

Este capítulo apresenta a descrição e a análise dos resultados obtidos pela aplicação dos modelos da Teoria Clássica dos Testes, do modelo unidimensional logístico de dois parâmetros (TRI) e do modelo de Homogeneidade Monótona (TRIN) ao conjunto de 10 itens dicotomizados que mensuram a condição econômica dos alunos nas populações de referência de Roraima e Rio de Janeiro.

9.1 TRIN

9.1.1 Hipóteses do modelo de Homogeneidade Monótona (HM)

Conforme o Capítulo 3, a qualidade do ajuste do modelo HM depende do grau com que as respostas dadas ao conjunto de itens satisfazem às seguintes hipóteses: *unidimensionalidade*, *independência local* e *monotonicidade não decrescente* da Curva Característica do Item (CCI).

A verificação da dimensionalidade no conjunto de itens pode ser avaliada pela *library mokken*, através do comando *search.normal*, disponível no pacote estatístico R.18.1. Neste sentido, observa-se, em cada população de referência, a presença de uma única escala.

Para cada item, a Figura 9.1 ilustra o comportamento monótono não decrescente da função que descreve a curva CCI na população de referência de Roraima. Tal resultado é observado também na população de referência do Rio de Janeiro (Figura 9.6).

Em cada população de referência, diante da não violação das três hipóteses, conclui-se que o modelo HM se ajusta bem aos dados empíricos.

9.1.2 Os coeficientes H_{ij} , H_i e H em cada população de referência

As *popularidades* observadas (P_i) no conjunto de itens é um aspecto inerente ao conteúdo do instrumento de medida e do contexto econômico de cada população de referência que participou da Prova Brasil 2007. Além disso, estas proporções desempenham papel importante no cálculo dos coeficientes de escalonabilidade: H_{ij} , H_i e H , embora as

popularidades muito próximas entre si (P_1 , P_2 e P_6)¹ ou as inferiores a 0.20 (P_5 e P_{10}) exigem cautela na análise dos resultados destes coeficientes.

Quadro 9.1: Popularidades dos itens do capital econômico da 8ª série - Prova Brasil 2007
Roraima e Rio de Janeiro

Item i	Questionário	Tipo	P_i	P_i
	Na sua casa tem...?		RR	RJ
1	banheiro		0.95	0.98
2	televisão em cores		0.94	0.978
3	rádio		0.77	0.94
4	carro		0.34	0.36
5	alguma empregada doméstica	1: sim	0.13	0.09
6	geladeira	0: não	0.92	0.98
7	computador		0.28	0.48
8	vídeo cassete ou DVD		0.76	0.86
9	máquina de lavar roupa		0.61	0.70
10	aspirador de pó		0.06	0.18

De modo geral, em ambas as populações de referências, nota-se que os itens 1 e 2 são os mais populares, ou seja, a maioria dos alunos possui em suas residências pelo menos um banheiro e uma televisão em cores (Quadro 9.1). Como estes itens apresentam popularidades muito elevadas (próximas de 1), isto é traduzido em pouca informação para a mensuração do nível econômico dos respondentes. Assim, ao critério do pesquisador, estes itens poderiam ser excluídos.

Por outro lado, a presença de empregada doméstica ou a posse de um aspirador de pó são bens de consumo raros nos domicílios dos respondentes (Quadro 9.1), já que as *popularidades* estão muito próximas de zero.

Na análise das *popularidades* dos itens entre as escolas participantes observa-se que estas não são homogêneas em termos da posse de alguns bens de consumo como: deste aspecto (Quadro 6.3). Isto implica que no processo de composição de uma réplica da amostra conglomerada de alunos, em ambos os planos amostrais, as escolas selecionadas contribuem com a respectiva popularidade do item.

Quanto ao aspecto de discriminação, os *itens 1, 2, 3 e 5* devem ser removidos do processo de mensuração do traço latente na população de referência do RJ enquanto que o *item 3* deve ser excluído deste processo na população de referência de Roraima. Estas exclusões são sugeridas, já que estes itens possuem baixo poder de discriminação. Isto

¹ Em ambas as populações de referência, as distâncias entre as popularidades são inferiores a 0.02.

significa que eles contribuem muito pouco para a ordenação precisa dos respondentes pelo escore total do teste (Quadro 9.2).

