



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

O Modelo de Samejima no estudo da relação $NSE \times$ Desempenho
Via Teoria da Resposta ao Item

Fabrício Martins da Costa

Orientador: Prof. Dr. Héilton Ribeiro Tavares

Belém
2010

Fabrcio Martins da Costa

O Modelo de Samejima no estudo da relaçaõ NSE \times Desempenho
Via Teoria da Resposta ao Item

Dissertaçaõ de Conclusãõ de
Curso apresentada ao Pro-
grama de Pós-Graduaçaõ em
Matemática e Estatística da
Universidade Federal do Pará
para a obtençaõ do Título de
Mestre em Estatística.

Área de Concentraçaõ: Teoria da Resposta ao Item

Orientador: Prof. Dr. Héilton Ribeiro Tavares

Belém
2010

Fabrício Martins da Costa

**O Modelo de Samejima no estudo da relação NSE \times Desempenho
Via Teoria da Resposta ao Item**

Dissertação apresentada para a obtenção do Título de Mestre em Estatística no Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará.

Belém, 5 de Julho de 2010

Prof. Dr. Giovany Malcher Figueiredo
Coordenador do Programa de Pós-Graduação em Matemática e Estatística

Banca Examinadora

Prof. Dr. Héilton Ribeiro Tavares
Universidade Federal do Pará
Orientador

Prof. Dr. Antônio Pereira Junior
Universidade Federal do Rio Grande do Norte
Examinador

Prof. Dr. Marcus Pinto da Costa da Rocha
Universidade Federal do Pará
Examinador

*A minha mãe Ana Maria,
por tudo que me ensinou.*

A minha irmã Fernanda pelo apoio.

E em especial a meu pai José Maria.

E principalmente a minha esposa Brenda e minha filha Beatriz.

AGRADECIMENTOS

A Deus, que sempre me amparou nos momentos mais difíceis de minha vida, e me concedeu uma família maravilhosa.

À Nossa Senhora que sempre me cubriu com seu manto sagrado.

À minha esposa Brenda Tainã, pelo amor, carinho, e dedicação.

À minha filha Beatriz que é a alegria da casa.

Ao meu orientador, Prof. Dr. Héilton Ribeiro Tavares, por elucidar todas as dúvidas que surgiram no decorrer deste trabalho de pesquisa, pois seu conhecimento científico colaborou bastante para conclusão desta dissertação.

À Profa Dra. Regina Madruga Tavares que sempre ajudou-me nas horas de dificuldades e demonstrando paciência e conhecimento profundo.

À Universidade do Estado do Pará.

À Universidade Federal do Pará.

À Coordenação do Curso de Pós-Graduação em Matemática e Estatística, representado pelo Prof. Dr. Giovany Malcher Figueiredo.

E aos meus colegas de turma: Aílton Corecha, Francisco Fialho, Natália Noronha, Márcio Cruz, Fábio José, Dennison Carvalho, Adriana Moraes, Ana Garcêz.

E em especial a minha colega de trabalho Michelle Borges pela ajuda no inglês.

E em especial ao Fábio dos Santos Silva por me ajudar na sintaxe do Multilog.

E também a todos os Legionários e Legionárias que pertencem a Legião de Maria, que acreditaram e oraram por mim.

RESUMO

COSTA, Fabrício Martins. O Modelo de Samejima no estudo da relação NSE \times Desempenho via Teoria da Resposta ao Item. Dissertação de Mestrado - (Mestrado em Matemática e Estatística - UFPA, Belém - PA, Brasil).

O presente trabalho visa à criação de uma escala de Nível Socioeconômico (NSE) para que escolas similares possam ser comparadas. Ela será construída a partir do Questionário Socioeconômico respondidos pelos alunos inscritos nas provas do Enem nos anos de 2006 a 2008, via Teoria de Resposta ao Item (TRI). Esta metodologia para variáveis latentes tem sua origem entre os anos de 1935 a 1940 e propõe modelos que representem a relação entre a probabilidade de uma certa resposta a um item (questão) e a habilidade ou traço latente de um indivíduo. Mais recentemente, a TRI vem tornando-se a técnica predominante no campo de testes em vários países. No Brasil, a TRI foi usada pela primeira vez em 1995 na análise dos dados do Sistema Nacional de Ensino Básico - SAEB. A partir dos resultados obtidos no SAEB, outras avaliações, como, por exemplo, o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP, também foram implementadas através da TRI. O modelo de resposta ao item adotado neste trabalho para estimação do Nível socioeconômico é o Modelo de Resposta Gradual (Samejima, 1969), implementado no software MULTILOG (THISSEN, 1991). Obtidos os NSE dos indivíduos, que serão agregados para a construção dos NSE das Escolas, poderemos comparar as notas médias de escolas com mesmo NSE, o que elimina muitas das críticas até então de comparação de desempenhos de escolas com estruturas tão distintas.

Palavras Chave: Enem, Nível socioeconômico e Modelo gradual.

ABSTRACT

COSTA, FABRICIO MARTINS. The Model Samejima in study of the relation SEL \times performance by Item Response Theory. Dissertation Masters - (Masters in Mathematics and Statistics - UFPA, Belém - PA, Brazil).

The present work aims at to the creation of a scale of Socioeconomic Level (NSE) so that similar schools can be compared. It will be constructed from the Socioeconomic Questionnaire answered by the enrolled pupils in the tests of the Enem in the years of 2006 to 2008 by Theory of Reply to Item (TRI). This methodology for latent variable has its origin enters the years of 1935 to 1940 and considers models that represent the relation between the probability of a certain reply to an item (question) and the ability or latent trace of an individual. More recently, the TRI has been become it predominant technique in the field of tests in some countries. In Brazil, the TRI was used for the first time in 1995 in the analysis of the data of the National System of Basic Education - SAEB. From results gotten in the SAEB, other evaluations, as for example, the System of Evaluation of School Income of the State of São Paulo - SARESP, had been also implemented through the TRI. The model of reply to the item adopted in this work to estimate the Socioeconomic Level is the Model of Gradual Reply (Samejima, 1969) implemented in MULTILOG (THISSEN, 1991) software. The average notes of schools with same NSE could be compared of the NSE of the individuals that will be added for the construction of the NSE of the Schools. It eliminates many of the critical until then of comparison of performances of schools with so distinct structures.

Keywords: Enem, Socioeconomic Level and Gradual Model.

SUMÁRIO

RESUMO	vi
ABSTRACT	vii
LISTA DE TABELAS	x
LISTA DE FIGURAS	xi
1 Introdução	1
1.1 Aspectos gerais	1
1.2 Justificativa e Importância do Trabalho	2
1.3 Objetivos	3
1.3.1 Objetivo Geral	3
1.3.2 Objetivos Específicos	3
1.4 Estrutura do Trabalho	3
2 Teoria de Resposta ao Item	4
2.1 Introdução	4
2.2 Definição	5
2.3 Modelos de Resposta ao Item	6
2.3.1 Modelos Logísticos para itens dicotômicos	6
2.3.2 O Modelo Logístico de 3 parâmetros (ML3)	7
2.3.3 Curva Característica do Item	7
2.3.4 O Modelo Logístico de 2 parâmetros (ML2)	9
2.3.5 Modelos Politômicos	9
2.4 Modelo de Resposta Gradual de Samejima	10
2.4.1 Interpretação	12
2.4.2 Função de Informação do Item	12
2.4.3 Função de Informação do Teste	13
3 Metodologia	15
3.1 Valor agregado	15
3.1.1 Introdução	15
3.1.2 Conceitos Fundamentais	15
3.1.3 Medidas de Valor agregado versus Escores de crescimento	15
3.1.4 Determinação dos efeitos da escola	16

3.1.5 Sistema de avaliação do Valor agregado	17
3.1.6 Aplicações do sistema de avaliação do Valor agregado	18
3.2 Modelagem do Valor agregado	18
3.3 A Função de Produção Educacional – FPE	19
3.4 Metodologia do Nível socioeconômico	21
3.4.1 Introdução	21
3.5 Nível socioeconômico—NSE	22
3.5.1 Descrição dos dados	22
3.5.2 Construção dos níveis	22
3.6 Itens utilizados na estimação do NSE	27
3.6.1 Análise dos dados	28
3.6.2 Parâmetros dos Itens	31
3.7 Alocação das escolas	32
3.7.1 Distribuição de Frequências do NSE	32
3.7.2 Intervalos de Desempenho	34
3.7.3 Distribuição de Frequências do Desempenho	34
3.8 Estudo comparativo	36
3.9 Análise da Associação das variáveis	36
4 Conclusões	45
4.1 Considerações finais	45
4.2 Recomendações para trabalhos futuros	46
BIBLIOGRAFIA	48
A SINTAXE	52
A.1 Sintaxe do Multilog para Estimação dos Parâmetros do Itens	52
A.2 Sintaxe do Multilog para Estimação dos Escores(NSE)	53
A.3 Sintaxe comentada – Estimação dos Parâmetros dos Itens	54
A.4 Sintaxe comentada – Estimação dos Escores (NSE)	55
A.5 Sintaxe do SPSS – Ranqueamento	55
B FIGURAS	57

LISTA DE TABELAS

3.1	<i>Totais de alunos participantes do ENEM 2006, 2007 e 2008</i>	23
3.2	<i>Forma final do indicador de Escolaridade dos pais</i>	23
3.3	<i>Forma final do indicador de Área de Ocupação dos pais</i>	24
3.4	<i>Forma Final do indicador de Cargo de Ocupação dos pais</i>	24
3.5	<i>Forma final do indicador de Renda Familiar</i>	24
3.6	<i>Codificação de oito itens de Conforto Doméstico</i>	25
3.7	<i>Forma final do indicador Televisão</i>	26
3.8	<i>Forma final dos indicadores Máquina de Lavar e Televisão a Cabo</i>	26
3.9	<i>Forma final do indicador Trabalho do Candidato</i>	26
3.10	<i>Rótulo dos indicadores do questionário socioeconômico usados na construção do Indicador</i>	27
3.11	<i>Indicadores utilizados na construção do Indicador</i>	28
3.12	<i>Medidas descritivas do Parâmetro de Discriminação, nos anos de 2006 a 2008</i>	28
3.13	<i>Parâmetro de Discriminação por ano e a diferença absoluta dos parâmetros entre os anos</i>	29
3.14	<i>Parâmetros de Dificuldade por ano</i>	30
3.15	<i>Parâmetros dos Itens estimados</i>	31
3.16	<i>Estimativas do NSE</i>	31
3.17	<i>Distribuição de frequências do NSE das escolas</i>	32
3.18	<i>Distribuição de frequências do Desempenho das escolas</i>	35
3.19	<i>Ranqueamento por Desempenho das 100 escolas com menor NSE</i>	42
3.20	<i>Ranqueamento por Desempenho das 100 escolas com maior NSE</i>	43

LISTA DE FIGURAS

2.1	<i>Exemplo de uma Curva Característica do Item – CCI</i>	8
2.2	<i>Curva Característica do Item — Modelo de Resposta Gradual</i>	12
2.3	<i>Curva de Informação do Item</i>	13
3.1	<i>Box plot dos parâmetros de Discriminação</i>	30
3.2	<i>Gráfico em Colunas do Nível socioeconômico– NSE</i>	33
3.3	<i>Gráfico em Colunas do Desempenho</i>	35
3.4	<i>Diagrama de Dispersão NSE vs Desempenho</i>	38
3.5	<i>Diagrama de Dispersão das 100 primeiras escolas</i>	39
3.6	<i>Diagrama de Dispersão das 100 últimas escolas</i>	40
3.7	<i>Diagrama de Dispersão de 100 escolas similares</i>	41
B.1	<i>Curva Característica do Item Escolaridade do pai</i>	57
B.2	<i>Curva Característica do Item Escolaridade da mãe</i>	57
B.3	<i>Curva Característica do Item Área de ocupação do pai</i>	58
B.4	<i>Curva Característica do Item Área de ocupação da mãe</i>	58
B.5	<i>Curva Característica do Item Renda Familiar</i>	59
B.6	<i>Curva Característica do Item Televisão</i>	59
B.7	<i>Curva Característica do Item Dvd</i>	60
B.8	<i>Curva Característica do Item Rádio</i>	60
B.9	<i>Curva Característica do Item Computador</i>	61
B.10	<i>Curva Característica do Item Automóvel</i>	61
B.11	<i>Curva Característica do Item Máquina de Lavar</i>	62
B.12	<i>Curva Característica do Item Geladeira</i>	62
B.13	<i>Curva Característica do Item Telefone Fixo</i>	63
B.14	<i>Curva Característica do Item Celular</i>	63
B.15	<i>Curva Característica do Item Internet</i>	64
B.16	<i>Curva Característica do Item Televisão a cabo</i>	64

Capítulo 1

Introdução

1.1 Aspectos gerais

Atualmente, vemos muitos debates a respeito da avaliação educacional. Muitos esforços já foram empreendidos para melhorá-la, podemos citar o SAEB (Sistema Nacional de Avaliação da Educação Básica), que é coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O SAEB trabalha na coleta, na sumarização e análise de dados sobre o ensino fundamental e médio, articulando-se com um conjunto de ações voltadas para a melhoria da qualidade do ensino e da aprendizagem.

Há vários mecanismos que visam analisar a situação da educação no Brasil. Por exemplo, o ENCEJA (Exame Nacional para Certificação de Competências de Jovens e Adultos) que tem como principal objetivo avaliar as habilidades e competências básicas de jovens e adultos que não tiveram oportunidade de acesso à escolaridade regular na idade apropriada. Dessa forma, o participante se submete a uma prova e, alcançando a média mínima exigida, obtém a certificação de conclusão daquela etapa educacional.

Destacamos também que, o ENEM (Exame Nacional do Ensino Médio) foi reformulado pelo Ministério da Educação e Cultura (MEC) e desse modo ganhou uma elevada importância no contexto educacional, pois, sua utilização seria uma forma de seleção unificada nos processos seletivos das universidades públicas federais.

É importante salientar que, vários estudos empíricos foram realizados. Nos Estados Unidos, tivemos o Relatório Coleman (1966), a partir dele houve uma grande reflexão a respeito da avaliação educacional, pois, os resultados deste estudo diziam que as características das escolas e dos professores não tinham nenhuma importância no desempenho escolar dos alunos, e sim as características socioeconômicas.

Em geral, quando se objetiva avaliar as escolas envolvidas em um processo, primeiro analisa-se os seus alunos, e com base nos resultados deles avalia-se as escolas. Uma metodologia muito utilizada para analisar o desempenho dos alunos no processo de apren-

dizagem é o valor agregado ou valor adicionado. Segundo Goldstein e Thomas (1996), o valor agregado é definido como o quantitativo acrescido por cada instituição de ensino(escolas), por intermédio de suas práticas, políticas e processos internos, ao desempenho acadêmico dos alunos.

O termo valor adicionado está intimamente ligado as Ciências Econômicas, pois, Santos (2003) considera que o valor adicionado, em determinada etapa da produção, é a diferença entre o valor bruto da produção e os consumos intermediários. Para De Luca (1998) o valor adicionado é a remuneração dos esforços desenvolvidos para a criação da riqueza da empresa.

Na área educacional, o cálculo do valor agregado consiste na comparação da situação do aluno ao ingressar na escola com o grau atingido por ele no final do processo de aprendizagem. Calcula-se dessa forma quanto o aluno se desenvolveu durante a permanência na escola.

Neste trabalho não utilizaremos a metodologia do valor adicionado, construiremos um Índice do nível socioeconômico a partir dos dados disponíveis no questionário socioeconômico respondido pelos inscritos no ENEM nos anos 2006, 2007 e 2008. Pois, a ideia é propor uma nova forma para analisar as escolas.

1.2 Justificativa e Importância do Trabalho

Justificamos a realização deste trabalho como alternativa à metodologia de valor adicionado, com eficiência desejada, mas com a simplicidade necessária à comparação de desempenhos de escolas similares. A proposta baseia-se em informações socioeconômicas dos estudantes, que têm se apresentado constante ao longo das edições do Enem.

É importante ressaltar que a execução deste trabalho trará uma grande contribuição para meio científico, mais especificamente na área de avaliação educacional, pois, até então não há um consenso na literatura sobre como medir o nível socioeconômico nas pesquisas Alves & Soares (2009).

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo deste trabalho é criar uma escala a partir do nível socioeconômico das escolas, e assim comparar escolas similares.

1.3.2 Objetivos Específicos

1. Investigar o nível socioeconômico dos alunos;
2. Obter o nível socioeconômico das escolas;
3. Propor uma metodologia alternativa para avaliação das escolas;
4. Propor um procedimento para comparar o nível socioeconômico de cada escola com as 100 escolas similares em termo de NSE.

1.4 Estrutura do Trabalho

A dissertação está organizada em 4 capítulos. Neste primeiro capítulo temos a motivação para o presente trabalho de pesquisa, mostrando o quanto a avaliação educacional ganhou destaque no Brasil no decorrer dos últimos anos.

No capítulo dois apresenta-se a Teoria de Resposta ao Item (TRI), com ênfase nos Modelos Politômicos, pois, a partir deles serão construídos os níveis socioeconomicos dos alunos e por conseguinte das escolas.

No capítulo três mostra-se a metodologia utilizada para a análise dos dados.

No capítulo quatro temos as conclusões e as considerações finais deste estudo.

Capítulo 2

Teoria de Resposta ao Item

2.1 Introdução

Em 1936, Richardson derivou o relacionamento entre os parâmetros do modelo da TRI e os parâmetros do modelo clássico, criando um caminho inicial para obtenção das estimativas dos parâmetros para os modelos da TRI.

O trabalho de Lawley (1943) marcou o início da teoria baseada nos itens de testes. Lawley mostrou como obter as estimativas de máxima verossimilhança para os parâmetros do modelo da curva característica do item, definiu o escore verdadeiro em termos dos itens e do teste e mostrou que o coeficiente de confiança fidedignidade pode ser expresso como função dos parâmetros dos itens.

Os primeiros modelos de resposta ao item surgiram na década de 50 e eram modelos em que se considerava que uma única habilidade, de um único grupo, estava sendo medida por um teste onde os itens eram corrigidos de maneira dicotômica.

Supostamente o primeiro a utilizar a expressão curva característica do item, em inglês *Item Characteristic Curve, ICC*, foi Tucker (1946).

Lazarsfeld (1950), foi quem introduziu o conceito de traço latente, ainda que no contexto da medida das atitudes, conceito novamente que se constituiu num parâmetro chave da nova TRI.

