cdot \sigma_{TV^2}^2}{\sigma_{TV^1}^2}$. Assim:

$$\begin{aligned}
\sigma_{x_2}^2 &= \sigma_{x_1}^2 - \sigma_{TX^1}^2 + \sigma_{TX^2}^2, \\
&= \sigma_{x_1}^2 - \sigma_{TX^1}^2 + \frac{\sigma_{TX^1}^2 \cdot \sigma_{TV^2}^2}{\sigma_{TV^1}^2}, \\
&= \sigma_{x_1}^2 - \sigma_{TX^1}^2 \left[1 - \frac{\sigma_{TV^2}^2}{\sigma_{TV^1}^2} \right].
\end{aligned}$$

Assim:

$$\sigma_{x_2}^2 = \sigma_{x_1}^2 - \frac{\sigma_{TX^1}^2}{\sigma_{TV^1}^2}[\sigma_{TV^1}^2 - \sigma_{TV^2}^2]. \quad (3.33)$$

De modo análogo calculamos a variância desconhecida do teste A_2 para o grupo 1 como:

$$\sigma_{y_1}^2 = \sigma_{y_2}^2 + \frac{\sigma_{TY^2}^2}{\sigma_{TV^2}^2}[\sigma_{TV^1}^2 - \sigma_{TV^2}^2]. \quad (3.34)$$

Reescrevendo as expressões dos parâmetros desconhecidos, temos:

$$\mu_{x2} = \mu_{x1} - \gamma_1[\mu_{v1} - \mu_{v2}], \quad (3.35)$$

$$\mu_{y1} = \mu_{y2} + \gamma_2[\mu_{v1} - \mu_{v2}], \quad (3.36)$$

$$\sigma_{x2}^2 = \sigma_{x1}^2 - \gamma_1^2[\sigma_{tv1}^2 - \sigma_{tv2}^2], \quad (3.37)$$

$$\sigma_{y1}^2 = \sigma_{y2}^2 + \gamma_2^2[\sigma_{tv1}^2 - \sigma_{tv2}^2], \quad (3.38)$$

com $\gamma_1 = \frac{\sigma_{Tx1}}{\sigma_{Tv1}}$ e $\gamma_2 = \frac{\sigma_{Ty2}}{\sigma_{Tv2}}$.

Os resultados encontrados são válidos para ambos casos em que o teste âncora seja um conjunto interno ou externo.

No entanto, ainda temos quantidades não observadas diretamente. São as variâncias dos escores verdadeiros, ou seja, σ_{Tx1}^2 , σ_{Ty2}^2 , σ_{Tv2}^2 e σ_{Tv1}^2 .

Para isso utilizamos o modelo congênérico clássico, o qual considera que os escores verdadeiros sobre o teste A_1 e sobre os itens comuns são linearmente relacionados, o que é consistente com a expressão 3.26. Então, temos $T_x = \lambda_x T + \theta_x$ e $T_v = \lambda_v T + \theta_v$ onde λ é a inclinação e θ os interceptos. Portanto, das equações 3.23 e 3.25, utilizando o modelo congênérico clássico, temos:

$$X = \theta_x + \lambda_x T + E_x, e \quad (3.39)$$

$$V = \theta_v + \lambda_v T + E_v. \quad (3.40)$$

Neste modelo adicionam-se as seguintes suposições apresentadas por Kolen(1995):

$$\sigma_{Ex}^2 = \lambda_x \sigma_E^2, \quad (3.41)$$

$$\sigma_{Ev}^2 = \lambda_v \sigma_E^2, \quad (3.42)$$

onde λ_x e λ_v correspondem ao comprimento efetivo do teste.

E, das equações (3.39), (3.40), (3.41), (3.42), obtemos as variâncias e covariância de X e V:

$$\sigma_X^2 = \sigma_{T_X}^2 + \sigma_{E_X}^2 = \sigma_{(\lambda_X T + \theta_X)}^2 + \sigma_{E_X}^2 = \lambda_X^2 \sigma_T^2 + \lambda_X \sigma_E^2, \quad (3.43)$$

$$\sigma_V^2 = \lambda_V^2 \sigma_T^2 + \lambda_V \sigma_E^2, \quad e \quad (3.44)$$

$$\sigma_{XV} = \lambda_X \lambda_V \sigma_T^2 + \sigma_{E_X E_V}, \quad (3.45)$$

onde temos a razão $\frac{\sigma_{TX}}{\sigma_{TV}} = \frac{\lambda_X \sigma_T}{\lambda_V \sigma_T} = \frac{\lambda_X}{\lambda_V}$ como sendo a razão do comprimento efetivo do teste para X e V, respectivamente. Portanto, aqui devemos considerar os casos em que V é interno ou externo.

Para o caso em que V é interno temos $X = W + V$, sendo V os itens âncora do teste, ou os itens comuns, e W os itens não comuns do teste A_1 . Sendo assim, $\rho(W, V) = 0$ pois constituem itens diferentes. Neste caso, $\sigma_{E_X E_V} = \sigma_{E_{W+V} E_V} = \sigma_{E_V E_V} = \sigma_{E_V}^2 = \lambda_V \sigma_E^2$.

Desse modo, substituindo o resultado para o caso em que o teste âncora é interno, e pelas equações (3.43) e (3.45), respectivamente, temos:

$$\sigma_X^2 = \lambda_X^2 \sigma_T^2 + \lambda_X \sigma_E^2 = \lambda_X [\lambda_X \sigma_T^2 + \sigma_E^2], \quad e \quad (3.46)$$

$$\sigma_{XV} = \lambda_X \lambda_V \sigma_T^2 + \lambda_V \sigma_E^2 = \lambda_V [\lambda_X \sigma_T^2 + \sigma_E^2]. \quad (3.47)$$

Das equações acima definimos os termos multiplicadores nas estimativas dos parâmetros desconhecidos para cada grupo, e dadas a seguir:

$$\gamma_1 = \frac{\lambda_X}{\lambda_V} = \frac{\sigma_{X1}^2}{\sigma_{XV1}} = \frac{1}{\beta_{(V|X)1}}, \quad (3.48)$$

$$\gamma_2 = \frac{\lambda_X}{\lambda_V} = \frac{\sigma_{Y2}^2}{\sigma_{YV2}} = \frac{1}{\beta_{(V|Y)2}}, \quad (3.49)$$

que correspondem ao inverso da inclinação da regressão para cada grupo.

Nosso interesse se destaca para o caso em V seja interno ao escore, portanto aqui não iremos considerar o caso em que V seja um conjunto externo ao escore (Kolen, 1995).

Kolen(1995) fornece alguns resultados diretos para a aplicação do modelo congênico clássico, os quais são listadas a seguir:

$$\gamma = \frac{\lambda_x}{\lambda_v} = \frac{\sigma_x^2}{\sigma_{xv}}, \quad (3.50)$$

$$\sigma_{T_X}^2 = \frac{\gamma^2[\sigma_{xv} - \sigma_v^2]}{\gamma - 1}, \quad (3.51)$$

$$\sigma_{T_V}^2 = \frac{\sigma_{xv} - \sigma_v^2}{\gamma - 1}, \quad (3.52)$$

$$\sigma_{E_X}^2 = \frac{\gamma^2\sigma_v^2 - \gamma\sigma_{xv}}{\gamma - 1}, \quad (3.53)$$

$$\sigma_{E_V}^2 = \frac{\gamma\sigma_v^2 - \sigma_{xv}}{\gamma - 1}. \quad (3.54)$$

Substituindo os resultados encontrados nas equações (3.19),(3.20),(3.21) e (3.22) obtemos os seguintes resultados:

$$\mu_X = \mu_{X1} - w_2\gamma_1[\mu_{V1} - \mu_{V2}], \quad (3.55)$$

$$\mu_Y = \mu_{Y2} + w_1\gamma_2[\mu_{V1} - \mu_{V2}], \quad (3.56)$$

$$\sigma_X^2 = \sigma_{X1}^2 - w_2\gamma_1^2[\sigma_{TV1}^2 - \sigma_{TV2}^2] + w_1w_2\gamma_1^2[\mu_{V1} - \mu_{V2}]^2, \quad (3.57)$$

$$\sigma_Y^2 = \sigma_{Y2}^2 + w_1\gamma_2^2[\sigma_{TV1}^2 - \sigma_{TV2}^2] + w_1w_2\gamma_2^2[\mu_{V1} - \mu_{V2}]^2. \quad (3.58)$$

Os resultados acima são substituídos na equação (3.1) para determinarmos a equalização linear de transformação de escala.

3.1.3 Método do Escore Verdadeiro de Levine

Esse método equipara através da equação linear inicial os escores verdadeiros sobre X na escala dos escores verdadeiros sobre Y, o qual temos:

$$l_Y = \frac{\sigma_{T_Y}}{\sigma_{T_X}}[t_x - \mu_{T_X}] + \mu_{T_Y}.$$

Como visto anteriormente, temos que $\mu_{T_X} = \mu_X$ e $\mu_{T_Y} = \mu_Y$, e assim:

$$l_Y(t_X) = \frac{\sigma_{T_Y}}{\sigma_{T_X}}[t_x - \mu_X] + \mu_Y.$$

As equações para μ_x e μ_y encontradas anteriormente são adequadas para todos os métodos desde que os termos correspondentes ao desvio padrão do teste em relação ao desvio padrão dos itens comuns sejam compatíveis com o estudo e com o método.

Vamos mostrar que $\sigma_{T_X}^2 = \gamma_1^2 \sigma_{T_V}^2$. A demonstração de $\sigma_{T_Y}^2 = \gamma_2^2 \sigma_{T_V}^2$ é análoga.

Foi demonstrado anteriormente que:

$$\sigma_{T_X}^2 = w_1 \sigma_{T_{X1}}^2 + w_2 \sigma_{T_{X2}}^2 + w_1 w_2 [\mu_{T_{X1}} - \mu_{T_{X2}}]^2, \quad (3.59)$$

com $\mu_{T_{X1}} = \mu_{x1}$ e $\mu_{T_{X2}} = \mu_{x2}$ substituídos em (3.59), obtém-se:

$$\sigma_{T_X}^2 = w_1 \sigma_{T_{X1}}^2 + w_2 \sigma_{T_{X2}}^2 + w_1 w_2 (\mu_{x1} - \mu_{x2})^2, \quad (3.60)$$

onde $\mu_{x2} = \mu_{x1} - \gamma_1 (\mu_{v1} - \mu_{v2})$ com $\gamma_1 = \frac{\sigma_{T_{X1}}}{\sigma_{T_{V1}}}$. Assim, a equação (3.60) será reformulada da seguinte forma:

$$\begin{aligned} \sigma_{T_X}^2 &= w_1 \sigma_{T_{X1}}^2 + w_2 \sigma_{T_{X2}}^2 + w_1 w_2 \left[\mu_{x1} - \mu_{x1} + \frac{\sigma_{T_{X1}}}{\sigma_{T_{V1}}} (\mu_{v1} - \mu_{v2}) \right]^2 \\ &= w_1 \sigma_{T_{X1}}^2 + w_2 \sigma_{T_{X2}}^2 + w_1 w_2 \frac{\sigma_{T_{X1}}^2}{\sigma_{T_{V1}}^2} (\mu_{v1} - \mu_{v2})^2 \\ &= \frac{\sigma_{T_{X1}}^2}{\sigma_{T_{V1}}^2} \left[w_1 \sigma_{T_{V1}}^2 + w_2 \sigma_{T_{X2}}^2 \frac{\sigma_{T_{V1}}^2}{\sigma_{T_{X1}}^2} + w_1 w_2 (\mu_{v1} - \mu_{v2})^2 \right]. \end{aligned}$$

Porém, fazemos uso da seguinte igualdade:

$$\frac{\sigma_{T_{X1}}^2}{\sigma_{T_{V1}}^2} = \frac{\sigma_{T_{X2}}^2}{\sigma_{T_{V2}}^2} \Rightarrow \sigma_{T_{V2}}^2 = \sigma_{T_{X2}}^2 \frac{\sigma_{T_{V1}}^2}{\sigma_{T_{X1}}^2}.$$

Assim:

$$\begin{aligned} \sigma_{T_X}^2 &= \frac{\sigma_{T_{X1}}^2}{\sigma_{T_{V1}}^2} \left[w_1 \sigma_{T_{V1}}^2 + w_2 \sigma_{T_{V2}}^2 + w_1 w_2 (\mu_{v1} - \mu_{v2})^2 \right] \\ &= \gamma_1^2 \sigma_{T_V}^2, \end{aligned}$$

onde $\sigma_{T_V}^2 = w_1 \sigma_{T_{V1}}^2 + w_2 \sigma_{T_{V2}}^2 + w_1 w_2 (\mu_{v1} - \mu_{v2})^2$.

A inclinação e o intercepto são dados, respectivamente, por:

$$\beta = \frac{\sigma_{TY}}{\sigma_{TX}} = \frac{\gamma_2 \sigma_{TV}}{\gamma_1 \sigma_{TV}} = \frac{\gamma_2}{\gamma_1},$$

$$\alpha = \mu_Y - \frac{\gamma_2}{\gamma_1} \mu_X = \mu_{Y2} - \frac{\gamma_2}{\gamma_1} \mu_{X1} + \gamma_2 [\mu_{V1} - \mu_{V2}].$$

Assim, determinamos a seguinte expressão para a equalização para esse método:

$$\begin{aligned} l_Y &= \frac{\gamma_2}{\gamma_1} t_x - \frac{\gamma_2}{\gamma_1} \mu_{X1} + \mu_{Y2} + \gamma_2 [\mu_{V1} - \mu_{V2}] \\ &= \frac{\gamma_2}{\gamma_1} [t_x - \mu_{X1}] + \mu_{Y2} + \gamma_2 [\mu_{V1} - \mu_{V2}]. \end{aligned}$$

O método requer a estrutura conceitual de população sintética, porém ele é invariante em relação aos pesos w_1 e w_2 .

3.1.4 Comparação entre os métodos

Como exemplo ilustrativo, podemos considerar dois testes A_1 e A_2 com 20 itens sendo 5 comuns (4,8,12,16,20). O teste A_1 foi aplicado ao grupo 1 e o teste A_2 ao grupo 2, sendo $n_1 = 2324$ e $n_2 = 2362$, respectivamente, o número de examinados em cada teste A_1 e A_2 . A princípio temos a Tabela 3.1 contendo os dois primeiros momentos para cada grupo, os quais consideram os itens comuns internos. Temos $\hat{\mu}$ e $\hat{\sigma}$ como os dois primeiros momentos relativos aos grupos 1 e 2, de forma que estamos trabalhando com amostras da população.

Tabela 3.1: Sumário estatístico dos testes A_1 e A_2 para grupos não equivalentes

Grupo	Escore	$\hat{\mu}$	$\hat{\sigma}$	Covariância	Correlação
1	X	8,3810	2,7462	2,5421	0,8329
1	V_1	2,3703	1,1114		
2	Y	8,8893	2,5615	2,3254	0,8199
2	V_2	2,4538	1,1073		

Utilizando os dados da Tabela 3.1 podemos determinar os valores correspondentes aos métodos descritos através da população sintética, aplicando-os para pesos diferentes dos grupos, nos quais os resultados obtidos se encontram na Tabela 3.2.

Tabela 3.2: Resultados da equalização linear do exemplo ilustrativo (* representam valores que não necessitam ser calculados.)

w_1	w_2	Método	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\mu}_{xs}$	$\hat{\mu}_{ys}$
1	0	Tucker	2,0580	1,8966	8,3810	8,7309
1	0	Lev.Obs.	2,9667	2,8216	8,3810	8,6537
0,5	0,5	Tucker	2,0580	1,8966	8,5049	8,8101
0,5	0,5	Lev.Obs.	2,9667	2,8216	8,4013	8,7715
0,2	0,8	Tucker	2,0580	1,8966	8,5185	8,8576
0,2	0,8	Lev.Obs.	2,9667	2,8216	8,5792	8,8422
*	*	Lev.Verd.	2,9667	2,8216	*	*

w_1	w_2	Método	$\hat{\sigma}_{xs}$	$\hat{\sigma}_{ys}$	$\hat{inc}$	$\hat{int}$
1	0	Tucker	2,7462	2,5679	0,9351	0,8941
1	0	Lev.Obs.	2,7462	2,5756	0,9379	0,7934
0,5	0,5	Tucker	2,7440	2,5659	0,9351	0,8928
0,5	0,5	Lev.Obs.	2,7417	2,5713	0,9378	0,7954
0,2	0,8	Tucker	2,7414	2,5636	0,9351	0,8919
0,2	0,8	Lev.Obs.	2,7363	2,5661	0,9378	0,7968
*	*	Lev.Verd.	*	*	0,9511	0,6827

Podemos perceber que os resultados variantes aos pesos dos grupos nos métodos apresentados na Tabela 3.2 não apresentam muita diferença. Porém, avaliando as estimativas para o caso de $w_1 = 1$ e $w_2 = 0$, e com a finalidade de compararmos o desempenho dos grupos que realizaram os testes A_1 e A_2 temos, para alguns escores do teste A_1 , na Tabela 3.3, os escores equivalentes no teste A_2 . Analisando essa tabela vemos que para $x=20$ os escores equivalentes no teste A_2 são menores que 20, indicando que o teste A_1 é mais fácil que o teste A_2 , e portanto, o grupo de indivíduos que realiza o teste A_1 tende a ter desempenho um pouco melhor e, dessa forma, o grupo 1 possui habilidade razoavelmente menor que o grupo 2 para responder ao teste A_2 .

Tabela 3.3: Teste A_2 equivalente para o exemplo ilustrativo

x	Tucker	Lev.Obs.	Lev.Verd.
0	0,8941	0,7934	0,6827
10	10,2448	10,1721	10,1935
15	14,9201	14,8615	14,9489
20	19,5955	19,5509	19,7044

Podemos verificar quanto das diferenças observadas em médias (ou variâncias)

são atribuídas às diferenças do grupo e às diferenças de forma. Se as diferenças dos grupos ou das formas forem muito grandes, esse pode ser um indício de que os resultados obtidos podem ser pouco confiáveis.

Utilizando a diferença entre as médias nas populações temos:

$$\begin{aligned}\mu_{x_1} - \mu_{y_2} &= w_1[\mu_{x_1} - \mu_{y_2}] \text{ diferença de forma para o grupo 1} \\ &+ w_2[\mu_{x_1} - \mu_{y_2}] \text{ diferença de forma para o grupo 2} \\ &+ w_2[\mu_{x_1} - \mu_{x_2}] \text{ diferença entre os grupos na escala do teste } A_1 \\ &+ w_1[\mu_{y_1} - \mu_{y_2}] \text{ diferença entre os grupos na escala do teste } A_2.\end{aligned}$$

No caso de $w_1 = 1$ e $w_2 = 0$ temos somente a diferença de forma para o grupo 1 e a diferença entre os grupos na escala do teste A_2 , onde $\mu_{y_1} = \mu_{y_2} + \gamma_2[\mu_{v_1} - \mu_{v_2}]$.

Aplicando o Método de Tucker quando $w_1 = 1$ e $w_2 = 0$, e considerando os grupos 1 e 2, temos:

$$\begin{aligned}\hat{\mu}_{x_1} - \hat{\mu}_{y_2} &= w_1[\hat{\mu}_{x_1} - \hat{\mu}_{y_2}] + w_1[\hat{\mu}_{y_1} - \hat{\mu}_{y_2}] \\ &= 1.[\hat{\mu}_{x_1} - \hat{\mu}_{y_2}] + 1.[\hat{\mu}_{y_2} + \hat{\gamma}_2[\hat{\mu}_{v_1} - \hat{\mu}_{v_2}] - \hat{\mu}_{y_2}] \\ &= [\hat{\mu}_{x_1} - \hat{\mu}_{y_2}] + \hat{\gamma}_2[\hat{\mu}_{v_1} - \hat{\mu}_{v_2}] \\ &= [8,3810 - 8,8893] + 1,8966[2,3703 - 2,4538] \\ &= -0,3499 - 0,1584.\end{aligned}$$

O grupo 1 tem menor realização que o grupo 2 por 0,1584 unidades sobre a escala do teste A_2 , e para o grupo 2 o teste A_1 é mais difícil que o teste A_2 por 0,3499 unidades.

Utilizando as diferenças entre as variâncias e no caso de $w_1 = 1$ e $w_2 = 0$, e considerando o exemplo, temos:

$$\begin{aligned}\hat{\sigma}_{x_1}^2 - \hat{\sigma}_{y_2}^2 &= \hat{\sigma}_{x_1}^2 - \hat{\sigma}_{y_2}^2 - \hat{\gamma}_2^2[\hat{\sigma}_{v_1}^2 - \hat{\sigma}_{v_2}^2] + \hat{\gamma}_2^2[\hat{\sigma}_{v_1}^2 - \hat{\sigma}_{v_2}^2] \\ &= 7,5416 - 6,5613 - 1,8966^2[1,2352 - 1,2261] + 1,8966^2[1,2352 - 1,2261] \\ &= 0,9476 + 0,0327.\end{aligned}$$

Sobre a escala do teste A_2 , o grupo 1 teve maior variância que o grupo 2 por

0,0327 unidades, e para o grupo 1, o teste A_1 teve maior variância que o teste A_2 por 0,9476 unidades.

Se $\sigma_{xv_1} > 0$ temos que os γ -termos para o Método de Tucker, na relação em (3.9), são menores que para o Método de Levine pelo Modelo Congenérico Clássico obtidos na expressão (3.48), os quais são fatores de expansão devido ao produto pelas diferenças dos grupos. Assim, temos uma diferença maior pelo Método de Levine.

Quando os grupos diferem substancialmente, o Método de Levine parece ser mais apropriado, porém o pesquisador tem a liberdade de decidir qual será o melhor.

3.2 Métodos de Equipercenil para grupos não equivalentes

Quando temos grupos não equivalentes considera-se as distribuições do escore total e dos escores sobre os itens comuns. Nestes casos, a aplicação do método está associado ao uso do conceito de população sintética, assim como no Método de Tucker para modelos lineares.

Utilizando a distribuição condicional do escore definimos f_{xv} como sendo a distribuição conjunta do escore total e do escore sobre os itens comuns $[P(X = x, V = v)]$, f_x como a distribuição marginal do escore sob o teste A_1 $[P(X = x)]$, g_y como a distribuição marginal do escore sob o teste A_2 $[P(Y = y)]$, h_v como a distribuição marginal dos escores dos itens comuns $[P(V = v)]$, $f_{x|v} = \frac{f_{xv}}{h_v}$ como a distribuição condicional dos escores sob o teste A_1 para examinados que receberam um escore particular sobre os itens comuns $[P(X = x | V = v)]$.

3.2.1 Método de Estimação de Frequência

Nesse método existe a necessidade de expressarmos a distribuição em termos da população sintética, ponderando de acordo com cada grupo da seguinte forma:

$$\begin{aligned} f_x &= w_1 f_{x_1} + w_2 f_{x_2} \text{ para o teste } A_1, \text{ e} \\ g_y &= w_1 g_{y_1} + w_2 g_{y_2} \text{ para o teste } A_2, \end{aligned}$$

onde o grupo 1 realiza o teste A_1 e o grupo 2 realiza o teste A_2 . Da mesma forma temos $w_1 + w_2 = 1$ e $w_1, w_2 \geq 0$. Temos f_{x_1} como a distribuição marginal dos escores sobre o teste A_1 para o grupo 1, f_{x_2} como a distribuição marginal dos escores sobre o teste A_1 para o grupo 2, g_{y_1} como a distribuição marginal dos escores sobre o teste A_2 para o grupo 1 e g_{y_2} como a distribuição marginal dos escores sobre o teste A_2 para o grupo 2.