Quadro 9.2: Coeficiente H_i da escala que mensura o capital econômico da 8ª série - Prova Brasil 2007
Roraima e Rio de Janeiro

População referencia	Item									
	1	2	3	4	5	6	7	8	9	10
RR	0.34	0.36	0.25	0.48	0.37	0.37	0.49	0.35	0.41	0.53
RJ	0.28	0.28	0.31	0.42	0.29	0.36	0.44	0.42	0.45	0.47

Com o intuito de avaliar a *escalabilidade* deste conjunto de itens dicotomizados, em cada *população de referência*, empregando a abordagem clássica da TRIN; Mokken (1971) e Sijtsma e Molenaar (2002) sugerem o uso do *coeficiente H*.

Na população de referência de Roraima, a escala construída com estes itens é classificada como fraca, já que o respectivo coeficiente H pertence ao intervalo $[0.30; 0.40)$. Por outro lado, na população de referência do Rio de Janeiro, a escala obtida é estabelecida por estes autores como razoável². Apesar do valor populacional do coeficiente H para esta segunda população de referência assumir o valor aproximado de 0.41, alguns coeficientes H_i (itens 1, 2, 3 e 5) são inferiores a 0.30 e, além disto, existir pouca variabilidade das popularidades no conjunto de itens (40% dos itens são classificados como muito populares); esta escala pode ser considerada como fraca (Quadro 9.3). De modo geral, as escalas em ambas as populações apresentam um padrão bem similar.

Quadro 9.3: Coeficiente H da escala que mensura o capital econômico da 8ª série - Prova Brasil 2007 - Roraima e Rio de Janeiro

Coeficiente H	
Roraima	Rio de Janeiro
0.39	0.41

9.2 TCT

Com o intuito de avaliar a qualidade de cada item sob o enfoque da Teoria Clássica dos Testes foram analisados os seguintes aspectos: a dificuldade e a discriminação, supondo que a unidimensionalidade esteja satisfeita no conjunto de itens em estudo, em cada população de referência.

Esta hipótese foi necessária, já que o escore total observado é definido como a soma das respostas (dicotômicas) dadas aos itens e estes, por sua vez, devem medir a única dimensão (ou o fator dominante) do traço latente.

² O coeficiente $H \in [0.40; 0.50)$.

9.2.1 Dificuldade do item

A dificuldade do item é definida como a proporção de indivíduos que acertaram o item. Esta definição coincide com o significado de *popularidade* estabelecido pela TRIN. Deste modo, a análise deste aspecto no conjunto de itens, em cada população de referência, já foi discutida na Seção 9.1.

9.2.2 Poder de discriminação do item

Como já foi visto na seção 2.3.1.2, a discriminação do item é estabelecida como a capacidade de diferenciar respondentes com escores observados mais altos de respondentes com escores observados baixos. Este conceito pode ser avaliado estatisticamente por várias medidas, dentre elas, a *correlação ponto-bisserial*. Além disso, um item apresenta baixo poder de discriminação quando o valor da *correlação ponto-bisserial* é inferior a 0.30. Sob este critério, na população de referência de Roraima, os itens 1, 2, 3, 5, 6 e 10 apresentam baixo poder de discriminação. Estes avaliam a posse de pelo menos um banheiro, uma televisão em cores, um rádio, uma empregada doméstica, uma geladeira e um aspirador de pó, respectivamente (Quadro 9.4).

Ao considerar este aspecto do item na população de referência do RJ, os resultados apontam este grupo de itens, exceto o item 10 (Quadro 9.5).

Cabe ressaltar que apesar de itens muito populares ou fáceis (itens 1 e 2) ou muito difíceis (item 10), a conclusão acima pode ser mantida, já que o valor da *correlação ponto-bisserial* destes itens não se aproxima do valor nulo (Pasquali, 2009).