A principal extensão desse trabalho se deve a Lord (1952), que mostrou que as várias medidas da teoria clássica poderiam ser expressas como funções dos parâmetros da função característica do item. Lord foi o primeiro a desenvolver o modelo unidimensional de dois parâmetros, baseado na distribuição normal acumulada. Posteriormente, Lord (1968) desenvolveu o modelo unidimensional de 3 parâmetros tendo em vista a necessidade da incorporação de um parâmetro que tratasse do problema do acerto casual.

Birnbaum (1968) substituiu a função ogiva normal pela função logística, matemati-

camente mais conveniente, pois é uma função explícita dos parâmetros dos itens e da habilidade e não envolve integração.

Rasch (1960 apud VALLE, 1999) propôs o modelo de resposta gradual de 1 parâmetro, expresso também como modelo de ogiva normal, e também mais tarde descrito por um parâmetro logístico.

Samejima (1969) propôs o modelo de resposta gradual. O objetivo era obter maiores informações acerca das respostas dos indivíduos e não simplesmente se foram dadas respostas corretas ou incorretas aos itens.

Existem vários outros modelos de respostas ao item, que se diferenciam na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo. Todos os modelos contêm um ou mais parâmetros relacionados ao item e um ou mais parâmetros relacionados ao indivíduo. Detalhes dos diversos modelos existentes podem ser vistos em Linden e Hambleton (1997).

Recentemente, Bock e Zimowski (1997) introduziram os modelos logísticos de 1, 2 e 3 parâmetros para duas ou mais populações de respondentes. A introdução desses modelos trouxe novas possibilidades para as comparações de rendimentos de duas ou mais populações submetidas a diferentes testes com itens comuns.

Segundo Pasquali (1997, 1999 e 2003) a TRI é atualmente a maneira mais adequada para analisar os itens de um instrumento de medida e essa teoria tornou os métodos de análises tradicionais dos itens obsoletos.

2.2 Definição

A Teoria de Resposta ao Item - TRI é um conjunto de modelos matemáticos que representam a relação entre a probabilidade de um indivíduo escolher uma determinada resposta a um item e seu traço latente. Os vários modelos propostos na literatura dependem fundamentalmente de três fatores:

- (i) da natureza do item — dicotômicos ou não dicotômicos;
- (ii) do número de populações envolvidas — apenas uma ou mais de uma;
- (iii) e da quantidade de traços latentes que está sendo medida — apenas um ou mais de um.

Neste trabalho a construção um Índice do nível socioeconômico, NSE, será feita por meio de um dos modelos da Teoria de Resposta ao Item.

Assumiremos que o NSE é uma variável latente, isto é, característica do indivíduo, cuja observação não se dá de forma direta, no qual cuja medida é feita pela agregação de informações sobre: a educação, a ocupação e a riqueza ou rendimento dos indivíduos ALVES & SOARES (2009).

2.3 Modelos de Resposta ao Item

A TRI proporciona uma vasta quantidade de modelos que permitem trabalhar com diferentes formatos de resposta, como por exemplo, itens dicotômicos e itens politômicos.

As principais distinções entre os modelos de resposta ao item estão no número de parâmetros usados para descrever e avaliar os itens e na forma como os itens estão estruturados de forma a permitir respostas dicotômicas ou politômicas. Nos casos de itens onde se considera apenas uma resposta correta e todas as demais erradas, define-se o item como dicotômico (0 ou 1). Por outro lado, quando se atribuem créditos aos acertos parciais dos alunos, classifica-se o item como politômico.

2.3.1 Modelos Logísticos para itens dicotômicos

Na prática, os modelos logísticos para itens dicotômicos são os modelos de resposta ao item mais utilizados, sendo que há basicamente três tipos, que se diferenciam pelo número de parâmetros que utilizam para descrever o item. Eles são conhecidos como os modelos logísticos de 1, 2 e 3 parâmetros, que consideram, respectivamente:

- (i) somente a dificuldade do item;
- (ii) a dificuldade e a discriminação;
- (iii) a dificuldade, a discriminação e a probabilidade de resposta correta dada por indivíduos de baixa habilidade.

2.3.2 O Modelo Logístico de 3 parâmetros (ML3)

Definição

Dos modelos propostos pela TRI, o *modelo logístico unidimensional de 3 parâmetros (ML3)* é atualmente o mais utilizado e é dado por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (2.1)$$

com $i = 1, 2, \dots, I$, e $j = 1, 2, \dots, n$, onde:

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

θ_j representa a habilidade (traço latente) do j -ésimo indivíduo.

$P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item – FRI.

b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da habilidade.

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

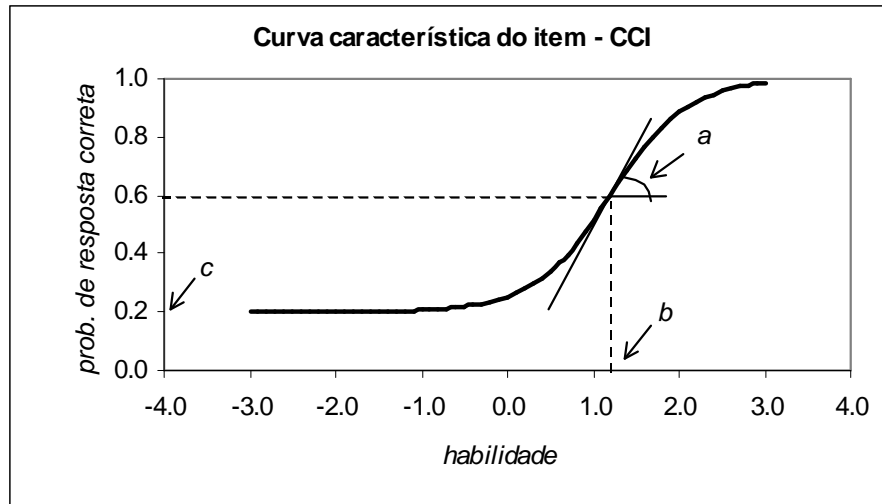
c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i (muitas vezes referido como a probabilidade de acerto casual).

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

2.3.3 Curva Característica do Item

A TRI proporciona uma descrição detalhada do comportamento individual dos itens ao longo dos níveis de habilidade. Em geral, esta Teoria assume que a relação entre a probabilidade de dar uma resposta correta a um item e o nível de habilidade dos sujeitos é uma função chamada Curva Característica do Item (CCI), conforme a Figura 2.1 abaixo.

Representação gráfica— Modelo Logístico—ML3

Figura 2.1 *Exemplo de uma Curva Característica do Item – CCI*

O modelo proposto baseia-se no fato de que indivíduos com maior habilidade possuem maior probabilidade de acertar o item e que esta relação não é linear. De fato, pode-se perceber a partir do gráfico acima que a CCI tem forma de “S” com inclinação e deslocamento na escala de habilidade definidos pelos parâmetros do item.

A escala gerada é arbitrária, e medida na mesma unidade que o parâmetro b , o parâmetro c não depende da escala Baker (2001), pois trata-se de uma probabilidade, e como tal, assume sempre valores entre 0 e 1. E o que interessa neste caso são as relações ordinais existentes e não necessariamente sua magnitude. Normalmente é pré-suposto que a escala assumira uma métrica normal (0, 1) ou seja, media 0 (zero) e desvio padrão 1 (um).

Na realidade, o parâmetro b representa a habilidade necessária para uma probabilidade de acerto igual a $(1 + c)/2$. Assim, quanto maior o valor de b , mais difícil é o item, e vice-versa.

O parâmetro c representa a probabilidade de um aluno com baixa habilidade responder corretamente o item e é muitas vezes referido como a probabilidade de acerto ao acaso.

O parâmetro a está relacionado com a inclinação da CCI. Logo, não se espera que os itens tenham a negativo, pois indicariam que a probabilidade de se responder corretamente àquele item diminui com o aumento da competência. Quanto maior o valor de a maior será o poder de discriminação daquele item, isto é, o a relaciona-se com a capacidade do

item para separar os alunos que sabem aquilo que está sendo perguntado daqueles que não sabem.

2.3.4 O Modelo Logístico de 2 parâmetros (ML2)

A partir do modelo logístico de 3 parâmetros, podemos obter outros.

Por exemplo, quando não existe possibilidade de acerto ao acaso, pode-se considerar $c = 0$ no modelo anterior e tem-se o chamado *modelo logístico unidimensional de 2 parâmetros (ML2)*, dado por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (2.2)$$

com $i = 1, 2, \dots, I$, e $j = 1, 2, \dots, n$.

2.3.5 Modelos Politômicos

Os modelos para itens politômicos dependem da natureza das categorias de resposta. Em testes de múltipla escolha em que as categorias não são ordenadas, o modelo é denominado Modelo Nominal e, nos casos em que as categorias são ordenadas, o modelo é denominado Modelo Ordinal, por exemplo, quando as categorias dos itens são dadas conforme o modelo da escala de Likert (EMBRETSON & REISE, 2000). Desse modo leva-se em consideração não só a resposta certa ou errada, mas também qual a resposta dada pelo respondente (ANDRADE, TAVARES E VALLE, 2000).

O modelo desenvolvido por Bock (1972) de resposta nominal assume que não há nenhuma ordenação a priori das opções de resposta. Com este modelo, é possível calcular a probabilidade de resposta para cada uma das k categorias existentes. Mais detalhes sobre este modelo podem ser encontrados em Azevedo (2003).

O modelo de resposta gradual de Samejima (1969) assume que as categorias de resposta de um item podem ser ordenadas entre si. Um caso particular deste modelo é o modelo de escala gradual, em que se supõe que os escores das categorias são igualmente espaçados (ANDRICH, 1978). E este modelo será mais explorado na próxima seção, pois, é a partir dele que será estimado o nível socioeconômico – NSE.

Muraki (1992) desenvolveu um modelo de crédito parcial generalizado, em que a suposição de igualdade do poder de discriminação dos itens foi relaxada.

O Modelo de crédito parcial foi desenvolvido por Andrich em 1978 e reformulado por

Masters em 1982 (MURAKI E BOCK, 1997) para análise de itens que não apresentavam apenas as categorias certo e errado ou falso e verdadeiro. O item é pontuado parcialmente, e o respondente ganha créditos à medida que sua resposta se aproxima da completa.

2.4 Modelo de Resposta Gradual de Samejima

Este modelo, como o modelo de Bock, tenta obter mais informação das respostas dos indivíduos do que simplesmente se eles deram respostas corretas ou incorretas.

Suponha que os escores das categorias de um item i são arranjados em ordem do menor para o maior e denotados por $k = 0, 1, \dots, m_i$ onde $(m_i + 1)$ é o número de categorias do i -ésimo item. A probabilidade de um indivíduo j escolher *uma particular categoria ou outra mais alta* do item i pode ser dada por uma extensão do modelo logístico de 2 parâmetros:

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}}, \quad (2.3)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$, e $k = 0, 1, \dots, m_i$, onde:

$P_{i,k}^+(\theta_j)$: é a probabilidade de um indivíduo j escolher a categoria k ou outra mais alta do item i ;

$b_{i,k}$: é o parâmetro de dificuldade da k -ésima categoria do item i .

θ_j : representa a habilidade, proficiência (traço latente) do j -ésimo indivíduo.

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

D : é um fator de escala constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultado semelhante ao da função ogiva normal.

O parâmetro de discriminação a varia a cada item, mas é constante dentro dos itens. Essa restrição de igual inclinação em cada categoria tem a finalidade de evitar probabilidades negativas.

Para o parâmetro de dificuldade $b_{i,k}$, por definição, tem-se, sem perda de generalidade que:

$$b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i},$$

ou seja, devemos ter necessariamente uma *ordenação* entre o nível de dificuldade das categorias de um dado item, de acordo com a classificação de seus escores.

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

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

Samejima também define $P_{i,0}^+(\theta_j)$ e $P_{i,m_i+1}^+(\theta_j)$ de modo que:

$$P_{i,0}^+(\theta_j) = 1$$

e

$$P_{i,m_i+1}^+(\theta_j) = 0.$$

Portanto,

$$P_{i,0}(\theta_j) = P_{i,0}^+(\theta_j) - P_{i,1}^+(\theta_j) = 1 - P_{i,1}^+(\theta_j)$$

e

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

Então, temos que:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}}. \quad (2.4)$$

Note que em um item com $(m_i + 1)$ categorias, m_i valores de dificuldade necessitam ser estimados, além do parâmetro de discriminação do item. Assim, para cada item, o número de parâmetros a ser estimado será dado pelo seu número de categorias de resposta. Se, por exemplo, tivermos um teste com I itens, cada um com $(m_i + 1)$ categorias de resposta, teremos então $\left[\sum_{i=1}^I m_i + I \right]$ parâmetros de item a serem estimados.

O símbolo Theta (θ) representa o valor da variável latente para cada aluno. Este valor é estimado pelo modelo e tão maior quanto maiores forem os valores atribuídos às respostas dos itens (MAMBRINI, 2009). É importante ressaltar que, devido ao método utilizado, são considerados apenas os itens para os quais os indivíduos forneceram resposta, desconsiderando aquelas em branco.

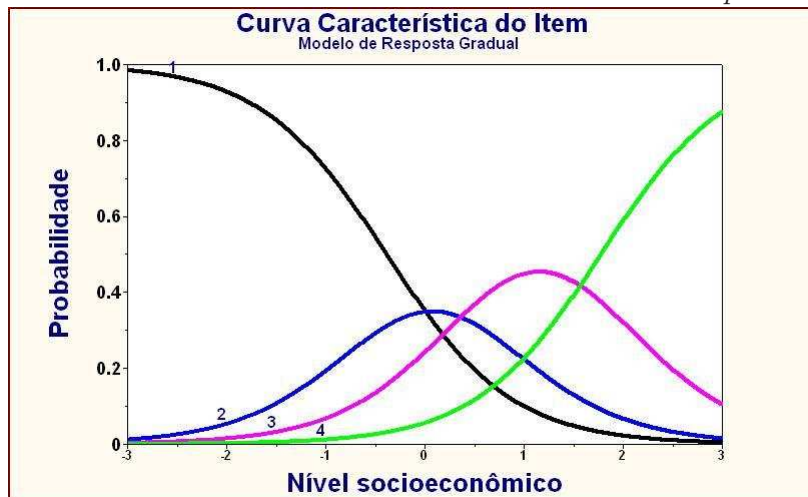
2.4.1 Interpretação

Na TRI cada item é caracterizado por um ou mais parâmetros. O parâmetro de dificuldade b é medido na mesma escala da proficiência(θ), e são representadas no eixo horizontal das figuras. Quando $b=\theta$ a probabilidade de um indivíduo escolher uma das categorias do item é igual a 50%.

O parâmetro de discriminação a descreve a capacidade de discriminação do item entre indivíduos com alto e baixo NSE (θ).

Representação gráfica—Modelo de Resposta Gradual

Figura 2.2 *Curva Característica do Item — Modelo de Resposta Gradual*



2.4.2 Função de Informação do Item

A Função de Informação do Item permite analisar quanto o item contém de informação para estimar a habilidade do indivíduo. Para o modelo dicotômico, esta função é dada

por:

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta} P_i(\theta) \right]^2}{P_i(\theta) Q_i(\theta)},$$

Em que:

$I_i(\theta)$ é a “informação” fornecida pelo item i no nível de habilidade θ ;

$P_i(\theta) = P(X_{ij} = 1|\theta)$ e $Q_i(\theta) = 1 - P_i(\theta)$.

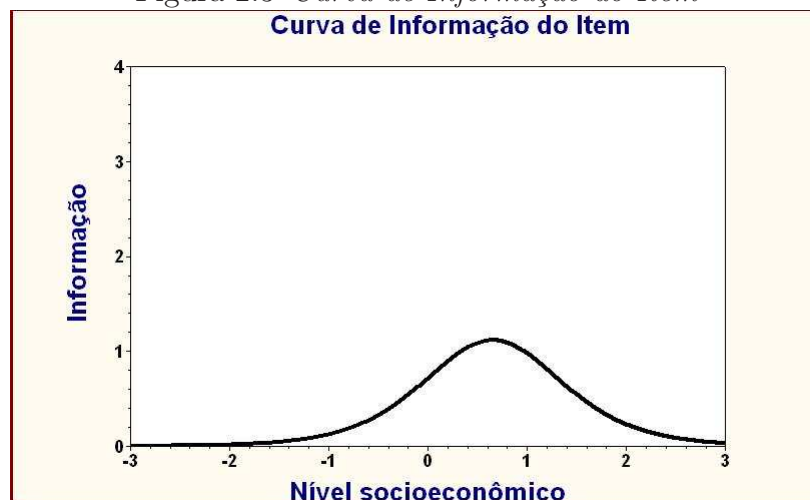
No modelo logístico de 3 parâmetros, a função de informação é dada por (BIRNBAUM, 1968)

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2.$$

Esta equação permite relacionar a função de informação com os parâmetros do item. Uma análise da equação mostra que a informação é maior quando:

- (i) quando b_i se aproxima de θ ;
- (ii) quanto maior for o a_i ;
- (iii) e quanto mais c_i se aproximar de 0.

Figura 2.3 *Curva de Informação do Item*



2.4.3 Função de Informação do Teste

A informação fornecida pelo teste é simplesmente a soma das informações fornecidas por cada item que compõe o mesmo:

$$I(\theta) = \sum_{i=1}^I I_i(\theta).$$

Pode-se também representar a função de informação do teste pelo erro-padrão de estimação, que é dado por:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}}.$$

É importante notar que essas medidas de informação dependem do valor de θ . Assim, a amplitude do intervalo de confiança para θ dependerá também do seu valor.

A função de informação apresentada anteriormente para o modelo dicotômico é um caso particular de uma expressão genérica para m categorias de resposta (SAMEJIMA, 1969). A informação de um item i é dada pela soma das informações associadas a cada categoria de resposta, ponderada pela probabilidade de escolha da categoria. Portanto, a função de informação do item é dada por

$$I_i(\theta) = \sum_{k=1}^{m_1} I_{ik}(\theta) P_{ik}(\theta).$$

em que:

$I_{ik}(\theta)$ é uma medida de informação da k -ésima categoria de resposta do item i .

$P_{ik}(\theta)$ é a probabilidade da k -ésima categoria de resposta do item i ser escolhida.

A quantidade $I_{ik}(\theta)P_{ik}(\theta)$ é o quanto a k -ésima categoria de resposta contribui para a informação do item, e é dada por:

$$I_{ik}(\theta)P_{ik}(\theta) = \frac{\left[\frac{\partial P_{ik}(\theta)}{\partial(\theta)}\right]^2}{P_{ik}(\theta)} - \frac{\partial^2 P_{ik}(\theta)}{\partial^2(\theta)}$$

Capítulo 3

Metodologia

3.1 Valor agregado

Os Modelos de Valor agregado são utilizados para estimar os efeitos da escola e do professor na aprendizagem dos alunos.