Do mesmo modo temos que f_{x_2} e g_{y_1} não podem ser determinados diretamente, e também serão obtidas expressões de forma a encontrar esses resultados através do uso de quantidades conhecidas.

Uma suposição feita neste método é que para ambos testes A_1 ou A_2 , a distribuição condicional do escore total, dado cada escore dos itens comuns, é a mesma para ambas populações. Então:

$$\begin{aligned} f_{(x|v)1} &= f_{(x|v)2}, \text{ e} \\ g_{(y|v)1} &= g_{(y|v)2}, \end{aligned}$$

mas

$$\begin{aligned} f_{xv2} &= f_{(x|v)2} h_{v2}, \text{ e} \\ g_{yv1} &= g_{(y|v)1} h_{v1}, \end{aligned}$$

portanto:

$$\begin{aligned} f_{xv2} &= f_{(x|v)1} h_{v2}, \text{ e} \\ g_{yv1} &= g_{(y|v)2} h_{v1}, \end{aligned}$$

as quais podem ser determinadas em função de quantidades observadas diretamente. Temos f_{xv2} como a distribuição conjunta do escore total do teste A_1 e o escore sobre os itens comuns para o grupo 2, g_{yv1} como a distribuição conjunta do escore total do teste A_2 e o escore sobre os itens comuns para o grupo 1, $f_{(x|v)1}$ como a distribuição condicional dos escores do teste A_1 dados os escores sobre os itens comuns no grupo 1, $f_{(x|v)2}$ como a distribuição condicional dos escores do teste A_1 dados os escores sobre os itens comuns no grupo 2, $g_{(y|v)1}$ como a distribuição condicional dos escores do teste A_2 dados os escores

sobre os itens comuns no grupo 1, $g_{(Y|V)2}$ como a distribuição condicional dos escores do teste A_2 dados os escores sobre os itens comuns no grupo 2, h_{v1} como a distribuição marginal dos escores sobre os itens comuns no grupo 1, e h_{v2} como a distribuição marginal dos escores sobre os itens comuns no grupo 2.

No entanto, a distribuição marginal se associa à soma sobre todos os escores dos itens comuns da seguinte forma:

$$\begin{aligned} f_{x2} &= \sum_V f_{(X|V)1} h_{v2}, \quad e \\ g_{y1} &= \sum_V g_{(Y|V)2} h_{v1}. \end{aligned}$$

Portanto, as expressões para a população sintética são dadas por:

$$f_x = w_1 f_{x1} + w_2 \sum_V f_{(X|V)1} h_{v2}, \quad e \quad (3.61)$$

$$g_y = w_1 \sum_V g_{(Y|V)2} h_{v1} + w_2 g_{y2}. \quad (3.62)$$

Desta forma, podemos determinar F_x e G_y correspondentes às distribuições acumuladas para os testes A_1 e A_2 , respectivamente, em relação à população sintética. Assim, determinamos também as funções ordem percentil dos testes, denominadas por P_x e Q_y , como definido anteriormente. As funções de percentil são P_x^{-1} e Q_y^{-1} . Portanto, a função de equipercntil será:

$$e_y = Q_y^{-1}[P_x].$$

Vamos considerar um exemplo ilustrativo cujos grupos 1 e 2 realizaram um teste cada um contendo 15 itens, dos quais 5 são comuns. O grupo 1 realizou o teste A_1 e o grupo 2 o teste A_2 .

Nas tabelas 3.4 e 3.5 temos no centro, respectivamente, os valores correspondentes a f_{xv1} e g_{yv2} . Na Tabela 3.4, por exemplo, podemos verificar que um escore 8 pode ser obtido de várias formas, como tendo escore 3 sobre os itens não comuns e 5 sobre os itens comuns (0,012), ou escore 4 sobre os itens não comuns e escore 4 sobre os itens comuns (0,028), ou escore 5 sobre os itens não comuns e escore 3 sobre os itens comuns

Tabela 3.4: Distribuição das probabilidades relativas ao teste A_1 e aos itens comuns do exemplo ilustrativo para o grupo 1

X	Escore sobre os itens comuns								
	0	1	2	3	4	5			
							f_x	F_x	P_x
0	0,001	0,002	0,003	0,000	0,002	0,004	0,012	0,012	1,20
1	0,006	0,008	0,003	0,005	0,008	0,005	0,035	0,047	2,95
2	0,012	0,014	0,010	0,013	0,014	0,008	0,071	0,118	8,25
3	0,018	0,016	0,012	0,019	0,020	0,012	0,097	0,215	16,65
4	0,024	0,025	0,016	0,020	0,028	0,013	0,126	0,341	27,80
5	0,025	0,032	0,022	0,026	0,032	0,020	0,157	0,498	41,95
6	0,021	0,028	0,025	0,028	0,040	0,024	0,166	0,644	58,10
7	0,015	0,023	0,020	0,025	0,035	0,018	0,136	0,800	73,20
8	0,012	0,016	0,015	0,020	0,027	0,010	0,100	0,900	85,00
9	0,009	0,011	0,010	0,018	0,012	0,004	0,064	0,964	93,20
10	0,005	0,006	0,002	0,013	0,009	0,001	0,036	1,000	98,20
h_{v_1}	0,148	0,181	0,138	0,187	0,227	0,119			

Tabela 3.5: Distribuição das probabilidades relativas ao teste A_2 e aos itens comuns do exemplo ilustrativo para o grupo 2

Y	Escore sobre os itens comuns								
	0	1	2	3	4	5			
							g_y	G_y	Q_y
0	0,002	0,000	0,001	0,003	0,005	0,002	0,013	0,013	0,65
1	0,004	0,006	0,004	0,008	0,016	0,003	0,041	0,054	3,35
2	0,010	0,008	0,007	0,015	0,018	0,008	0,066	0,120	8,70
3	0,014	0,015	0,012	0,020	0,021	0,012	0,094	0,214	16,70
4	0,021	0,020	0,014	0,024	0,022	0,015	0,116	0,330	27,20
5	0,025	0,025	0,020	0,030	0,025	0,018	0,143	0,473	40,15
6	0,022	0,027	0,021	0,027	0,032	0,020	0,149	0,622	54,75
7	0,018	0,030	0,023	0,020	0,028	0,017	0,136	0,758	69,00
8	0,015	0,024	0,017	0,018	0,025	0,014	0,113	0,871	81,45
9	0,011	0,016	0,013	0,014	0,020	0,008	0,082	0,953	91,20
10	0,006	0,010	0,006	0,008	0,015	0,002	0,047	1,000	97,65
h_{v_2}	0,148	0,181	0,138	0,187	0,227	0,119			

(0,026), ou escore 6 sobre os itens não comuns e escore 2 sobre os itens comuns (0,025), ou escore 7 sobre os itens não comuns e escore 1 sobre os itens comuns (0,023) e ou escore 8 sobre os itens não comuns e escore 0 sobre os itens comuns (0,012). Assim, em cada 1000 examinados, por volta de 126 indivíduos devem obter escore 8 ($0,012 + 0,028 + 0,026 +$

0,025 + 0,023 + 0,012 = 0,126). O mesmo raciocínio vale para a Tabela 3.5.

Com esses dados podemos determinar as probabilidades condicionais $f_{(x|V)1} = \frac{f_{xV1}}{h_{V1}}$ e $g_{(Y|V)2} = \frac{g_{YV2}}{h_{V2}}$. Consideramos os itens comuns externos ao escore total e, em relação à população sintética consideraremos $w_1 = 1$ e $w_2 = 0$. Portanto:

$$f_x = f_{x1}, \text{ e}$$

$$g_Y = \sum_V g_{(Y|V)2} h_{V1} = g_{Y2}.$$

Assim, os resultados quanto ao equipercntil equivalente do teste A_1 na escala do teste A_2 por esse método se encontram na Tabela 3.6. Os métodos de suavização podem ser utilizados com os resultados obtidos de forma a melhorar o ajuste da equalização realizada (Kolen,1995).

Tabela 3.6: Equipercntil equivalente da população sintética para $w_1 = 1$ e $w_2 = 0$

Escore	$e_Y(x)$
0	0,423
1	0,902
2	1,932
3	2,995
4	4,052
5	5,126
6	6,225
7	7,309
8	8,314
9	9,244
10	10,117

Neste exemplo, a equalização é realizada somente até o escore 10 devido ao fato de estarmos tratando de um caso em que os escores sobre os itens comuns são externos ao escore total.

3.2.2 Método Linear de Braun Holland

O método linear de Braun-Holland é considerado uma generalização do Método de Tucker, porém, por ser mais complicado em termos computacionais, é considerado somente quando se suspeita de uma regressão não linear (Kolen,1995,p.274). Apesar de se apresentar como sendo um método linear, utiliza definições apresentadas no método de equipercenil de estimação de frequência.

Considera a média e a variância da população sintética expressas, respectivamente, por:

$$\mu_{xs} = \sum_X x f_{xs}, \quad (3.63)$$

e

$$\sigma_{xs}^2 = \sum_X [x - \mu_{xs}]^2 f_{xs}, \quad (3.64)$$

onde f_x é dado pelo Método de Estimação de Frequência. O processo é análogo para o teste A_2 . Utilizamos a média e o desvio padrão na equação (3.63) e (3.64) para obtermos a equalização através da população sintética.

Ainda utilizando o exemplo ilustrativo anterior, podemos determinar a equalização através da população sintética utilizando as medidas resumo onde $\mu_{x1} = 5,441$, $\mu_{y1} = 5,592$, $\sigma_{x1} = 2,314$ e $\sigma_{y1} = 2,432$.

A equação é obtida com $\beta = \frac{\sigma_{y1}}{\sigma_{x1}} = \frac{2,432}{2,314} = 1,051$ e $\alpha = \mu_{y1} - \beta \cdot \mu_{x1} = 5,592 - 1,051(5,441) = -0,126$, e dada por:

$$l_Y(x) = \alpha + \beta x = -0,126 + 1,051x.$$

Os resultados obtidos com a equalização se encontram na Tabela 3.7, onde $l_Y(x)$ determina a equalização linear de transformação da escala do teste A_1 para a escala do teste A_2 , utilizando as frequências relativas de ambos testes, as quais foram obtidas através das tabelas 3.4 e 3.5.

Tabela 3.7: Método linear de Braun-Holland para o exemplo ilustrativo

Escore	f_{x1}	g_{y1}	$l_Y(x)$
0	0,012	0,013	-0,126
1	0,035	0,041	0,925
2	0,071	0,066	1,976
3	0,097	0,094	3,027
4	0,126	0,116	4,078
5	0,157	0,143	5,129
6	0,166	0,149	6,180
7	0,136	0,136	7,231
8	0,100	0,113	8,282
9	0,064	0,082	9,333
10	0,036	0,047	10,384

3.3 Três grupos ou mais

Ao trabalharmos com grupos não equivalentes, consideramos a existência de um teste âncora entre os testes para podermos fazer uma comparação entre o desempenho dos grupos.

Consideraremos, por exemplo, que três testes A_1 , A_2 e A_3 , cada um com 30 itens, tenham sido aplicados, respectivamente, à 4ª e 8ª séries do ensino fundamental e ao 3º ano do ensino médio. Por terem sido aplicados a grupos não equivalentes os testes são constituídos de um grupo de itens comuns, sendo 6 comuns aos testes A_1 e A_2 e outros 6 comuns aos testes A_2 e A_3 . Para compararmos o desempenho dos grupos é necessário todo o processo de equalização para grupos não equivalentes de itens comuns para transformação de escala, recordando que a análise é realizada aos pares.

Consideraremos X como sendo a variável aleatória dos escores do teste A_1 , x como uma realização de X , Y como sendo a variável aleatória dos escores do teste A_2 , y como uma realização de Y e Z como sendo a variável aleatória dos escores do teste A_3 , com z como uma realização de Z .

Podemos considerar a comparação dos grupos que realizaram os testes A_2 e A_3 com o grupo que realizou o teste A_1 , porém, somente o teste A_2 possui itens comuns ao teste A_1 . Levando-se em consideração os itens comuns e, considerando que tenha sido aplicado um método linear para grupos não equivalentes, temos a equalização linear entre os testes A_1 e A_2 determinada por $l_X(y)$, e a equalização linear entre os testes A_2 e A_3

determinada por $l_Y(z)$. Dessa forma, para compararmos o desempenho dos grupos que realizaram os testes A_1 e A_3 , temos a seguinte equalização:

$$l_X[l_Y(z)],$$

onde temos uma junção das equalizações lineares anteriores que farão a transformação dos escores do teste A_3 diretamente para o teste A_1 .

O mesmo raciocínio pode ser utilizado no caso de trabalharmos com a equalização do equipercntil entre os testes. Apesar da existência de um teste âncora em todas as provas ainda temos a ocorrência de várias equalizações, o que torna o processo bastante trabalhoso, dependendo da quantidade de testes com itens comuns aplicados.

Capítulo 4

Equalização na Teoria da Resposta ao Item

Como podemos prever sobre a resposta de um examinado a um determinado item sem que ele tenha sido administrado de forma similar anteriormente? Para tanto, é necessário que se tenha uma estrutura que descreva os itens através de algum parâmetro de item e os examinados por um parâmetro de examinados de tal forma que possamos determinar a probabilidade da resposta de algum examinado a um determinado item sem que esse tenha sido apresentado anteriormente, mesmo que similarmente.

Existem algumas limitações na Teoria Clássica, as quais se referem às estatísticas do item (discriminação e dificuldade), pois são dependentes dos grupos examinados. Hambleton(1993) e Pinheiro(2006) descrevem estas estatísticas na Teoria Clássica. Além disso considera que os escores observados e verdadeiros são dependentes do teste, os quais podem aumentar ou diminuir em função da dificuldade dos testes. Também parece ilógica a suposição de igualdade dos erros de medida para todos os examinados. Por fim, a suposição de testes equivalentes (paralelos) é bastante subjetiva, sugerindo uma estimativa não muito precisa da fidedignidade dos testes, já que muitas vezes podem ser considerados testes não equivalentes (não paralelos) ao invés de testes equivalentes.

Hambleton(1993) afirma que o modelo desejado em estudo deva ter estatísticas do item não dependentes do grupo, escores que descrevam a proficiência do examinado que não dependa da dificuldade do teste, modelos que melhorem a comparabilidade dos itens do teste aos níveis de habilidade, modelos que não se encontrem baseados em suposições implausíveis, e que os modelos não suponham que os testes devam ser estritamente paralelos. Esses argumentos se encontram satisfeitos pela Teoria da Resposta ao Item

(TRI).

A TRI - Teoria da Resposta ao Item é composta de modelos matemáticos construídos com o objetivo de representar essa probabilidade, de tal modo que venha especificar uma relação entre a performance dos examinados ao item/teste e as habilidades não observadas diretamente. Esses modelos se fundamentam em três fatores: a natureza do item (dicotômico ou não), o número de populações envolvidas, e a quantidade de traços latentes (habilidades) que está se medindo.

Os modelos da TRI podem ser unidimensionais, onde os testes medem somente uma habilidade, ou multidimensionais, medem mais de uma habilidade. Também tem sido utilizados escores dicotômicos (0,1) e, respectivamente, entendidos por errado ou certo; ou policotômicos, onde possa, por exemplo, receber escores classificados em 0, 1, 2 ou 3.

Nos modelos da TRI supõe-se uma independência local, a qual a probabilidade, para uma dada habilidade, de resposta correta a dois itens ($p_{1,2}$) é dada pelo produto da probabilidade de resposta correta ao primeiro (p_1) e a probabilidade de resposta correta ao segundo (p_2).

$$p_{1,2} = p_1 p_2,$$

sendo 1 e 2 dois itens de um teste.

No estudo em questão será dada ênfase aos modelos dicotomizados unidimensionais. São utilizados modelos logísticos de até três parâmetros. No caso de termos somente um parâmetro é avaliada a dificuldade do item. No caso de dois parâmetros avalia-se a dificuldade e a discriminação do item. E, no caso de termos três parâmetros, observa-se a dificuldade, a discriminação e a probabilidade de resposta correta para examinados com pouca habilidade. O modelo escolhido no estudo será o logístico com três parâmetros pelo fato dos outros modelos serem consequência deste.

4.1 Modelo Logístico Unidimensional de 3 Parâmetros

O Modelo Logístico unidimensional de três parâmetros (ML3) é o modelo mais utilizado atualmente, e é dado pela seguinte função logística:

$$P(U_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (4.1)$$

com $i=1,\dots,I$ correspondente aos itens e $j=1,\dots,n$ correspondente aos examinados, onde:

$P(U_{ij} = 1|\theta_j)$ é a probabilidade do examinado j responder corretamente o item i dada sua habilidade θ_j . É denominada FRI - Função de Resposta ao Item;

U_{ij} corresponde à variável dicotômica que equivale a 1 quando o examinado j responde corretamente o item i , e 0 caso contrário;

θ_j corresponde ao traço latente (habilidade) do examinado j . A habilidade θ se define no intervalo $-\infty < \theta < \infty$, e muitas vezes é escalonado por ser distribuído normalmente com $\mu = 0$ e $\sigma = 1$. É interessante notar que examinados que tenham recebido o mesmo escore, não necessariamente, possuem a mesma habilidade, pois isso depende dos itens que responderam corretamente;

a_i representa o parâmetro de discriminação do item, o grau cuja resposta varia com o nível de habilidade. Não se espera que a seja negativo, pois indicaria que a probabilidade de responder um item corretamente é inversamente proporcional ao aumento da habilidade. Valores baixos de a indicam que, seja qual for o nível de habilidade, a probabilidade de responder corretamente um item é praticamente a mesma, ou seja, tem pouco poder de discriminação. Para valores altos de a_i temos a CCI - Curva Característica do Item muito acentuada, separando os examinados em dois grupos: aqueles que possuem habilidades abaixo e acima do parâmetro b ;

b_i é um parâmetro de locação do item i . É também conhecido como a *difículdade do item* e representa o nível de habilidade tal que a probabilidade de resposta correta será igual a $\frac{1+c}{2}$. Quanto maior o valor de b , maior a habilidade exigida para responder corretamente o item e vice-versa. Na escala de proficiência o ponto zero não significa que o examinado não tenha nenhuma habilidade. Quando não se permite acerto casual, $c = 0$, temos b como sendo o nível de habilidade em que a probabilidade de resposta correta é 0,5;

c_i é a probabilidade de que examinados com baixa habilidade respondam de forma correta o item i , e portanto, assume valores entre 0 e 1;

D é denominado fator de escala. Quando se pretende resultados aproximados ao da função ogiva normal utiliza-se o valor 1,7. Caso contrário, o valor será 1. A função ogiva normal é dada por:

$$P \equiv P(\theta) = c + (1 - c) \int_{-\infty}^a (\theta - b) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (4.2)$$

Os modelos (4.1) e (4.2) produzem resultados muito similares na prática, e a diferença entre eles não chega a 0,01 para todos os parâmetros quando utilizamos (4.1) considerando $D = 1,7$. O estudo da TRI também pode ser encontrado em Andrade et al(2000), Lord(1980), Hambleton(1991), entre outros.

No ML3 a *curva característica do item* relaciona a *função de resposta ao item* com os parâmetros do modelo.

A estimação dos parâmetros e habilidades, em geral, é realizada pelo Método de Máxima Verossimilhança (MMV) através de processos iterativos tais como Newton Raphson ou Scoring de Fisher. Também podem ser utilizados processos bayesianos nessa estimação (Azevedo,2003).

Definiremos por $\varsigma = (a_i, b_i, c_i)$ o vetor de parâmetros do item i , com $i = 1, \dots, I$ conhecidos, $U_j = (U_{j1}, U_{j2}, \dots, U_{jI})$ o vetor aleatório de respostas do examinado j , com $j = 1, \dots, n$, $U_{..} = (U_{1.}, U_{2.}, \dots, U_{n.})$ o vetor correspondente ao conjunto integral de respostas, $\theta = (\theta_1, \dots, \theta_n)$ representa o vetor de habilidades dos n respondentes, e u_{ji} , u_j e $u_{..}$ representam as observações, conforme descrito anteriormente. Andrade (2000), Baker(1992), entre outros, apresentam as demonstrações em detalhes.

4.1.1 Parâmetros dos itens

Dado o modelo logístico unidimensional de 3 parâmetros como apresentado na expressão (4.1), considerando os parâmetros populacionais conhecidos, a independência entre as respostas dos respondentes e a independência local, podemos determinar a verossimilhança como::

$$\begin{aligned}
L(\varsigma) &= \prod_{j=1}^n P(U_{j\cdot} = u_{j\cdot} \mid \theta_j, \varsigma) \\
&= \prod_{j=1}^n \prod_{i=1}^I P(U_{ji} = u_{ji} \mid \theta_j, \varsigma_i) \\
&= \prod_{j=1}^n \prod_{i=1}^I P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}},
\end{aligned}$$

onde $P_{ji} = P(U_{ji} = 1 \mid \theta_j, \varsigma_i)$ e $Q_{ji} = 1 - P_{ji}$. Dessa forma, temos a log-verossimilhança dada por:

$$\log L(\varsigma) = \sum_{j=1}^n \sum_{i=1}^I u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}. \quad (4.3)$$

Igualando a zero as derivadas da log-verossimilhança em função de cada parâmetro do item determinaremos as estimativas de máxima verossimilhança, que são os valores dos parâmetros que maximizam a verossimilhança. Assim:

$$\frac{\partial \log L(\varsigma)}{\partial \varsigma_i} = 0, \quad i = 1, \dots, I.$$

Temos:

$$\frac{\partial \log L(\varsigma)}{\partial \varsigma_i} = \sum_{j=1}^n \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^* Q_{ji}^*} \right\} \left(\frac{\partial P_{ji}}{\partial \varsigma_i} \right),$$

onde $W_{ji} = \frac{P_{ji}^* Q_{ji}^*}{P_{ji} Q_{ji}}$ corresponde a uma ponderação, $P_{ji}^* = 1 + e^{-Da_i(\theta_j - b_i)^{-1}}$, e $Q_{ji}^* = 1 - P_{ji}^*$.

Em relação a cada parâmetro dos itens, temos determinadas as seguintes expressões das equações de estimação:

$$\begin{aligned}
\frac{\partial \log L(\varsigma)}{\partial a_i} &= D(1 - c_i) \sum_{j=1}^n (u_{ji} - P_{ji})(\theta_j - b_i) W_{ji}, \\
\frac{\partial \log L(\varsigma)}{\partial b_i} &= -Da_i(1 - c_i) \sum_{j=1}^n (u_{ji} - P_{ji}) W_{ji}, \quad e \\
\frac{\partial \log L(\varsigma)}{\partial c_i} &= \sum_{j=1}^n (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^*}.
\end{aligned}$$

4.1.2 Parâmetros dos respondentes

Quando temos itens já calibrados em outros testes, podemos definir a estimação das habilidades de um grupo de respondentes. Considerando a notação e as propriedades utilizadas para a estimação dos parâmetros dos itens, temos a log-verossimilhança em função da habilidade dada por:

$$\log L(\theta) = \sum_{j=1}^n \sum_{i=1}^I u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}.$$

O estimador de máxima verossimilhança da habilidade θ_j que maximiza a verossimilhança é dado por:

$$\frac{\partial \log L(\theta)}{\partial \theta_j} = 0, \quad j = 1, \dots, n,$$

onde

$$\frac{\partial \log L(\theta)}{\partial \theta_j} = D \sum_{i=1}^I a_i (1 - c_i) (u_{ji} - P_{ji}) W_{ji},$$

e correspondente à equação de estimação da habilidade de cada respondente.

