

ALEXANDRE WEBER ARAGÃO VELOSO

**DETERMINANTES DE DESEMPENHO EM CONCURSOS PÚBLICOS:
UM ESTUDO DE CASO**

FORTALEZA/CEARÁ
2004

ALEXANDRE WEBER ARAGÃO VELOSO

**DETERMINANTES DE DESEMPENHO EM CONCURSOS PÚBLICOS:
UM ESTUDO DE CASO**

Dissertação de Mestrado apresentada ao Curso de Pós-Graduação em Economia da Universidade Federal do Ceará, CAEN/UFC, como requisito parcial para obtenção do título de Mestre.

Área de Concentração: Teoria Econômica

Orientador:
Prof. Dr. Luiz Ivan de Melo Castelar

FORTALEZA/CEARÁ
2004

Esta dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de mestre em economia, outorgado pela Universidade Federal do Ceará, e encontra-se a disposição dos interessados na Biblioteca do Curso de Mestrado em Economia da referida Universidade.

A citação de qualquer trecho desta dissertação é permitida, desde que feita em conformidade com as normas científicas.

Alexandre Weber Aragão Veloso

BANCA EXAMINADORA

Prof. Luiz Ivan de Melo Castelar
Orientador

Prof. José Raimundo Carvalho
Membro da Banca Examinadora

Prof. Manoel Bosco de Almeida
Membro da Banca Examinadora

FORTALEZA/CEARÁ
2004

Veloso, Alexandre Weber Aragão

Determinantes de desempenho em concursos públicos: um estudo de caso/Alexandre Weber Aragão Veloso.

80p.

Dissertação (Mestrado). Universidade Federal do Ceará. CAEN

1.Avaliação de desempenho 2. Econometria 3. Mercado de trabalho I. Título

CDD – 330.0151

AGRADECIMENTOS

Aos que fizeram parte da minha turma, sempre companheiros com suas presenças e atitudes.

Ao Alan, ao Francis e ao Paulo, pela amizade.

À Maria Águeda pelo carinho e apoio.

À Virna Weber pelo carinho e compreensão.

Ao Rutênio Sampaio, ao Fábio Bastos, à Maria Salete, ao Roberto Smith e ao Valdemar Bandeira, pessoas que estiveram presentes, em diferentes momentos, me apoiando.

Aos professores que ministraram as disciplinas que cursei e aos funcionários e prestadores de serviços do CAEN, que diariamente criam, com competência, condições para o bom andamento das atividades nessa instituição.

Aos membros da banca que colaboraram com sugestões e contribuíram para uma melhor qualidade desse trabalho.

Ao Ilton Soares que ajudou de forma decisiva no aprimoramento dessa dissertação.

Ao Prof. Ivan Castelar que mais uma vez se mostrou extremamente generoso com sua forma atenciosa, compreensiva e dedicada de orientar, assim como o fez na minha graduação, apoiando-me durante o transcorrer do curso, principalmente nos momentos mais delicados. (*Fortiter in re, suaviter in modo*)