Quadro 9.4: Popularidade, correlação ponto-bisserial e coeficiente H_i dos itens que mensuram o capital econômico dos alunos da 8ª série - Prova Brasil - Roraima - 2007

Item i	Na sua casa tem...?	P_i	Correlação ponto bisserial	H_i
1	banheiro	0.95	0.21	0.34
2	televisão em cores	0.94	0.24	0.36
3	rádio	0.77	0.23	0.25
4	carro	0.34	0.44	0.48
5	alguma empregada doméstica	0.13	0.24	0.37
6	geladeira	0.92	0.26	0.37
7	computador	0.28	0.42	0.49
8	vídeo cassete ou DVD	0.76	0.33	0.35
9	máquina de lavar roupa	0.61	0.39	0.41
10	aspirador de pó	0.06	0.27	0.53

Quadro 9.5: Popularidade, correlação ponto-biserial e coeficiente H_i dos itens que mensuram o capital econômico dos alunos da 8ª série - Prova Brasil - Rio de Janeiro - 2007

Item i	Na sua casa tem...?	P_i	Correlação ponto biserial	H_i
1	banheiro	0.98	0.14	0.28
2	televisão em cores	0.98	0.15	0.28
3	rádio	0.94	0.18	0.31
4	carro	0.36	0.37	0.42
5	alguma empregada doméstica	0.09	0.15	0.29
6	geladeira	0.98	0.18	0.36
7	computador	0.48	0.40	0.44
8	vídeo cassete ou DVD	0.86	0.30	0.42
9	máquina de lavar roupa	0.70	0.37	0.45
10	aspirador de pó	0.18	0.33	0.47

De modo geral, o conjunto de itens dicotomizados para mensurar a condição econômica dos alunos, em ambas as populações de referência, apresentou baixa fidedignidade; já que os respectivos valores do *alfa de Cronbach* são inferiores a 0.70. Este resultado aponta que alguns dos fatores podem estar influenciando a baixa consistência interna: o número pequeno de itens ou a qualidade dos itens em termos de dificuldade e poder de discriminação.

Estes resultados permitiram concluir que os itens muito populares, muito difíceis e, além disso, com baixo poder de discriminação podem ser removidos do processo de mensuração. No entanto, como um dos objetivos deste capítulo é avaliar, sob o enfoque das três teorias, a qualidade da escala construída com este conjunto de itens, optou-se por mantê-los.

9.3TRI

O modelo unidimensional logístico de dois parâmetros apresentado na Seção 2.4.2.2 permite que se construa uma medida estatística (escala) da condição econômica familiar do aluno quando se pressupõe uma única dimensão latente e as variáveis (itens) envolvidas sejam todas dicotômicas.

Para uma dada população de referência, o modelo é definido por:

$$\Pr(U_{ik} = 1 | \beta_k) = \frac{e^{Da_i(\beta_k - b_i)}}{1 + e^{Da_i(\beta_k - b_i)}} \quad (9.1)$$

com $i = 1, 2, \dots, J$ itens e $k = 1, 2, \dots, N$ representando os k alunos avaliados pela Prova Brasil em 2007 na população de referência.

Classificando-se como possuidor ou não do bem de consumo, o modelo (9.1) é interpretado como:

- U_{ik} , variável aleatória dicotômica que assume o valor 1 quando o k -ésimo aluno, na dada população de referência, possui o bem de consumo representado pelo item i , ou assume o valor 0 quando o k -ésimo aluno não possui o bem de consumo representado pelo item i .
- β_k , é o traço latente do k -ésimo aluno, ou seja, a condição econômica deste aluno estimada pelo modelo acima; $\Pr(U_{ik} = 1 | \beta_k)$: probabilidade do k -ésimo aluno possui o bem de consumo representado pelo item i (acertar o item i), dada sua condição econômica β_k (traço latente); a_i : parâmetro que representa o poder de discriminação do item i , ou seja, discrimina os alunos com condição econômica elevada daqueles que possuem baixa condição em relação ao item i ; b_i : parâmetro que representa a dificuldade do item i , ou seja, a abscissa na escala da variável latente β para o qual existe a probabilidade de 0.50 do aluno acertar o item i ; D : uma constante que assume os valores 1.7 ou 1.

Neste sentido, para o ajuste deste modelo com $D=1$, em cada população de referência, foi empregada a *library ltm* disponível no pacote R (versão 18.1). Com os resultados obtidos, foi possível avaliar os seguintes aspectos de cada item: dificuldade (parâmetro a), discriminação (parâmetro b), curva característica do item (CCI), curva de informação do item e curva de informação do teste.

Para a população de referência de Roraima, os parâmetros: dificuldade (b) e discriminação (a) e os respectivos erros-padrão - referentes a cada um dos 10 itens - são apresentados no Quadro 9.6 e também representados graficamente na curva característica de cada item (CCI) (Figura 9.2).