Como essas equações não possuem solução explícita, aplicamos algoritmos como o de Newton-Raphson e "Scoring" de Fisher, os quais utilizam processos iterativos.

4.2 Equalização na TRI

4.2.1 Equalização a posteriori

Ao fazermos a calibração dos itens para cada grupo separadamente, cada um se encontra em uma métrica. Por exemplo, se tivermos dois grupos não equivalentes de respondentes, um pertencente à 4ª série do ensino fundamental e outro à 6ª série do ensino fundamental, considerando que cada grupo tenha sido calibrado numa métrica de média 0 e desvio 1. No entanto, uma habilidade 0 no grupo da 4ª série não corresponde à mesma habilidade 0 no grupo da 6ª série. Dessa forma, realizamos uma equalização a posteriori, no qual após a calibração dos itens para cada grupo, fazemos a equalização em função dos parâmetros determinados, considerando o princípio de invariância dos grupos.

O objetivo da equalização na TRI é colocar os parâmetros dos itens de duas ou mais provas, ou as habilidades dos respondentes de diferentes grupos numa escala que seja comum, de modo a tornar possível a comparação, do ponto de vista da interpretação. A equalização pode ser feita de duas formas: via população ou via itens comuns. Quando se trata da equalização na TRI não se faz diferença entre o tipo de grupo que está sendo utilizado, portanto a regra é a mesma para grupos equivalentes e não equivalentes, o que facilita bastante em relação à Teoria Clássica.

Na equalização via população, um único grupo realiza provas distintas, ou seja, grupos equivalentes de uma mesma população que realizam provas distintas, no qual, dado que os itens tenham sido calibrados em conjunto, existe a garantia de que estejam numa escala comum. Por exemplo, se forem aplicados dois testes, cada um com 30 itens, porém compostos de itens diferentes, desde que estejam avaliando a mesma habilidade para um certo conteúdo, estes podem ser comparados diretamente. Se esses testes forem aplicados para a 3ª série do ensino fundamental, trabalharemos com dois grupos equivalentes, onde o primeiro grupo responde aos itens de 1 a 30 e o outro grupo responde aos itens de 31 a 60. Como se trata do mesmo grupo de respondentes, esses itens podem ser calibrados conjuntamente, considerando assim que os itens 31 a 60 não tenham sido apresentados ao grupo 1. No caso da equalização via itens comuns, a escala comum será dada justamente pelo elo entre as populações envolvidas. Em geral, a calibração, processo de estimação dos parâmetros dos itens, é feita utilizando-se uma escala com média 0 e desvio padrão 1.

No caso de grupos equivalentes, segundo a forma a ser administrado o teste, os parâmetros são calibrados simultaneamente na mesma escala, não necessitando de equalização.

Para grupos não equivalentes as estimativas dos parâmetros de habilidade não são considerados estarem na mesma escala. Nesse caso é necessária a transformação de escala correspondente aos itens comuns para os grupos. Podemos pensar na situação em que desejamos comparar uma dada habilidade em matemática dos alunos da escola pública de 5ª série com alunos da 6ª série do ensino fundamental. De certa forma, a habilidade média dos alunos de 6ª série, deve ser maior que a habilidade média dos alunos de 5ª série, e que podem ser colocados na mesma métrica quando comparamos os itens comuns em ambas provas.

Para os grupos não equivalentes podemos calibrar os dados simultaneamente,

onde cada examinado resolve o teste âncora, e assim teremos uma combinação dos dados sobre esses itens para ambos testes, no qual já é realizada a equalização na TRI diretamente através dos modelos de grupos múltiplos. Também podemos fazer a calibração separadamente, e depois a equalização através de métodos de equalização da TRI, os quais serão tratados mais adiante.

Quanto à quantidade de itens comuns numa prova, é sugerido por Andrade(2001) e Kolen(1995,p.248) que ao menos 20% do total de itens seja comum, tanto na Teoria Clássica assim como na TRI. Porém a quantidade exata depende do método de equalização a ser utilizado, e da boa qualidade desses itens. Ou seja, pode-se realizar a equalização durante o processo de calibração ou a posteriori; ou, se tivermos itens comuns com pouco poder de discriminação, ou níveis de dificuldade muito baixos ou muito altos, existe a necessidade de um número maior de itens comuns.

A equalização na TRI segue três passos:

1. Estimação conjunta dos parâmetros dos itens e das habilidades dos examinados.
2. As estimativas dos parâmetros devem ser escalonadas para uma métrica comum através de uma transformação linear.
3. Faz-se a conversão dos escores obtidos para os escores escalonados.

Os parâmetros dos itens devem se encontrar numa mesma escala que represente, respectivamente, a média e o desvio das habilidades dos respondentes da população ou populações envolvidas. Os programas computacionais utilizados, em geral, escalam as habilidades dos grupos como tendo $\mu = 0$ e $\sigma = 1$, mesmo que haja diferença e, nesse caso, a transformação de escala é necessária.

Aqui podemos considerar a situação em que tenhamos três testes A_1 , A_2 e A_3 respondidos, respectivamente, por três grupos não equivalentes, no qual os testes A_1 e A_2 tenham itens comuns, os testes A_2 e A_3 tenham itens comuns, e também os testes A_1 e A_3 tenham itens comuns. Podemos pensar em colocar todos os parâmetros dos itens na métrica do grupo que tenha realizado o teste A_1 . Dessa forma, temos duas maneiras de transformar a métrica do teste A_3 na métrica do teste A_1 , transformando diretamente os parâmetros dos itens do teste A_3 para a métrica dos parâmetros do teste A_1 , ou transformar os parâmetros dos itens do teste A_3 para a métrica dos parâmetros do teste A_2 e então transformar na métrica dos parâmetros do teste A_1 . Assim, podemos obter

resultados diferentes na equalização para todos os parâmetros, e então podemos considerar um dos resultados como sendo o "verdadeiro", ou fazemos uma média dos resultados dos parâmetros obtidos em ambas equalizações de modo a obter um valor único.

Sendo assim, sejam M e N correspondentes a duas escalas de habilidade do ML3. Ambas escalas são equiparadas linearmente da seguinte forma:

$$\theta_{Nj} = A\theta_{Mj} + B, \quad (4.4)$$

com A e B constantes, e θ_{Nj} e θ_{Mj} como sendo as habilidades correspondentes a cada examinado j .

Os parâmetros do item possuem as seguintes relações:

$$a_{Ni} = \frac{a_{Mi}}{A} \Rightarrow A = \frac{a_{Mi}}{a_{Ni}}, \quad (4.5)$$

$$b_{Ni} = Ab_{Mi} + B \Rightarrow B = b_{Ni} - Ab_{Mi}, \quad (4.6)$$

$$c_{Ni} = c_{Mi}, \quad (4.7)$$

onde temos os parâmetros do item para cada item i de ambas escalas M e N .

Podemos demonstrar a relação linear (4.4) considerando o modelo ML3. Podemos igualar a probabilidade de um indivíduo j responder corretamente um item i dadas suas habilidades θ_{Nj} e θ_{Mj} , pois estamos comparando o mesmo indivíduo cuja habilidade é representada em escalas diferentes. Portanto, temos:

$$c_{Ni} + \frac{1 - c_{Ni}}{1 + e^{-a_{Ni}(\theta_{Nj} - b_{Ni})}} = c_{Mi} + \frac{1 - c_{Mi}}{1 + e^{-a_{Mi}(\theta_{Mj} - b_{Mi})}}, \quad (4.8)$$

mas, como $c_{Ni} = c_{Mi}$, temos:

$$\begin{aligned}
\frac{1}{1 + e^{-a_{Ni}(\theta_{Nj} - b_{Ni})}} &= \frac{1}{1 + e^{-a_{Mi}(\theta_{Mj} - b_{Mi})}} \\
1 + e^{-a_{Ni}(\theta_{Nj} - b_{Ni})} &= 1 + e^{-a_{Mi}(\theta_{Mj} - b_{Mi})} \\
e^{-a_{Ni}(\theta_{Nj} - b_{Ni})} &= e^{-a_{Mi}(\theta_{Mj} - b_{Mi})} \\
a_{Ni}(\theta_{Nj} - b_{Ni}) &= a_{Mi}(\theta_{Mj} - b_{Mi}) \\
\theta_{Nj} - b_{Ni} &= \frac{a_{Mi}}{a_{Ni}}\theta_{Mj} - \frac{a_{Mi}}{a_{Ni}}b_{Mi} \\
\theta_{Nj} &= \frac{a_{Mi}}{a_{Ni}}\theta_{Mj} - \frac{a_{Mi}}{a_{Ni}}b_{Mi} + b_{Ni} \\
&= A\theta_{Mj} - Ab_{Mi} + b_{Ni} \\
&= A\theta_{Mj} + B
\end{aligned}$$

onde A e B estão definidos, respectivamente, em (4.5) e (4.6).

Para quaisquer dois examinados j e j^* , ou quaisquer dois itens i e i^* temos que as constantes A e B podem ser expressas da seguinte forma:

$$\begin{aligned}
A &= \frac{\theta_{Nj} - \theta_{Nj^*}}{\theta_{Mj} - \theta_{Mj^*}}, \\
&= \frac{b_{Ni} - b_{Ni^*}}{b_{Mi} - b_{Mi^*}}, \\
&= \frac{a_{Mi}}{a_{Ni}}, \\
B &= b_{Ni} - Ab_{Mi}, \\
&= \theta_{Nj} - A\theta_{Mj}.
\end{aligned}$$

Existe uma propriedade denominada *indeterminância de locação e extensão de escala* cujos valores de $P(U_{ij}|\theta_j)$ são idênticos para ambas escalas relacionadas linearmente.

Em termos de grupos de itens e/ou examinados temos:

$$A = \frac{\sigma(b_N)}{\sigma(b_M)}, \quad (4.9)$$

$$= \frac{\mu(a_M)}{\mu(a_N)}, \quad (4.10)$$

$$= \frac{\sigma(\theta_N)}{\sigma(\theta_M)}, \quad (4.11)$$

$$B = \mu(b_N) - A\mu(b_M), \quad (4.12)$$

$$= \mu(\theta_N) - A\mu(\theta_M), \quad (4.13)$$

onde as médias são definidas sobre dois ou mais itens ou examinados com parâmetros expressos por ambas escalas, e os desvios padrões estão definidos para dois ou mais itens ou examinados com parâmetros expressos por ambas escalas.

As equações correspondentes aos traços latentes podem somente ser usadas se os parâmetros para o mesmo grupo de examinados são expressos em ambas escalas.

Tabela 4.1: Exemplo hipotético para determinação de A e B

item	a_{Mi}	b_{Mi}	c_{Mi}	a_{Ni}	b_{Ni}	c_{Ni}
i=1	0,400	-1,100	0,100	0,444	-1,390	0,100
i=2	1,700	0,900	0,200	1,889	0,410	0,200
i=3	1,200	2,200	0,200	1,333	1,580	0,200
i=4	1,000	2,500	0,100	1,111	1,850	0,100
média	1,075	1,125	0,150	1,194	0,613	0,150
desvio	0,538	1,638	0,058	0,597	1,474	0,058

Podemos pensar num exemplo hipotético de um teste que contenha 4 itens comuns, cujas estimativas dos parâmetros para ambas escalas M e N são dadas pela Tabela 4.1. Dadas as estimativas obtidas, e substituindo nas equações para determinar A e B, encontramos, em termos de grupos, os seguintes resultados:

$$A = \frac{\sigma(b_N)}{\sigma(b_M)} = \frac{1,474}{1,638} = 0,900, \quad e$$

$$B = \mu(b_N) - A\mu(b_M) = 0,613 - 0,900.(1,125) = -0,400.$$

Utilizando esses valores podemos comparar a habilidade de um indivíduo na escala M com a escala N dada a equação (4.4). Por exemplo, um indivíduo j com habilidade

$\theta_{Mj} = -1,800$ na escala M terá habilidade na escala N equivalente a $\theta_{Nj} = A\theta_{Mj} + B = 0,900(-1,800) - 0,400 = -2,020$.

Como esses valores se encontram linearmente relacionados, temos pela equação (4.8) que a probabilidade de um indivíduo j responder corretamente o item i, após a transformação da escala M na escala N, são equivalentes. Por exemplo, podemos determinar, para o item 2, as probabilidades em ambas escalas utilizando a função (4.1) e os resultados da Tabela 4.1. Então,

$$P(U_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta_j - b_i)}} = 0,200 + \frac{1 - 0,200}{1 + e^{-1,700 \cdot 1,700 \cdot (-1,800 - 0,900)}} = 0,20,$$

correspondente à probabilidade do item 2 na escala M, e,

$$P(U_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta_j - b_i)}} = 0,200 + \frac{1 - 0,200}{1 + e^{-1,889 \cdot 1,700 \cdot (-2,020 - 0,410)}} = 0,20,$$

correspondente à probabilidade do item 2 na escala N, e cujos valores são equivalentes, o que ocorre com todos os itens.

Podemos aqui pensar na situação em que tenhamos grupos equivalentes onde cada um deles realiza testes totalmente distintos. Por exemplo, se os alunos do 4º ano do ciclo II do período diurno pertencentes à escola pública realizarem dois testes distribuídos aleatoriamente contendo 35 itens cada, porém nenhum item comum. Como estamos tratando da mesma população, os itens podem ser calibrados em conjunto, pois os parâmetros certamente se encontrarão na mesma escala, porém considerando os 35 últimos itens como não apresentados ao grupo que realizou os outros 35 itens.

Quando trabalhamos com grupos não equivalentes, os grupos se encontram em escalas diferentes, e portanto, é necessário realizarmos essa transformação. Existem dois métodos que podem ser aplicados a essa transformação linear. O **Método Média/Média**, o qual determina A e B somente com o uso da média das estimativas dadas pelos parâmetros de discriminação dos itens comuns em ambas escalas M e N. Já o **Método Média/Desvio** determina A e B através da média e do desvio das estimativas dos parâmetros de dificuldade dos itens comuns, também dados pelas escalas M e N. Ambos métodos utilizam as mesmas relações descritas anteriormente, no entanto,

enquanto uma trabalha com as estimativas do parâmetro a , a outra se refere às estimativas do parâmetro b .

No entanto, os métodos descritos anteriormente trabalham com cada parâmetro separadamente na determinação de A e B . Porém, também existem métodos que consideram os parâmetros a , b e c dos itens simultaneamente. Esse método é denominado **Método de transformação da curva característica**, e aplicado à medida que as estimativas dos parâmetros do item vêm a produzir curvas características quase idênticas no intervalo de habilidades.

Como vimos anteriormente, quando utilizamos a transformação linear temos a seguinte igualdade:

$$p_{ij}(\theta_{Nj}, a_{Ni}, b_{Ni}, c_{Ni}) = p_{ij}\left(A\theta_{Mj} + B, \frac{a_{Mi}}{A}, Ab_{Mi} + B, c_{Mi}\right),$$

onde i corresponde ao item e j corresponde ao examinado.

Da mesma forma, temos estimativas sendo utilizadas no lugar dos parâmetros, não garantindo totalmente a igualdade, a qual será minimizada pela aplicação de algum método de aproximação. Os métodos conhecidos são: **Aproximação de Haebara** e **Aproximação de Stocking e Lord**.

A aproximação de Haebara se dá pela soma sobre os itens comuns da diferença quadrada entre as CCI para examinados com uma dada habilidade em particular. Isto é, podemos pensar na diferença dada por:

$$Hdiff(\theta_j) = \sum_{i:V} \left[p_{ij}(\theta_{Nj}, \hat{a}_{Ni}, \hat{b}_{Ni}, \hat{c}_{Ni}) - p_{ij}\left(\theta_{Mj}, \frac{\hat{a}_{Mi}}{A}, A\hat{b}_{Mi} + B, \hat{c}_{Mi}\right) \right]^2.$$

O processo de estimação utilizado para determinar A e B é dado de forma a minimizar o critério a seguir, cuja soma é realizada sobre os examinados:

$$H_{crit} = \sum_j Hdiff(\theta_j).$$

A aproximação de Stocking e Lord é dada pela diferença quadrada da soma das probabilidades sobre os itens comuns, ou seja,

$$SLdiff(\theta_j) = \left[\sum_{i:V} p_{ij}(\theta_{Nj}, \hat{a}_{Ni}, \hat{b}_{Ni}, \hat{c}_{Ni}) - \sum_{i:V} p_{ij} \left(\theta_{Mj}, \frac{\hat{a}_{Mi}}{A}, A\hat{b}_{Mi} + B, \hat{c}_{Mi} \right) \right]^2.$$

Da mesma forma, o processo de estimação utilizado para determinar A e B é dado de forma a minimizar o critério a seguir, cuja soma é realizada sobre os examinados:

$$SL_{crit} = \sum_j SLdiff(\theta_j).$$

Suponhamos a existência de um teste que contenha 4 itens comuns, e cujos parâmetros se encontram na Tabela 4.1, juntamente com as respectivas média e desvio de cada parâmetro do item.

Com esses resultados podemos determinar as constantes A e B para os métodos Média/Desvio e Média/Média. Pelo método Média/Desvio utilizamos os resultados obtidos com o parâmetro b e para o método Média/Média os resultados do parâmetro a . Na Tabela 4.2 apresentamos os resultados obtidos utilizando ambas transformações, e cujos resultados foram idênticos.

Tabela 4.2: Resultados obtidos para as constantes A e B utilizando os métodos Média/Média e Média/Desvio (os verdadeiros valores das constantes para o método Média/Média são $A=0,900147$ e $B=-0,40016$, os quais foram arredondados, e por isso apresentam-se iguais na Tabela.)

Constantes	Média/Desvio	Média/Média
A	0,900	0,900
B	-0,400	-0,400

Com os resultados obtidos podemos fazer a conversão dos parâmetros que se encontram na escala M para a escala N utilizando as constantes A e B determinadas para cada método e as equações (4.5) e (4.6). Os resultados se encontram na Tabela 4.3.

Podemos então estimar a probabilidade de resposta correta dada uma habilidade através das estimativas obtidas dos parâmetros após a transformação para podermos comparar os resultados. Utilizando a equação (4.1) e considerando a habilidade $\theta = 0$, as probabilidades correspondentes se encontram na tabela (4.4).

Tabela 4.3: Transformação da escala M para a escala N utilizando os métodos Média/Desvio e Média/ Média

Item	Média/Desvio			Média/Média		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
1	0,444	-1,390	0,100	0,444	-1,390	0,100
2	1,889	0,410	0,200	1,889	0,410	0,200
3	1,333	1,580	0,200	1,333	1,580	0,200
4	1,111	1,850	0,100	1,111	1,850	0,100
Média	1,194	0,613	0,150	1,194	0,613	0,150
Desvio	0,597	1,474	0,058	0,597	1,474	0,058

Tabela 4.4: Probabilidade estimada de responder corretamente o item *i* dada a habilidade $\theta = 0$

Item	Escala N	Média/Desvio	Média/Média
1	0,766554	0,766736	0,766749
2	0,369103	0,369114	0,369156
3	0,221687	0,221668	0,221672
4	0,126533	0,126524	0,126528
Soma	1,483877	1,484041	1,484106

Fazendo a aplicação aos métodos de transformação da curva característica podemos verificar a diferença existente entre os métodos Média/Desvio e Média/Média segundo as aproximações de Haebara e Stocking & Lord.

Para o método Média/Desvio temos:

$$\begin{aligned}
 Hdiff(\theta_i = 0) &= (0,766554 - 0,766736)^2 + (0,369103 - 0,369114)^2 + \\
 &\quad (0,221687 - 0,221668)^2 + (0,126533 - 0,126524)^2 \\
 &= 3,34976 \cdot 10^{-8}
 \end{aligned}$$

$$SLdiff(\theta_i = 0) = (1,483877 - 1,484041)^2 = 2,68893 \cdot 10^{-8}.$$

Para o método Média/Média temos:

$$\begin{aligned}
Hdiff(\theta_i = 0) &= (0,766554 - 0,766749)^2 + (0,369103 - 0,369156)^2 + \\
&\quad (0,221687 - 0,221672)^2 + (0,126533 - 0,126528)^2 \\
&= 4,08783 \cdot 10^{-8} \\
SLdiff(\theta_i = 0) &= (1,483877 - 1,484106)^2 = 5,25857 \cdot 10^{-8}.
\end{aligned}$$

Os métodos da curva característica fornecem resultados similares. Porém, H_{diff} poderá ter a diferença igual a 0 somente se as CCI forem idênticas para θ_j , o que pode indicar maior confiança, pois SL_{diff} pode ter a diferença entre as curvas características do teste igual a 0 mesmo que sejam diferentes. O método SL_{diff} pode ser preferido teoricamente devido ao fato de trabalhar com a diferença existente entre as curvas características do teste, porém H_{diff} parece ser mais rigoroso.

Os resultados para $Hdiff(\theta_i = 0)$ e $SLdiff(\theta_i = 0)$ são menores para o método Média/Desvio se comparado ao método Média/Média, indicando assim que o primeiro método fornece resultados mais estáveis. Porém, os critérios tem o objetivo de determinar A e B que minimize-os, utilizando como valores iniciais os obtidos pelos métodos.

Uma limitação desses métodos se encontra no fato de não se ter uma explicação clara do erro na estimação dos parâmetros do item.

Os métodos de Stocking e Lord e Média/Média, na maioria das comparações, definem resultados muito similares. O método de aproximação de Haebara pode ser mais difícil implementar, apesar de produzir uma função simétrica. Quando temos problemas com a estimação dos parâmetros na TRI, o método de aproximação de Stocking e Lord parece fornecer melhores resultados.

O mais indicado na análise é a implementação de todos os métodos com o objetivo de termos comparação. Também, o uso de métodos gráficos auxilia na eliminação de itens. Quando itens são retirados por serem considerados "ruins" e, quando obtemos resultados análogos ao utilizarmos os métodos de equalização a posteriori, podemos então considerar os métodos da curva característica.

Na equalização através da TRI não comparamos os escores obtidos em dois testes A_1 e A_2 como era feito na Teoria Clássica, mesmo porque indivíduos que obtiveram os mesmos escores não significa necessariamente que possuem a mesma habilidade θ no modelo logístico, pois diferentes conjuntos de respostas aos itens poderão ter diferentes

habilidades. No entanto, as habilidades estimadas podem ser convertidas em escores, utilizando métodos que trabalham com escores verdadeiros ou observados (Kolen, 1995). Porém, como tais métodos não são muito utilizados na prática, estes não serão tratados no desenvolvimento deste trabalho.

4.3 Equalização via modelos de grupos múltiplos

A TRI consegue satisfazer o que na Teoria Clássica seria bastante trabalhoso. O chamado **Modelo para Grupos Múltiplos** trabalha com a estimação dos parâmetros dos itens em testes distintos constituídas também de itens comuns e de forma a englobar as várias populações em estudo, de modo que se encontrem já calibradas e determinadas em uma mesma escala. Uma generalização do ML3 é feita para o caso de grupos múltiplos. A grande vantagem desse método é que a equalização é realizada no processo de estimação, não necessitando posteriormente de uma transformação de escala como foi feito anteriormente. O modelo para grupos múltiplos é dado por:

$$P(U_{ijk} = 1 | \theta_{jk}) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta_{jk} - b_i)}}, \quad (4.14)$$

onde $i = 1, 2, \dots, I$ corresponde aos itens, $j = 1, 2, \dots, n_k$ correspondente aos examinados e $k = 1, 2, \dots, K$ relativo às populações independentes envolvidas.