3.1.1 Introdução

O “Valor agregado” é um termo originalmente utilizado no mundo dos negócios e da economia, tornou-se amplamente utilizado para descrever certas avaliações educacionais. Por exemplo, temos o Tennessee Value-Added Assessment System (TVAAS)— Sistema de Avaliação do Valor Agregado do Tennessee, desenvolvido por William Sanders em uso desde 1996 (MCCAFFREY, LOCKWOOD, KORETZ, e HAMILTON, 2003), é considerado um dos líderes em implementação do valor agregado em sistemas de avaliação nos estados, municípios e consórcios em todo o país (CAREY 2004; KUPERMINTZ, 2003).

3.1.2 Conceitos Fundamentais

O sistema de avaliação do Valor agregado se refere a qualquer um dos vários modelos que são usados para interpretar os resultados dos testes, de modo que avalia o desenvolvimento ou o progresso de um aluno ao longo do tempo, geralmente ao longo de vários anos letivos (RUBIN, STUART, e ZANUTTO, 2004).

O conceito de uma avaliação em que as medidas de um estudante são observadas por vários anos, popularmente conhecida como *avaliação longitudinal*, existe há muito tempo na área educacional (GOLDSCHMIDT, CHOI, & MARTINEZ, 2004).

3.1.3 Medidas de Valor agregado versus Escores de crescimento

Os modelos de Valor agregado são frequentemente comparados com os *Escores de crescimento*, que há muito tempo foram utilizados na avaliação educacional. Um *Escore*

de crescimento é geralmente calculado como a diferença entre os escores do aluno no ano corrente e no ano anterior.

Os modelos de Valor agregado diferem acentuadamente dos simples cálculos de *escores de crescimento*, através da utilização de fórmulas estatísticas sofisticadas que se visam isolar os fatores não educacionais, tais como, situação socioeconômica e demográfica dos estudantes.

O escore do Valor agregado é interpretado como uma medida do efeito direto dos fatores educacionais, tais como, o professor, escola ou distrito onde o aluno estudou. (DRURY e DORAN, 2004; MCCAFFREY et al., 2003).

3.1.4 Determinação dos efeitos da escola

Um conceito fundamental da avaliação do Valor agregado é a afirmação de que, as escolas são responsáveis pela aprendizagem de cada aluno com o equivalente a um ano de crescimento, independentemente do nível de educação com que o aluno começou o ano letivo (CALLENDER, 2004; CAREY, 2004).

Especificamente, não se pode afirmar que o conhecimento adquirido pelos alunos durante qualquer uma das fases da aprendizagem é unicamente devido à contribuição da escola (MEYER, 1997; LADD e WALSH, 2002).

Se a medida do Valor agregado reflete o verdadeiro efeito de professores e das escolas em um aluno, então seria possível determinar se a contribuição para o crescimento do aluno foi suficiente.

Cada modelo de Valor agregado especifica o que constitui um ano de crescimento para um estudante. No caso de TVAAS, um ano de crescimento é estimado usando o histórico de realizações individuais de cada aluno (BIANCHI, 2003).

No final do ano, se o aluno conseguiu atingir o que foi estipulado para um ano de crescimento, o aluno é considerado como tendo recebido uma educação eficaz. Se o aluno atingir o equivalente a mais de um ano de crescimento, o estudante recebeu uma educação eficaz elevada. Os alunos que apresentam rendimento inferior a um ano de crescimento, ou seja, estão abaixo da meta, então receberam uma educação menos eficaz (HERSHBERG et al., 2004).

Baseados nas conclusões independentes do estudo de TVAAS, partidários da avaliação do Valor agregado creditam que a eficácia dos professores é muito mais importante de que qualquer outro fator não educacional para a aprendizagem dos alunos neste momento e

no futuro (BIANCHI, 2003; MCCAFFREY et al., 2003). Eles afirmam que, em relação ao nível de escolaridade, os alunos que recebem educação menos eficaz continuarão abaixo dos níveis de aprendizagem, enquanto que, os alunos que recebem educação eficaz atingirão níveis mais elevados de aprendizagem. Alguns pesquisadores educacionais apontam para a avaliação valor agregado como suporte para mostrar o papel preponderante dos professores na concretização do crescimento do aluno (VAUGHAN, 2002).

3.1.5 Sistema de avaliação do Valor agregado

Outra diferença do sistema avaliação por valor agregado para o sistema de crescimento tradicional, é que os escores são os recursos necessários. Para implementar um sistema de avaliação do valor agregado em uma sistema de ensino, várias considerações devem ser feitas:

1. Deve haver um sistema para monitorar os escores do desempenho e informações demográficas de cada aluno ao longo do tempo.
2. Os dados devem ser armazenados em um banco de dados capaz de lidar com de tamanho população de milhões de estudantes.
3. Os recursos de computacionais devem estar disponíveis para lidar com cálculos estatísticos complexos.

Embora a avaliação do Valor agregado é independente de qualquer teste de desempenho, há uma concordância de que os escores usados na análise devem produzir a realização de um teste válido e confiável, que será administrado pelo menos uma vez por ano e para poder se expresso em uma escala comum, que são os níveis de ensino (STONE, J.E., 1999; MEYER, 1996 GOLDSHMIDT; CHOI, MARTINEZ, 2004; HERDMAN et al., 2002; DRURY e DORAN, 2003). Ou seja, a avaliação deve produzir uma *escala de escores*, onde seja possível colocar os escores dos alunos em uma escala vertical que corresponde a um continuum da aprendizagem e assim poder mostrar o progresso ao longo do tempo (NITKO, 2004).

Em um estudo recente de modelos de valor agregado, os investigadores basearam-se no escore da escala *Stanford Achievement Test Series*, Nona Edição (MCCAFFREY, KORETZ, LOUIS, e HAMILTON, 2004).

3.1.6 Aplicações do sistema de avaliação do Valor agregado

Os defensores da avaliação do Valor agregado acreditam que esse novo tipo de metodologia possibilita avaliar e comparar a qualidade das escolas que tem diferentes populações de estudantes (MEYER, 1996). Esta afirmação, foi feita das conclusões retiradas de estudos, especialmente o conhecido “Relatório de Coleman”, constatou que os fatores socioeconômicos e demográficos dos alunos, tem maior impacto no desempenho escolar, em vez das escolas e professores (BRACEY, 2004). Desde a publicação desses estudos, o avaliação das escolas tem sido normalmente conhecidas como medidas de insumos educacionais, como o financiamento e as certificações de professores, ao invés de resultados de testes (MEYER, 1996).

3.2 Modelagem do Valor agregado

Uma abordagem para estimar os determinantes do valor agregado para escola deriva da fundamentação teórica da função de produção. Após o trabalho de Hanushek (1979) e Hanushek e Taylor (1990), a educação pode ser tratada como um processo de produção, em que a escolarização um indivíduo é determinado por suas características pessoais, antecedentes familiares e insumos escolares. Após Hanushek (1992), podemos expressar a função de produção educacional como um modelo linear:

$$A_{i,t} = \alpha A_{i,t-1} + \mathbf{S}_{i,t}\boldsymbol{\beta} + \mathbf{P}_{i,t}\boldsymbol{\gamma} + \epsilon_{i,t}, \quad (i = \text{escola}) \quad (3.1)$$

onde $A_{i,t}$ é o nível de educação dos alunos no final de um período letivo de tempo (t), $A_{i,t-1}$ é o rendimento anterior do aluno, $\mathbf{S}_{i,t}$ são os fatores relacionados a escola, $\mathbf{P}_{i,t}$ são os fatores relacionados aos alunos (incluindo a renda familiar) que determinam o rendimento educacional durante o período de tempo t , α , $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ são os parâmetros a serem estimados e $\epsilon_{i,t}$, é o termo do erro residual.

A base deste raciocínio é a função de produção educacional, que analisa o ganho de conhecimento do aluno ao longo do tempo. A inclusão do nível anterior de rendimento, destina-se a verificar os efeitos sobre o nível de conhecimento atual para todos os insumos passados (entradas), incluindo características hereditárias. Subtraindo o rendimento anterior em ambos os lados da equação 3.1, obtém-se o Valor agregado para função de produção educacional:

$$A_{i,t} - A_{i,t-1} = (\alpha - 1)A_{i,t-1} + \mathbf{S}_{i,t}\boldsymbol{\beta} + \mathbf{P}_{i,t}\boldsymbol{\gamma} + \epsilon_{i,t}$$

A característica fundamental da formulação do Valor agregado para a função de produção educacional é que ele enfoca os fatores que influenciam o ganho de conhecimento durante um período de tempo especificado.

Quais os fatores que podem influenciar na medida em que os alunos de uma escola adquirem conhecimentos específicos durante etapas chaves? Um grande número de estudos estão sendo realizados nos E.U.A., vem tentando estimar o nível de rendimento. Estes estudos vão desde análises de conjuntos de dados longitudinais, com amostras representativas de indivíduos (por exemplo, High School and Beyond; and the National Educational Longitudinal Study) para aqueles com base em dados de nível de escola (veja Haveman e Wolfe (1995) para uma revisão dos estudos anteriores.)

O modelo valor agregado está presente na discussão teórica feita por Todd e Wolpin (2003) sobre a função de produção educacional, na qual será abordada na próxima seção.

3.3 A Função de Produção Educacional – FPE

A Função de Produção Educacional – FPE, estuda a relação de produtividade entre os insumos e o produto final, a exemplo da função de produção que é utilizada nas Ciências Econômicas. É importante salientar que, o rendimento dos alunos, é medido por meio do seu desempenho em um dado teste, é o resultado de um processo cumulativo de aquisição de conhecimento (MACEDO 2004). Como conceito para a aquisição de conhecimento, os autores assumem a semelhança com um processo de produção, no qual os insumos presentes e passados são combinados com as características genéticas dos indivíduos para produzir o um rendimento cognitivo. A função de produção é dada por:

$$Y = f(A, F, E) + \epsilon \quad (3.2)$$

onde: Y representa o rendimento escolar dos alunos, que é função de três grupos distintos de fatores associados ao desempenho escolar: (A) características individuais dos alunos, (F) características familiares desses alunos e (E) características das escolas, incluindo professores e diretores.

Temos um problema que dificulta o processo de estimação da função de produção, que é o fato de que os bancos de dados existentes não apresentam certas informações importantíssimas a respeito dos alunos, por exemplo os insumos aos quais o aluno teve acesso, entretanto há informações sobre o momento presente.

Todd e Wolpin (2003) sugerem duas especificações para contornar este problema, que são os modelos contemporâneos e os modelos valor agregado.

Modelo de especificação contemporânea

Nestes modelos o rendimento dos alunos está relacionado apenas às medidas contemporâneas (do tempo t analisado, corresponde neste caso a uma determinada série s) dos insumos educacionais e familiares. Assumindo que:

- (i) os insumos correntes são importantes para a produção do rendimento corrente;
- (ii) os insumos não variam através do tempo, o insumo corrente é capaz de captar a “história” dos insumos;
- (iii) os insumos correntes não estão relacionados com a capacidade mental de aquisição de conhecimento.

Os modelos contemporâneos seguem a mesma especificação apresentada pela equação 3.2 , exceto pela inclusão dos fatores relativos aos insumos passados no termo de erro, e são dados por:

$$Y_{ijs} = \mathbf{y}_s(\mathbf{E}_{ijs}, \mathbf{F}_{ijs}) + \epsilon_{js} \quad (3.3)$$

onde Y_{ijs} é o rendimento do aluno i , do domicílio j , da série s ; F_{ijs} são os insumos da família do mesmo aluno; E_{ijs} os insumos educacionais e ϵ_{ijs} é o termo de erro residual que inclui todas as variáveis omitidas (insumos passados e capacidade mental de aquisição de conhecimento).

Os pressupostos assumidos para a estimação deste modelo podem ser considerados fortes. Entretanto, quando nos deparamos com dados relativamente limitados, assumir estes pressupostos são condições necessárias para que se possa estimar um modelo de especificação contemporânea.

Modelo de especificação do Valor agregado

Os modelos de Valor agregado incluem na especificação do modelo o escore de teste anterior, assumindo que este corresponde a um instrumento suficiente para todo o histórico de insumos e capacidades inatas “carregadas” pelo aluno.

Alguns estudiosos utilizam a abordagem do Valor agregado como forma de contornar a dificuldade imposta por esses fatores não observados (TODD;WOLPIN, 2003; HANUSHEK, 2003). O modelo valor adicionado é dado por:

$$Y_{ijs} = (Y_{sji-1}\lambda, X_{ijs}\alpha) + \epsilon_{js} \quad (3.4)$$

onde Y_{ijs} é o rendimento do aluno i , do domicilio j , da série s ; X_{ijs} são os insumos familiares e educacionais desse mesmo aluno, Y_{sji-1} é o rendimento do mesmo aluno na série anterior e ϵ_{ijs} é o termo de erro residual.

Ao incluir a nota do teste anterior, as demais variáveis incluídas no lado direito da equação passam a explicar apenas a variação no desempenho ocorrida entre as duas provas. Os coeficientes estimados são medidas do efeito marginal de seus fatores sobre o desempenho escolar, uma vez que o efeito acumulado destes insumos está contido no escore passado. Para tanto, é preciso que as seguintes hipóteses sejam assumidas:

- (i) Toda informação de insumos passados está representada na equação dentro do escore do teste anterior.
- (ii) A função 3.4 não varia com o tempo, ao menos dentro do intervalo considerado na implementação do modelo Valor agregado.

3.4 Metodologia do Nível socioeconômico

3.4.1 Introdução

Anteriormente para podermos comparar escolas similares, a metodologia do Valor agregado era utilizada, agora mostraremos uma nova metodologia para este problema.

Qual a motivação para implatação desta nova metodologia? Há muitas reclamações por parte das escolas no tocante a comparação, pois escolas com pouca estrutura estão sendo comparadas com outras que tem mais estrutura, um exemplo disso é a nota geral ENEM por escola, uma escola com péssima estrutura terá uma nota menor do que aquela que possui uma ótima estrutura. Então, a ideia é construir uma medida que seja livre da estrutura que escola detém. Para fazer isso, é necessário construir um modelo a partir das variáveis da escola e assim contornar o efeito escola. Entretanto, precisamos ter grupos de escolas similares e depois aplicarmos o modelo.

Adotaremos neste trabalho a estimação do nível socioeconômico — NSE, para criar os

grupos similares e depois realizar as comparações entre as escolas. O modelo adotado para estimar o NSE, será o modelo de Samejima.

3.5 Nível socioeconômico—NSE

O que é NSE? Conforme Alves e Soares (2009), a medida do NSE é, conseqüentemente, uma tentativa de se agregar estatisticamente esses estratos sociais, produzindo, assim, um resultado que possa contextualizar a medida de desempenho cognitivo dos alunos do sistema de educação básica brasileiro . Entretanto é bom salientar que não há um consenso na literatura a respeito do conceituação e nem na forma de medir o NSE em pesquisas empíricas.

3.5.1 Descrição dos dados

Para construirmos o Índice do nível socioeconômico, NSE, selecionamos as variáveis do questionário socioeconômico, no qual foi respondido pelos inscritos. Critérios para a composição do NSE:

1. Ser aluno concluinte, ou seja, terminado o Ensino Médio no ano em que se inscreveu para o ENEM.
2. Ter respondido todas as perguntas selecionadas para a composição do NSE.
3. E as escolas que tiveram o código identificado.

O estudo baseou-se nas base de dados (microdados) do ENEM nos anos 2006, 2007, e 2008 que foram disponibilizadas pelo INEP/MEC.

A Tabela 3.1 dada abaixo apresenta o número de inscritos no ENEM anos de 2006 a 2008 e o número de inscritos considerados na estimação do NSE. Ressaltamos que o mesmo questionário foi usado no ENEM 2006, 2007, e 2008.

3.5.2 Construção dos níveis

Neste trabalho adotou-se a metodologia proposta por Soares (2009) para a escolha das variáveis que seriam consideradas na estimação do NSE. Os indicadores utilizados para o cálculo do NSE foram:

Tabela 3.1 *Totais de alunos participantes do ENEM 2006, 2007 e 2008*

ANO	Alunos inscritos	Alunos considerados na estimação do NSE
2006	3.743.370	825064
2007	3.584.577	807071
2008	4.018.070	876584
Total	11.346.017	2508719

- Educação dos pais
- Ocupação dos pais
- Renda familiar
- Itens de conforto doméstico

Com relação ao indicador educação dos pais temos que, as questões Q17 e Q18 do questionário socioeconômico são referentes nível de escolaridade do pai e da mãe.

Das nove opções oferecidas nestas questões agregou-se em quatro níveis conforme vemos na Tabela 3.2 dada abaixo:

Tabela 3.2 *Forma final do indicador de Escolaridade dos pais*

Níveis	Rótulo	Opções do questionário
1	Até a 4ª série	‘Não estudou’ e ‘1ª e 4ª série do ensino fundamental’
2	Fundamental completo	‘5ª a 8ª série do ensino fundamental’ e ‘Ensino médio incompleto’
3	Médio completo	‘Ensino médio completo’ e ‘Ensino superior incompleto’
4	Superior completo	‘Ensino superior completo’ e ‘Pós-graduação’
9	NR/NS/B	‘Não sei’ ou em branco

Observando o indicador ocupação dos pais temos que, as questões Q19 a Q22, nos dão essas informações sobre a área de atuação e do cargo exercido.

A caracterização para o indicador área de atuação foi a seguinte, as opções foram agrupadas em dois níveis, e enquanto para o indicador cargo exercido as informações foram agregadas em três níveis.