De modo geral, os valores encontrados para o parâmetro de dificuldade de cada item pertencem ao intervalo $[-3; 3]$; sendo que os itens 1 (possuir banheiro) e 2 (possuir televisão em cores) têm os menores valores de b ($b \cong -3$), o item 6 ($b \cong -2$) e os itens 5 e 10 ($b > 2$) (Quadro 9.6). Isto significa que:

- Para os alunos com estimativas de traço latente superior ao valor -2.62 , a probabilidade de possuírem banheiro e televisão em cores em suas residências supera a probabilidade de 0.50.
- Para alunos com condição econômica que se afasta cerca de 2 desvios-padrão da média, a probabilidade de possuírem em seus domicílios a presença de alguma empregada doméstica e aspirador de pó ultrapassa 0.50.

De acordo com Andrade et al (2000), os valores b mais apropriados devem estar situados no intervalo $[-2; 2]$. Esta recomendação permite ao pesquisador refletir se itens muito fáceis (1 e 2) ou itens com dificuldade elevada (5 e 10) devem permanecer no instrumento de medida.

Segundo Baker (2001), os itens podem ser classificados quanto ao poder de discriminação. Assim, de acordo com o Quadro 9.6, nota-se a seguinte classificação: moderada discriminação (itens 3, 5 e 8), alta discriminação (itens 6 e 9) e muito alta discriminação (itens 4, 7 e 10). Vale destacar que os itens 1 e 2 não foram incluídos nesta classificação, pois são itens muito fáceis.

Além disso, o item 3, apesar de moderada discriminação ($a = 0.72 < 1$) contribui pouco para a mensuração do capital econômico, já que tanto alunos com baixa condição econômica quanto alunos com alta condição econômica têm praticamente a mesma probabilidade de possuir pelo menos um rádio em suas residências (Figura 9.2).

Na Figura 9.3 é possível visualizar, por exemplo, a curva de informação dos itens 3 e 10. No primeiro item, a maior contribuição para a medida estimada da condição econômica ocorre no intervalo $(-4; 0)$ e atinge seu valor máximo em 0.12 quando o parâmetro $b = \beta$. Por outro lado, o item 10 atinge o máximo de informação em 0.8 quando os alunos possuem condição econômica no intervalo $(1; 3)$. Desta forma, o item 10 apresenta maior quantidade de informação para a mensuração da condição econômica do que o item 3.

Na Figura 9.4 é apresentada a curva de informação do teste construída a partir do conjunto dos 10 itens dicotomizados, sob o ajuste do modelo 2PL. Como se pode observar, os maiores níveis de informação da medida estimada ocorre para os alunos com condição econômica no intervalo $(0,1]$ (aproximadamente 30% dos casos para uma distribuição normal). Este conjunto de itens, portanto, contribui muito pouco para discriminar os alunos nesta população de referência com níveis na escala obtida compreendidos nos intervalos $[-4; 0)$ e $(1; 4]$ (Figura 9.5). Desta forma, conclui-se que este conjunto de itens é pouco informativo (muito restritivo) para a mensuração do construto de capital econômico, sob o modelo especificado, nesta *população de referência*.

Ao considerar a população de referência do Rio de Janeiro, os resultados mostram que os aspectos de dificuldade e discriminação de cada item bem como a qualidade da escala obtida sob o modelo logístico unidimensional de dois parâmetros são distintos.

Quanto ao aspecto de dificuldade, os itens 1, 2, 3 e 6 são os mais fáceis (b_i próximo de -3) enquanto que o item 5 é o mais difícil ($b > 3$). Vale destacar que estes itens não se enquadram no intervalo $[-2; 2]$ conforme a recomendação de Andrade et al (2000) (Figura 9.7 e Quadro 9.7).