Os parâmetros a_i , b_i e c_i , e a constante D já foram apresentados e discutidos anteriormente. Segue também as suposições de independência local, considerando que as respostas dos examinados são independentes. Sejam:

$P(U_{ijk} = 1 | \theta_{jk})$ se refere à probabilidade do examinado j da população k com habilidade θ responder corretamente o item i .

U_{ijk} correspondente à variável dicotômica 0 e 1. Assume valor 1 quando o examinado j da população k responde corretamente o item i , e assume valor 0 caso contrário;

θ_{jk} correspondente à habilidade do examinado j pertencente à população k .

Para grupos múltiplos os examinados das várias populações não são submetidos exatamente aos mesmos itens comuns, porém, a nível de comparação existe a necessidade de que sejam comuns a pelo menos duas populações. Dessa forma, a quantidade total de itens é representada por I .

Na população k consideramos que a distribuição das habilidades dos examinados segue uma distribuição normal com $\eta_k = (\mu_k, \sigma_k^2)'$, representando especificamente sua média e sua variância. Porém esse processo está sujeito a erros, mas existe a vantagem de se trabalhar com menos itens comuns comparado a outros métodos.

Para o conjunto dos K testes definiremos por $\varsigma = (\varsigma_1, \dots, \varsigma_I)$ a ordenação dos I itens, $\mathbf{I}_k$ o conjunto dos índices dos itens dados ao grupo k , I_k a quantidade de itens no teste k , n_k a quantidade de examinados do grupo k , n o número total de examinados, $U_{.jk} = (U_{1jk}, U_{2jk}, \dots, U_{I_kjk})$ o vetor aleatório de respostas do examinado j do grupo k , $U_{..k} = (U_{.1k}, U_{.2k}, \dots, U_{.n_kk})$ o vetor aleatório de respostas do grupo k e $U_{...} = (U_{.1}, U_{.2}, \dots, U_{.n})$ o vetor total de respostas. Dadas as suposições, temos a probabilidade relativa ao vetor de respostas dada por:

$$P(u_{.jk} | \theta_{jk}, \varsigma) = \prod_{i \in \mathbf{I}_k} P(u_{ijk} | \theta_{jk}, \varsigma_i). \quad (4.15)$$

Pela relação na expressão (4.15) definimos os parâmetros de interesse pela utilização dos métodos de máxima verossimilhança marginal ou bayesiano, lembrando que quando estamos tratando de várias populações o novo conjunto de parâmetros a ser estimado é $\eta = (\eta_1, \dots, \eta_k)$. Porém existe a necessidade de se fixar os parâmetros para uma das populações de modo a estabelecer uma métrica. Portanto, se fixarmos $\mu_1 = 0$ e $\sigma_1 = 1$ devemos estimar $\eta_2, \dots, \eta_k$, os quais se encontrarão já transformados para essa métrica inicial, de modo que não existe a necessidade da equalização na TRI da mesma forma que foi feito com o programa computacional Bilog. Andrade et al(2000) apresenta as demonstrações em detalhes.

4.2.2.1 Parâmetros dos itens

Podemos considerar os parâmetros populacionais conhecidos, cuja escala se encontra definida, desde que tenham sido estimados em outra análise, não havendo a necessidade da reestimação. Assim, temos a probabilidade marginal de $U_{.jk}$ dada por:

$$\begin{aligned}
P(u_{.jk}|\varsigma, \eta_k) &= \int_{\mathbb{R}} P(u_{.jk}|\theta, \varsigma, \eta_k) g(\theta|\eta_k) d\theta \\
&= \int_{\mathbb{R}} P(u_{.jk}|\theta, \varsigma) g(\theta|\eta_k) d\theta,
\end{aligned}$$

não sendo função de η_k na segunda igualdade.

Dada a independência entre as respostas de todos os examinados, a probabilidade relativa ao vetor de respostas $U...$ é dada por:

$$P(u_{...}|\varsigma, \eta) = \prod_{k=1}^K \prod_{j=1}^{n_k} P(u_{.jk}|\varsigma, \eta_k).$$

Consideramos aqui o que chamamos de *Padrões de Resposta*, isto é, duas possibilidades de resposta para cada item de um teste com I_k itens e classificados como certo ou errado. Chamamos r_{jk} o número de ocorrências distintas do padrão de resposta j no grupo k , sendo $s_k \leq \min(\eta_k, S_k)$ o número de padrões de respostas com $r_{jk} > 0$ onde $S_k = 2^{I_k}$. Assim:

$$\sum_{j=1}^{s_k} r_{jk} = n_k.$$

Considerando novamente a independência entre as respostas de todos os examinados, temos a distribuição da verossimilhança dada pelo seguinte produto multinomial:

$$L(\varsigma, \eta) = \prod_{k=1}^K \left\{ \frac{n_k!}{\prod_{j=1}^{s_k} r_{jk}!} \prod_{j=1}^{s_k} [P(u_{.jk}|\varsigma, \eta_k)]^{r_{jk}} \right\}.$$

Dessa maneira, a log-verossimilhança será:

$$\log L(\varsigma, \eta) = \sum_{k=1}^K \log \left\{ \frac{n_k!}{\prod_{j=1}^{s_k} r_{jk}!} \right\} + \sum_{k=1}^K \sum_{j=1}^{s_k} r_{jk} \log P(u_{.jk}|\varsigma, \eta_k). \quad (4.16)$$

Igualando a zero as derivadas da log-verossimilhança em função de cada item determinaremos a estimação de seus parâmetros. Consideramos $Q_i = 1 - P_i$ sendo P_i

a probabilidade de resposta ao item i , a ponderação $W_i = \frac{P_i^* Q_i^*}{P_i Q_i}$ com $Q_i^* = 1 - P_i^*$ e $P_i^* = \{1 + e^{-Da_i(\theta - b_i)}\}^{-1}$. As estimativas são dadas a seguir:

$$\frac{\partial \log L(\varsigma, \eta)}{\partial \varsigma_i} = \sum_{k=1}^K \sum_{j=1}^{s_k} r_{jk} \int_{\mathbb{R}} \left[(u_{ijk} - P_i) \left(\frac{\partial P_i}{\partial \varsigma_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_{jk}^*(\theta) d\theta,$$

onde

$$g_{jk}^*(\theta) \equiv g(\theta | u_{jk}, \varsigma, \eta_k) = \frac{P(u_{jk} | \theta, \varsigma) g(\theta | \eta_k)}{P(u_{jk} | \varsigma, \eta_k)},$$

de tal forma que $\varsigma_i = (a_i, b_i, c_i)'$. Assim, para cada parâmetro, temos as seguintes equações:

$$\begin{aligned} \frac{\partial \log L(\varsigma, \eta)}{\partial a_i} &= D(1 - c_i) \sum_{k=1}^K \sum_{j=1}^{s_k} r_{jk} \int_{\mathbb{R}} [(u_{ijk} - P_i)(\theta - b_i) W_i] g_{jk}^*(\theta) d\theta, \\ \frac{\partial \log L(\varsigma, \eta)}{\partial b_i} &= -Da_i(1 - c_i) \sum_{k=1}^K \sum_{j=1}^{s_k} r_{jk} \int_{\mathbb{R}} [(u_{ijk} - P_i) W_i] g_{jk}^*(\theta) d\theta, e \\ \frac{\partial \log L(\varsigma, \eta)}{\partial c_i} &= \sum_{k=1}^K \sum_{j=1}^{s_k} r_{jk} \int_{\mathbb{R}} \left[(u_{ijk} - P_i) \frac{W_i}{P_i^*} \right] g_{jk}^*(\theta) d\theta. \end{aligned}$$

4.2.2.1 Parâmetros populacionais

Considerando os parâmetros dos itens conhecidos, a log-verossimilhança em (4.16), e a métrica $\mu_1 = 0$ e $\sigma_1 = 1$, igualamos a zero as derivadas da log-verossimilhança em função da média e variância das populações para $k = 2, \dots, K$. Porém:

$$\frac{\partial \log L(\varsigma, \eta)}{\partial \eta_k} = \sum_{j=1}^{s_k} r_{jk} \int_{\mathbb{R}} \left(\frac{\partial \log g(\theta | \eta_k)}{\partial \eta_k} \right) g_{jk}^*(\theta) d\theta.$$

Utilizando-se a distribuição normal com média μ_k e variância σ_k^2 para o conjunto das habilidades θ_k tem-se que:

$$\begin{aligned} \frac{\partial \log g(\theta | \eta_k)}{\partial \mu_k} &= \frac{\theta - \mu_k}{\sigma_k^2}, e \\ \frac{\partial \log g(\theta | \eta_k)}{\partial \sigma_k^2} &= -\frac{\sigma_k^2 - (\theta - \mu_k)^2}{\sigma_k^4}. \end{aligned}$$

Sendo assim, temos as seguintes equações:

$$\begin{aligned} (\sigma_k^2)^{-1} \sum_{j=1}^{s_k} r_{jk} \int_{\mathbb{R}} (\theta - \mu_k) g_{jk}^*(\theta) d\theta & \quad \text{para a média de cada grupo,} \\ -(2\sigma_k^4)^{-1} \sum_{j=1}^{s_k} r_{jk} \int_{\mathbb{R}} [\sigma_k^2 - (\theta - \mu_k)^2] g_{jk}^*(\theta) d\theta & \quad \text{para a variância de cada grupo.} \end{aligned}$$

A estimação conjunta é realizada através da aplicação do algoritmo EM. A estimação das habilidades pode ser feita por máxima verossimilhança marginal utilizando métodos iterativos. A estimação dos parâmetros dos itens e das habilidades também pode ser feita utilizando o método bayesiano (Andrade et al, 2000).

Em Andrade(2001) é verificado que o método de grupos múltiplos possui superioridade em relação ao método de transformação de escala Média/Desvio, apesar de apresentarem desempenhos bastante similares. Um exemplo de aplicação desse método pode ser verificado em Andrade(2000), no qual foram utilizados dados do SARESP (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo).

O pacote computacional BILOG-MG trabalha com a equalização para múltiplos grupos, não necessitando assim da transformação de escala, enquanto o pacote computacional BILOG necessita da utilização da transformação.

Capítulo 5

Recursos computacionais e aplicação dos métodos

5.1 Recursos computacionais

Para a realização de todos os tipos de equalização contidos nesse trabalho foram utilizados vários pacotes computacionais, os quais são listados a seguir.

RAGE Este programa realiza as equalizações linear e do equipercartil na Teoria Clássica para grupos equivalentes, e faz a implementação do método de suavização a posteriori. O programa foi criado por L. Zeng, M. J. Kolen & B. A. Hanson, e fornecido por Kolen.

(<http://www.education.uiowa.edu/casma/computer-programs.htm>)

CIPE Este programa faz a implementação da Teoria Clássica para grupos não equivalentes cujos testes possuem itens comuns, apresentando as equalizações dos métodos lineares e dos métodos do equipercartil apresentados (Método de Tucker, Método do escore observado de Levine, Método de estimação de frequência e Método linear de Braun Holland), utilizando a suavização a posteriori. O programa foi criado por M. J. Kolen, e fornecido pelo mesmo.

(<http://www.education.uiowa.edu/casma/computer-programs.htm>)

BILOG Este programa realiza todo o processo de calibração dos itens de determinados testes quando utiliza-se a TRI (Teoria da Resposta ao Item) a fim dos resultados serem utilizados na equalização a posteriori.

Bilog-MG Este programa realiza o processo de calibração dos parâmetros dos itens e o processo de equalização colocando as várias populações envolvidas numa única

métrica.

R 2.1.1 Este programa foi utilizado para a construção das figuras inseridas neste trabalho.

Uma característica dos programas utilizados na equalização da Teoria Clássica está no fato de serem válidos somente quando tratamos de testes que possuam a mesma quantidade total de itens em cada teste (RAGE e CIPE), o que não acontece quando trabalhamos com a TRI, pois para o pacote computacional Bilog a quantidade total de itens dos testes não faz diferença. O trabalho de mestrado de Pinheiro(2006) faz as implementações de vários métodos de equalização para grupos não equivalentes onde, mesmo na Teoria clássica, admite que os testes tenham comprimentos diferentes.

O trabalho recente de Pinheiro(2006) apresenta programas que foram desenvolvidos em conjunto com um aluno de mestrado em ciência da computação do INE-UFSC criando um ambiente de software que contém todos os programas criados para a equalização na teoria clássica e na TRI para grupos não equivalentes, e que se encontram disponíveis no endereço <http://www.inf.ufsc.br/~dandrade/Avaliacao-Educacional>.

5.2 Aplicação dos métodos

A aplicação será realizada com base em um conjunto de dados reais, os quais foram obtidos do SARESP – Sistema de avaliação de Rendimento Escolar do Estado de São Paulo implantado pela SEE/SP – Secretaria de Estado da Educação de São Paulo. Em abril de 1996, foi aplicado um teste contendo 28 itens de 2ª série de Língua Portuguesa em alunos da 3.ª série e, em abril de 1997, foi aplicado um teste de 30 itens de 3ª série também de Língua Portuguesa para a 4.ª série. Porém, como ambos testes não possuíam itens comuns, foi aplicado, em novembro de 1997, para uma amostra de alunos da 3.ª série, um teste contendo 11 itens comuns ao teste dado para a 3.ª série em 1996, e com 21 itens comuns ao teste aplicado na 4.ª série em 1997. Assim, o teste de ligação era composto de 32 itens. A aplicação deste capítulo será realizada com base nesse teste de ligação dado em 1997, e o teste dado aos alunos da 4.ª série em 1997, trabalhando, portanto, com grupos não equivalentes. Os itens se encontram distribuídos nos testes como apresentado na Tabela 5.1. O teste da 3.ª série foi respondido por 640 indivíduos, e o teste da 4.ª série

por 1989 indivíduos.

Tabela 5.1: Distribuição e posição dos itens nos testes da 3.^a e 4.^a séries

Item	4. ^a série	3. ^a série	Item	4. ^a série	3. ^a série
1	1	29	17	17	27
2	2	23	18	18	39
3	3	31	19	19	40
4	4	32	20	20	41
5	5	33	21	21	17
6	6	34	22	22	18
7	7	35	23	23	19
8	8	36	24	24	20
9	9	24	25	25	21
10	10	11	26	26	1
11	11	13	27	27	3
12	12	14	28	28	6
13	13	37	29	29	7
14	14	38	30	30	28
15	15	25	31	—	15
16	16	26	32	—	16

5.2.1 Equalização pela Teoria Clássica

Os resultados referentes à aplicação da Teoria Clássica para grupos não equivalentes foram obtidos utilizando o programa computacional CIPE (Kolen,1995), o qual realiza os cálculos para os métodos apresentados no Capítulo 3, utilizando, para o método de equipercentil de estimação de frequência, a suavização a posteriori. Um problema com este programa ocorreu, no entanto, pelo fato de poder ser utilizado somente no caso em que os testes tenham a mesma quantidade de itens, o que não ocorre no nosso caso. Para aplicação dos métodos da Teoria Clássica não tive acesso a nenhum outro programa computacional que trabalhasse com testes que tivessem comprimentos diferentes (Pinheiro,2006). Assim, para podermos utilizá-lo, é necessário retirarmos dois itens não comuns no teste de ligação da 3.^a série. O critério para a eliminação dos itens foi realizado observando-se os itens não comuns que possuísem menor correlação bisserial, cujos resultados foram obtidos da fase 1 do programa computacional Bilog (Andrade,2000) utilizado na TRI. Os itens que possuem menor correlação bisserial são os itens 37 (0,349) e 38

(0,290).

Dessa forma, os resultados correspondentes à frequência relativa de ambos grupos são considerados como tendo 30 itens cada, e dados pela tabelas 5.2 e 5.3 sobre o escore total e sobre os itens comuns, respectivamente.

Tabela 5.2: Frequência relativa dos examinados para cada escore da 3.^a e 4.^a séries.

Escore	3. ^a série	4. ^a série
0	0,0000	0,0000
1	0,0000	0,0005
2	0,0000	0,0005
3	0,0016	0,0005
4	0,0031	0,0020
5	0,0031	0,0045
6	0,0078	0,0075
7	0,0156	0,0186
8	0,0188	0,0251
9	0,0234	0,0287
10	0,0313	0,0392
11	0,0375	0,0417
12	0,0688	0,0382
13	0,0672	0,0463
14	0,0594	0,0598
15	0,0672	0,0618
16	0,0609	0,0603
17	0,0719	0,0734
18	0,0531	0,0618
19	0,0563	0,0548
20	0,0516	0,0694
21	0,0641	0,0588
22	0,0734	0,0558
23	0,0438	0,0543
24	0,0313	0,0357
25	0,0344	0,0377
26	0,0234	0,0246
27	0,0078	0,0191
28	0,0156	0,0116
29	0,0047	0,0060
30	0,0031	0,0015

Tabela 5.3: Frequência relativa aos itens comuns para os testes da 3.^a e 4.^a séries.

Escore	3. ^a série	4. ^a série
0	0,0031	0,0000
1	0,0047	0,0015
2	0,0141	0,0070
3	0,0219	0,0161
4	0,0375	0,0261
5	0,0688	0,0427
6	0,0750	0,0563
7	0,0797	0,0759
8	0,0781	0,0754
9	0,0891	0,0825
10	0,0859	0,0825
11	0,0750	0,0804
12	0,0734	0,0684
13	0,0672	0,0794
14	0,0703	0,0709
15	0,0438	0,0689
16	0,0469	0,0573
17	0,0281	0,0483
18	0,0109	0,0282
19	0,0156	0,0191
20	0,0078	0,0085
21	0,0031	0,0045

Seja X a variável aleatória dos escores do teste de ligação aplicado para a 3.^a série, x uma realização de X , Y a variável aleatória dos escores do teste aplicado para a 4.^a série e y uma realização de Y . O intuito é colocar a escala de desempenho da 3.^a série na escala de desempenho da 4.^a série. Portanto, para o método linear para grupos não equivalentes determinaremos a equalização linear $l_Y(x)$, e pelo método do equipercentil da frequência estimada encontraremos $e_Y(x)$.

Precisamos testar as hipóteses dos métodos lineares da Teoria Clássica para verificar qual será mais adequado para a análise. Podemos, para uma possível comparação, fazer uma regressão dos escores nos testes sobre os escores dos itens comuns.

As regressões nas figuras 5.1 e 5.2 demonstram produzir relações lineares, dando indícios de que as hipóteses sobre o método linear de Tucker, apresentadas no Capítulo 3, pode ser considerado. Através das medidas resumo dos testes na Tabela 5.4, observamos que as correlações entre cada teste e seus itens comuns se encontram bastante próximas de 1, justificando também que a hipótese de igualdade das correlações sobre o método linear

Figura 5.1: Gráfico de dispersão dos escores do teste da 3.^a série e dos escores dos itens comuns.

de escores observados de Levine, mostrado no Capítulo 3, também pode ser considerado. E portanto, o método linear de Braun-Holland descrito no Capítulo 3 também é válido. Dessa forma, na Teoria Clássica, mostraremos os resultados dos vários métodos lineares e o método de equipercantil através da estimação de frequência.

Tabela 5.4: Medidas resumo dos testes

Grupo	Teste	μ	σ	<i>ass</i>	<i>curt</i>	<i>Cov</i>	<i>Corr</i>
3. ^a série	X	17,1297	5,2930	0,0037	2,3920	21,3810	0,9675
	V	10,0500	4,1752	0,1781	2,4331		
4. ^a série	Y	17,3112	5,4435	-0,1022	2,3440	21,7466	0,9598
	V	11,0583	4,1622	0,0559	2,2539		

Apresentamos na Figura 5.3 o comportamento da 3.^a e 4.^a séries em relação a seu respectivo teste através da frequência relativa de cada um. Percebemos que o desempenho de ambos grupos são bastante parecidos, porém a curva da 4.^a série, como se encontra um

Figura 5.2: Gráfico de dispersão dos escores do teste da 4.^a série e dos escores dos itens comuns.

pouco deslocada para a direita, tende a obter escores um pouco maiores do que o grupo da 3.^a série.

Utilizando as medidas resumo da Tabela 5.4, e os resultados obtidos pela saída do programa CIPE, pelo método linear de Tucker obtemos a equalização dada por $l_Y(x) = -1,6216 + 1,0314x$, onde $l_Y(x)$ transforma os escores da escala da 3.^a série para a escala da 4.^a série. Conforme o método linear do escore observado de Levine temos $l_Y(x) = -1,7387 + 1,0319x$, e pelo método linear de Braun-Holland definimos a equalização linear como sendo $l_Y(x) = -1,7787 + 1,0398x$.

Na Figura 5.4 observando o comportamento das equalizações quando fazemos o escore da 4.^a série equalizado menos o escore da 3.^a série, temos as equalizações lineares bastante próximas, indicando que não devem fornecer resultados muito discrepantes. Esses resultados das equalizações lineares, para alguns escores se encontra dentro das bandas de confiança obtida pela equalização do equipercil da frequência estimada,

Figura 5.3: Frequência relativa dos escores obtidos pelos testes aplicados a 3.^a e 4.^a séries.

não sofrendo tantas variações. Isso parece indicar que, nesse caso, todos os métodos são relativamente bons nas transformações. A equalização suavizada a posteriori com $S=0,10$ se encontra fora das bandas de confiança somente para escores abaixo de 5.

Em geral, todos os métodos apresentaram resultados bastante aproximados, indicando que possuem confiabilidade nos resultados. Os resultados referentes aos quatro primeiros momentos para cada método de equalização aplicado se encontram na Tabela 5.5. Para todos os métodos, o desempenho médio da 3.^a série na escala do teste da 4.^a série se encontra próximo de 16, com um desvio aproximado de 5,5. O grupo possui uma fraca assimetria positiva, e curtose abaixo da normal.

5.2.2 Equalização pela TRI

Utilizando esse mesmo conjunto de dados simulados podemos verificar como ocorre a equalização através da TRI através do ML3 (Modelo Logístico Unidimensional de 3 parâmetros).

Figura 5.4: Y equalizado menos o escore do teste X comparando as equalizações lineares e da frequência estimada.

Tabela 5.5: Momentos dos escores da 3.^a série convertidos para a escala dos escores da 4.^a série.

Método	μ	σ	<i>ass</i>	<i>curt</i>
Tucker	16,0455	5,4591	0,0037	2,3920
Levine Observado	15,9373	5,4618	0,0037	2,3920
Braun-Holland	16,0325	5,5036	0,0037	2,3920
Equipercartil				
Não suavizado	16,0325	5,5036	0,0114	2,4154
S=0,10	16,0430	5,4926	0,0337	2,3735

5.2.2.1 Equalização a posteriori

Na TRI, após a estimação dos parâmetros, será utilizado o método Média/Desvio para a transformação de escala dos grupos. Apresentamos na Tabela 5.6 os parâmetros estimados dos itens, utilizando o programa computacional Bilog, considerando os mesmos testes aplicados à 3.^a e 4.^a séries, respectivamente, da mesma forma como foi

descrito anteriormente. Os parâmetros foram estimados através do Método de Máxima Verossimilhança Marginal(MVM), onde todos esses parâmetros obtidos se encontram na escala de média 0 e desvio 1, de cada grupo. Os testes possuem os mesmos itens comuns analisados na Teoria Clássica. No entanto, na TRI não é necessário que os testes tenham a mesma quantidade de itens, portanto, o teste da 3.^a série estará composta por 32 itens e o teste da 4.^a série por 30 itens. Podemos visualizar na Tabela 5.6 os resultados obtidos com a calibração com relação aos itens comuns em cada teste.