Podemos observar isso através das Tabelas 3.3 e 3.4 dadas abaixo:

A questão Q23 do questionário, refere-se a renda familiar que é um indicador necessário para estimar o NSE. A partir da análise das frequências, pôde-se agregar as opções em quatro categorias, conforme Tabela 3.5 dada abaixo:

Tabela 3.3 *Forma final do indicador de Área de Ocupação dos pais*

Níveis	Rótulo	Opções do questionário
1	Nível 1	‘Agricultura, campo, fazenda ou pesca’, ‘Trabalhador do setor informal (sem carteira assinada)’, ‘Trabalha em casa, em serviços: costura, cozinha, etc.’, ‘Funcionário público: gov. fed., est. ou munic., ou militar’ e ‘Profissional liberal, professor ou técnico de nível superior’
2	Nível 2	‘Indústria’ e ‘Comércio, banco, transporte ou outros serviços’
9	NR/NS/B	‘No lar’, ‘Não trabalha’, ‘Não sei’ e em branco

Tabela 3.4 *Forma Final do indicador de Cargo de Ocupação dos pais*

Níveis	Rótulo	Opções do questionário
1	Nível 1	‘Trabalho temporário informal sem carteira assinada’ e ‘Trabalho por conta própria’
2	Nível 2	‘Empregado no setor privado com carteira assinada’, ‘Funcionário público sem função de direção’ e ‘Militar sem posto de comando’
3	Nível 3	‘Gerente, administrador ou diretor de empresa privada’, ‘Funcionário público com funções de diretor’ e ‘Militar com posto de comando’
9	NR/NS/B	‘Desempregado’, ‘Aposentado’, ‘Outra situação’ e em branco

Tabela 3.5 *Forma final do indicador de Renda Familiar*

Níveis	Rótulo	Opções do questionário
1	Até um salário mínimo	‘Até 1 salário mínimo’
2	De 1 a 2 salários mínimos	‘De 1 a 2 salários mínimos’
3	De 2 a 5 salários mínimos	‘De 2 a 5 salários mínimos’
4	Mais de 5 salários mínimos	‘De 5 a 10 salários mínimos’, ‘De 10 a 30 salários mínimos’ e ‘De 30 a 50 salários mínimos’ e ‘Mais de 50 salários mínimos’
9	NR/NS/B	‘Nenhuma renda’ e em branco

Sabemos que quando uma pesquisa indaga qual a renda familiar, as respostas não são tão fidedignas, seja por desconhecimento por parte do membro da família que respondeu ao questionário, ou por decisão de não fornecer os valores reais (ALVES & SOARES, 2009). Uma alternativa para contornar isso é a inclusão de itens de conforto doméstico

como indicadores de renda, juntamente com a renda familiar. No questionário utilizado, esses itens compreendem em 11 questões, da Q24 a Q34. Para algumas dessas variáveis, a saber, DVD, rádio, computador, automóvel, geladeira, telefone fixo, celular, internet, foram criadas três opções de respostas, conforme se vê na Tabela 3.6 abaixo:

Tabela 3.6 *Codificação de oito itens de Conforto Doméstico*

Nível	Rótulo	Opções do questionário
1	Não tem	‘Não tem’
2	Um	‘Um’
3	Dois ou mais	‘Dois’ e ‘Três ou mais’
9	NR/NS/B	‘Em branco’

Para os itens de conforto doméstico: televisão, máquina de lavar e televisão a cabo a recodificação foi obtida depois da análise das frequências e da constatação de que dois níveis seriam suficientes para caracterizar tais itens, um nível que registra a presença e um segundo que registra a ausência do respectivo item no domicílio do aluno.

Tabela 3.7 *Forma final do indicador Televisão*

Níveis	Rótulo	Opções do questionário
1	Até uma TV	‘Não tem’ e ‘Um’
2	Duas ou mais TV	‘Dois’ e ‘Três ou mais’
9	NR/NS/B	‘Em branco’

Tabela 3.8 *Forma final dos indicadores Máquina de Lavar e Televisão a Cabo*

Níveis	Rótulo	Opções do questionário
1	Não tem	‘Não tem’
2	Tem	‘Um’, ‘Dois’ e ‘Três ou mais’
9	NR/NS/B	‘Em branco’

Por fim, a opção que averigua se o aluno trabalhou durante o ensino médio, a Q43, foi considerada para a construção do indicador socioeconômico. As opções foram elencadas de forma que o nível mais baixo representa o aluno que trabalhou durante todo o período e o nível mais alto refere-se àquele que nunca trabalhou.

Tabela 3.9 *Forma final do indicador Trabalho do Candidato*

Níveis	Rótulo	Opções do questionário
1	Sim, todo o tempo	‘Sim, todo o tempo’
2	Sim, entre 1 e 3 anos	‘Sim, de 1 a 2 anos’ e ‘Sim, de 2 a 3 anos’
3	Sim, menos de 1 ano	‘Sim, menos de um ano’
4	Não trabalhou	‘Não’
9	NR/NS/B	Em branco

A Tabela 3.10 dada abaixo mostra a correspondência entre as questões do questionário e seus novos nomes usados na nossa análise.

Tabela 3.10 *Rótulo dos indicadores do questionário socioeconômico usados na construção do Indicador*

Questionário	Nome utilizado na análise
MASC_INSCRITO	ID Aluno
Q17	Escolaridade do pai
Q18	Escolaridade da mãe
Q19	Área de Ocupação do pai
Q20	Cargo de Ocupação do pai
Q21	Área de Ocupação da mãe
Q22	Cargo de Ocupação da mãe
Q23	Renda Familiar
Q24	Televisão
Q25	DVD ou Vídeo
Q26	Rádio
Q27	Computador
Q28	Automóvel
Q29	Máquina de Lavar
Q30	Geladeira
Q31	Telefone Fixo
Q32	Celular
Q33	Internet
Q34	Televisão a Cabo
Q43	Trabalho do Candidato

3.6 Itens utilizados na estimação do NSE

Depois da escolha das variáveis que seriam utilizadas para estimação do NSE, o modelo de Samejima (1969) foi aplicado utilizando-se a implementação do software MULTLOG (THISSEN, 1991). E a partir desta análise verificou-se que a ocupação da mãe e do pai (Q20 e Q22) e o trabalho do aluno (Q43) não fornecem informações consistentes para a construção do nível socioeconômico, sendo por isso, desconsideradas do indicador final. Na tabela 3.11 dada abaixo temos, as variáveis que foram adotadas na estimação final do NSE.

Tabela 3.11 *Indicadores utilizados na construção do Indicador*

Questionário	Nome utilizado na análise
MASC_INSCRITO	ID Aluno
Q17	Escolaridade do pai
Q18	Escolaridade da mãe
Q19	Área de Ocupação do pai
Q21	Área de Ocupação da mãe
Q23	Renda Familiar
Q24	Televisão
Q25	DVD ou Vídeo
Q26	Rádio
Q27	Computador
Q28	Automóvel
Q29	Máquina de Lavar
Q30	Geladeira
Q31	Telefone Fixo
Q32	Celular
Q33	Internet
Q34	Televisão a Cabo

3.6.1 Análise dos dados

Inicialmente a estimação dos parâmetros dos itens foi feita ano a ano, isto é, primeiro para o ano de 2006, e depois para os demais anos, 2007 e 2008. Após a estimação dos mesmos verificou-se que os valores não mudavam muito de um ano para outro, e isso foi de fundamental importância, pois, como um dos objetivos é estimar o NSE das escolas, então teremos uma análise mais consistente em virtude dos parâmetros não oscilarem muito nos três anos de estudos.

Tabela 3.12 *Medidas descritivas do Parâmetro de Discriminação, nos anos de 2006 a 2008*

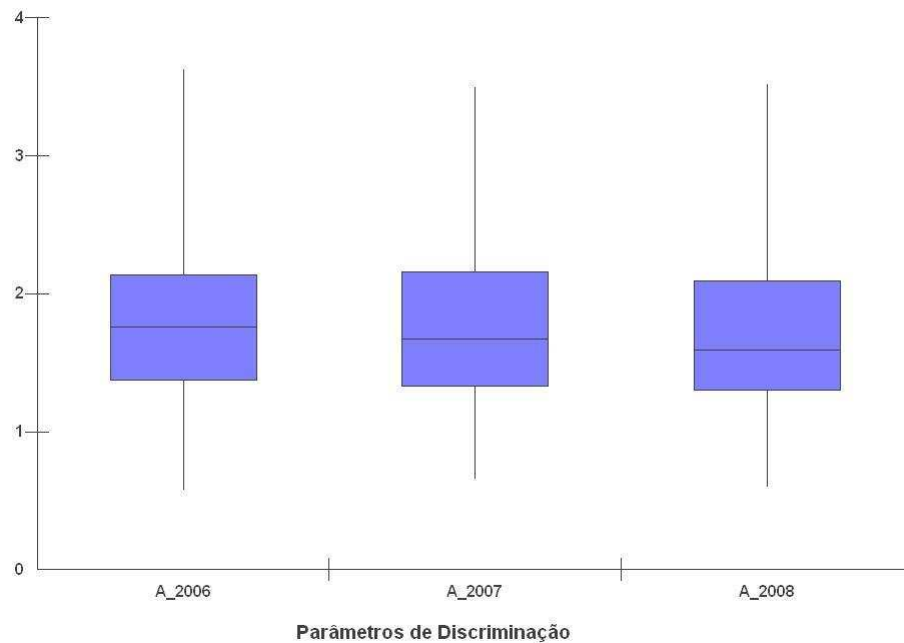
Ano	Média	Desvio padrão	Primeiro quartil	Terceiro quartil	Coefficiente de variação
2006	1.849389	0.817664	1.3784	2.1400	44.21%
2007	1.805914	0.783520	1.3326	2.1555	43.39%
2008	1.747522	0.790118	1.3008	2.0961	45.21%

Na Tabela 3.13 dada abaixo temos os valores estimados para os parâmetros de Inclinação ou discriminação dos itens:

Tabela 3.13 *Parâmetro de Discriminação por ano e a diferença absoluta dos parâmetros entre os anos*

Item	A(2006)	A(2007)	A(2008)	A(2006)-A(2007)	A(2006)-A(2008)	A(2007)-A(2008)
1	1.58378	1.60039	1.56341	0.01661	0.02037	0.00376
2	1.37075	1.34256	1.32851	0.02819	0.04224	0.01405
3	0.57962	0.65664	0.60074	0.07702	0.02112	0.05590
4	0.62219	0.66931	0.60466	0.04712	0.01753	0.02959
5	2.50944	2.57395	2.52269	0.06451	0.01325	0.05126
6	1.99872	1.88127	1.75719	0.11745	0.24153	0.12408
7	1.96117	1.74066	1.56626	0.22051	0.39491	0.17440
8	1.24710	1.15851	1.09267	0.08859	0.15443	0.06584
9	3.62500	3.49523	3.51818	0.12977	0.10682	0.02295
10	1.87780	1.81183	1.76304	0.06597	0.11476	0.04879
11	1.64676	1.60802	1.62057	0.03874	0.02619	0.01255
12	1.57857	1.55229	1.52456	0.02628	0.05401	0.02773
13	2.18371	2.19456	2.08240	0.01085	0.10131	0.09046
14	1.38099	1.30268	1.21748	0.07831	0.16351	0.08520
15	3.29925	3.16426	3.06080	0.13499	0.23845	0.10346
16	2.12537	2.14247	2.13719	0.01710	0.01182	0.00528

De acordo com os parâmetros de Discriminação estimados, que estão na Tabela 3.13, observamos que, as diferenças absolutas dos parâmetros por ano são pequenas. A Figura 3.1 dada abaixo reforça o que já foi dito anteriormente, aparentemente não há diferenças significativas entre os parâmetros de Discriminação.

Figura 3.1 *Box plot dos parâmetros de Discriminação*

Com relação aos parâmetros de dificuldade dos itens, temos que a Tabela 3.14, apresenta os valores dos parâmetros de dificuldade dos itens por ano.

Tabela 3.14 *Parâmetros de Dificuldade por ano*

Item	B1(2006)	B1(2007)	B1(2008)	B2(2006)	B2(2007)	B2(2008)	B3(2006)	B3(2007)	B3(2008)
1	-0.38600	-0.35221	-0.48188	0.53534	0.55548	0.46361	1.7723	1.81238	1.75989
2	-0.64815	-0.64832	-0.79769	0.39946	0.39192	0.27634	1.75555	1.78279	1.68657
3	0.33970	0.46285	0.41653						
4	2.24462	2.18334	2.21088						
5	-1.17781	-1.11811	-1.11874	-0.00813	0.00373	0.01765	1.05221	1.10492	1.14122
6	0.20725	0.22615	0.12894						
7	-0.64957	-1.00856	-1.4246	1.49048	1.54369	1.47687			
8	-2.25876	-2.13369	-2.19692	1.08872	1.35137	1.38285			
9	0.24705	0.12211	-0.14038	1.87565	1.80086	1.58751			
10	0.06726	0.05306	-0.05494	1.70851	1.7484	1.69163			
11	-0.64405	-0.62464	-0.72644						
12	-2.70425	-2.63183	-2.83767	2.31992	2.44002	2.36236			
13	-0.56237	-0.40323	-0.39917	1.79129	1.89379	1.82958			
14	-1.41973	-1.62882	-2.20098	0.14635	-0.00189	-0.38827			
15	0.44459	0.35348	0.10968	2.17023	2.14875	1.97863			
16	1.53036	1.49979	1.34668						

3.6.2 Parâmetros dos Itens

Depois de observar que os parâmetros dos itens não mudam de um ano para outro, as bases de dados foram agregadas e assim o NSE de todos os alunos foram estimados. Nesta seção são apresentados os valores da estimação dos parâmetros dos itens.

Tabela 3.15 *Parâmetros dos Itens estimados*

Item	Nome utilizado na análise	A	B(1)	B(2)	B(3)
1	Escolaridade do pai	1.55654	0.39117	0.5427	1.82854
2	Escolaridade da mãe	1.31751	-0.68491	0.36373	1.79068
3	Área de Ocupação do pai	0.60611	0.4636		
4	Área de Ocupação da mãe	0.63496	2.24084		
5	Renda Familiar	2.45826	-1.14093	0.02064	1.12427
6	Televisão	1.88431	0.22813		
7	DVD ou Vídeo	1.75010	-0.97926	1.52895	
8	Rádio	1.14657	-2.18789	1.31185	
9	Computador	3.52162	0.09949	1.75709	
10	Automóvel	1.83549	0.05241	1.74115	
11	Máquina de Lavar	1.60061	-0.62597		
12	Geladeira	1.59015	-2.66389	2.35744	
13	Telefone Fixo	2.09509	-0.44079	1.87434	
14	Celular	1.31743	-1.68748	-0.05484	
15	Internet	3.15078	0.32245	2.09528	
16	Televisão a Cabo	2.09004	1.48936		

A partir da estimação destes parâmetros, foi realizada a estimação dos escores, ou seja, do nível socioeconômico — NSE, para posteriormente serem agrupados em grupos similares. Abaixo temos a Tabela 3.16, que nos apresenta os Escores (NSE) estimados de alguns alunos e a medidas do Erro padrão.

Tabela 3.16 *Estimativas do NSE*

NSE	Erro Padrão
-1.403	0.461
-0.045	0.426
0.915	0.416
-1.217	0.476
-0.067	0.402
0.416	0.402

3.7 Alocação das escolas

Para efeito de comparação entre as escolas similares, agregou-se as mesmas e intervalos de nível socioeconômico, neste caso adotamos 10 intervalos. Pois, desse modo teremos o agrupamento das escolas similares em uma classe.

3.7.1 Distribuição de Frequências do NSE

Tabela 3.17 *Distribuição de frequências do NSE das escolas*

NSE	f_i	Percentual
-2,5 -2,0	0	0,00%
-2,0 -1,5	133	0,74%
-1,5 -1,0	2701	14,93%
-1,0 -0,5	6073	33,57%
-0,5 0,0	5352	29,59%
0,0 0,5	1917	10,60%
0,5 1,0	1630	9,01%
1,0 1,5	276	1,53%
1,5 2,0	8	0,04%
2,0 2,5	0	0,00%
Soma	18090	100%

A estimação do NSE foi realizada pelo software MULTILOG (THISSEN, 1991), que trabalha com a métrica ou escala (0,1), isto é, Média 0 e Desvio Padrão 1, por isso os intervalos variam de -2,5 a 2,5. A Tabela de frequências dada acima, nos mostra que, há uma certa concentração em torno dos limites de -1,5 a 1,0.

É bom enfatizar que, na métrica (0,1) a presença de um valor negativo significa que o mesmo está abaixo da média, e se o valor for positivo ele está acima da média. Então, partindo deste princípio, os limites mencionados anteriormente nos revelam que, a maioria está abaixo da média, ou seja, o NSE dessas escolas é pequeno, este NSE foi obtido por meio da média do NSE dos alunos, logo, o NSE da escola será um reflexo do NSE dos alunos, mas será se o desempenho destas escolas está acima ou abaixo da média? A partir da construção desses intervalos de nível socioeconômico essas indagações serão respondidas, pois teremos escolas agrupadas com NSE bem próximos, então estudos comparativos poderão ser realizados .

Para melhor apresentar as medidas do NSE e do desempenho que será mostrado futu-

ramente, adotaremos uma escala mais objetiva para as medidas citadas anteriormente. A escala será a (500, 100), ou seja, Média 500 e Desvio Padrão 100.

A transformação da escala (0,1) para a escala (500,100) é dada por:

$$N = NSE * DP + M$$

onde, N é o NSE dado na escala (500,100), M é a Média da escala (500,100), e DP é o Desvio Padrão da escala (500,100).

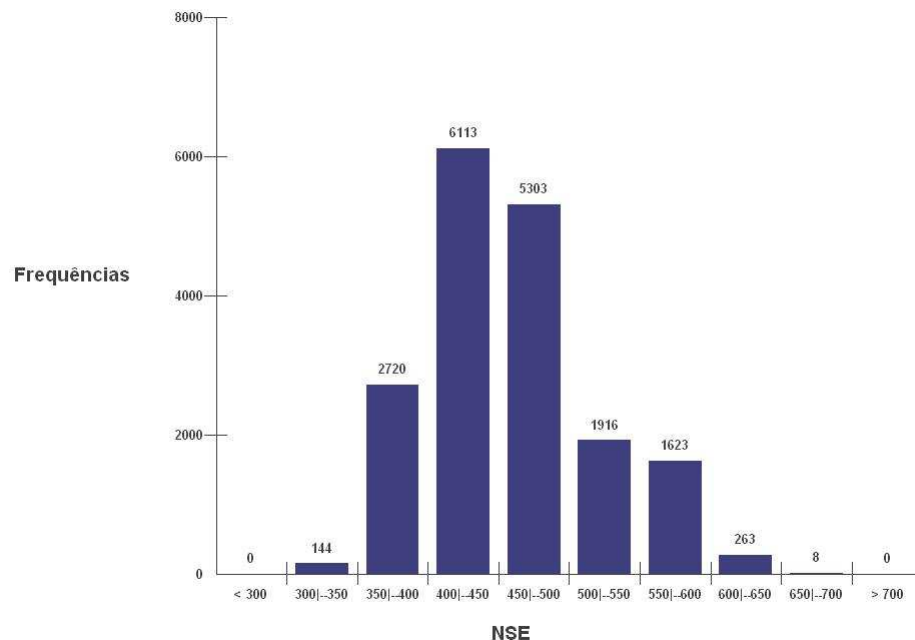
Por exemplo, para transformar o $NSE = -1,45$, para a escala (500,100), faz-se o seguinte:

$$N = -1,45 * 100 + 500 = 355$$

De agora em diante adotaremos esta métrica para o NSE e o Desempenho dos alunos na prova do ENEM.