Quadro 9.6: Estimativas dos coeficientes do modelo logístico 2PL - Prova Brasil - Roraima - 2007

Item	Na sua casa tem...?	Dificuldade	Discriminação
		(b)	(a)
1	banheiro	-2.76	1.33
2	televisão em cores	-2.62	1.31
3	rádio	-1.88	0.72
4	carro	0.54	2.01
5	alguma empregada doméstica	2.14	1.05
6	geladeira	-2.24	1.41
7	computador	0.76	2.10
8	vídeo cassete ou DVD	-1.21	1.20
9	máquina de lavar roupa	-0.42	1.40
10	aspirador de pó	2.14	1.79

Quanto ao poder de discriminação, os itens 4, 7, 8, 9 e 10 são classificados como os mais discriminativos, já que o valor do parâmetro a associado ao item é superior a 1. Cabe enfatizar que os itens muito fáceis, com valor de $a > 1$, não foram considerados nesta análise (Quadro 9.7).

Quadro 9.7: Estimativas dos coeficientes do modelo logístico 2PL - Prova Brasil Rio de Janeiro - 2007

Item	Na sua casa tem...?	Dificuldade	Discriminação
		(b)	(a)
1	banheiro	-3.71	1.24
2	televisão em cores	-3.65	1.20
3	rádio	-3.07	1.05
4	carro	0.57	1.40
5	alguma empregada doméstica	3.30	0.77
6	geladeira	-3.15	1.64
7	computador	0.06	1.59
8	vídeo cassete ou DVD	-1.78	1.46
9	máquina de lavar roupa	-0.77	1.56
10	aspirador de pó	1.31	1.70

Ao avaliar o intervalo nos quais estes itens têm maior poder de discriminação, observa-se que os maiores níveis de informação são obtidos em torno do parâmetro de dificuldade de cada item (Figura 9.8).

De acordo com a curva de informação total, nota-se que os maiores níveis da informação (entre 1.5 e 2) captados pela medida estimada, alcança indivíduos com traço latente no intervalo $[-4; 1]$. Esta conclusão é sustentada com a análise da curva que descreve o

comportamento do erro-padrão da escala obtida pela estimação do modelo especificado com o conjunto de itens (Figuras 9.9 e 9.10).

9.4 As possíveis relações entre TRIN, TCT e TRI

Apesar do conceito de dificuldade do item ser medido como uma proporção (TCT e TRIN) ou como o valor do traço latente para o qual a probabilidade do respondente acertar (possuir o bem de consumo) é 0.50 (TRI); a ordenação dos itens por este aspecto é preservada nos resultados obtidos pelas três teorias, em ambas as populações de referências.

Na população de referência de Roraima, ao comparar o aspecto de discriminação do item pelo critério: correlação ponto bisserial (TCT) versus coeficiente H_i (TRIN); observa-se, no enfoque da segunda teoria, que os itens 1, 2, 4, 5, 6, 7, 8, 9 e 10 são classificados como *discriminativos* ($H_i > 0.30$). Mas esta conclusão não é sustentada pela primeira teoria. A concordância entre estas duas abordagens acontece apenas para os itens 4, 7, 8 e 9.

Ao analisar este aspecto pela TRI, somente os itens 4 e 7 apresentam tal classificação.

Na população de referência do Rio de Janeiro, quanto ao aspecto discriminativo do item, observa-se a concordância das três teorias nos itens 4, 7, 8, 9 e 10. Mas, a análise da região de discriminação de cada item pode ser complementada com a curva característica do item que é objeto apenas da TRI.

Em ambas as populações de referência, tanto pela TRIN quanto pela TRI, a escala obtida para mensurar a condição socioeconômica do aluno pode ser considerada fraca. No entanto, no enfoque da segunda teoria, é possível visualizar a região de pouca informação da escala produzida pelo ajuste do modelo especificado (curva de informação do teste) para em análises futuras incluir novos itens para uma maior cobertura (melhor mensuração) do construto.

Na população de referência de Roraima, apesar da correlação alta³ entre o escore total e a estimativa do traço latente obtida sob o modelo 2PL (Gráfico 9.11), existem limitações no ordenamento dos indivíduos pelo escore produzido pela TRIN. Por exemplo, existem indivíduos que obtiveram o escore de 2 pontos⁴ sem possuir em suas residências banheiro e rádio (itens 1 e 3). Para estes, as estimativas de traço latente ($\hat{\beta}_{2PL}$) variam de -1.66 a -1.33. Este achado é observado também na população de referência do Rio de Janeiro (Gráfico 9.12).

³ Valor próximo de 0.98.

⁴Existem nove combinações de respostas (acertos) entre cada par de itens que correspondem ao escore de 2 pontos.