Faremos a transformação dos parâmetros do grupo da 4.^a série para a métrica do grupo da 3.^a série utilizando o método Média/Desvio descrito no capítulo anterior. Pela Tabela 5.7 temos as medidas resumo dos itens comuns em ambos grupos em função do parâmetro de dificuldade do item. Podemos verificar que a média do parâmetro de dificuldade para a 4.^a série é menor em relação à média da 3.^a série, indicando que os itens comuns em ambos testes foram, em média, mais fáceis para a 4.^a série do que para a 3.^a série.

Figura 5.5: Parâmetro de discriminação dos itens comuns entre os testes da 3.^a e 4.^a séries.

Tabela 5.6: Estimação dos parâmetros dos itens para a 3.^a e 4.^a séries.

Item	3. ^a série			4. ^a série		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
1	1,1443	-0,4586	0,0000	1,3098	-1,1529	0,0000
2	—	—	—	1,3323	-1,4565	0,0000
3	1,6281	-0,1577	0,1204	1,1673	-0,8570	0,0974
4	—	—	—	0,3586	-3,5971	0,0003
5	—	—	—	0,5809	-0,5625	0,0004
6	1,0667	-0,0260	0,0000	1,0022	-0,8924	0,0000
7	1,6145	0,4401	0,0448	1,6281	0,2662	0,1232
8	—	—	—	0,9927	-0,6667	0,1253
9	—	—	—	0,5816	-0,9773	0,0004
10	—	—	—	0,6082	-1,1038	0,0005
11	1,3000	-0,2290	0,1717	1,3044	-0,6183	0,1319
12	—	—	—	1,5271	-1,5274	0,0000
13	0,7957	-0,2588	0,0000	1,6331	-0,0856	0,2317
14	1,9913	1,6842	0,1279	1,4499	1,6739	0,1142
15	2,6580	1,2603	0,1238	2,0635	1,0979	0,1005
16	1,4890	0,4013	0,1152	1,8682	0,1196	0,2093
17	2,8089	1,7081	0,1248	0,7671	1,8123	0,0552
18	1,4333	-0,5270	0,3744	1,0739	-0,7442	0,3437
19	1,2834	-0,1715	0,1664	1,7066	0,0004	0,3804
20	0,3380	-1,1447	0,0002	0,3956	-1,5855	0,0004
21	3,4957	1,3494	0,3156	1,7505	1,0300	0,2509
22	—	—	—	0,8925	-0,0294	0,1240
23	0,7926	-0,7967	0,1769	1,7786	-0,0108	0,2569
24	0,5446	0,9213	0,1476	0,8900	0,5131	0,1775
25	0,8642	-0,1450	0,0000	1,0557	-0,1288	0,0182
26	1,4064	-0,2327	0,0000	2,3203	0,1680	0,2304
27	1,6961	1,0417	0,1756	2,2046	1,6452	0,2576
28	2,0979	1,7145	0,1177	1,9225	1,2301	0,1358
29	1,8707	1,6934	0,1865	1,4628	1,8603	0,1589
30	—	—	—	0,9790	-0,3553	0,3015
31	1,4860	-0,9554	0,5000	—	—	—
32	1,7965	-0,3779	0,2486	—	—	—
33	1,2211	-0,0924	0,0317	—	—	—
34	0,9917	-3,6248	0,0000	—	—	—
35	1,0710	-2,7053	0,5000	—	—	—
36	1,0396	-2,7542	0,5000	—	—	—
37	0,6025	-0,0906	0,0000	—	—	—
38	0,5456	0,1216	0,0452	—	—	—
39	1,9575	0,8789	0,2863	—	—	—
40	1,0414	-1,4942	0,0000	—	—	—
41	1,4354	-1,1074	0,0000	—	—	—

Tabela 5.7: Medidas Resumo dos parâmetros de dificuldade dos itens para a 3.^a e 4.^a séries.

Momentos	3. ^a série	4. ^a série
Média	0,3841	0,2544
Desvio	0,9177	1,0341

Figura 5.6: Parâmetro de dificuldade dos itens comuns entre os testes da 3.^a e 4.^a séries.

Na Figura 5.5 podemos verificar a relação entre os parâmetros de discriminação dos itens comuns aos testes da 3.^a e 4.^a séries. Os pontos se encontram um pouco dispersos, não apresentando uma relação linear entre os dois grupos. A Figura 5.6 apresenta a relação entre os parâmetros de dificuldade dos dois testes, o qual parece fornecer uma relação linear entre ambos. Em geral, o parâmetro de dificuldade do teste da 3.^a série é maior em relação à 4.^a série, a não ser para parâmetros de dificuldade mais altos.

Para definirmos a transformação dos parâmetros da 4.^a série para a métrica da 3.^a série, necessitamos, para utilizar o método Média/Desvio, determinarmos os valores de A e B. Assim, utilizando os resultados da Tabela 5.7, temos:

$$A = \frac{\sigma_{3a}}{\sigma_{4a}} = \frac{0,9177}{1,0341} = 0,8875, \quad e$$

$$B = \mu_{3a} - A \cdot \mu_{4a} = 0,3841 - 0,8875 \cdot (0,2544) = 0,1584.$$

Dessa forma, temos as seguintes relações de transformação dos parâmetros:

$$b_{3a} = A \cdot b_{4a} + B = 0,8875 \cdot b_{4a} + 0,1584,$$

$$a_{3a} = \frac{a_{4a}}{A} = \frac{a_{4a}}{0,8875}, \quad e$$

$$\theta_{3a} = A \cdot \theta_{4a} + B = 0,8875 \cdot \theta_{4a} + 0,1584.$$

Ao transformarmos uma habilidade $\theta_{4a} = 1,2520$, por exemplo, para a métrica da 3ª série, calculamos θ_{3a} pela relação dada acima, onde $\theta_{3a} = 0,8875 \cdot \theta_{4a} + 0,1584 = 0,8875 \cdot (1,2520) + 0,1584 = 1,2696$. Desta forma, qualquer parâmetro pode ser transformado na métrica de outro grupo utilizando o método Média/Desvio.

Podemos observar as estimativas das habilidades dos grupos através das figuras 5.7 e 5.8. Comparando os histogramas das habilidades estimadas vemos que a habilidade média da 4ª série equalizada para a escala da 3ª é um pouco maior com relação à habilidade média da 3ª série o que pode ser verificado também nas transformações da métrica realizadas anteriormente. Pela Tabela 5.8 vemos que a habilidade média da 3ª série é menor se comparada à habilidade média da 4ª série.

Tabela 5.8: Medidas Resumo das habilidades dos respondentes da 3.ª e 4.ª séries.

Momentos	3.ª série	4.ª série
Média	0	0,1113
Desvio	1	0,8952

Figura 5.7: Proficiência estimada da 3ª série.

5.2.2.2 Equalização via grupos múltiplos

O programa Bilog-MG realiza esse processo diretamente sem a necessidade de se fazer algum método de transformação de escala, colocando assim todos os grupos na mesma métrica de uma só vez.

Comparando os valores médios dos parâmetros de discriminação e de dificuldade dos itens comuns obtidos pelas equalizações, através dos programas computacionais Bilog-MG e Bilog utilizando o método de transformação Média/Desvio, transformando os parâmetros da 3ª série para a métrica da 4ª série, podemos verificar pela Tabela 5.9 que os valores encontrados são bastante aproximados.

Apresentamos na Tabela 5.10 os resultados obtidos com a calibração utilizando o método de grupos múltiplos. Os resultados foram obtidos utilizando o grupo da 4ª série como referência, e numa métrica com média 0 e desvio 1.

Os resultados provenientes dos programas computacionais utilizados neste capítu-

Habilidade da 4a. série equalizada para a escala da 3a. série

Figura 5.8: Proficiência estimada da 4ª série.

Tabela 5.9: Valores médios dos parâmetros dos itens na transformação de escala da 4.ª série para a métrica da 3.ª série utilizando os programas computacionais Bilog e Bilog-MG.

Momentos	Parâmetro a		Parâmetro b	
	Bilog	Bilog-MG	Bilog	Bilog-MG
Média	1,3658	1,8411	0,2544	0,3841
Desvio	0,6842	0,5975	1,0341	0,7239

lo de aplicação se encontram no Apêndice. Os programas computacionais disponíveis na Teoria Clássica trabalham com grupos que tenham realizado testes contendo a mesma quantidade de itens, o que nos fez admitir um critério de exclusão de itens na prova da 3ª série. No entanto, existe o trabalho em desenvolvimento de Pinheiro(2006) na implementação desses métodos de equalização na Teoria Clássica para grupos não equivalentes, que permite que os testes possuam quantidades diferentes de itens, e além disso, desenvolvidos em software livre.

Tabela 5.10: Estimação dos parâmetros dos itens para a 3.^a e 4.^a séries utilizando o programa computacional Bilog-MG.

Item	<i>a</i>	<i>b</i>	<i>c</i>
1	1,5271	-0,6479	0,1696
2	1,4805	-1,0860	0,1853
3	1,5312	-0,3786	0,2292
4	0,4494	-2,0708	0,2416
5	0,9286	0,3634	0,2517
6	1,2483	-0,3938	0,1251
7	2,0358	0,3378	0,1363
8	1,3740	-0,0980	0,3003
9	0,8210	-0,0661	0,2354
10	0,8647	-0,1681	0,2473
11	1,6323	-0,2411	0,2614
12	1,6124	-1,2470	0,1662
13	1,7501	0,0782	0,2595
14	2,0518	1,3912	0,1254
15	2,9716	0,9682	0,1137
16	2,0857	0,2120	0,2062
17	1,4956	1,5201	0,1217
18	1,2173	-0,5824	0,3906
19	1,8640	0,0292	0,3568
20	0,5063	-0,4418	0,2006
21	2,9223	0,9782	0,2892
22	1,2593	0,3393	0,2410
23	2,1942	0,2128	0,3751
24	1,2670	0,8018	0,2820
25	1,2599	0,1305	0,1335
26	2,3344	0,1566	0,2198
27	2,1669	1,3075	0,2397
28	2,5237	1,1273	0,1350
29	2,0774	1,5008	0,1738
30	1,1442	-0,1272	0,3475
31	1,0493	-2,0732	0,2802
32	1,6750	-0,5985	0,2863
33	1,3992	-0,0723	0,1645
34	1,0679	-3,6127	0,2216
35	1,7285	-3,2705	0,2132
36	1,7624	-3,5236	0,2173
37	0,8840	0,2767	0,1983
38	1,0686	0,7942	0,3109
39	2,2281	0,5965	0,2882
40	1,1317	-1,3171	0,2545
41	1,4859	-1,1664	0,1842

Conclusão

A equalização perfeita, provavelmente, na prática, nunca deverá ocorrer. Dessa forma, devemos sempre analisar os métodos de equalização e verificar o que melhor se ajusta quando desejamos comparar testes, ou o desempenho de grupos de indivíduos. A escolha do melhor modelo de equalização deve levar em consideração problemas que podem ocorrer na aplicação ou no desenvolvimento dos testes, e principalmente, se as suposições estatísticas estão satisfeitas.

Em geral, *métodos lineares* são considerados quando trabalhamos com amostras pequenas, e quando não existe muita diferença entre os testes (Kolen, 1995). Além disso, esses métodos são de fácil conversão, e suas análises também são relativamente simples de compreender.

Em contrapartida, os *métodos do equipercentil* são mais razoáveis quando temos amostras grandes, e principalmente, no caso de grupos não equivalentes. Tais métodos tem a vantagem de que os resultados das equalizações variam sobre toda a escala do escore. Porém, esses métodos são mais complexos que os métodos lineares, tanto nas tabelas, como nas equações e nas análises.

Os *métodos da TRI*, utilizando o ML3, são adequados quando temos pequenas ou grandes amostras, e também quando os testes diferem ou não quanto ao nível de dificuldade. Existe, no entanto, uma certa complexidade nas tabelas de conversão, na estimação dos parâmetros e na análise, porém, as suposições se encontram razoavelmente bem garantidas nesse modelo.

Porém, para que possamos realizar as equalizações, independentemente do modelo, é necessário que os testes sejam formulados com o intuito das mesmas especificações, passando assim, por um controle de qualidade, onde são verificadas as condições de administração do teste, se as respostas são corretamente especificadas, se os itens não possuem duplo sentido, o procedimento de equalização especificado, a consistência das distribuições e estatísticas do escore e a tabela correta de conversão.

Além disso, nos testes que possuem itens comuns, estes itens devem ser representativos do desempenho dos grupos que os tenham realizado.

Nos grupos não equivalentes com itens comuns, podemos verificar a linearidade entre os escores do teste e os escores sobre os itens comuns. Se for verificada uma relação linear, o método de Tucker deve ser levado em consideração. Se a regressão for não linear, podemos considerar o método de Braun-Holland. Considerando que os grupos sejam similares, podemos trabalhar com os métodos de Tucker e de Estimação de Frequência do equipercentil. Caso contrário, se os grupos se apresentarem muito diferentes, podemos considerar os métodos do escore observado de Levine e o ML3-TRI.

Como sugestão para estudos futuros podemos considerar a aplicação da Teoria Clássica e da TRI tendo em vista os outros grupos de respondentes que aqui não foram discutidos, e possivelmente compará-los com os resultados obtidos pelos grupos equivalentes e não equivalentes. Além disso podemos considerar o trabalho para modelos policotômicos e multidimensionais na TRI pelo fato de agregar uma quantidade maior de informações, e desde que trabalhada em conjunto com outras áreas de forma a colaborar na interpretação dos resultados obtidos, aumentando assim, o entendimento qualitativo e quantitativo da educação, bem como de outras áreas de interesse.

Goldstein(2001) afirma que *"a fim de descrever a complexa realidade que constitui os sistemas educacionais, precisamos de instrumentos de modelagem que envolvam um nível comparado de complexidade"*, e para que isso ocorra é necessário que os estudiosos da área qualitativa e quantitativa trabalhem em conjunto.

Uma dificuldade encontrada na realização deste trabalho é justamente quanto à bibliografia, pois existe pouco material disponível, principalmente atualizado, até a nível nacional sobre o assunto. Inclusive, por esse motivo, surgiu a idéia de compilarmos os variados métodos de equalização neste trabalho.

Referências Bibliográficas

- [1] Andrade, D. F., Tavares, H. R. & Valle, R. C.(2000). *Teoria da Resposta ao Item - Conceitos e Aplicações*. ABE – Associação Brasileira de Estatística. 14º SINAPE, Caxambu.
- [2] Andrade, D. F.(2001) *Comparando desempenho de grupos de alunos por intermédio da Teoria da Resposta ao Item*. Estudos em Avaliação Educacional, 23, 31-69.
- [3] Angoff, W. A.(1971). *Scales, norms and equivalent scores*. In R. L. Thorndike (Ed.), *Educational Measurement* (2ªed., p.508-600). Washington, DC: American Council on Education.
- [4] Azevedo, C. L. N.(2003). *Métodos de Estimação na Teoria da Resposta ao Item*. Dissertação de Mestrado. São Paulo: IME/USP.
- [5] Baker, F. B.(1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.
- [6] Bussab, W. O.(1988). *Análise de variância e de regressão: uma introdução*. 2ª.ed. São Paulo: Atual.
- [7] Goldstein, H.(2001). *Modelos da realidade: Novas abordagens para a compreensão de processos educacionais*. In Crespo Franco, *Avaliação, ciclos e promoção na educação* (p.85-99). Porto Alegre.
- [8] Gulliksen, H.(1950). *Theory of mental tests*. New York: Wiley.
- [9] Haberman, S. J.(1974). *Log-linear models for frequency tables with ordered classifications*. Biometrics, 30, 589-600.

- [10] Hambleton, R. K.(1993). *Principles and selected applications of item response theory*. In: R. L. Linn (Ed.), *Educational Measurement* (3ªed., p.147). American Council on Education.
- [11] Hambleton, R. K., Swaminathan, H. & Rogers, H. J.(1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- [12] Hanson, B. A., Zeng, L. & Colton, D.(1994). *A comparison of presmoothing and postsMOOTHING methods in equipercentile equating*. (ACT Research Report 94-4). Iowa City, IA: American College Testing.
- [13] James, B. A.(1996). *Probabilidade: um curso em nível intermediário*. 2ª ed. Rio de Janeiro: Instituto de Matemática Pura e Aplicada. Projeto Euclides.
- [14] Kolen, M. J. & Brennan, R. L.(1995). *Test Equating: Methods and Practices*. New York.
- [15] Kolen, M. J.(1981). *Comparison of traditional and item response theory methods for equating tests*. Journal of educational measurement, 18, nº1.
- [16] Lord, F. M.(1965). *A strong true score theory with applications*. Psychometrika, 30, p.239-270.
- [17] Lord, F. M.(1980). *Applications of item theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- [18] Lord, F. M. & Novick, M. R.(1968). *Statistical theories of mental test scores*. Resding, MA: Addison Wesley.
- [19] Muraki, E., Hombo, C. M. & Lee, Y. W.(2000). *Equating and Linking of Performance Assessments*. Educational Testing Service. Applied Psychological Measurement, 24, nº4, p.325-337.
- [20] Petersen, N. S., Kolen, M. J. & Hoover, H. D.(1989), *Scaling, Norming and Equating*. In: R. L. Linn (Ed.), *Educational Measurement* (3ªed, p.221-262). New York: Macmillan.
- [21] Pinheiro, C. E.(2006). *Implementação de Métodos Estatísticos para Avaliação Educacional no Software R*. Dissertação de Mestrado. São Paulo: IME-USP.

- [22] Thorndike, R. L.(1971). *Reliability*. In: R. L. Thorndike (Ed.), *Educational Measurement* (2^aed, p.560-620). Washington, DC: American Council on Education.
- [23] Vianna, H. M.(1987). *Testes em educação*. São Paulo: IBRASA.

Apêndice

Saídas dos programas CIPE e BILOG utilizados na aplicação de dados reais.

COMANDOS DE ENTRADA PARA UTILIZAÇÃO DO PROGRAMA CIPE

```

OUT1
mondatx
mondaty
OUT2
OUT3
EQUATE x to y  com it monograph example  w1=1
(30X,I2,I2)
    30    30    21    5    0    1.0000    0.005
      0.01      0.10      0.25      0.50      0.75

```

onde

mondatx corresponde à prova da 3a. série;
mondaty corresponde à prova da 4a. série;
OUT1, OUT2 e OUT3 são os arquivos de saída do programa.

O programa realiza a entrada dos dados em que ambos testes possuem 30 itens, sendo 21 comuns. As 30 primeiras colunas se referem aos 30 itens, cujos dados aparecem já corrigidos colocando 1 para acerto e 0 para erro. As outras duas colunas posteriores contém o escore total de cada respondente, e as últimas duas colunas o escore sobre os itens comuns. Nos métodos que utilizam a população sintética, os pesos das populações são, respectivamente, $w_1=1$ e $w_2=0$. A suavização a posteriori é realizada para 5 valores de S.

20	5	0.0078	0.9969	28.6000	0.4899	29.3338	1.3057	-0.7338
-0.8158								
21	2	0.0031	1.0000	30.0000	0.0000	30.5604	1.3057	-0.5604
-1.3057								

PRED = (1.2265)*V+(4.8029)
 ETASQUARE = 0.9391 SEE FOR ETASQUARE = 1.3057
 RSQUARE = 0.9361 SEE FOR RSQUARE = 1.3383

REGRESSION ANALYSIS GROUP 2 - MEANS AND STANDARD DEVIATIONS OF Y GIVEN V

V	F	RF	RCF	MN	SD	PRED MN	PRED SD	RES MN
RES SD								
0	0	0.0000	0.0000	0.0000	0.0000	3.4298	1.5138	0.0000
0.0000								
1	3	0.0015	0.0015	4.3333	2.4944	4.6851	1.5138	-0.3517
0.9806								
2	14	0.0070	0.0085	5.5000	1.7217	5.9404	1.5138	-0.4404
0.2079								
3	32	0.0161	0.0246	7.4063	1.4111	7.1957	1.5138	0.2106
-0.1027								
4	52	0.0261	0.0508	8.5192	1.8552	8.4510	1.5138	0.0683
0.3414								
5	85	0.0427	0.0935	9.2118	1.7895	9.7062	1.5138	-0.4945
0.2757								
6	112	0.0563	0.1498	10.6696	1.9475	10.9615	1.5138	-0.2919
0.4337								
7	151	0.0759	0.2257	12.1523	1.8655	12.2168	1.5138	-0.0645
0.3517								
8	150	0.0754	0.3012	13.5333	1.8098	13.4721	1.5138	0.0612
0.2960								
9	164	0.0825	0.3836	14.8902	1.5812	14.7274	1.5138	0.1628
0.0674								
10	164	0.0825	0.4661	16.2317	1.4634	15.9827	1.5138	0.2490
-0.0504								
11	160	0.0804	0.5465	17.5125	1.4662	17.2380	1.5138	0.2745
-0.0476								
12	136	0.0684	0.6149	18.5221	1.5290	18.4933	1.5138	0.0288
0.0152								
13	158	0.0794	0.6943	19.8354	1.3956	19.7486	1.5138	0.0869
-0.1182								
14	141	0.0709	0.7652	20.9787	1.2908	21.0039	1.5138	-0.0252
-0.2230								
15	137	0.0689	0.8341	22.2190	1.1508	22.2592	1.5138	-0.0402
-0.3630								
16	114	0.0573	0.8914	23.2807	1.2461	23.5145	1.5138	-0.2338
-0.2677								
17	96	0.0483	0.9397	24.7292	1.0751	24.7698	1.5138	-0.0406
-0.4387								
18	56	0.0282	0.9678	26.1071	0.9196	26.0250	1.5138	0.0821
-0.5942								
19	38	0.0191	0.9869	27.0000	1.0000	27.2803	1.5138	-0.2803
-0.5138								
20	17	0.0085	0.9955	28.3529	0.9037	28.5356	1.5138	-0.1827
-0.6101								
21	9	0.0045	1.0000	28.8889	0.8749	29.7909	1.5138	-0.9020
-0.6389								

PRED = (1.2553)*V+(3.4298)
 ETASQUARE = 0.9227 SEE FOR ETASQUARE = 1.5138
 RSQUARE = 0.9213 SEE FOR RSQUARE = 1.5274

RELATIVE CUMULATIVE FREQUENCY DISTRIBUTIONS FOR X AND Y

SCORE	GROUP 1 ON X	GROUP 2 ON Y	GROUP S ON X	GROUP S ON Y
0.	0.000000	0.000000	0.000000	0.000000
1.	0.000000	0.000503	0.000000	0.002604
2.	0.000000	0.001006	0.000000	0.003609
3.	0.001563	0.001508	0.001563	0.004613
4.	0.004688	0.003519	0.004688	0.007706
5.	0.007813	0.008044	0.007813	0.017546
6.	0.015625	0.015586	0.015625	0.028833
7.	0.031250	0.034188	0.031250	0.058405
8.	0.050000	0.059326	0.050000	0.094801

9.	0.073438	0.087984	0.073438	0.134334
10.	0.104688	0.127200	0.104688	0.186163
11.	0.142187	0.168929	0.142187	0.237265
12.	0.210938	0.207139	0.210938	0.280826
13.	0.278125	0.253394	0.278125	0.332074
14.	0.337500	0.313223	0.337500	0.395882
15.	0.404688	0.375063	0.404687	0.460985
16.	0.465625	0.435395	0.465625	0.523315
17.	0.537500	0.508798	0.537500	0.597818
18.	0.590625	0.570639	0.590625	0.658822
19.	0.646875	0.625440	0.646875	0.710436
20.	0.698437	0.694822	0.698437	0.775313
21.	0.762500	0.753645	0.762500	0.826577
22.	0.835938	0.809452	0.835938	0.870964
23.	0.879687	0.863751	0.879688	0.911456
24.	0.910937	0.899447	0.910937	0.936523
25.	0.945313	0.937154	0.945313	0.960043
26.	0.968750	0.961790	0.968750	0.973733
27.	0.976563	0.980895	0.976563	0.984273
28.	0.992188	0.992459	0.992188	0.993668
29.	0.996875	0.998492	0.996875	0.998958
30.	1.000000	1.000000	1.000000	1.000000