Abaixo temos o gráfico da distribuição de frequências do NSE das escolas, na métrica (500,100).

Figura 3.2 Gráfico em Colunas do Nível socioeconômico– NSE



3.7.2 Intervalos de Desempenho

Para efeito de comparação entre as escolas similares, agregou-se as mesmas por intervalos de nível socioeconômico. A construção da medida de Desempenho ocorreu da seguinte maneira, na base de dados foram observadas duas variáveis, nota da prova objetiva e nota da redação, o valor médio dessas duas notas originou a variável Desempenho.

Com relação ao uso da escala (500,100) para o Desempenho foram realizados os seguintes procedimentos:

O Desempenho deve ser padronizado, ou seja, os valores devem ser postos na escala (0,1), da seguinte maneira:

$$Z = \frac{\text{Desempenho} - \mu}{\sigma}$$

onde, Z é o Desempenho padronizado, μ é a Média do desempenho de todas as escolas, e σ é o Desvio padrão dos desempenho de todas as escolas.

Por Exemplo, o Desempenho de uma dada escola foi de 64,61. Qual será desempenho desta escola na escala (0,1)? A análise descritiva nos revela que, μ é igual a 47,78 e o valor de σ foi igual a 7,45. Então, teremos que:

$$Z = \frac{64,61 - 47,78}{7,45} = 2,2590$$

Depois de passar o Desempenho para a escala (0,1), aplica-se a escala (500,100), do seguinte modo:

$$D = Z * DP + M$$

onde, o D é o Desempenho das escolas na escala (500,100), M é a Média da escala (500,100), e DP é o Desvio Padrão da escala (500,100).

Por exemplo, para transformar o Desempenho padronizado, $Z=2,2590$, para a escala (500,100), faz-se o seguinte:

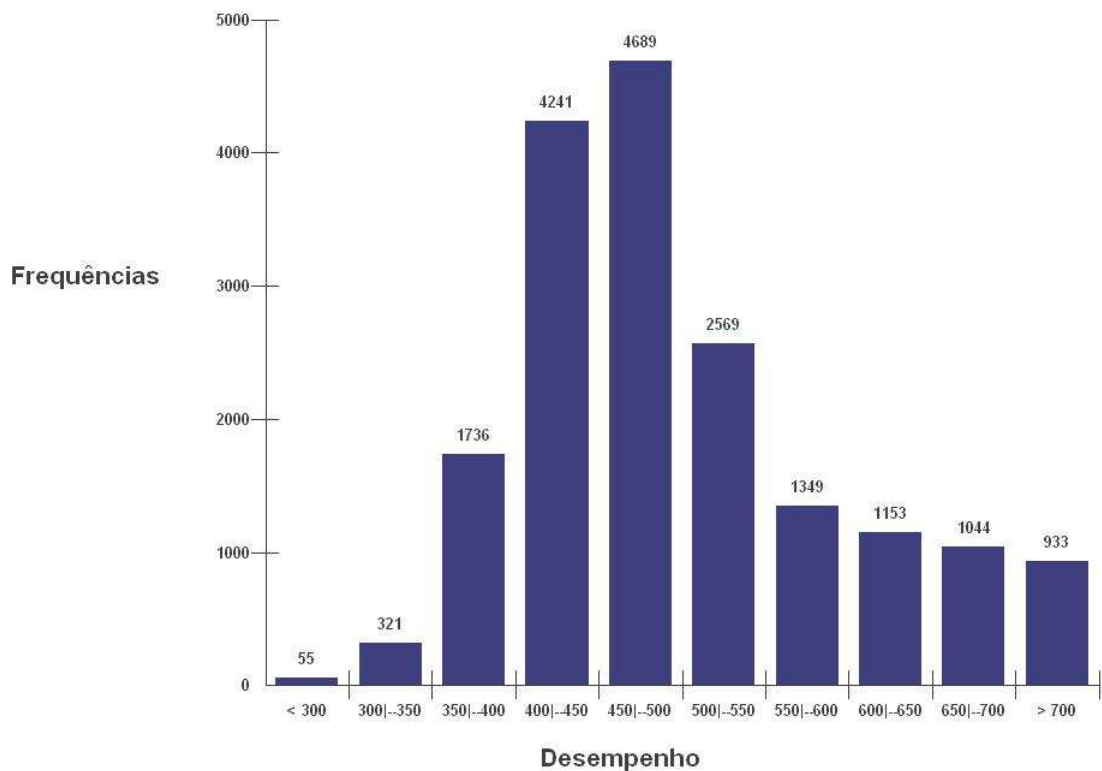
$$D = 2,2590 * 100 + 500 = 725,90$$

3.7.3 Distribuição de Frequências do Desempenho

Observando a Tabela 3.18 dada acima, observamos que há uma concentração bem expressiva em torno dos limites 354 a 701, também é salientar que o percentual de escolas que estiveram com o Desempenho acima da média foi baixo em relação as demais.

Tabela 3.18 *Distribuição de frequências do Desempenho das escolas*

Desempenho	Frequência	Percentual
<300	55	0.304%
300-350	321	1.774%
350-400	1.736	9.596%
400-450	4.241	23.444%
450-500	4.689	25.920%
500-550	2.569	14.201%
550-600	1.349	7.457%
600-650	1.153	6.374%
650-700	1.044	5.771%
≥ 700	933	5.158%
Soma	18.090	100%

Figura 3.3 *Gráfico em Colunas do Desempenho*

O gráfico em colunas dado acima nos revela que há um comportamento assimétrico positivo. Pode-se dizer que, o Desempenho médio das escolas nas provas ENEM dos anos 2006, 2007 e 2008 foi pequeno.

3.8 Estudo comparativo

Para se realizar um estudo comparativo é necessário estabelecer alguns critérios, por exemplo, termos algo que possa ser comparado ou comparável.

Por todo o mundo, temos diversos estudos sendo feitos, destacamos por exemplo, um que é realizado pelo Departamento de educação da Califórnia, estamos falando do **Índice de Desempenho Acadêmico**, em inglês **Academic Performance Index (API)**, que tem como objetivo medir anualmente o crescimento do desempenho acadêmico das escolas da Califórnia.

O API sumariza as escolas em um único número por meio de um teste de escores padronizados, que variam de 200 a 1000. A meta estadual do API é 800 para todas as escolas, os números mais altos geralmente indicam um melhor desempenho nos testes.

Este estudo visa a comparação entre as escolas similares, entretanto era preciso eliminar o efeito da escola, pois como já foi dito anteriormente, escolas com pouco estrutura se diziam prejudicadas, pois eram comparadas com escolas com ótima estrutura. Então, para contornar esse problema, o estudo comparativo foi feito a partir do nível socioeconômico – NSE das 100 escolas mais próximas com seus respectivos Desempenhos na prova do ENEM.

A análise comparativa entre as escolas similares foi realizada da seguinte forma:

- a) Ordenou-se o NSE das escolas na ordem crescente;
- b) Tomou-se as 100 escolas mais próximas, ou seja, aquelas que possuem o mesmo NSE ou NSE bem próximos;
- c) Obteve-se o ranqueamento das escolas através do Desempenho obtido na prova do ENEM.

3.9 Análise da Associação das variáveis

Antes de fazer qualquer afirmação comparativa a respeito das variáveis NSE e Desempenho, será estudada a associação entre o NSE das escolas e o Desempenho das mesmas

na prova do ENEM, pois torna-se interessante e útil estabelecer uma relação entre duas variáveis, e a partir dos coeficientes de Correlação de Pearson e de Spearman, teremos o comportamento exato desta associação.

O coeficiente de correlação de Pearson é dado por:

$$r = \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}\right) \left(\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n}\right)}} \quad (3.5)$$

onde,

r , é o coeficiente de Correlação de Pearson;

n , é o número de observações, ou seja, o número de pares (x_i, y_i) ;

X_i , é o NSE, ou seja, é a variável independente;

Y_i , é o Desempenho, ou seja, é a variável dependente.

O valor do coeficiente de correlação de Pearson, r , tem variação entre +1 e -1, ou seja, está limitado entre os valores do intervalo $[-1, +1]$.

Temos que,

- (i) Se, $r = +1$, a correlação é positiva entre as variáveis;
- (ii) Se, $r = -1$, a correlação é perfeita negativa entre as variáveis;
- (iii) Se, $r = 0$, não há correlação entre as variáveis ou, ainda, a correlação não é linear, caso exista.

O coeficiente ρ de Spearman mede a intensidade da relação entre variáveis ordinais. Nesta correlação o que importa é a ordem das observações e não o valor observado da variável.

Logo, este coeficiente não é sensível a assimetrias na distribuição, nem à presença de outliers (valores atípicos), não é exigido portanto que os dados tenham vindo de duas populações normais.

O coeficiente de correlação de Spearman é dado por:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n^2} \quad (3.6)$$

onde, em que n é o número de pares (x_i, y_i) ;

$d_i = dx_i - dy_i$, é a diferença entre os postos;

dx_i , são os postos de x_i dentre os valores de X ;

dy_i , são os postos de y_i dentre os valores de Y .

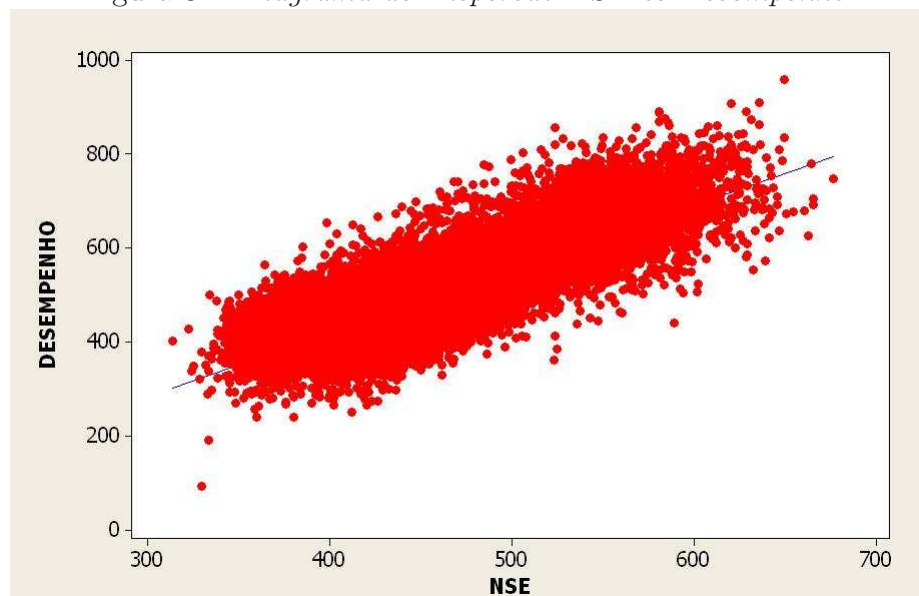
Se os postos de X são exatamente iguais aos pontos de Y , então todos os d_i serão zero e ρ será 1.

O coeficiente ρ de Spearman varia entre -1 e 1. Quanto mais próximo estiver destes extremos, maior será a associação entre as variáveis. O sinal negativo da correlação significa que as variáveis variam em sentido contrário, isto é, as categorias mais elevadas de uma variável estão associadas a categorias mais baixas da outra variável.

Neste estudo o coeficiente de correlação de Pearson, será utilizado para estudar a associação entre o NSE das escolas e os seus respectivos Desempenhos na prova do ENEM, enquanto que o coeficiente de Spearman, estudará a relação entre as ordens das variáveis, isto é, verificar se escolas que apresentam um NSE baixo, apresentarão um Desempenho baixo, ou verificar se escolas que apresentam um NSE alto, apresentarão um Desempenho alto.

A Figura 3.4 dada abaixo nos apresenta o gráfico de dispersão entre as variáveis NSE e Desempenho.

Figura 3.4 *Diagrama de Dispersão NSE vs Desempenho*



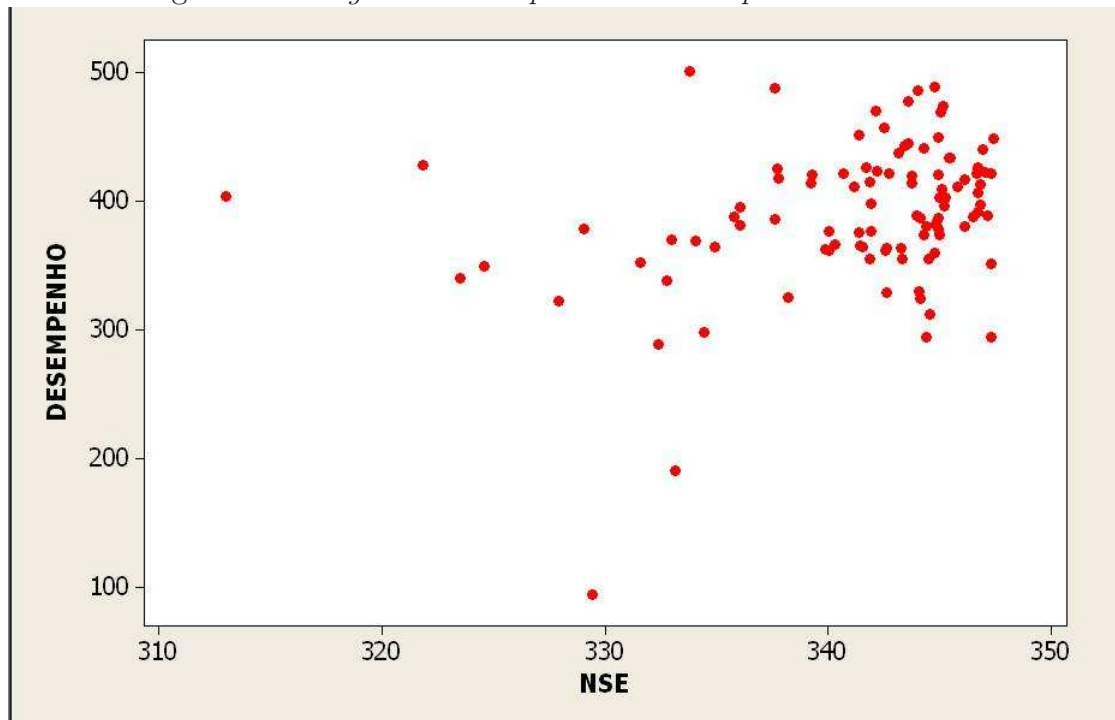
De acordo com a Figura 3.4 dada acima, percebe-se que, há evidências de uma associação linear.

O valor do coeficiente de correlação linear de Pearson foi igual a 0,890, revelando assim uma correlação positiva muito forte, apesar da grande variabilidade, e o de Spearman foi igual a 0,753, apresenta uma correlação positiva forte.

Tomando as 100 primeiras escolas, ou seja, aquelas que tiveram os menores NSE, e analisando os coeficientes de correlação de Pearson e Spearman, obtém-se respectivamente, 0,299, e 0,275, observa-se que, ambos os coeficientes apresentam valores pequenos, logo temos uma correlação positiva fraca em ambos os casos.

A Figura 3.5 abaixo apresenta o diagrama de dispersão do NSE e Desempenho das 100 primeiras escolas.

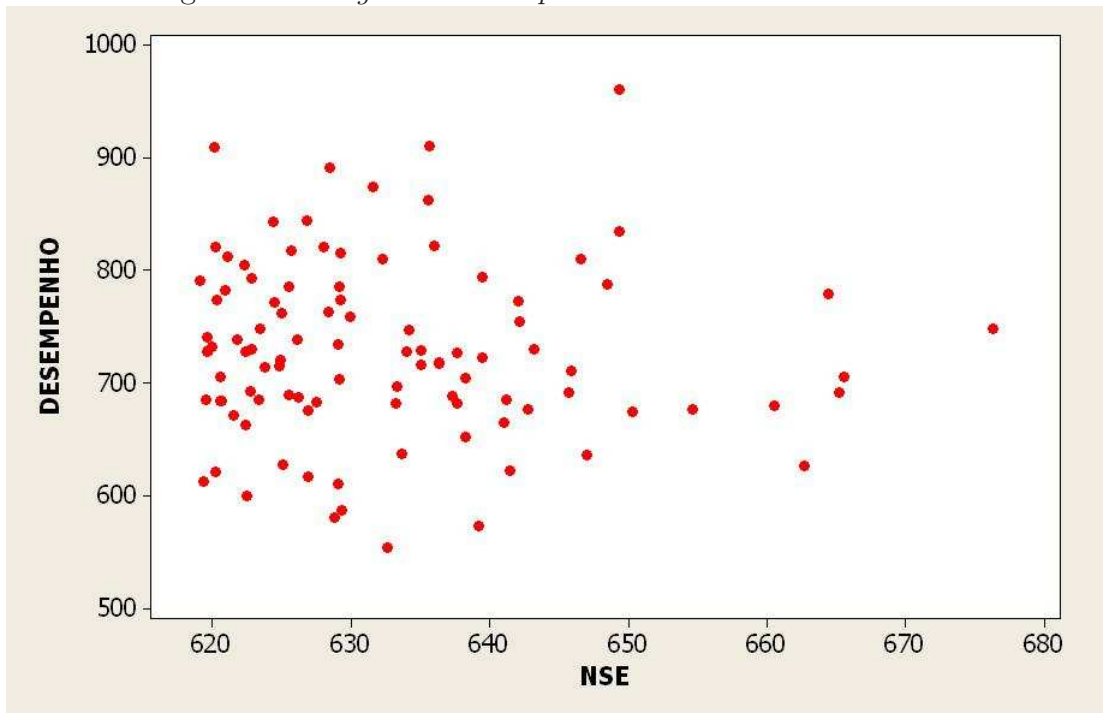
Figura 3.5 *Diagrama de Dispersão das 100 primeiras escolas*



Tomando as 100 últimas escolas, ou seja, aquelas que tiveram os maiores NSE, e analisando os coeficientes de correlação de Pearson e Spearman, obtém-se respectivamente, -0,0417, e -0,0766, observa-se que, ambos os coeficientes apresentam valores negativos, logo temos uma correlação positiva negativa muito fraca em ambos os casos.

A Figura 3.6 abaixo apresenta o diagrama de dispersão do NSE e Desempenho das 100 últimas escolas.

Figura 3.6 Diagrama de Dispersão das 100 últimas escolas



Foi realizado estudo comparativo com as 100 escolas posicionadas nos extremos, ou seja, as 100 escolas que apresentaram os menores NSE, e as 100 escolas que apresentaram os maiores NSE.

As Tabelas 3.19 e 3.20 dadas abaixo mostram uma visão parcial deste ranqueamento.