RELATIVE CUMULATIVE FREQUENCY DISTRIBUTIONS FOR V

SCORE	GROUP 1 ON V	GROUP 2 ON V	GROUP S ON V
0.	0.003125	0.000000	0.003125
1.	0.007813	0.001508	0.007813
2.	0.021875	0.008547	0.021875
3.	0.043750	0.024635	0.043750
4.	0.081250	0.050779	0.081250
5.	0.150000	0.093514	0.150000
6.	0.225000	0.149824	0.225000
7.	0.304688	0.225742	0.304688
8.	0.382813	0.301156	0.382813
9.	0.471875	0.383610	0.471875
10.	0.557813	0.466063	0.557813
11.	0.632813	0.546506	0.632813
12.	0.706250	0.614882	0.706250
13.	0.773438	0.694319	0.773438
14.	0.843750	0.765209	0.843750
15.	0.887500	0.834087	0.887500
16.	0.934375	0.891403	0.934375
17.	0.962500	0.939668	0.962500
18.	0.973437	0.967823	0.973437
19.	0.989062	0.986928	0.989062
20.	0.996875	0.995475	0.996875
21.	1.000000	1.000000	1.000000

FORM Y EQUIVALENTS OF FORM X INTEGER SCORES

X-SCORE	TMEAN	LMEAN	BMEAN	TLIN	LLIN	BLIN	UNSM	S=0.01	S=0.10	S=0.25	S=0.50	S=0.75
0.	-1.08	-1.19	-1.10	-1.62	-1.74	-1.78	-0.09	0.15	0.15	0.15	0.13	0.12
1.	-0.08	-0.19	-0.10	-0.59	-0.71	-0.74	0.74	0.94	0.96	0.94	0.90	0.87
2.	0.92	0.81	0.90	0.44	0.33	0.30	1.56	1.73	1.76	1.74	1.67	1.62
3.	1.92	1.81	1.90	1.47	1.36	1.34	2.38	2.52	2.57	2.53	2.44	2.37
4.	2.92	2.81	2.90	2.50	2.39	2.38	3.21	3.31	3.37	3.33	3.20	3.11
5.	3.92	3.81	3.90	3.54	3.42	3.42	4.03	4.10	4.17	4.12	3.97	3.86
6.	4.92	4.81	4.90	4.57	4.45	4.46	4.91	4.97	5.01	4.93	4.78	4.69
7.	5.92	5.81	5.90	5.60	5.48	5.50	6.02	5.98	5.88	5.78	5.66	5.58
8.	6.92	6.81	6.90	6.63	6.52	6.54	6.90	6.85	6.72	6.64	6.58	6.55
9.	7.92	7.81	7.90	7.66	7.55	7.58	7.59	7.59	7.54	7.51	7.51	7.53
10.	8.92	8.81	8.90	8.69	8.58	8.62	8.34	8.35	8.37	8.41	8.48	8.52
11.	9.92	9.81	9.90	9.72	9.61	9.66	9.22	9.23	9.30	9.40	9.49	9.56
12.	10.92	10.81	10.90	10.75	10.64	10.70	10.31	10.33	10.40	10.49	10.58	10.63
13.	11.92	11.81	11.90	11.79	11.68	11.74	11.67	11.68	11.67	11.68	11.72	11.74
14.	12.92	12.81	12.90	12.82	12.71	12.78	13.03	13.01	12.95	12.91	12.87	12.85
15.	13.92	13.81	13.90	13.85	13.74	13.82	14.11	14.11	14.10	14.05	13.99	13.94
16.	14.92	14.81	14.90	14.88	14.77	14.86	15.10	15.12	15.14	15.11	15.04	14.99
17.	15.92	15.81	15.90	15.91	15.80	15.90	16.15	16.13	16.12	16.08	16.03	15.99
18.	16.92	16.81	16.90	16.94	16.84	16.94	17.05	17.04	17.02	17.00	16.98	16.96
19.	17.92	17.81	17.90	17.97	17.87	17.98	17.84	17.86	17.87	17.89	17.91	17.92
20.	18.92	18.81	18.90	19.01	18.90	19.02	18.77	18.77	18.79	18.83	18.87	18.90
21.	19.92	19.81	19.90	20.04	19.93	20.06	19.81	19.80	19.82	19.86	19.89	19.92
22.	20.92	20.81	20.90	21.07	20.96	21.10	20.97	20.98	20.98	20.97	20.97	20.98

23.	21.92	21.81	21.90	22.10	21.99	22.14	22.20	22.19	22.14	22.10	22.07	22.07
24.	22.92	22.81	22.90	23.13	23.03	23.18	23.10	23.14	23.18	23.19	23.17	23.16
25.	23.92	23.81	23.90	24.16	24.06	24.22	24.16	24.18	24.25	24.28	24.27	24.26
26.	24.92	24.81	24.90	25.19	25.09	25.26	25.37	25.40	25.37	25.37	25.37	25.36
27.	25.92	25.81	25.90	26.23	26.12	26.30	26.42	26.46	26.47	26.47	26.47	26.46
28.	26.92	26.81	26.90	27.26	27.15	27.34	27.51	27.53	27.55	27.55	27.56	27.56
29.	27.92	27.81	27.90	28.29	28.19	28.38	28.66	28.64	28.63	28.63	28.65	28.65
30.	28.92	28.81	28.90	29.32	29.22	29.41	29.89	29.86	29.86	29.87	29.87	29.88
X-SCORE	TMEAN	LMEAN	BMEAN	TLIN	LLIN	BLIN	UNSM	S=0.01	S=0.10	S=0.25	S=0.50	S=0.75

STANDARD ERRORS OF EQUATING FOR SELECTED METHODS			
X-SCORE	TLIN	LLIN	EQ%TILE
0.	0.2428	0.2546	-999.0000
1.	0.2313	0.2421	-999.0000
2.	0.2198	0.2296	-999.0000
3.	0.2084	0.2172	-999.0000
4.	0.1970	0.2049	-999.0000
5.	0.1856	0.1926	0.5269
6.	0.1744	0.1805	0.5927
7.	0.1632	0.1685	0.4660
8.	0.1522	0.1566	0.2821
9.	0.1412	0.1449	0.2581
10.	0.1305	0.1334	0.2148
11.	0.1200	0.1223	0.2114
12.	0.1097	0.1115	0.2149
13.	0.0998	0.1012	0.2062
14.	0.0904	0.0916	0.1873
15.	0.0817	0.0829	0.1615
16.	0.0738	0.0755	0.1477
17.	0.0672	0.0697	0.1288
18.	0.0621	0.0659	0.0966
19.	0.0590	0.0646	0.0697
20.	0.0582	0.0658	0.0641
21.	0.0599	0.0695	0.1151
22.	0.0637	0.0752	0.1372
23.	0.0694	0.0826	0.1010
24.	0.0765	0.0913	0.1760
25.	0.0847	0.1008	0.2248
26.	0.0938	0.1111	0.2198
27.	0.1033	0.1219	0.1767
28.	0.1134	0.1330	0.4180
29.	0.1237	0.1445	0.6659
30.	0.1344	0.1562	-999.0000

SYNTHETIC GROUP SUMMARY STATISTICS

		MEAN	SD	SKEWNESS	KURTOSIS
FORM X	TLIN	17.1297	5.2930	0.0037	2.3920
	LLIN	17.1297	5.2930		
	BLIN	17.1297	5.2930		
	EQ%TILE	17.1297	5.2930		
FORM Y	TLIN	16.0455	5.4591	0.0114	2.4154
	LLIN	15.9373	5.4618		
	BLIN	16.0325	5.5036		
	EQ%TILE	16.0325	5.5036		

SLOPE AND INTERCEPT FOR MEAN AND LINEAR METHODS

	TMEAN	LMEAN	BMEAN	TLIN	LLIN	BLIN
SLOPE	1.000000	1.000000	1.000000	1.031370	1.031895	1.039784
INTERCEPT	-1.084213	-1.192381	-1.097231	-1.621574	-1.738728	-1.778716

THE FOLLOWING WAS WRITTEN BECUASE W1=1

GROUP 1 SUMMARY STATISTICS FOR SMOOTHED EQUI%TILE
FORM X SCORES CONVERTED TO FORM Y SCALE FOR GROUP 1

	MEAN	SD	SKEWNESS	KURTOSIS
S = 0.01	16.0407	5.4878	0.0326	2.3707
S = 0.10	16.0430	5.4926	0.0337	2.3735
S = 0.25	16.0421	5.4925	0.0334	2.3846
S = 0.50	16.0396	5.4921	0.0323	2.4006
S = 0.75	16.0369	5.4921	0.0318	2.4096

anexo ciipe out2

0.	-1.0842130666 -1.6215739425 -0.0882489012 0.1472877637	-1.1923809934 -1.7387277879 0.1450458508 0.1337043960	-1.0972309618 -1.7787158381 0.1521565678 0.1235935607
1.	-0.0842130666 -0.5902037919 0.7352532963 0.9418632912	-0.1923809934 -0.7068330558 0.9351375525 0.9011131881	-0.0972309618 -0.7389319906 0.9564697035 0.8707806821
2.	0.9157869334 0.4411663586 1.5587554939 1.7364388186	0.8076190066 0.3250616764 1.7252292542 1.6685219802	0.9027690382 0.3008518570 1.7607828391 1.6179678035
3.	1.9157869334 1.4725365092 2.3822576914 2.5310143460	1.8076190066 1.3569564085 2.5153209559 2.4359307723	1.9027690382 1.3406357046 2.5650959747 2.3651549250
4.	2.9157869334 2.5039066597 3.2057598890 3.3255898735	2.8076190066 2.3888511407 3.3054126576 3.2033395644	2.9027690382 2.3804195521 3.3694091104 3.1123420464
5.	3.9157869334 3.5352768103 4.0292620865 4.1201654009	3.8076190066 3.4207458728 4.0955043593 3.9707483564	3.9027690382 3.4202033997 4.1737222460 3.8595291678
6.	4.9157869334 4.5666469608 4.9078074166 4.9287961436	4.8076190066 4.4526406049 4.9689702099 4.7846733305	4.9027690382 4.4599872473 5.0080401436 4.6862878547
7.	5.9157869334 5.5980171114 6.0219708333 5.7824637450	5.8076190066 5.4845353371 5.9830832897 5.6640999589	5.9027690382 5.4997710948 5.8818405334 5.5828244417
8.	6.9157869334 6.6293872620 6.8987599971 6.6418292905	6.8076190066 6.5164300692 6.8451217017 6.5823149670	6.9027690382 6.5395549424 6.7231217309 6.5490891554
9.	7.9157869334 7.6607574125 7.5910586716 7.5099846839	7.8076190066 7.5483248014 7.5906711760 7.5139907070	7.9027690382 7.5793387900 7.5380105857 7.5258795745
10.	8.9157869334 8.6921275631 8.3423419700 8.4146634678	8.8076190066 8.5802195335 8.3509364693 8.4768208709	8.9027690382 8.6191226375 8.3742152613 8.5245863249
11.	9.9157869334 9.7234977136 9.2243721775 9.3953589126	9.8076190066 9.6121142657 9.2310612326 9.4925238918	9.9027690382 9.6589064851 9.3026337062 9.5577697155
12.	10.9157869334 10.7548678642 10.3147686141 10.4872109899	10.8076190066 10.6440089978 10.3303141701 10.5754586560	10.9027690382 10.6986903327 10.3985460060 10.6315644362
13.	11.9157869334 11.7862380147 11.6668116959 11.6844615145	11.8076190066 11.6759037300 11.6810196594 11.7156067314	11.9027690382 11.7384741802 11.6703362876 11.7371480482
14.	12.9157869334 12.8176081653	12.8076190066 12.7077984621	12.9027690382 12.7782580278

	13.0265853653 12.9054342649	13.0096276712 12.8699775441	12.9528746776 12.8490610807
15.	13.9157869334 13.8489783159 14.1115225729 14.0528996938	13.8076190066 13.7396931942 14.1088536276 13.9871754118	13.9027690382 13.8180418754 14.0968991633 13.9376536485
16.	14.9157869334 14.8803484664 15.1032669896 15.1071042120	14.8076190066 14.7715879264 15.1180204102 15.0401968649	14.9027690382 14.8578257229 15.1401823720 14.9851417315
17.	15.9157869334 15.9117186170 16.1510143935 16.0829948652	15.8076190066 15.8034826585 16.1326152735 16.0325936001	15.9027690382 15.8976095705 16.1162668643 15.9891360843
18.	16.9157869334 16.9430887675 17.0469291857 16.9961895159	16.8076190066 16.8353773907 17.0446189526 16.9779834251	16.9027690382 16.9373934181 17.0160863063 16.9590786137
19.	17.9157869334 17.9744589181 17.8431318083 17.8870745342	17.8076190066 17.8672721228 17.8572909623 17.9079168594	17.9027690382 17.9771772657 17.8704230444 17.9176108154
20.	18.9157869334 19.0058290686 18.7680353883 18.8261543245	18.8076190066 18.8991668550 18.7742869818 18.8689376208	18.9027690382 19.0169611132 18.7877317219 18.8967234959
21.	19.9157869334 20.0371992192 19.8087772941 19.8563526435	19.8076190066 19.9310615871 19.8042963881 19.8908164129	19.9027690382 20.0567449608 19.8249365104 19.9181367152
22.	20.9157869334 21.0685693698 20.9663279098 20.9689623514	20.8076190066 20.9629563192 20.9796308673 20.9690504765	20.9027690382 21.0965288084 20.9839836492 20.9805708723
23.	21.9157869334 22.0999395203 22.2037111470 22.0987267809	21.8076190066 21.9948510514 22.1850202499 22.0731211715	21.9027690382 22.1363126559 22.1388384111 22.0676449513
24.	22.9157869334 23.1313096709 23.1013170927 23.1890365035	22.8076190066 23.0267457835 23.1370688831 23.1717627414	22.9027690382 23.1760965035 23.1830057212 23.1607087709
25.	23.9157869334 24.1626798214 24.1649649861 24.2753416136	23.8076190066 24.0586405157 24.1761412013 24.2697744120	23.9027690382 24.2158803511 24.2454168391 24.2583533495
26.	24.9157869334 25.1940499720 25.3719405594 25.3747937245	24.8076190066 25.0905352478 25.3965015936 25.3702379176	24.9027690382 25.2556641986 25.3737763945 25.3593062007
27.	25.9157869334 26.2254201225 26.4213746190 26.4671079078	25.8076190066 26.1224299800 26.4628425411 26.4663517807	25.9027690382 26.2954480462 26.4738624657 26.4589238541
28.	26.9157869334 27.2567902731 27.5108129756 27.5496608876	26.8076190066 27.1543247121 27.5258379184 27.5577708488	26.9027690382 27.3352318938 27.5501176297 27.5564589609
29.	27.9157869334 28.2881604237 28.6631274131 28.6289159082	27.8076190066 28.1862194443 28.6361955831 28.6472745115	27.9027690382 28.3750157413 28.6264596853 28.6530742289

30.

28.9157869334	28.8076190066	28.9027690382
29.3195305742	29.2181141764	29.4147995889
29.8877091377	29.8623445071	29.8599848281
29.8654466771	29.8749397695	29.8780507812

anexo cipec out3

X-SCORE	RFX-1	RFX-2	RFX-S	RFX-S
0.000000	0.000000	0.000000	0.000000	0.000000
1.000000	0.000000	0.000503	0.000000	0.002604
2.000000	0.000000	0.000503	0.000000	0.001004
3.000000	0.001563	0.000503	0.001563	0.001004
4.000000	0.003125	0.002011	0.003125	0.003093
5.000000	0.003125	0.004525	0.003125	0.009840
6.000000	0.007813	0.007541	0.007813	0.011287
7.000000	0.015625	0.018602	0.015625	0.029572
8.000000	0.018750	0.025138	0.018750	0.036396
9.000000	0.023438	0.028658	0.023437	0.039533
10.000000	0.031250	0.039216	0.031250	0.051829
11.000000	0.037500	0.041730	0.037500	0.051102
12.000000	0.068750	0.038210	0.068750	0.043562
13.000000	0.067188	0.046254	0.067188	0.051247
14.000000	0.059375	0.059829	0.059375	0.063808
15.000000	0.067188	0.061840	0.067187	0.065103
16.000000	0.060937	0.060332	0.060938	0.062330
17.000000	0.071875	0.073404	0.071875	0.074503
18.000000	0.053125	0.061840	0.053125	0.061004
19.000000	0.056250	0.054801	0.056250	0.051615
20.000000	0.051562	0.069382	0.051562	0.064876
21.000000	0.064063	0.058824	0.064063	0.051264
22.000000	0.073438	0.055807	0.073438	0.044387
23.000000	0.043750	0.054299	0.043750	0.040492
24.000000	0.031250	0.035696	0.031250	0.025067
25.000000	0.034375	0.037707	0.034375	0.023520
26.000000	0.023438	0.024635	0.023437	0.013689
27.000000	0.007813	0.019105	0.007813	0.010541
28.000000	0.015625	0.011564	0.015625	0.009395
29.000000	0.004687	0.006033	0.004687	0.005290
30.000000	0.003125	0.001508	0.003125	0.001042
TUCKER	LEVINE	BRAUN-HOLLAND	UNSMOOTHED	SE TUCKER
-1.621574	-1.738728	-1.778716	-0.088249	0.242836
-1.590204	-1.706833	-1.738932	-0.264747	0.231305
-1.558834	-1.674938	-1.699148	-0.441245	0.219812
-1.527463	-1.643044	-1.659364	-0.617742	0.208365
-1.496093	-1.611149	-1.619580	-0.794240	0.196971
-1.464723	-1.579254	-1.579797	-0.970738	0.185639
-1.433353	-1.547359	-1.540013	-1.092193	0.174383
-1.401983	-1.515465	-1.500229	-0.978029	0.163218
-1.370613	-1.483570	-1.460445	-1.101240	0.152163
-1.339243	-1.451675	-1.420661	-1.408941	0.141245
-1.307872	-1.419780	-1.380877	-1.657658	0.130498
-1.276502	-1.387886	-1.341094	-1.775628	0.119967
-1.245132	-1.355991	-1.301310	-1.685231	0.109716
-1.213762	-1.324096	-1.261526	-1.333188	0.099831
-1.182392	-1.292202	-1.221742	-0.973415	0.090431
-1.151022	-1.260307	-1.181958	-0.888477	0.081684
-1.119652	-1.228412	-1.142174	-0.896733	0.073823
-1.088281	-1.196517	-1.102390	-0.848986	0.067159
-1.056911	-1.164623	-1.062607	-0.953071	0.062079
-1.025541	-1.132728	-1.022823	-1.156868	0.058995
-0.994171	-1.100833	-0.983039	-1.231965	0.058223
-0.962801	-1.068938	-0.943255	-1.191223	0.059853
-0.931431	-1.037044	-0.903471	-1.033672	0.063702
-0.900060	-1.005149	-0.863687	-0.796289	0.069401
-0.868690	-0.973254	-0.823903	-0.898683	0.076538
-0.837320	-0.941359	-0.784120	-0.835035	0.084750
-0.805950	-0.909465	-0.744336	-0.628059	0.093755
-0.774580	-0.877570	-0.704552	-0.578625	0.103347
-0.743210	-0.845675	-0.664768	-0.489187	0.113376
-0.711840	-0.813781	-0.624984	-0.336873	0.123736
-0.680469	-0.781886	-0.585200	-0.112291	0.134351
SE LEVINE	SE UNSMOOTHED	SM = 0.01	SM = 0.10	SM = 0.25
0.254560	0.000000	0.145046	0.152157	0.147288
0.242060	0.000000	-0.064862	-0.043530	-0.058137
0.229609	0.000000	-0.274771	-0.239217	-0.263561
0.217216	0.000000	-0.484679	-0.434904	-0.468986
0.204892	0.000000	-0.694587	-0.630591	-0.674410
0.192649	0.526933	-0.904496	-0.826278	-0.879835
0.180504	0.592713	-1.031030	-0.991960	-1.071204

0.168479	0.466028	-1.016917	-1.118159	-1.217536
0.156601	0.282059	-1.154878	-1.276878	-1.358171
0.144906	0.258077	-1.409329	-1.461989	-1.490015
0.133442	0.214750	-1.649064	-1.625785	-1.585337
0.122275	0.211357	-1.768939	-1.697366	-1.604641
0.111493	0.214901	-1.669686	-1.601454	-1.512789
0.101219	0.206185	-1.318980	-1.329664	-1.315538
0.091626	0.187277	-0.990372	-1.047125	-1.094566
0.082948	0.161547	-0.891146	-0.903101	-0.947100
0.075504	0.147692	-0.881980	-0.859818	-0.892896
0.069688	0.128788	-0.867385	-0.883733	-0.917005
0.065933	0.096559	-0.955381	-0.983914	-1.003810
0.064600	0.069663	-1.142709	-1.129577	-1.112925
0.065837	0.064065	-1.225713	-1.212268	-1.173846
0.069505	0.115111	-1.195704	-1.175063	-1.143647
0.075250	0.137231	-1.020369	-1.016016	-1.031038
0.082640	0.100982	-0.814980	-0.861162	-0.901273
0.091277	0.175986	-0.862931	-0.816994	-0.810963
0.100841	0.224757	-0.823859	-0.754583	-0.724658
0.111092	0.219781	-0.603498	-0.626224	-0.625206
0.121857	0.176716	-0.537157	-0.526138	-0.532892
0.133012	0.418044	-0.474162	-0.449882	-0.450339
0.144465	0.665915	-0.363804	-0.373540	-0.371084
0.156153	0.000000	-0.137655	-0.140015	-0.134553

SM = 0.50	SM = 0.75
0.133704	0.123594
-0.098887	-0.129219
-0.331478	-0.382032
-0.564069	-0.634845
-0.796660	-0.887658
-1.029252	-1.140471
-1.215327	-1.313712
-1.335900	-1.417176
-1.417685	-1.450911
-1.486009	-1.474120
-1.523179	-1.475414
-1.507476	-1.442230
-1.424541	-1.368436
-1.284393	-1.262852
-1.130022	-1.150939
-1.012825	-1.062346
-0.959803	-1.014858
-0.967406	-1.010864
-1.022017	-1.040921
-1.092083	-1.082389
-1.131062	-1.103277
-1.109184	-1.081863
-1.030950	-1.019429
-0.926879	-0.932355
-0.828237	-0.839291
-0.730226	-0.741647
-0.629762	-0.640694
-0.533648	-0.541076
-0.442229	-0.443541
-0.352725	-0.346926
-0.125060	-0.121949

anexo Bilog comandos

Prova da 4a. série de 1997 - SARESP

Prova com 30 questões de 5 alternativas - 1989 respondentes

>COMMENTS

Prova composta de 30 questões com 5 alternativas

```
>GLObal  NPArm=3, LOGistic,
          DFName='C:\TRI\PORT1.DAT',
          KFName='C:\TRI\PORT1.DAT',
          SAVe;
>SAVe    PARM='C:\TRI\PORT1.PAR',
          GRaph='C:\TRI\PORT1.PLT',
          SCOrE='C:\TRI\PORT1.SCO';
>LENGHT  NITems=30;
>INPut   NTotAl=30, SAMple=1989, NALt=5,  NIDch=13;
(13Al,30Al)
>TEST    TName='FOFAl',
          ITems=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,
                20,21,22,23,24,25,26,27,28,29,30);
>CALib   NOSprior, NOGprior, NQPT = 10, IDIst=0,
          CYCles=50, NEWton=1, CRIt=0.001, RIDge=0;
>SCOrE   METHOD=1, INFo= 2, NOPRINT;
```

Prova de ligação da 3a. série de 1997 - SARESP

Prova com 32 questões de 5 alternativas - 640 respondentes

>COMMENTS

Prova composta de 32 questões com 5 alternativas

```
>GLObal  NPArm=3, LOGistic,
          DFName='C:\TRI\PORT2.DAT',
          KFName='C:\TRI\PORT2.DAT',
          SAVe;
>SAVe    PARM='C:\TRI\PORT2.PAR',
          GRaph='C:\TRI\PORT2.PLT',
          SCOrE='C:\TRI\PORT2.SCO';
>LENGHT  NITems=32;
>INPut   NTotAl=32, SAMple=640, NALt=5,  NIDch=13;
(13Al,32Al)
>TEST    TName='FOFAl',
          ITems=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,
                20,21,22,23,24,25,26,27,28,29,30,31,32);
>CALib   NOSprior, NOGprior, NQPT = 10, IDIst=0,
          CYCles=50, NEWton=1, CRIt=0.001, RIDge=0;
>SCOrE   METHOD=1, INFo= 2, NOPRINT;
```

anexo PH1 PORT1

BILOG 3.11, WINDOWS VERSION

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: LOGISTIC MODEL

DISTRIBUTED BY

SCIENTIFIC SOFTWARE INTERNATIONAL

7383 N. LINCOLN AVE, SUITE 100
 CHICAGO, ILLINOIS 60646
 (800) 247-6113, (847) 675-0720
 WEBSITE: WWW.SSICENTRAL.COM

PROGRAM COPYRIGHT HELD BY SCIENTIFIC SOFTWARE INTERNATIONAL 1985-98
 ALL RIGHTS RESERVED.