Com relação ao estudo comparativo pode-se fazer vários tipos de abordagem entre as escolas similares, como as escolas foram previamente divididas em intervalos de nível socioeconômico, escolhe-se uma escola e compara-se a mesma com as 49 que estão à sua direita e as 50 que estão à esquerda.

E esta abordagem foi realizada e será mostrada a seguir.

Ordenou-se as escolas a partir do NSE e uma escola foi selecionada, que será chamada de escola central.

A análise comparativa deu-se do seguinte modo, tomou-se escolas que estão à direita da escola central e as escolas à esquerda, e realizou o ranqueamento pelo Desempenho das mesmas na prova do ENEM. Desse modo, podemos ter uma noção exata do posicionamento desta escola dita central em relação as demais escolas similares.

Pode-se observar, por exemplo, pelo fato dela ocupar a 50^a em relação ao NSE, implicará

em um desempenho superior as escolas que estão abaixo dela. Portanto, observa-se que a análise é simples, porém de grande relevância para o meio educacional em virtude da metodologia apresentada.

Depois de tomar uma escola conforme a metodologia fixada anteriormente, temos que, o desempenho da escola central é razoavelmente bom, pois ocupa a posição 26^a entre as 100 analisadas.

Observando o coeficiente de correlação de Pearson que foi igual a 0,081, muito fraca entre as variáveis NSE e Desempenho, já a correlação ordinal de Spearman foi igual a 0,106 que é um pouco superior a de Pearson, mesmo assim a correlação pode-se dizer muito fraca.

Portanto, pode-se expandir esta comparação para as demais escolas.

A Figura 3.7 abaixo apresenta o diagrama de dispersão do NSE e Desempenho de 100 escolas similares.

Figura 3.7 *Diagrama de Dispersão de 100 escolas similares*

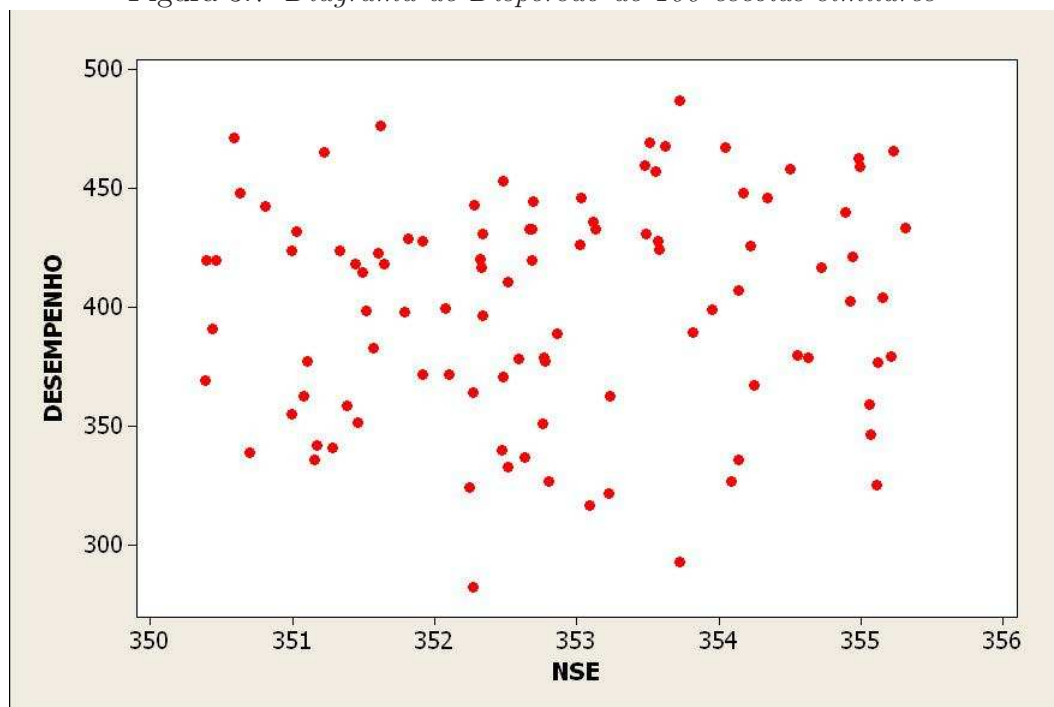


Tabela 3.19 *Ranqueamento por Desempenho das 100 escolas com menor NSE*

NSE	Desempenho	Ordem por Desempenho	Ordem por NSE
333.770	500.880	1 ^o	88 ^o
344.740	488.450	2 ^o	33 ^o
337.590	487.680	3 ^o	81 ^o
344.020	485.560	4 ^o	43 ^o
343.560	478.010	5 ^o	48 ^o
345.120	473.420	6 ^o	22 ^o
342.120	470.040	7 ^o	59 ^o
345.040	469.390	8 ^o	24 ^o
342.500	457.030	9 ^o	57 ^o
341.380	451.400	10 ^o	68 ^o
344.910	449.240	11 ^o	29 ^o
347.390	448.250	12 ^o	1 ^o
343.570	444.740	13 ^o	47 ^o
343.390	442.640	14 ^o	49 ^o
344.300	441.210	15 ^o	38 ^o
346.900	439.940	16 ^o	7 ^o
343.130	437.170	17 ^o	52 ^o
345.430	433.880	18 ^o	19 ^o
345.440	433.470	19 ^o	18 ^o
321.800	428.270	20 ^o	99 ^o
341.690	426.540	21 ^o	64 ^o
346.700	426.510	22 ^o	10 ^o
337.710	425.580	23 ^o	79 ^o
342.200	423.760	24 ^o	58 ^o
347.030	422.100	25 ^o	6 ^o
340.650	421.750	26 ^o	70 ^o
346.630	421.570	27 ^o	13 ^o
342.740	421.500	28 ^o	53 ^o
347.300	421.500	29 ^o	5 ^o
339.260	420.390	30 ^o	75 ^o
344.940	420.180	31 ^o	28 ^o
343.720	419.790	32 ^o	46 ^o
337.740	417.570	33 ^o	78 ^o
346.100	417.090	34 ^o	16 ^o
341.840	414.770	35 ^o	63 ^o
343.740	414.270	36 ^o	45 ^o
339.200	414.150	37 ^o	76 ^o
346.820	413.330	38 ^o	8 ^o
341.160	411.490	39 ^o	69 ^o
345.810	411.340	40 ^o	17 ^o

Tabela 3.20 *Ranqueamento por Desempenho das 100 escolas com maior NSE*

NSE	Desempenho	Ordem por Desempenho	Ordem por NSE
649.310	960.380	1 ^o	9 ^o
635.620	909.810	2 ^o	34 ^o
620.130	908.720	3 ^o	93 ^o
628.430	891.210	4 ^o	55 ^o
631.510	873.690	5 ^o	45 ^o
635.560	862.660	6 ^o	35 ^o
626.790	843.880	7 ^o	61 ^o
624.310	842.930	8 ^o	72 ^o
649.300	835.150	9 ^o	10 ^o
636.010	821.760	10 ^o	33 ^o
627.970	821.070	11 ^o	57 ^o
620.200	820.940	12 ^o	92 ^o
625.690	817.190	13 ^o	64 ^o
629.220	814.950	14 ^o	48 ^o
621.030	812.250	15 ^o	85 ^o
646.520	810.650	16 ^o	13 ^o
632.240	810.030	17 ^o	44 ^o
622.280	804.860	18 ^o	82 ^o
639.450	793.700	19 ^o	23 ^o
622.810	792.870	20 ^o	76 ^o
619.030	791.430	21 ^o	100 ^o
648.450	787.310	22 ^o	11 ^o
625.450	785.800	23 ^o	66 ^o
629.090	785.760	24 ^o	51 ^o
620.850	781.970	25 ^o	86 ^o
664.380	779.560	26 ^o	4 ^o
620.250	774.430	27 ^o	90 ^o
629.210	773.530	28 ^o	49 ^o
642.010	772.500	29 ^o	19 ^o
624.460	771.840	30 ^o	71 ^o
628.370	763.570	31 ^o	56 ^o
624.920	762.720	32 ^o	68 ^o
629.910	759.060	33 ^o	46 ^o
642.160	755.190	34 ^o	18 ^o
623.400	748.690	35 ^o	74 ^o
676.280	747.890	36 ^o	1 ^o
634.130	747.120	37 ^o	38 ^o
619.560	740.760	38 ^o	97 ^o
621.740	739.290	39 ^o	83 ^o
626.110	738.330	40 ^o	63 ^o

Com relação as 100 primeiras escolas, ou seja, aquelas que tiveram os menores NSE, temos que, aquela que obteve o NSE igual a 347,390 ficou com o maior valor em relação as demais escolas, isto é, 1^a posição, entretanto analisando o seu Desempenho na prova do ENEM, observa-se que esta escola ocupa a 12^a posição.

Observando a escola dentre as 100 primeiras que obteve o menor valor do NSE, isto é, aquela que ocupava a 100^a posição, com relação ao seu Desempenho na prova, observa-se que, esta escola ocupa a 43^a posição.

Agora, observando as 100 últimas escolas, ou seja, aquelas que tiveram os maiores NSE, temos que, aquela que obteve o NSE igual a 676,280 ficou com o maior valor em relação as demais escolas, ficando deste modo na 1^a posição. Entretanto, observando o ranqueamento pelo Desempenho na prova do ENEM, verificou-se que, esta escola ocupa a 36^a posição.

Observando a escola dentre as 100 últimas que obteve o menor valor do NSE, isto é, aquela que ocupava a 100^a posição, com relação ao seu Desempenho na prova, observa-se que, esta escola ocupa a 21^a posição.

Então, pode-se estudar as escolas similares a partir desta escala, e assim fazer diversos tipos de análise.

Capítulo 4

Conclusões

4.1 Considerações finais

Na área de avaliação educacional é frequente a apresentação nacional dos desempenhos de escolas em determinadas áreas, tais como Português e Matemática, e muitas vezes esses desempenhos são utilizados para diversas ações, inclusive de aspecto financeiro. Há uma série de críticas pela comparação de desempenhos de escolas com características tão distintas.

Dentre os objetivos propostos neste trabalho, destaca-se a elaboração de uma nova metodologia que sirva para comparar os desempenhos das escolas similares. Ela traz uma grande vantagem em relação à metodologia padrão, que envolve a construção de modelos de valor adicionado ou agregado, que é uma técnica bastante complicada matematicamente e computacionalmente. Sobretudo, a simplicidade do procedimento será um atrativo para que a mesma possa ter um grande nível de aplicação na prática.

O trabalho consistiu na criação de uma escala, na qual as escolas similares pudessem ser comparadas. A partir da estimação do nível socioeconômico - NSE dos alunos, pôde-se obter o NSE das escolas, através do valor médio do NSE dos alunos e assim comparar as escolas. Caberá ainda uma interpretação mais específica do que consiste cada nível da escala.

É bom salientar que, a construção do NSE, só foi possível em virtude da adoção do modelo de resposta gradual de Samejima, que é um modelo relativamente fácil de compreender. E essa construção foi obtida através do software MULTILOG.

Um ponto crucial foi a construção do NSE, realizada em etapas. Primeiro para os alunos que se submeteram ao ENEM de 2006, ou seja, estimou-se (calibrou-se) os parâmetros dos itens, e depois o NSE. Os mesmos itens foram calibrados com os dados de 2007, e verificou-se que os parâmetros dos itens mantiveram-se constantes, o mesmo ocorrendo com os dados de 2008. Este fato foi de extrema importância para nossa análise, pois um

dos critérios de seleção da amostra de alunos para calibração, e composição do NSE, era que o aluno fosse concluinte. Com isso, o aluno participou do estudo em apenas um ano, mas como as estimativas dos parâmetros dos itens se mantinham constantes ao longo dos três anos, podemos estimar todos os NSE's dos concluintes como se fossem em um mesmo ano. Deste modo, estimou-se o NSE de todos os alunos concluintes nas três edições do ENEM, e agregou-se o NSE dos alunos pelas escolas a que eles pertenciam.

O uso da métrica (500, 100) foi adotado, pois o NSE foi estimado na métrica (0,1), isto é, valores padronizados. Por outro lado o Desempenho dos estudantes também foi colocado na métrica (500, 100) e isso possibilita uma melhor entendimento por parte dos profissionais de outras áreas, especificamente da área de Educação.

Antes de estabelecer o procedimento para comparações entre as escolas, foi feita a análise de associação das variáveis. Para tal, adotou-se os coeficientes de correlação de Pearson e Spearman, considerando todas as escolas. A análise revelou que havia uma associação muito forte entre as variáveis, em ambos os coeficientes, entretanto quando a avaliação destes coeficientes deu-se nos extremos, ou seja, com as 100 primeiras escolas e as 100 últimas, os resultados obtidos mostraram que, a correlação passou a ser negativa muito fraca para as 100 últimas, e para as 100 primeiras escolas a correlação foi positiva fraca. Portanto, analisando as escolas a partir do ranqueamento, as diferenças são mais visíveis. E assim ao indicar essas diferenças damos uma contribuição bastante valioso para a área educacional, que fará suas devidas adequações a respeito dessas diferenças.

Por fim, tem-se a ideia intuitiva de que, se uma escola apresenta um valor do NSE pequeno, ao submeter os seus alunos a uma prova, teste ou exame, o desempenho será equivalente ao NSE. Entretanto, pôde-se observar no presente estudo que, há escolas que não seguem esta lógica, apresentam NSE baixo, mas o desempenho na prova foi bem superior e quando comparamos esta mesma escola com outras similares, ela tendem a ficar bem posicionada no tocante ao desempenho.

Portanto, a partir dessa nova metodologia, as escolas poderão ser comparadas de forma mais igualitária.

4.2 Recomendações para trabalhos futuros

Recomenda-se realizar um estudo comparativo entre as escolas similares por regiões, estados, municípios. Ainda, Fazer uma análise comparativa com resultados de outros exames

educacionais, tais como a Prova Brasil. Por fim, o estudo de novas propostas de modelos pode melhorar ainda mais a análise.

BIBLIOGRAFIA

- ALVES, M. T. G., SOARES, J. F. *Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. Opinião Pública, vol. 15, nº 1, p.1-30.* Campinas, junho de 2009.
- ANDRADE, D.F., TAVARES, H. R., VALLE, R.C. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações.* Associação Brasileira de Estatística: São Paulo.
- ANDRICH, D. *A rating formulation for ordered response categories.* Psychometrika, v. 43, p. 561-573, 1978.
- API. *Academic Perfomance Index. Disponível na internet via <http://www.cde.ca.gov/ta/ac/ap/>* Acesso em abril de 2010
- AZEVEDO, C. L. N. *Métodos de estimação na teoria de resposta ao item.* Dissertação de Mestrado. Instituto de Matemática e Estatística Universidade de São Paulo. São Paulo, 2003.
- BAKER, F. B. *The Basics of Item Response Theory. 2 ed. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001.*Disponível em: <http://edres.org/irt/> Acessado em 8 de fevereiro 2010.
- BIANCHI, A. B (2003). *A new look at accountability: Value - added assessment.Forecast: Emerging issues in public education,1(11), p.1-4.*
- BIRNBAUM, A. (1968). *Some Latent Trait Models and Their Use in Infering an Examinee's Ability.* In F.M. Lord & M.R. Novick. *Statistical Theories of Mental Test Scores.* Reading, MA:Addison-Wesley.
- BRACEY, G. W (2004). Serious questions about the Tennessee value-added assessment system. *Phi delta kappan, 85(9), pp. 716-717.*
- BOCK, R. D. and ZIMOWSKI, M. F. (1997). Multiple group IRT. In *Handbook of Modern Item response Theory.* W.J. van der Linder and R.K. Hambleton Eds. New York: Springer-Verlag.
- CALLENDER, J. (2004). Value - added student assessment. *Journal of educational and behavioral statistics, 29(1), p.5.*
- CAREY, K. (2004). *The real value of teachers: Using new information about teavher effectiveness to close the achievement gap.* Thinking K-16, 8(1), pp. 1-42.

- COLEMAN, J. S. *Equality of Educational Opportunity*. Washington, DC, US. Government Printing Office, 1966.
- DE LUCA, M. M. M. *Demonstração do valor adicionado: do cálculo da riqueza criada pela empresa ao valor do PIB*. São Paulo: Atlás, 1998.
- DRURY, D. & DORAN, H. C (2003). The value of value-added analysis. *National School Boards Association policy research brief 3(1)*, pp. 1-4.
- EMBRETSON, S. E., & REISE, S. P. *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers, 2000.
- GOLDSCHMIDT, P., CHOI, K., & MARTINEZ, F. (2004). *Using hierarchical growth models to monitor school performance: The effects of the model, metric and time on the validity of inferences*. Paper presented at the Council of Chief State School Officers annual conference on large-scale assessment, Boston.
- HANUSHEK, A. E. (1979). 'Conceptual and empirical issues in the estimation of educational production functions', *Journal of Human Resources*, Vol. 14, 351-388.
- HANUSHEK, E. A. and TAYLOR, L. L. (1990). 'Alternative assessments of the performance of schools: measurement of state variations in achievement', *Journal of Human resources*, Vol. 25, 179-201.
- HAVEMAN, R. H. and WOLFE, B. L. (1995). 'The determinants of children's attainments: A review of methods and findings', *Journal of Economic Literature*, Vol. 33, pp. 1829-78.
- HERDMAN, P. A, SMITH, N., & DORAN, H. (2002). *CFNN policy brief. Value-added analysis: A critical component of determining adequate yearly progress (AYP)*. St. Paul, MN: National Charter School Alliance.
- HERSHBERG, T., SIMON, V. A, & LEA-KRUGER, B. (2004). *Measuring what matters*. National School Boards Association. Obtido em http://www.cgp.upenn.edu/pdf/measuring_what_matters.pdf, Acesso em março de 2010
- INEP. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*. Disponível na internet via www.inep.gov.br. Acesso em janeiro e fevereiro de 2010.
- KUPERMINTZ, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational evaluation and policy analysis*, 25(3) pp. 287-298.
- LADD, H. F. and WALSH, R. P. (2002). "Implementing value-added measures of school effectiveness: getting the incentives right", *Economics of Education Review*, Vol. 21, 1-17.

-
- LAZARSELD, P. F. (1950). *The logical and mathematical foundation of latent structure analysis*. In S.A. Stauffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, & J.A. Clausen (Eds.), *Measurement and prediction*. Princeton, NJ: Princeton University Press, 1950.
- LAWLEY, D. N. (1943). On problems connected with item selection and tests construction. *Proceedings of the Royal Society of Edinburgh*, 62-A, 273-287.
- LORD, F. M. (1952). A theory of test scores (No. 7). *Psychometric Monograph*.
- LORD, F. M. (1968). *An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model*. *Educational and Psychological Measurement*, 28, 989-1020.
- LINDEM, W. J. & HAMBLETION, R. K. (1997). *Handbook of modern Item Response Theory*. New York: Springer-Verlag.
- MACEDO, G. A, *Fatores Associados ao Rendimento Escolar de Alunos da 5a série (2000) - uma abordagem do valor adicionado e da heterogeneidade* Dissertação (Mestrado em Demografia) Centro de Desenvolvimento e Planejamento Regional– Universidade Federal de Minas Gerais. Minas Gerais, 2004.
- MAMBRINI, J. V. Melo. *Desigualdade e saúde o Brasil: Medida e Avaliação*. Tese (Doutorado em Demografia). Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais. Minas Gerais, 2009.
- MASTERS, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- MCCAFFREY, D. F., LOCKWOOD, J. R., KORETZ, D. M., & HAMILTON, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: the Rand Corporation.
- MEYER, R. H. (1997). "Value added indicators of school performance: a primer", *Economics of Education Review*, Vol. 16, 283-301.
- MURAKI, E. (1992). A generalized partial credit model : Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- MURAKI, E. and Bock, R. D. (1997). *PARSCALE : IRT Based Test Scoring and Item Analysis for Graded Open-Ended Exercises and Performance Tasks*. Chicago : Scientific Software, Inc.
- NITKO, A. J (2004). *Educational assessments of students* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- PASQUALI, L. (1997). *Psicometria: Teoria e Aplicações*. Brasília: Editora Universidade de Brasília.

-
- PASQUALI, L. (1999). *Instrumentos Psicológicos: Manual Prático de Elaboração*. Brasília: Editora Universidade de Brasília.
- PASQUALI, L. (2003). *Teoria dos Testes na Psicologia e na Educação*. Petropolis: Vozes.
- RASCH, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen : Danish Institute for Educational Research.
- RICHARDSON, M. W. (1936). *The relationship between difficulty and the differential validity of a test*. *Psychometrika*, 1, 33-49.
- RUBIN, D. B., STUART, E. A., & ZANUTTO, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of educational and behavioral statistics*, 29(1), pp.103-116.
- SAMEJIMA, F. A. (1969). *Estimation of latent ability using a response pattern of graded scores*. *Psychometric Monograph*, 17.
- SANTOS, A. dos. *Demonstração do valor adicionado: como elaborar e analisar DVA* São Paulo: Atlas, 2003.
- STONE, J. E (1999). Value-added assessment: An accountability revolution. Em Kanstoroom M. and Chester, EF Jr. (Eds.). *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation. Obtido em http://education-consumers.com/articles/value_added_assessment.shtm acessado em 2 de Março, 2010.
- THISSEN, D. *Multilog: multiple, categorical item analysis and test scoring: Using Item Response Theory*. Chicago: Scientific Software, Inc., 1991.
- TODD, P., WOLPIN, K. *On the especification and estimation of the production function for cognitive achievement*, In.: *The Economic Journal*. Oxford, v.113, n.485, p. F3-F33, fevereiro, 2003.
- VAUGHAN, A. C (2002). Standards, accountability, and the determination of success. *The educational forum*, 66(3), p. 206- 213.

Apêndice A

SINTAXE

A.1 Sintaxe do Multilog para Estimação dos Parâmetros do Itens

```
>PROBLEM RANDOM,  
    INDIVIDUAL,  
    DATA = 'G:\MY FILES\Dissertacao\Multilog\ENEM060708.DAT',  
    NITEMS = 16,  
    NGROUPS = 1,  
    NEXAMINEES = 99999,  
    NCHARS = 8;  
>TEST ALL,  
    GRADED,  
    NC = (4, 4, 2, 2, 4, 2, 3(0)4, 2, 3(0)4, 2);  
>SAVE ;  
>END ;  
4  
1234  
1111111111111111  
2222222222222222  
3300303333033330  
4400400000000000  
(8A1,16A1)
```

A.2 Sintaxe do Multilog para Estimação dos Escores(NSE)

```
>PROBLEM SCORE,  
    INDIVIDUAL,  
    DATA = 'G:\MY FILES\Dissertacao\Multilog\ENEM060708.DAT',  
    NITEMS = 16,  
    NGROUPS = 1,  
    NEXAMINEES = 99999,  
    NCHARS = 8;  
>TEST ALL,  
    GRADED,  
    NC = (4, 4, 2, 2, 4, 2, 3(0)4, 2, 3(0)4, 2);  
>SAVE ;  
>END ;  
4  
1234  
1111111111111111  
2222222222222222  
3300303333033330  
4400400000000000  
(8A1,16A1)
```


A.3 Sintaxe comentada – Estimação dos Parâmetros dos Itens

>PROBLEM RANDOM, → Estima os Parâmetros dos itens
INDIVIDUAL, → Vetores individuais de resposta dos itens
DATA = 'G:\MY FILES \Dissertacao\Multilog\ENEM060708.DAT', → Caminho do diretório.
NITEMS = 16, → Número de itens
NGROUPS = 1, → Número de grupos envolvidos na análise
NEXAMINEES = 99999, → Número de examinados ou indivíduos.
NCHARS = 8; → Número de caracteres de identificação do indivíduo ou examinados
>TEST ALL, → Todos os itens serão analisados
GRADED, → Modelo adotado é o Gradual
NC = (4, 4, 2, 2, 4, 2, 3(0)4, 2, 3(0)4, 2); → Mostra o número de categorias de cada item. Por exemplo, os dois primeiros itens possuem 4 categorias
>SAVE ; → Salva os Parâmetros dos Itens em um arquivo com formato.par
>END ;
4 → Número de categorias
1234 → Código das Respostas
1111111111111111
2222222222222222 → Disposição das categorias em todos os itens.
3300303333033330
4400400000000000
(8A1,16A1) → Formato da leitura dos dados

A.4 Sintaxe comentada – Estimação dos Escores (NSE)

>PROBLEM SCORE, → Estima os Escores (NSE)
 INDIVIDUAL, → Vetores individuais de resposta dos itens
 DATA = 'G:\MY FILES \Dissertacao\Multilog\ENEM060708.DAT', → Caminho do diretório.
 NITEMS = 16, → Número de itens
 NGROUPS = 1, → Número de grupos envolvidos na análise
 NEXAMINEES = 99999, → Número de examinados ou indivíduos.
 NCHARS = 8; → Número de caracteres de identificação do indivíduo ou examinados
 >TEST ALL, → Todos os itens serão analisados
 GRADED, → Modelo adotado é o Gradual
 NC = (4, 4, 2, 2, 4, 2, 3(0)4, 2, 3(0)4, 2); → Mostra o número de categorias de cada item. Por exemplo, os dois primeiros itens possuem 4 categorias
 >SAVE ; → Salva os Escores em um arquivo com formato.sco
 >END ;
 4 → Número de categorias
 1234 → Código das Respostas
 1111111111111111
 2222222222222222 → Disposição das categorias em todos os itens.
 3300303333033330
 4400400000000000
 (8A1,16A1) → Formato da leitura dos dados

A.5 Sintaxe do SPSS – Ranqueamento

GET

FILE='C:\NSE\FULL\NSE E DESEMPENHO\NSE_MEAN.sav'.

DATASET CLOSE ALL.

SORT CASES BY CODESC(A).

GET

FILE='C:\NSE\FULL\NSE E DESEMPENHO\DESEMPENHO_MEAN.sav'.

DATASET CLOSE ALL.

SORT CASES BY CODESC(A).

```
MATCH FILES /TABLE=*
  /FILE='C:\NSE\FULL\NSE E DESEMPENHO\NSE_MEAN.sav'
  /BY CODESC.
EXECUTE .
SAVE OUTFILE='C:\NSE\FULL\NSE E DESEMPENHO\FINAL.sav'
  /COMPRESSED.
COMPUTE M=500.
COMPUTE S=100.
COMPUTE DP=7.44778.
COMPUTE MEDIA=47.7823.
COMPUTE DESEMPENHO_PADRONIZADO=(Desempenho_mean - MEDIA)/DP *100+500.
COMPUTE NSE_2 = (NSE_mean*S)+M.
EXECUTE.
SORT CASES BY NSE_2(A).
FILTER OFF.
USE 150 thru 249.
EXECUTE.
RANK VARIABLES=NSE_2 (D)
  /RANK
  /PRINT=YES
  /TIES=MEAN.
RANK VARIABLES=DESEMPENHO_PADRONIZADO (D)
  /RANK
  /PRINT=YES
  /TIES=MEAN.
FORMATS RDESEMPE (F2.0).
FORMATS RNSE_2 (F2.0).
```

Apêndice B

FIGURAS

Figura B.1 *Curva Característica do Item Escolaridade do pai*

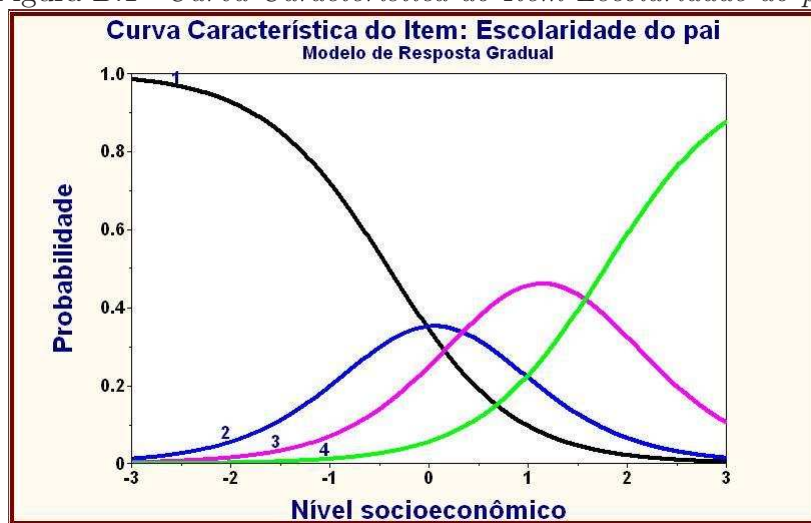


Figura B.2 *Curva Característica do Item Escolaridade da mãe*

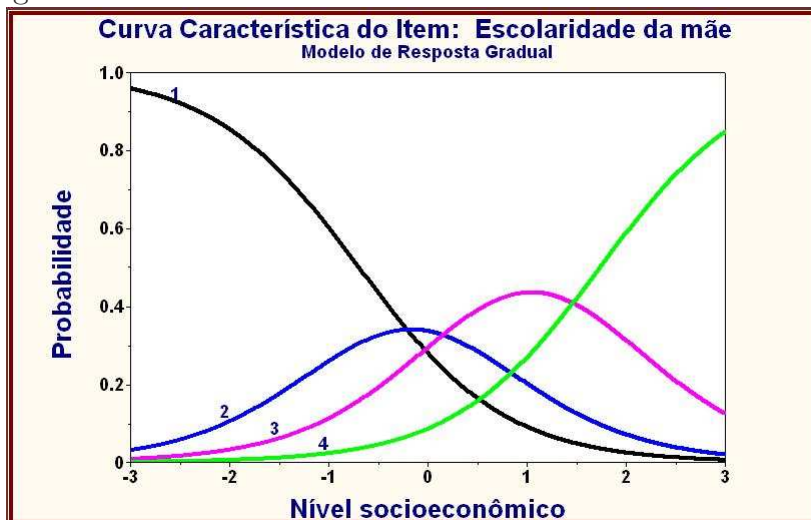


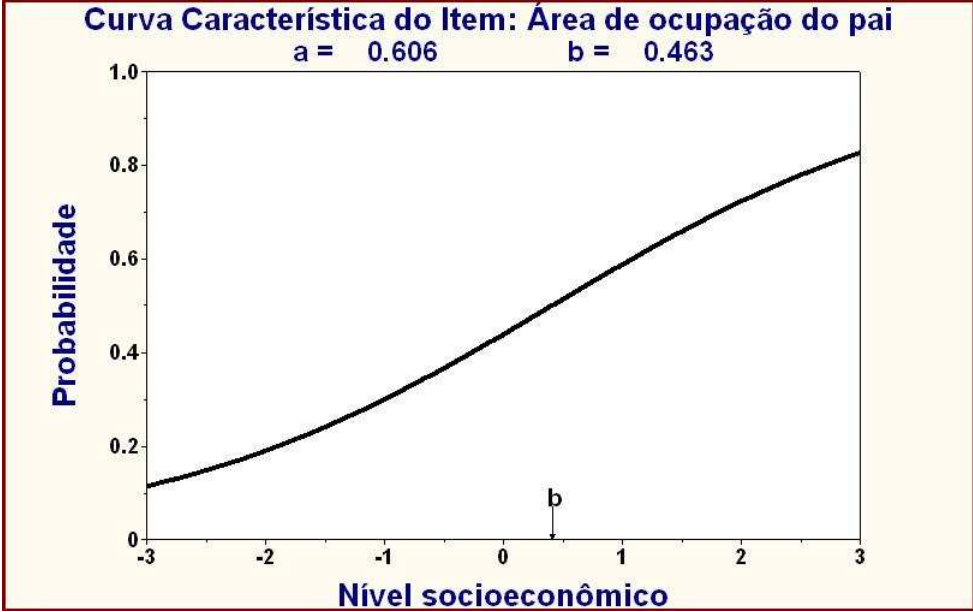
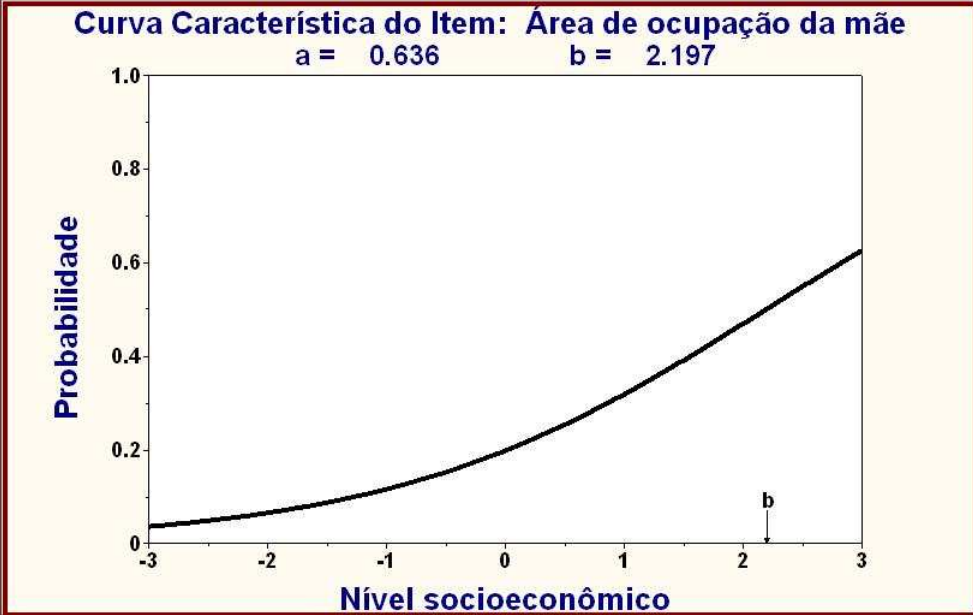
Figura B.3 *Curva Característica do Item Área de ocupação do pai*Figura B.4 *Curva Característica do Item Área de ocupação da mãe*

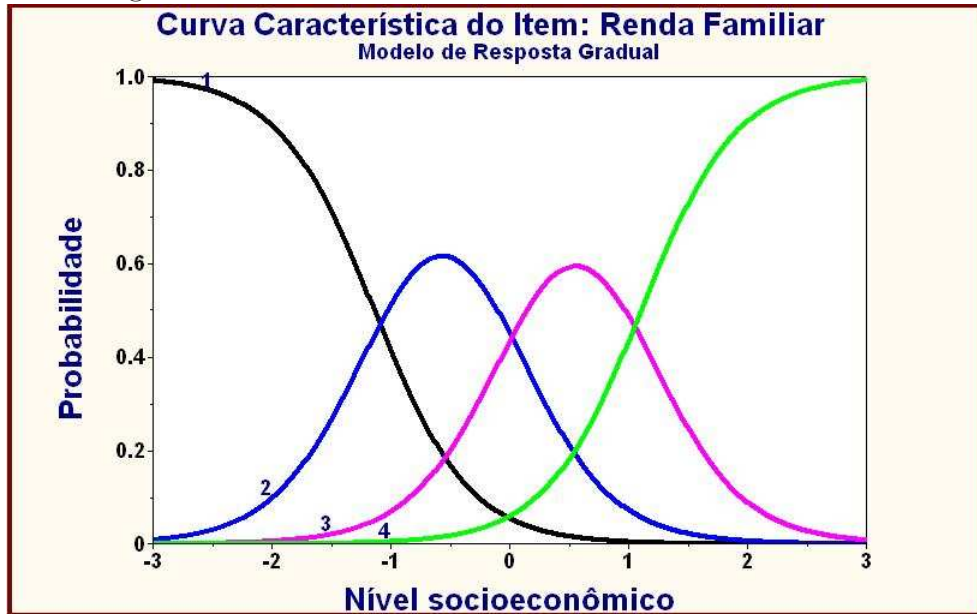
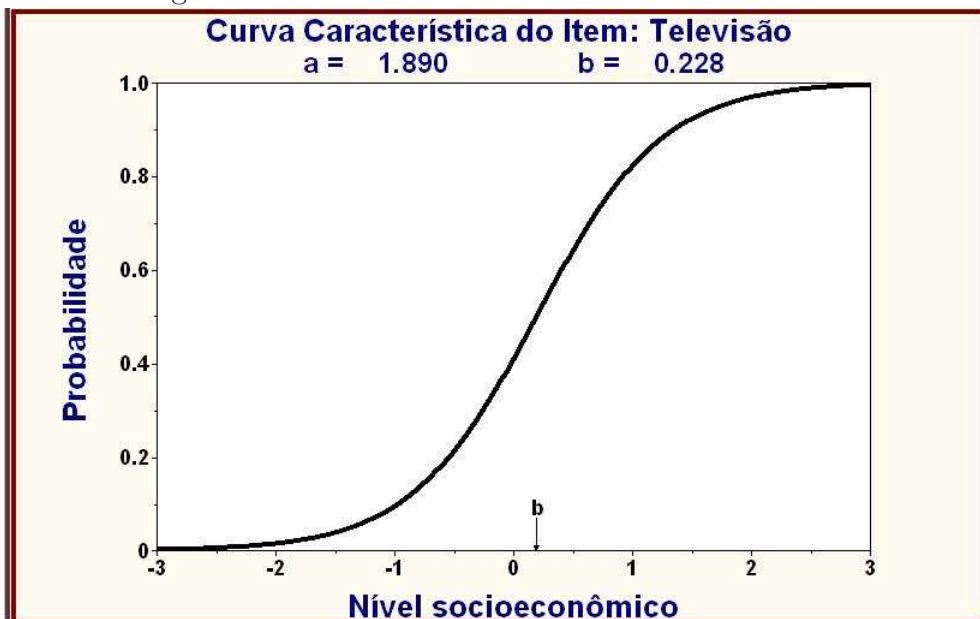
Figura B.5 *Curva Característica do Item Renda Familiar*Figura B.6 *Curva Característica do Item Televisão*