1

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 1 ***

Exemplo de Programa da Teoria da Resposta ao Item
 Prova com 30 questões de 5 alternativas - 1500 respondentes

>COMMENTS

Prova composta de 30 questões com 5 alternativas

>Global NPArm=3, LOGistic,

DFName='C:\TRI\PORT1.DAT',

KFName='C:\TRI\PORT1.DAT',

SAVe;

GLOBAL PARAMETERS

=====

NUMBER OF SUBTESTS

1

CASE WEIGHTING

NONE EMPLOYED

ITEM RESPONSE MODEL

3 PARAMETER LOGISTIC

LOGIT METRIC (I.E., D = 1.0)

>SAVe PARm='C:\TRI\PORT1.PAR',

GRaph='C:\TRI\PORT1.PLT',

SCore='C:\TRI\PORT1.SCO';

>LENGHT NITems=30;

SUBTEST LENGTHS:

30

>INPut NTotal=30, SAMple=1989, NAlt=5, NIDch=13;

DATA INPUT SPECIFICATIONS

=====

NUMBER OF FORMAT CARDS

1

NUMBER OF ITEMS IN INPUT STREAM

30

NUMBER OF RESPONSE ALTERNATIVES

5

NUMBER OF SUBJECT ID CHARACTERS

13

SUBJECT DATA INPUT OPTION

SINGLE-SUBJECT DATA, NO CASE WEIGHTS

MAXIMUM SAMPLE SIZE FOR ITEM CALIBRATION 1989

ALL SUBJECTS INCLUDED IN RUN

RESPONSE CODES: 0 1

FORMAT CARD FOR INPUT IS

(13A1,30A1)

FILE ASSIGNMENT AND DISPOSITION

=====

[INPUT FILES]

SUBJECT DATA INPUT FILE C:\TRI\PORT1.DAT
SINGLE-SUBJECT DATA, NO CASE WEIGHTS

CORRECT-RESPONSE KEY FILE C:\TRI\PORT1.DAT

[OUTPUT FILES]

ITEM PARAMETERS FILE C:\TRI\PORT1.PAR

SUBJECT SCALE-SCORE FILE C:\TRI\PORT1.SCO

GRAPHIC PARAMETER FILE C:\TRI\PORT1.PLT

[SCRATCH FILES]

BILOG SYSTEM BINARY DATA FILE PORT1.MFL

CALIBRATION BINARY DATA FILE PORT1.CFL

ESTIMATED COVARIANCE FILE PORT1.VFL

TEMPORARY FILE PORT1.T02

TEMPORARY FILE PORT1.T03

TEMPORARY FILE PORT1.T14

TEMPORARY FILE PORT1.T99

>TEST TName='FOFA1',

ITEMs=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,

20,21,22,23,24,25,26,27,28,29,30);

ANSWER KEY:

1 FOFA1 BCDABDCABABDCDBDCADABACBCDADCA

OBSERVATION 1 WEIGHT: 1.0000 ID : 0000244103081

SUBTEST 1 FOFA1

TRIED RIGHT
30.000 15.000

ITEM	1	2	3	4	5	6	7	8	9	10
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	1.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0
ITEM	11	12	13	14	15	16	17	18	19	20
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0
ITEM	21	22	23	24	25	26	27	28	29	30
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0

OBSERVATION 2 WEIGHT: 1.0000 ID : 0000244103121

SUBTEST 1 FOFA1

TRIED RIGHT
30.000 22.000

ITEM	1	2	3	4	5	6	7	8	9	10
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0

ITEM	11	12	13	14	15	16	17	18	19	20
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0

ITEM	21	22	23	24	25	26	27	28	29	30
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0

1989 OBSERVATIONS READ FROM FILE: C:\TRI\PORT1.DAT

1989 OBSERVATIONS WRITTEN TO FILE: PORT1.MFL

1

CLASSICAL ITEM STATISTICS FOR SUBTEST FOFA1

ITEM	NAME	NUMBER		PERCENT	LOGIT	ITEM*TEST CORRELATION	
		TRIED	RIGHT			PEARSON	BISERIAL
1	0001	1989.0	1513.0	0.761	1.16	0.405	0.556
2	0002	1989.0	1620.0	0.814	1.48	0.379	0.551
3	0003	1989.0	1430.0	0.719	0.94	0.359	0.478
4	0004	1989.0	1550.0	0.779	1.26	0.138	0.193
5	0005	1989.0	1151.0	0.579	0.32	0.252	0.318
6	0006	1989.0	1354.0	0.681	0.76	0.368	0.481
7	0007	1989.0	996.0	0.501	0.00	0.437	0.548
8	0008	1989.0	1354.0	0.681	0.76	0.340	0.443
9	0009	1989.0	1257.0	0.632	0.54	0.244	0.312
10	0010	1989.0	1299.0	0.653	0.63	0.254	0.328
11	0011	1989.0	1383.0	0.695	0.83	0.390	0.512
12	0012	1989.0	1676.0	0.843	1.68	0.389	0.588
13	0013	1989.0	1266.0	0.637	0.56	0.404	0.518
14	0014	1989.0	478.0	0.240	-1.15	0.249	0.342
15	0015	1989.0	564.0	0.284	-0.93	0.372	0.496
16	0016	1989.0	1154.0	0.580	0.32	0.429	0.542
17	0017	1989.0	533.0	0.268	-1.00	0.232	0.312
18	0018	1989.0	1542.0	0.775	1.24	0.283	0.394
19	0019	1989.0	1377.0	0.692	0.81	0.350	0.460
20	0020	1989.0	1292.0	0.650	0.62	0.173	0.222
21	0021	1989.0	848.0	0.426	-0.30	0.302	0.381
22	0022	1989.0	1130.0	0.568	0.27	0.310	0.390
23	0023	1989.0	1260.0	0.633	0.55	0.409	0.524
24	0024	1989.0	1016.0	0.511	0.04	0.277	0.347
25	0025	1989.0	1070.0	0.538	0.15	0.389	0.489
26	0026	1989.0	1150.0	0.578	0.32	0.451	0.569
27	0027	1989.0	661.0	0.332	-0.70	0.187	0.242
28	0028	1989.0	585.0	0.294	-0.88	0.326	0.431
29	0029	1989.0	511.0	0.257	-1.06	0.197	0.268
30	0030	1989.0	1396.0	0.702	0.86	0.280	0.370

5244 BYTES OF NUMERICAL WORKSPACE USED OUT OF 900000 IN PHASE-1

2784 BYTES OF CHARACTER WORKSPACE USED OUT OF 80000 IN PHASE-1

anexo PH2 PORT1

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 2 ***

Exemplo de Programa da Teoria da Resposta ao Item
 Prova com 30 questões de 5 alternativas - 1500 respondentes

>CALib NOSprior, NOGprior, NQPT = 10, IDist=0,
 CYCles=50, NEWton=1, CRIt=0.001, RIDge=0;

CALIBRATION PARAMETERS
 =====

MAXIMUM NUMBER OF EM CYCLES: 50
 MAXIMUM NUMBER OF NEWTON CYCLES: 1
 CONVERGENCE CRITERION: 0.0010

 SUBJECT DISTRIBUTION: NORMAL PRIOR
 PLOT EMPIRICAL VS. FITTED ICC'S: NO
 DATA HANDLING: RIGHT/WRONG/NOT PRESENTED DATA,
 IN CORE
 PRIOR DISTRIBUTION ON ASYMPTOTES: NO
 PRIOR DISTRIBUTION ON SLOPES: NO
 PRIOR DISTRIBUTION ON THRESHOLDS: NO

1

 CALIBRATION OF SUBTEST
 FOFAL

METHOD OF SOLUTION
 =====

EM CYCLES (MAXIMUM OF 50)
 FOLLOWED BY NEWTON-RAPHSON STEPS (MAXIMUM OF 1)

QUADRATURE POINTS AND PRIOR WEIGHTS:

	1	2	3	4	5
POINT	-0.4000E+01	-0.3111E+01	-0.2222E+01	-0.1333E+01	-0.4444E+00
WEIGHT	0.1190E-03	0.2805E-02	0.3002E-01	0.1458E+00	0.3213E+00
	6	7	8	9	10
POINT	0.4444E+00	0.1333E+01	0.2222E+01	0.3111E+01	0.4000E+01
WEIGHT	0.3213E+00	0.1458E+00	0.3002E-01	0.2805E-02	0.1190E-03

DATA FOR FIRST TWO OBSERVATIONS (TT=RIGHT, TF=WRONG, FF=NOT PRESENTED)

OBSERVATION 1 WEIGHT = 1.000
 TT TF TT TF TT TT TT TT TT TF TT TT TF TF TF TF TT TT TF TF TF TT TF TF
 TT TF TT TF TT

OBSERVATION 2 WEIGHT = 1.000
 TT TT TT TF TT TT TT TT TT TT TT TT TF TF TF TF TT TT TT TT TT TT TT TT
 TF TT TF TF TT

[EM STEP]

-2 LOG LIKELIHOOD = 68395.0034
 CYCLE 1: LARGEST CHANGE = 1.23313

```

-2 LOG LIKELIHOOD = 68147.2880
CYCLE 2:  LARGEST CHANGE = 0.40991
-2 LOG LIKELIHOOD = 68115.8491
CYCLE 3:  LARGEST CHANGE = 0.24715
-2 LOG LIKELIHOOD = 68108.0291
CYCLE 4:  LARGEST CHANGE = 0.38953
-2 LOG LIKELIHOOD = 68107.3111
CYCLE 5:  LARGEST CHANGE = 0.07535
-2 LOG LIKELIHOOD = 68103.5923
CYCLE 6:  LARGEST CHANGE = 0.08309
-2 LOG LIKELIHOOD = 68104.3113
CYCLE 7:  LARGEST CHANGE = 0.06456
-2 LOG LIKELIHOOD = 68102.6081
CYCLE 8:  LARGEST CHANGE = 0.09504
-2 LOG LIKELIHOOD = 68102.6771
CYCLE 9:  LARGEST CHANGE = 0.21302
-2 LOG LIKELIHOOD = 68103.1425
CYCLE 10: LARGEST CHANGE = 0.13439
-2 LOG LIKELIHOOD = 68102.5292
CYCLE 11: LARGEST CHANGE = 0.00693
-2 LOG LIKELIHOOD = 68102.5231
CYCLE 12: LARGEST CHANGE = 0.00108
-2 LOG LIKELIHOOD = 68102.5226
CYCLE 13: LARGEST CHANGE = 0.00078
[NEWTON STEP]
-2 LOG LIKELIHOOD = 68102.4700
CYCLE 14: LARGEST CHANGE = 0.00234

```

1

SUBTEST FOFA1 : ITEM PARAMETERS AFTER CYCLE 14

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	DISPERSN S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
0001	1.510	1.310	-1.153	0.763	0.000	21.9	8.0
	0.253*	0.163*	0.314*	0.095*	0.180*	(0.0053)	
0002	1.941	1.332	-1.457	0.751	0.000	21.0	8.0
	0.303*	0.174*	0.387*	0.098*	0.241*	(0.0072)	
0003	1.000	1.167	-0.857	0.857	0.097	8.6	9.0
	0.270*	0.169*	0.338*	0.124*	0.152*	(0.4715)	
0004	1.290	0.359	-3.597	2.789	0.000	13.6	9.0
	3.164*	0.282*	11.607*	2.195*	2.444*	(0.1347)	
0005	0.327	0.581	-0.563	1.721	0.000	8.2	9.0
	0.625*	0.179*	1.241*	0.531*	0.339*	(0.5140)	
0006	0.894	1.002	-0.892	0.998	0.000	14.6	9.0
	0.328*	0.161*	0.456*	0.160*	0.205*	(0.1008)	
0007	-0.433	1.628	0.266	0.614	0.123	16.4	9.0
	0.175*	0.178*	0.086*	0.067*	0.038*	(0.0585)	
0008	0.662	0.993	-0.667	1.007	0.125	10.4	9.0
	0.342*	0.181*	0.454*	0.183*	0.171*	(0.3173)	
0009	0.568	0.582	-0.977	1.720	0.000	6.7	9.0
	0.674*	0.176*	1.441*	0.520*	0.405*	(0.6746)	
0010	0.671	0.608	-1.104	1.644	0.000	14.8	9.0
	0.698*	0.184*	1.465*	0.497*	0.435*	(0.0946)	
0011	0.807	1.304	-0.618	0.767	0.132	14.0	9.0
	0.207*	0.157*	0.219*	0.092*	0.098*	(0.1200)	
0012	2.332	1.527	-1.527	0.655	0.000	29.3	7.0
	0.287*	0.182*	0.333*	0.078*	0.237*	(0.0002)	
0013	0.140	1.633	-0.086	0.612	0.232	14.9	8.0
	0.196*	0.200*	0.128*	0.075*	0.054*	(0.0600)	
0014	-2.427	1.450	1.674	0.690	0.114	11.7	9.0
	0.366*	0.253*	0.108*	0.120*	0.021*	(0.2329)	
0015	-2.265	2.063	1.098	0.485	0.100	17.7	9.0
	0.278*	0.244*	0.055*	0.057*	0.016*	(0.0385)	
0016	-0.223	1.868	0.120	0.535	0.209	15.0	8.0
	0.195*	0.224*	0.093*	0.064*	0.041*	(0.0587)	
0017	-1.390	0.767	1.812	1.304	0.055	7.4	9.0
	0.375*	0.198*	0.175*	0.337*	0.059*	(0.5986)	
0018	0.799	1.074	-0.744	0.931	0.344	5.7	9.0
	0.385*	0.205*	0.484*	0.178*	0.149*	(0.7672)	
0019	-0.001	1.707	0.000	0.586	0.380	8.9	8.0
	0.252*	0.251*	0.148*	0.086*	0.052*	(0.3508)	
0020	0.627	0.396	-1.586	2.528	0.000	13.8	9.0
	1.505*	0.230*	4.703*	1.467*	0.953*	(0.1304)	
0021	-1.803	1.751	1.030	0.571	0.251	4.9	9.0
	0.368*	0.301*	0.082*	0.098*	0.028*	(0.8445)	
0022	0.026	0.893	-0.029	1.120	0.124	19.9	9.0

		0.317*		0.168*		0.360*		0.211*		0.118*		(0.0184)	
0023		0.019		1.779		-0.011		0.562		0.257		10.2	8.0
		0.202*		0.216*		0.114*		0.068*		0.048*		(0.2531)	
0024		-0.457		0.890		0.513		1.124		0.177		9.2	9.0
		0.332*		0.181*		0.282*		0.228*		0.087*		(0.4156)	
0025		0.136		1.056		-0.129		0.947		0.018		14.1	9.0
		0.199*		0.135*		0.202*		0.121*		0.082*		(0.1190)	
0026		-0.390		2.320		0.168		0.431		0.230		17.1	8.0
		0.205*		0.275*		0.073*		0.051*		0.033*		(0.0292)	
0027		-3.627		2.205		1.645		0.454		0.258		8.1	9.0
		0.842*		0.575*		0.105*		0.118*		0.018*		(0.5287)	
0028		-2.365		1.923		1.230		0.520		0.136		12.1	9.0
		0.354*		0.285*		0.066*		0.077*		0.020*		(0.2075)	
0029		-2.721		1.463		1.860		0.684		0.159		9.0	9.0
		0.482*		0.312*		0.139*		0.146*		0.021*		(0.4371)	
0030		0.348		0.979		-0.355		1.021		0.301		10.0	9.0
		0.402*		0.223*		0.484*		0.232*		0.141*		(0.3470)	

LARGEST CHANGE = 0.002

* STANDARD ERROR
389.4 261.0

PARAMETER	MEAN	STN DEV
ASYMPTOTE	0.128	0.115
SLOPE	1.287	0.535
LOG(SLOPE)	0.150	0.490
THRESHOLD	-0.164	1.209

QUADRATURE POINTS AND POSTERIOR WEIGHTS:

	1	2	3	4	5
POINT	-0.4000E+01	-0.3111E+01	-0.2222E+01	-0.1333E+01	-0.4444E+00
WEIGHT	0.2037E-03	0.3922E-02	0.3458E-01	0.1415E+00	0.3111E+00

	6	7	8	9	10
POINT	0.4444E+00	0.1333E+01	0.2222E+01	0.3111E+01	0.4000E+01
WEIGHT	0.3244E+00	0.1537E+00	0.2814E-01	0.2453E-02	0.9577E-04

527600 BYTES OF NUMERICAL WORKSPACE USED OUT OF 800000 IN PHASE-2

2628 BYTES OF CHARACTER WORKSPACE USED OUT OF 80000 IN PHASE-2

anexo PH3 PORT1

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 3 ***

Exemplo de Programa da Teoria da Resposta ao Item
 Prova com 30 questões de 5 alternativas - 1500 respondentes

>Score METHOD=1, INFO= 2, NOPRINT;

PARAMETERS FOR SCORING, RESCALING, AND TEST AND ITEM INFORMATION

METHOD OF SCORING SUBJECTS: MAXIMUM LIKELIHOOD
 SCORES WRITTEN TO FILE C:\TRI\PORT1.SCO
 TYPE OF RESCALING: NONE REQUESTED
 ITEM AND TEST INFORMATION: TEST INFORMATION CURVES
 TABLE OF ITEM INFORMATION INDICES

TEST	NAME	QUAD POINTS	PRIOR MEAN	PRIOR STN DEV	RESCALING SCALING	CONSTANTS LOCATION
1	FOFA1	10	0.000	1.000	1.000	0.000

1

SCORING

1

SUBJECT IDENTIFICATION

WEIGHT	SUBTEST	TRIED	RIGHT	PERCENT	ABILITY	S.E.
0000244103081						
1.00	FOFA1	30	15	0.5000	-0.0934	0.4083
0000244103121						
1.00	FOFA1	30	22	0.7333	0.6235	0.3893
0000614101211						
1.00	FOFA1	30	15	0.5000	-0.3376	0.4307

CORRELATIONS AMONG SUBTEST SCORE ESTIMATES:

	FOFA1
FOFA1	1.000

MEANS AND STANDARD DEVIATIONS OF SCORE ESTIMATES:

TEST: FOFA1
 MEAN: -0.007
 S.D.: 1.200

1

ITEM INFORMATION STATISTICS FOR SUBTEST FOFA1

ITEM	MAXIMUM INFORMATION STANDARD ERROR *	POINT OF MAX INFORMATION STANDARD ERROR *	MAXIMUM EFFECTIVENESS POINT OF MAX EFFECTIVENESS *	AVERAGE INFORMATION INDEX OF RELIABILITY *
0001	0.4289	-1.1529	0.0000	0.0000
	0.1068*	0.3139*	0.0000*	0.0000*
0002	0.4438	-1.4565	0.0000	0.0000
	0.1159*	0.3870*	0.0000*	0.0000*
0003	0.2821	-0.7248	0.0000	0.0000
	0.0533*	0.3243*	0.0000*	0.0000*
0004	0.0321	-3.5953	0.0000	0.0000
	0.1083*	11.8180*	0.0000*	0.0000*
0005	0.0843	-0.5611	0.0000	0.0000
	0.0180*	1.2591*	0.0000*	0.0000*
0006	0.2511	-0.8924	0.0000	0.0000
	0.0805*	0.4559*	0.0000*	0.0000*
0007	0.5226	0.3805	0.0000	0.0000
	0.0799*	0.0789*	0.0000*	0.0000*
0008	0.1935	-0.4768	0.0000	0.0000
	0.0364*	0.4260*	0.0000*	0.0000*
0009	0.0845	-0.9758	0.0000	0.0000
	0.0261*	1.4619*	0.0000*	0.0000*

0010	0.0924		-1.1021		0.0000		0.0000
	0.0329*		1.4867*		0.0000*		0.0000*
0011	0.3299		-0.4679		0.0000		0.0000
	0.0476*		0.2053*		0.0000*		0.0000*
0012	0.5830		-1.5274		0.0000		0.0000
	0.1386*		0.3326*		0.0000*		0.0000*
0013	0.4275		0.0957		0.0000		0.0000
	0.0620*		0.1119*		0.0000*		0.0000*
0014	0.4215		1.7949		0.0000		0.0000
	0.1267*		0.1201*		0.0000*		0.0000*
0015	0.8764		1.1746		0.0000		0.0000
	0.1816*		0.0565*		0.0000*		0.0000*
0016	0.5841		0.2673		0.0000		0.0000
	0.0915*		0.0820*		0.0000*		0.0000*
0017	0.1320		1.9370		0.0000		0.0000
	0.0518*		0.1809*		0.0000*		0.0000*
0018	0.1478		-0.3866		0.0000		0.0000
	0.0266*		0.4251*		0.0000*		0.0000*
0019	0.3455		0.2401		0.0000		0.0000
	0.0552*		0.1225*		0.0000*		0.0000*
0020	0.0391		-1.5835		0.0000		0.0000
	0.0322*		4.7768*		0.0000*		0.0000*
0021	0.4732		1.2086		0.0000		0.0000
	0.1206*		0.0812*		0.0000*		0.0000*
0022	0.1568		0.1802		0.0000		0.0000
	0.0259*		0.3281*		0.0000*		0.0000*
0023	0.4829		0.1678		0.0000		0.0000
	0.0709*		0.0995*		0.0000*		0.0000*
0024	0.1409		0.7886		0.0000		0.0000
	0.0296*		0.2407*		0.0000*		0.0000*
0025	0.2687		-0.0958		0.0000		0.0000
	0.0373*		0.2007*		0.0000*		0.0000*
0026	0.8652		0.2951		0.0000		0.0000
	0.1429*		0.0642*		0.0000*		0.0000*
0027	0.7408		1.7895		0.0000		0.0000
	0.3173*		0.1294*		0.0000*		0.0000*
0028	0.7115		1.3345		0.0000		0.0000
	0.1786*		0.0704*		0.0000*		0.0000*
0029	0.3943		2.0147		0.0000		0.0000
	0.1431*		0.1614*		0.0000*		0.0000*
0030	0.1339		0.0054		0.0000		0.0000
	0.0229*		0.4124*		0.0000*		0.0000*

52 BYTES OF NUMERICAL WORKSPACE USED OUT OF 800000 IN PHASE-3

52 BYTES OF CHARACTER WORKSPACE USED OUT OF 80000 IN PHASE-3

anexo PH1 PORT2

BILOG 3.11, WINDOWS VERSION

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: LOGISTIC MODEL

DISTRIBUTED BY

SCIENTIFIC SOFTWARE INTERNATIONAL

7383 N. LINCOLN AVE, SUITE 100
 CHICAGO, ILLINOIS 60646
 (800) 247-6113, (847) 675-0720
 WEBSITE: WWW.SSICENTRAL.COM

PROGRAM COPYRIGHT HELD BY SCIENTIFIC SOFTWARE INTERNATIONAL 1985-98
 ALL RIGHTS RESERVED.