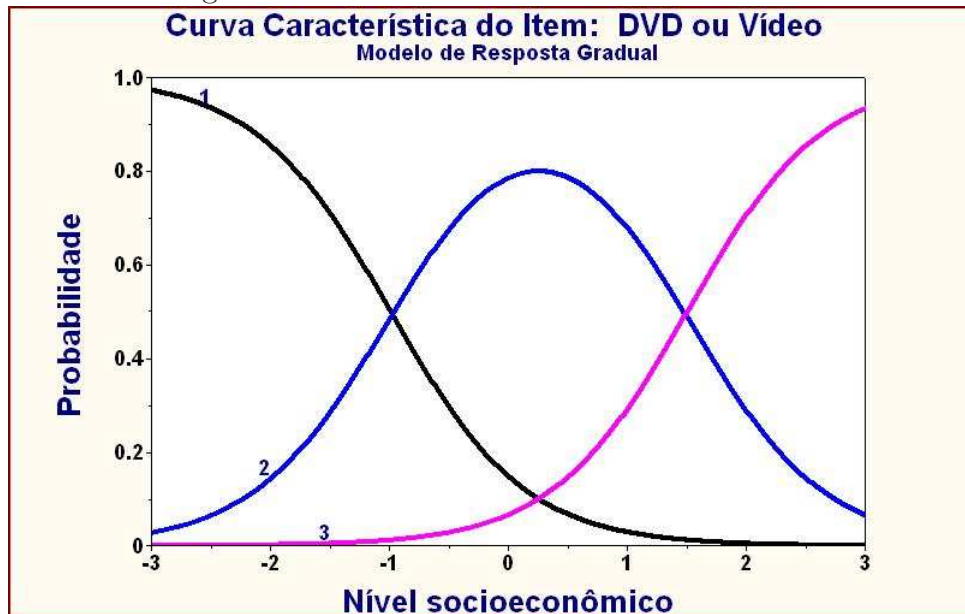
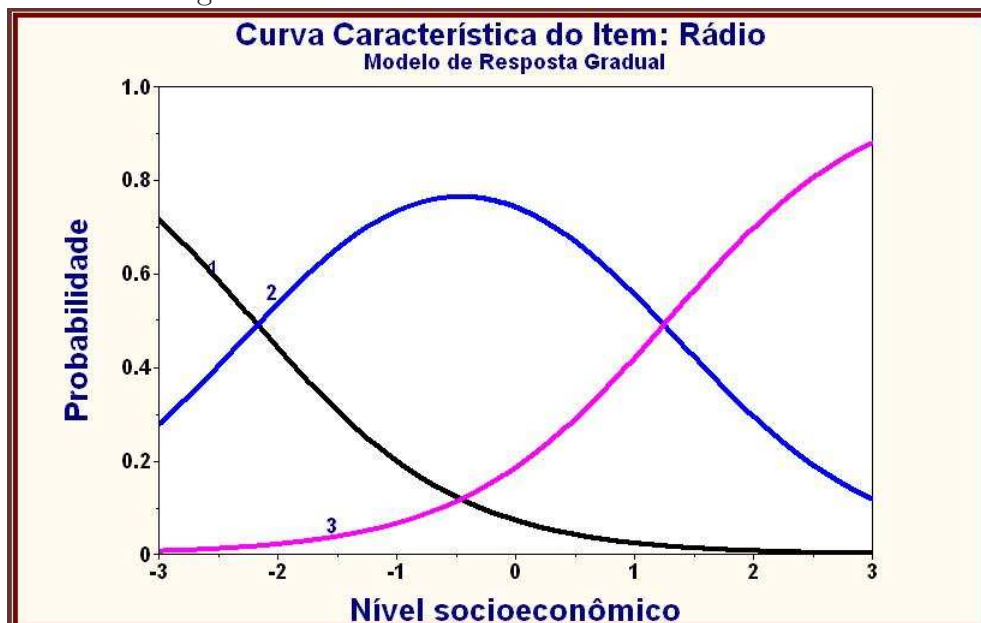
Figura B.7 *Curva Característica do Item Dvd*Figura B.8 *Curva Característica do Item Rádio*

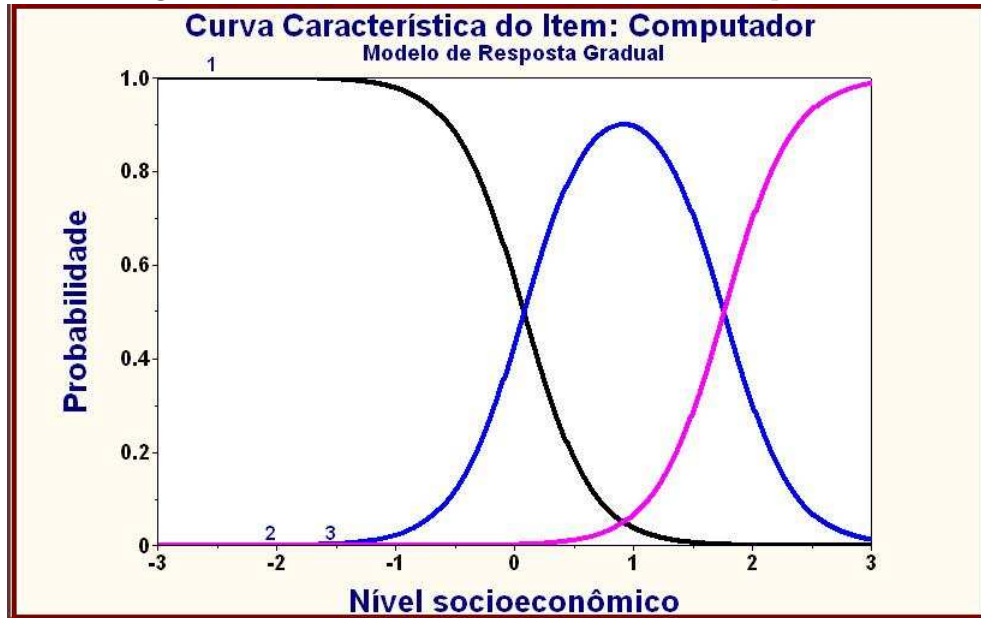
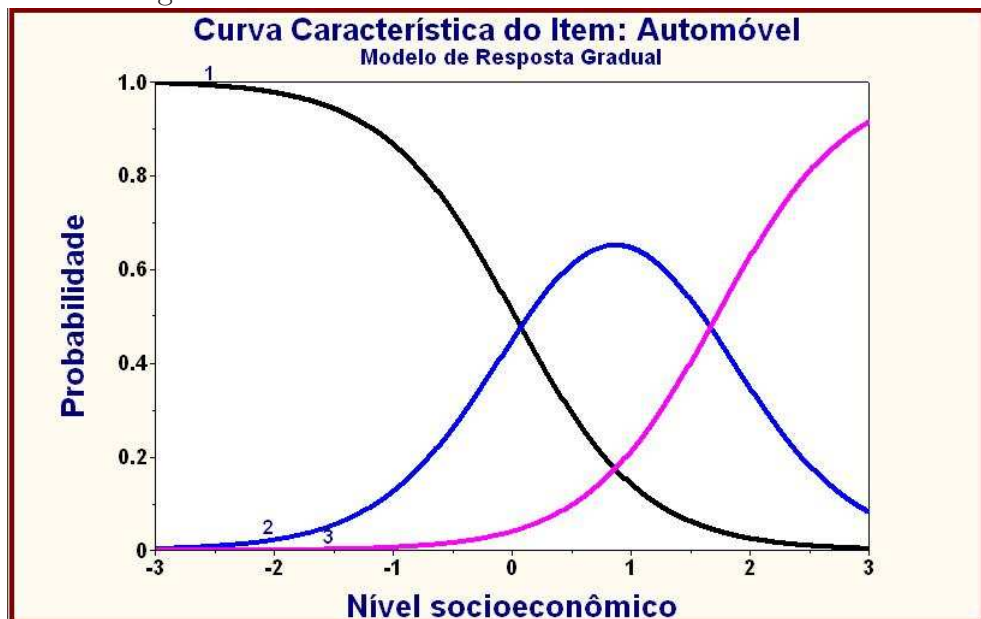
Figura B.9 *Curva Característica do Item Computador*Figura B.10 *Curva Característica do Item Automóvel*

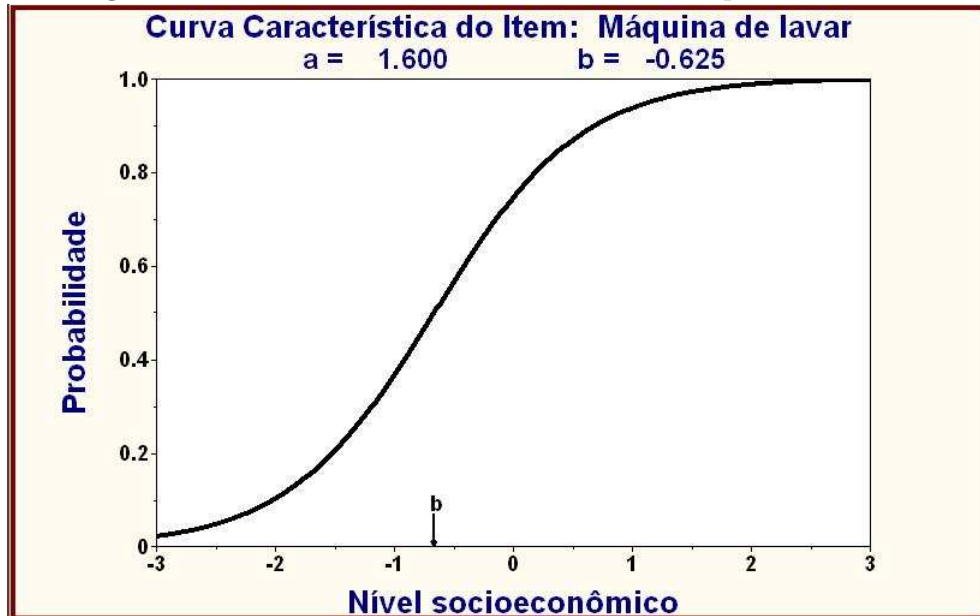
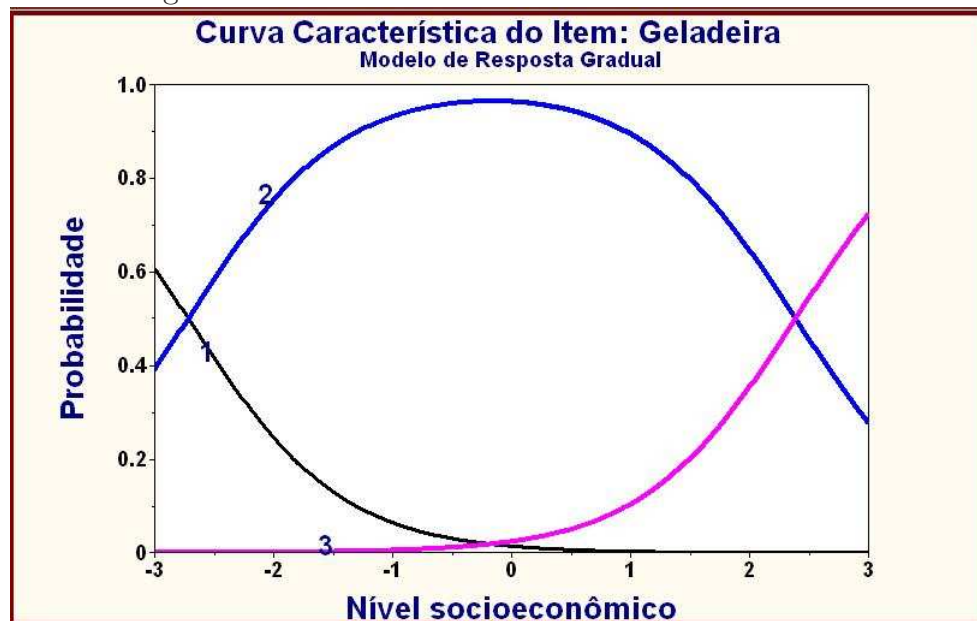
Figura B.11 *Curva Característica do Item Máquina de Lavar*Figura B.12 *Curva Característica do Item Geladeira*

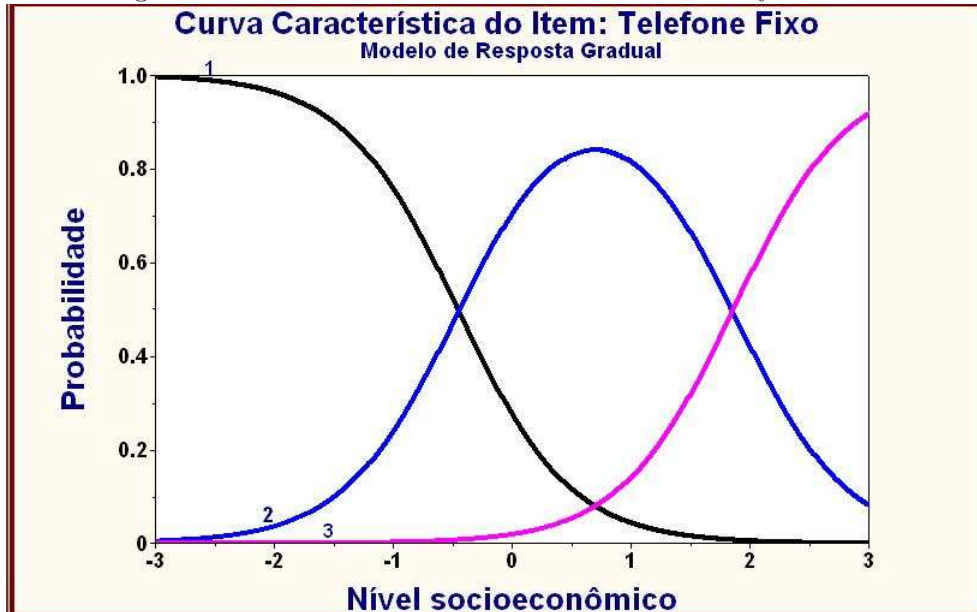
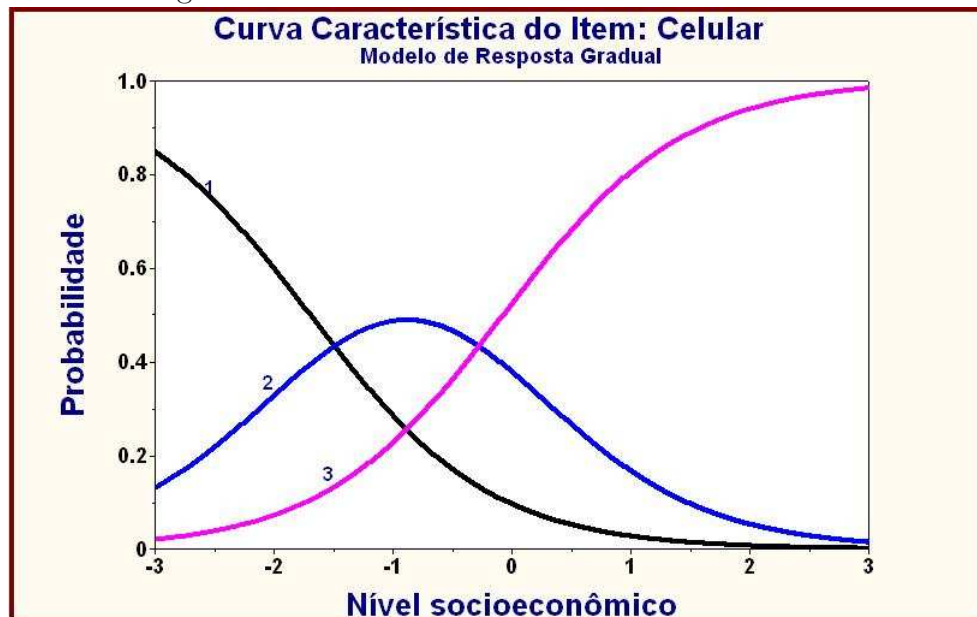
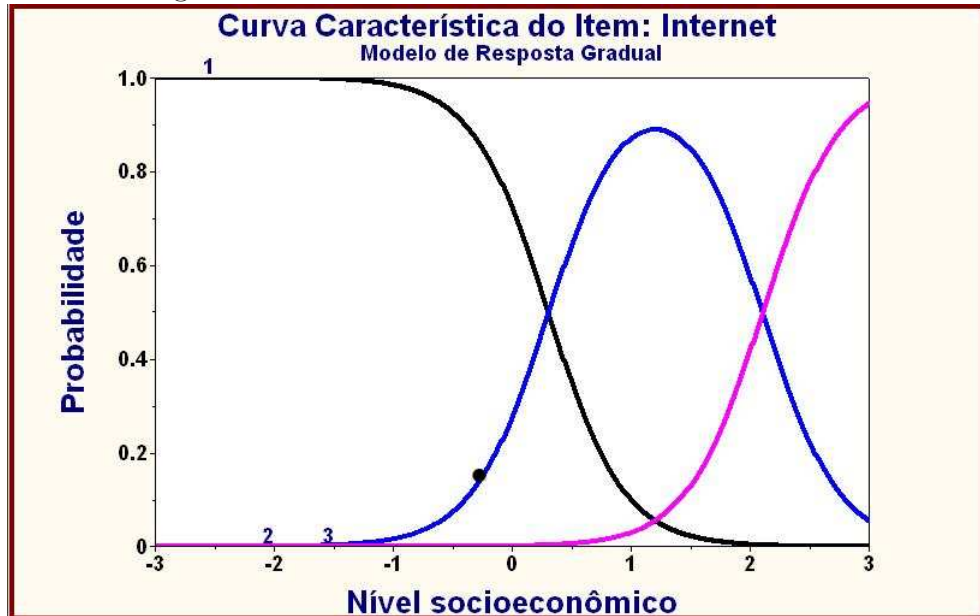
Figura B.13 *Curva Característica do Item Telefone Fixo*Figura B.14 *Curva Característica do Item Celular*

Figura B.15 *Curva Característica do Item Internet*Figura B.16 *Curva Característica do Item Televisão a cabo*