1

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 1 ***

Exemplo de Programa da Teoria da Resposta ao Item
 Prova com 32 questões de 5 alternativas - 640 respondentes

>COMMENTS

Prova composta de 32 questões com 5 alternativas

>Global NPArm=3, LOGistic,
 DFName='C:\TRI\PORT2.DAT',
 KFName='C:\TRI\PORT2.DAT',
 SAVe;

GLOBAL PARAMETERS
 =====

NUMBER OF SUBTESTS	1
CASE WEIGHTING	NONE EMPLOYED
ITEM RESPONSE MODEL	3 PARAMETER LOGISTIC
	LOGIT METRIC (I.E., D = 1.0)

>SAVe PARM='C:\TRI\PORT2.PAR',
 GRaph='C:\TRI\PORT2.PLT',
 SCOrE='C:\TRI\PORT2.SCO';

>LENGHT NITems=32;

SUBTEST LENGTHS:
 32

>INPut NTotAl=32, SAMple=640, NALt=5, NIDch=13;

DATA INPUT SPECIFICATIONS
 =====

NUMBER OF FORMAT CARDS	1
NUMBER OF ITEMS IN INPUT STREAM	32
NUMBER OF RESPONSE ALTERNATIVES	5
NUMBER OF SUBJECT ID CHARACTERS	13
SUBJECT DATA INPUT OPTION	SINGLE-SUBJECT DATA, NO CASE WEIGHTS
MAXIMUM SAMPLE SIZE FOR ITEM CALIBRATION	640
ALL SUBJECTS INCLUDED IN RUN	
RESPONSE CODES:	0 1

OBSERVATION 2 WEIGHT: 1.0000 ID : 0000013118021

SUBTEST 1 FOFA1

TRIED RIGHT
32.000 13.000

ITEM	1	2	3	4	5	6	7	8	9	10
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	0.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0

ITEM	11	12	13	14	15	16	17	18	19	20
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0

ITEM	21	22	23	24	25	26	27	28	29	30
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0

ITEM	31	32
TRIED	1.0	1.0
RIGHT	0.0	0.0

640 OBSERVATIONS READ FROM FILE: C:\TRI\PORT2.DAT
640 OBSERVATIONS WRITTEN TO FILE: PORT2.MFL

1

CLASSICAL ITEM STATISTICS FOR SUBTEST FOFA1

ITEM	NAME	NUMBER		PERCENT	LOGIT	ITEM*TEST CORRELATION	
		TRIED	RIGHT			PEARSON	BISERIAL

1	0001	640.0	179.0	0.280	-0.95	0.205	0.273
2	0002	640.0	451.0	0.705	0.87	0.285	0.376
3	0003	640.0	557.0	0.870	1.90	0.294	0.467
4	0004	640.0	457.0	0.714	0.92	0.415	0.552
5	0005	640.0	350.0	0.547	0.19	0.411	0.516
6	0006	640.0	614.0	0.959	3.16	0.170	0.383
7	0007	640.0	624.0	0.975	3.66	0.180	0.475
8	0008	640.0	629.0	0.983	4.05	0.177	0.529
9	0009	640.0	307.0	0.480	-0.08	0.192	0.241
10	0010	640.0	410.0	0.641	0.58	0.402	0.516
11	0011	640.0	353.0	0.552	0.21	0.341	0.429
12	0012	640.0	143.0	0.223	-1.25	0.261	0.364
13	0013	640.0	332.0	0.519	0.08	0.278	0.349
14	0014	640.0	328.0	0.512	0.05	0.231	0.290
15	0015	640.0	340.0	0.531	0.13	0.350	0.439
16	0016	640.0	365.0	0.570	0.28	0.472	0.595
17	0017	640.0	241.0	0.377	-0.50	0.334	0.426
18	0018	640.0	306.0	0.478	-0.09	0.319	0.400
19	0019	640.0	504.0	0.788	1.31	0.328	0.463
20	0020	640.0	490.0	0.766	1.18	0.425	0.587
21	0021	640.0	124.0	0.194	-1.43	0.241	0.347
22	0022	640.0	497.0	0.777	1.25	0.362	0.504
23	0023	640.0	401.0	0.627	0.52	0.390	0.498
24	0024	640.0	382.0	0.597	0.39	0.150	0.190
25	0025	640.0	257.0	0.402	-0.40	0.199	0.252
26	0026	640.0	387.0	0.605	0.43	0.420	0.534
27	0027	640.0	390.0	0.609	0.44	0.451	0.573
28	0028	640.0	330.0	0.516	0.06	0.399	0.501
29	0029	640.0	270.0	0.422	-0.32	0.468	0.591
30	0030	640.0	132.0	0.206	-1.35	0.248	0.352
31	0031	640.0	168.0	0.262	-1.03	0.363	0.490
32	0032	640.0	305.0	0.477	-0.09	0.421	0.528

5588 BYTES OF NUMERICAL WORKSPACE USED OUT OF 800000 IN PHASE-1

2952 BYTES OF CHARACTER WORKSPACE USED OUT OF 80000 IN PHASE-1

anexo PH2 PORT2

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 2 ***

Exemplo de Programa da Teoria da Resposta ao Item
 Prova com 32 questões de 5 alternativas - 640 respondentes

>CALib NOSprior, NOGprior, NQPT = 10, IDist=0,
 CYCles=50, NEWton=1, CRIt=0.001, RIDge=0;

CALIBRATION PARAMETERS
 =====

MAXIMUM NUMBER OF EM CYCLES: 50
 MAXIMUM NUMBER OF NEWTON CYCLES: 1
 CONVERGENCE CRITERION: 0.0010

 SUBJECT DISTRIBUTION: NORMAL PRIOR
 PLOT EMPIRICAL VS. FITTED ICC'S: NO
 DATA HANDLING: RIGHT/WRONG/NOT PRESENTED DATA,
 IN CORE
 PRIOR DISTRIBUTION ON ASYMPTOTES: NO
 PRIOR DISTRIBUTION ON SLOPES: NO
 PRIOR DISTRIBUTION ON THRESHOLDS: NO

1

 CALIBRATION OF SUBTEST

 FOFA1

METHOD OF SOLUTION
 =====

EM CYCLES (MAXIMUM OF 50)
 FOLLOWED BY NEWTON-RAPHSON STEPS (MAXIMUM OF 1)

QUADRATURE POINTS AND PRIOR WEIGHTS:

	1	2	3	4	5
POINT	-0.4000E+01	-0.3111E+01	-0.2222E+01	-0.1333E+01	-0.4444E+00
WEIGHT	0.1190E-03	0.2805E-02	0.3002E-01	0.1458E+00	0.3213E+00
	6	7	8	9	10
POINT	0.4444E+00	0.1333E+01	0.2222E+01	0.3111E+01	0.4000E+01
WEIGHT	0.3213E+00	0.1458E+00	0.3002E-01	0.2805E-02	0.1190E-03

DATA FOR FIRST TWO OBSERVATIONS (TT=RIGHT, TF=WRONG, FF=NOT PRESENTED)

OBSERVATION 1 WEIGHT = 1.000
 TT TT TT TT TT TT TT TT TT TF TF TF TF TT TT TT TT TT TT TF TF TT TT TT
 TT TT TT TT TF TT TF

 OBSERVATION 2 WEIGHT = 1.000
 TF TT TT TF TF TT TT TF TT TF TF TF TF TT TF TF TT TF TF TT TF TT TT
 TT TF TF TT TF TF TF

[EM STEP]
 -2 LOG LIKELIHOOD = 21943.2722
 CYCLE 1: LARGEST CHANGE = 1.83680
 -2 LOG LIKELIHOOD = 21765.7436
 CYCLE 2: LARGEST CHANGE = 0.78146

```

      -2 LOG LIKELIHOOD =      21750.4486
CYCLE 3:  LARGEST CHANGE =      0.41023
      -2 LOG LIKELIHOOD =      21747.2653
CYCLE 4:  LARGEST CHANGE =      0.52328
      -2 LOG LIKELIHOOD =      21748.1776
CYCLE 5:  LARGEST CHANGE =      0.21149
      -2 LOG LIKELIHOOD =      21746.2729
CYCLE 6:  LARGEST CHANGE =      0.11883
      -2 LOG LIKELIHOOD =      21746.2470
CYCLE 7:  LARGEST CHANGE =      0.06092
      -2 LOG LIKELIHOOD =      21745.7763
CYCLE 8:  LARGEST CHANGE =      0.06538
      -2 LOG LIKELIHOOD =      21745.7612
CYCLE 9:  LARGEST CHANGE =      0.12125
      -2 LOG LIKELIHOOD =      21745.8455
CYCLE 10: LARGEST CHANGE =      0.06822
      -2 LOG LIKELIHOOD =      21745.7297
CYCLE 11: LARGEST CHANGE =      0.02593
      -2 LOG LIKELIHOOD =      21745.7125
CYCLE 12: LARGEST CHANGE =      0.02034
      -2 LOG LIKELIHOOD =      21745.7132
CYCLE 13: LARGEST CHANGE =      0.01647
      -2 LOG LIKELIHOOD =      21745.7054
CYCLE 14: LARGEST CHANGE =      0.00307
      -2 LOG LIKELIHOOD =      21745.7052
CYCLE 15: LARGEST CHANGE =      0.00297
      -2 LOG LIKELIHOOD =      21745.7050
CYCLE 16: LARGEST CHANGE =      0.00131
      -2 LOG LIKELIHOOD =      21745.7049
CYCLE 17: LARGEST CHANGE =      0.00159
      -2 LOG LIKELIHOOD =      21745.7048
CYCLE 18: LARGEST CHANGE =      0.00209
      -2 LOG LIKELIHOOD =      21745.7048
CYCLE 19: LARGEST CHANGE =      0.00107
      -2 LOG LIKELIHOOD =      21745.7047
CYCLE 20: LARGEST CHANGE =      0.00078
[NEWTON STEP]
      -2 LOG LIKELIHOOD =      21745.7047
CYCLE 21: LARGEST CHANGE =      0.00054

```

1

SUBTEST FOFAL : ITEM PARAMETERS AFTER CYCLE 21

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	DISPERSN S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
0001	-3.168	1.871	1.693	0.535	0.186	5.5	9.0
	0.986*	0.668*	0.181*	0.191*	0.032*	(0.7893)	
0002	0.631	0.793	-0.797	1.262	0.177	14.3	8.0
	0.836*	0.322*	1.352*	0.512*	0.406*	(0.0736)	
0003	1.420	1.486	-0.955	0.673	0.500	3.2	6.0
	0.625*	0.463*	0.661*	0.210*	0.210*	(0.7873)	
0004	0.679	1.796	-0.378	0.557	0.249	4.0	7.0
	0.336*	0.370*	0.245*	0.115*	0.113*	(0.7758)	
0005	0.113	1.221	-0.092	0.819	0.032	6.9	8.0
	0.320*	0.244*	0.277*	0.164*	0.120*	(0.5498)	
0006	3.595	0.992	-3.625	1.008	0.000	2.2	4.0
	4.260*	0.486*	5.897*	0.494*	4.160*	(0.7024)	
0007	2.897	1.071	-2.705	0.934	0.500	8.8	2.0
	2.580*	0.734*	4.046*	0.640*	1.243*	(0.0120)	
0008	2.863	1.040	-2.754	0.962	0.500	10.8	0.0
	3.367*	0.855*	5.295*	0.791*	1.559*	(1.0000)	
0009	-0.502	0.545	0.921	1.836	0.148	3.6	9.0
	1.087*	0.371*	1.412*	1.252*	0.320*	(0.9334)	
0010	0.298	1.300	-0.229	0.769	0.172	3.6	8.0
	0.406*	0.307*	0.357*	0.182*	0.145*	(0.8929)	
0011	0.206	0.796	-0.259	1.257	0.000	10.7	9.0
	0.749*	0.314*	1.037*	0.496*	0.366*	(0.2952)	
0012	-3.354	1.991	1.684	0.502	0.128	5.3	9.0
	1.013*	0.712*	0.174*	0.180*	0.029*	(0.8090)	
0013	0.055	0.603	-0.091	1.660	0.000	21.6	9.0
	0.808*	0.268*	1.379*	0.740*	0.379*	(0.0105)	
0014	-0.066	0.546	0.122	1.833	0.045	6.5	9.0
	1.144*	0.343*	2.023*	1.152*	0.490*	(0.6927)	
0015	0.125	0.864	-0.145	1.157	0.000	20.4	9.0
	0.618*	0.266*	0.756*	0.356*	0.282*	(0.0158)	
0016	0.327	1.406	-0.233	0.711	0.000	7.2	8.0

	0.251*	0.232*	0.208*	0.117*	0.099*	(0.5117)	
0017	-1.767	1.696	1.042	0.590	0.176	7.2	9.0
	0.537*	0.440*	0.126*	0.153*	0.045*	(0.6213)	
0018	-1.720	1.958	0.879	0.511	0.286	12.2	8.0
	0.563*	0.466*	0.133*	0.122*	0.045*	(0.1413)	
0019	1.556	1.041	-1.494	0.960	0.000	6.9	7.0
	0.558*	0.239*	0.826*	0.221*	0.417*	(0.4433)	
0020	1.590	1.435	-1.107	0.697	0.000	21.7	7.0
	0.389*	0.259*	0.430*	0.126*	0.264*	(0.0030)	
0021	-4.798	2.809	1.708	0.356	0.125	5.9	9.0
	1.647*	1.137*	0.161*	0.144*	0.021*	(0.7471)	
0022	0.755	1.433	-0.527	0.698	0.374	11.1	7.0
	0.477*	0.339*	0.433*	0.165*	0.153*	(0.1351)	
0023	0.220	1.283	-0.172	0.779	0.166	4.4	8.0
	0.430*	0.322*	0.371*	0.196*	0.151*	(0.8216)	
0024	0.387	0.338	-1.145	2.958	0.000	5.8	9.0
	2.954*	0.430*	10.167*	3.764*	1.717*	(0.7632)	
0025	-4.717	3.496	1.349	0.286	0.316	16.4	8.0
	1.916*	1.374*	0.094*	0.112*	0.027*	(0.0373)	
0026	0.525	1.144	-0.459	0.874	0.000	7.3	8.0
	0.375*	0.237*	0.408*	0.181*	0.187*	(0.5075)	
0027	0.257	1.628	-0.158	0.614	0.120	10.1	7.0
	0.266*	0.277*	0.182*	0.105*	0.085*	(0.1799)	
0028	0.028	1.067	-0.026	0.937	0.000	14.6	9.0
	0.407*	0.255*	0.387*	0.224*	0.161*	(0.1031)	
0029	-0.711	1.615	0.440	0.619	0.045	4.1	8.0
	0.238*	0.244*	0.105*	0.094*	0.042*	(0.8450)	
0030	-3.597	2.098	1.715	0.477	0.118	8.5	9.0
	0.970*	0.657*	0.154*	0.149*	0.024*	(0.4885)	
0031	-3.350	2.658	1.260	0.376	0.124	9.1	9.0
	0.836*	0.671*	0.087*	0.095*	0.024*	(0.4248)	
0032	-0.598	1.489	0.401	0.672	0.115	4.7	8.0
	0.349*	0.311*	0.170*	0.140*	0.072*	(0.7857)	

LARGEST CHANGE = 0.001

* STANDARD ERROR
284.6 244.0
(0.0380)

PARAMETER	MEAN	STN DEV
ASYMPTOTE	0.144	0.155
SLOPE	1.422	0.688
LOG(SLOPE)	0.236	0.506
THRESHOLD	-0.129	1.305

QUADRATURE POINTS AND POSTERIOR WEIGHTS:

	1	2	3	4	5
POINT	-0.4000E+01	-0.3111E+01	-0.2222E+01	-0.1333E+01	-0.4444E+00
WEIGHT	0.3904E-03	0.4901E-02	0.3138E-01	0.1394E+00	0.3003E+00
	6	7	8	9	10
POINT	0.4444E+00	0.1333E+01	0.2222E+01	0.3111E+01	0.4000E+01
WEIGHT	0.3228E+00	0.1706E+00	0.2776E-01	0.2394E-02	0.1205E-03

211452 BYTES OF NUMERICAL WORKSPACE USED OUT OF 800000 IN PHASE-2

2636 BYTES OF CHARACTER WORKSPACE USED OUT OF 80000 IN PHASE-2

anexo PH3 PORT2

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 3 ***

Exemplo de Programa da Teoria da Resposta ao Item
 Prova com 32 questões de 5 alternativas - 640 respondentes

>SCORE METHOD=1, INFO= 2, NOPRINT;

PARAMETERS FOR SCORING, RESCALING, AND TEST AND ITEM INFORMATION

METHOD OF SCORING SUBJECTS:	MAXIMUM LIKELIHOOD
SCORES WRITTEN TO FILE	C:\TRI\PORT2.SCO
TYPE OF RESCALING:	NONE REQUESTED
ITEM AND TEST INFORMATION:	TEST INFORMATION CURVES
	TABLE OF ITEM INFORMATION INDICES

TEST	NAME	QUAD POINTS	PRIOR MEAN	PRIOR STN DEV	RESCALING SCALING	CONSTANTS LOCATION
1	FOFAl	10	0.000	1.000	1.000	0.000

1

SCORING

1

SUBJECT IDENTIFICATION

WEIGHT	SUBTEST	TRIED	RIGHT	PERCENT	ABILITY	S.E.
0000013118031						
1.00	FOFAl	32	24	0.7500	1.2836	0.3298
0000013118021						
1.00	FOFAl	32	13	0.4063	-0.7694	0.4587
0000013118011						
1.00	FOFAl	32	15	0.4688	-0.5122	0.4344

CORRELATIONS AMONG SUBTEST SCORE ESTIMATES:

	FOFAl
FOFAl	1.000

MEANS AND STANDARD DEVIATIONS OF SCORE ESTIMATES:

TEST: FOFAl
 MEAN: 0.020
 S.D.: 1.189

1

ITEM INFORMATION STATISTICS FOR SUBTEST FOFAl

ITEM	MAXIMUM INFORMATION STANDARD ERROR *	POINT OF MAX INFORMATION STANDARD ERROR *	MAXIMUM EFFECTIVENESS POINT OF MAX EFFECTIVENESS *	AVERAGE INFORMATION INDEX OF RELIABILITY *
0001	0.6119	1.8292	0.0000	0.0000
	0.3717*	0.2120*	0.0000*	0.0000*
0002	0.1119	-0.4881	0.0000	0.0000
	0.0497*	1.2468*	0.0000*	0.0000*
0003	0.1991	-0.6316	0.0000	0.0000
	0.0609*	0.5761*	0.0000*	0.0000*
0004	0.5005	-0.2049	0.0000	0.0000
	0.1198*	0.2182*	0.0000*	0.0000*
0005	0.3502	-0.0448	0.0000	0.0000
	0.0805*	0.2721*	0.0000*	0.0000*
0006	0.2459	-3.6248	0.0000	0.0000
	0.2409*	5.8969*	0.0000*	0.0000*
0007	0.1034	-2.2559	0.0000	0.0000
	0.2289*	3.7798*	0.0000*	0.0000*
0008	0.0975	-2.2913	0.0000	0.0000
	0.2698*	4.9612*	0.0000*	0.0000*
0009	0.0558	1.3140	0.0000	0.0000

	0.0335*	1.1909*	0.0000*	0.0000*
0010	0.3039	-0.0449	0.0000	0.0000
	0.0737*	0.3232*	0.0000*	0.0000*
0011	0.1583	-0.2588	0.0000	0.0000
	0.1250*	1.0370*	0.0000*	0.0000*
0012	0.7748	1.7804	0.0000	0.0000
	0.4927*	0.1975*	0.0000*	0.0000*
0013	0.0908	-0.0906	0.0000	0.0000
	0.0809*	1.3786*	0.0000*	0.0000*
0014	0.0681	0.2686	0.0000	0.0000
	0.0256*	1.9603*	0.0000*	0.0000*
0015	0.1867	-0.1450	0.0000	0.0000
	0.1148*	0.7560*	0.0000*	0.0000*
0016	0.4945	-0.2327	0.0000	0.0000
	0.1633*	0.2076*	0.0000*	0.0000*
0017	0.5135	1.1851	0.0000	0.0000
	0.2091*	0.1251*	0.0000*	0.0000*
0018	0.5519	1.0533	0.0000	0.0000
	0.1907*	0.1235*	0.0000*	0.0000*
0019	0.2711	-1.4942	0.0000	0.0000
	0.1247*	0.8259*	0.0000*	0.0000*
0020	0.5151	-1.1074	0.0000	0.0000
	0.1857*	0.4296*	0.0000*	0.0000*
0021	1.5509	1.7751	0.0000	0.0000
	1.1622*	0.1811*	0.0000*	0.0000*
0022	0.2468	-0.2444	0.0000	0.0000
	0.0603*	0.3793*	0.0000*	0.0000*
0023	0.2992	0.0107	0.0000	0.0000
	0.0743*	0.3353*	0.0000*	0.0000*
0024	0.0286	-1.1437	0.0000	0.0000
	0.0309*	10.3249*	0.0000*	0.0000*
0025	1.6592	1.4535	0.0000	0.0000
	1.0446*	0.0990*	0.0000*	0.0000*
0026	0.3273	-0.4586	0.0000	0.0000
	0.1355*	0.4079*	0.0000*	0.0000*
0027	0.5254	-0.0454	0.0000	0.0000
	0.1195*	0.1698*	0.0000*	0.0000*
0028	0.2845	-0.0260	0.0000	0.0000
	0.1358*	0.3870*	0.0000*	0.0000*
0029	0.5968	0.4894	0.0000	0.0000
	0.1461*	0.1025*	0.0000*	0.0000*
0030	0.8768	1.8001	0.0000	0.0000
	0.4961*	0.1711*	0.0000*	0.0000*
0031	1.3912	1.3306	0.0000	0.0000
	0.6289*	0.0917*	0.0000*	0.0000*
0032	0.4438	0.5199	0.0000	0.0000
	0.1265*	0.1551*	0.0000*	0.0000*

52 BYTES OF NUMERICAL WORKSPACE USED OUT OF 800000 IN PHASE-3

52 BYTES OF CHARACTER WORKSPACE USED OUT OF 80000 IN PHASE-3