RESUMO

O concurso público para provimento de cargos do aparelho do Estado brasileiro é um forte atrator de recursos humanos, e no qual se observa uma grande demanda de pessoas buscando uma carreira através da aprovação num conjunto de provas. A principal preocupação deste trabalho foi a de investigar os determinantes da aprovação de um candidato num concurso público. Para tanto fez-se um estudo de caso com base em dados e resultados de um concurso para provimento de cargo administrativo de nível médio para o Banco do Nordeste do Brasil, do qual participaram 232.771 candidatos. Discute-se ainda o problema paralelo da super-educação como forma de ilustrar uma possível má alocação de recursos no mercado de trabalho. Na base de dados utilizada, além do resultado do concurso, constam informações referentes à renda, educação e outros atributos mais gerais como faixa de idade, estado civil e local de origem do candidato. Na regressão o resultado do candidato no concurso (aprovado ou reprovado) é a variável a ser explicada, e o vetor de variáveis explicativas é formado pelos atributos dos candidatos. Utiliza-se inicialmente o modelo *logit* de escolha binária e a estimação é baseada no método de máxima verossimilhança. Resultados padrões foram obtidos nos testes de razão de máxima verossimilhança e de significância estatística de cada um dos parâmetros estimados, onde somente uma variável, referente ao estado civil, não se mostrou estatisticamente significativa. As que se mostraram significativas apresentaram os sinais esperados quanto ao efeito marginal sobre a chance de aprovação. Quanto à magnitude dos coeficientes, o dado relevante é que a condição do candidato ter o grau de escolaridade requerido para a função a ser exercida (nível médio) foi o que apresentou maior impacto (negativo) na chance do candidato ser aprovado. Para avaliar problemas de especificação do modelo dois tipos de teste foram realizados. O primeiro foi um teste de heterocedasticidade, cujo objetivo foi detectar erro de especificação para o caso de heterogeneidade não observada. Detectou-se ausência de erro de especificação. O segundo teste foi quanto a forma funcional, onde confrontou-se os modelos *logit e probit* e, em seguida, cotejou-se o melhor dentre esses modelos simétricos versus um com distribuição *gompit*, como representativa da classe assimétrica. Como resultado obteve-se uma razoável confiança de que o modelo *logit* é bem especificado. As principais conclusões do trabalho apontam o ensino médio público como um possível *locus* de ação e políticas que atuem para reduzir as desigualdades de condições iniciais de cada candidato a concursos públicos. Por outro lado, o preenchimento de postos de trabalho que requerem apenas uma educação de nível médio com detentores de título universitário, e até mesmo com pós-graduação, aponta para uma má alocação dos recursos públicos conjugada com uma possível deficiência estrutural do mercado de trabalho regional.

ABSTRACT

The main objective of this thesis was to investigate the factors determining success in an aptitude test for admission in a public bank in the Northeast of Brazil. The data set is based in a questionnaire for 232,771 Candidates which is composed of the result of the test and a number of questions about some attributes of the candidate. Attributes selected were personal monthly income; household monthly income; education level; type of school attended (public or private); marital status; occupation and age. A *logit* model was used to estimate a regression of success in the test against the attributes defined above. Results of the estimated model show that all the coefficients but marital status were significant and with the expected sign. The marginal effects inform that a high school level (a requirement to take the test) had the strongest and negative impact on the probability of success. Specification tests for functional form and unobserved heterogeneity show that the model was correctly specified.

SUMÁRIO

LISTA DE TABELAS	vi
LISTA DE FIGURAS	vii
CAPÍTULO 1.0 - INTRODUÇÃO	01
CAPÍTULO 2.0 - REVISÃO DA LITERATURA SOBRE AVALIAÇÃO DE DESEMPENHO	03
CAPÍTULO 3.0 - A BASE DE DADOS DISPONÍVEL E SUA UTILIZAÇÃO PARA DETECTAR OS DETERMINANTES DE DESEMPENHO	08
3.1. A BASE DE DADOS	08
3.1.1. Análise Descritiva da Base de Dados	11
3.2. IDENTIFICANDO DETERMINAÇÕES RELEVANTES QUANTO AO DESEMPENHO DOS CANDIDATOS NO CONCURSO	19
3.2.1. Características do Exame do Concurso do Banco do Nordeste do Brasil.....	19
3.2.2. Informações do Exame Nacional do Ensino Médio/ENEM	21
3.2.3. Digressão sobre o Caso da Educação Formal do Candidato ser Superior ao Exigido no Concurso	24
3.3. DEFINIÇÃO DAS VARIÁVEIS EXPLICATIVAS	30
3.3.1. Relevância e Definição do Formato de cada Variável Explicativa ..	31
3.3.2. Síntese da Definição das Variáveis Explicativas	40
CAPÍTULO 4.0 - METODOLOGIA	42
4.1. ESPECIFICAÇÃO DO MODELO ECONOMETRICO	42
4.1.1. O Modelo de Regressão	42
4.1.2. Estimação e Inferência em Modelos de Escolha Binária	48

CAPÍTULO 5.0 - RESULTADOS E CONCLUSÕES	59
5.1. RESULTADOS	59
5.2. CONCLUSÕES	69
BIBLIOGRAFIA	71
APÊNDICE	74

LISTA DE TABELAS

TABELA 01 - Quesitos e respostas possíveis do Questionário Socioeconômico Respondido pelos Candidatos ao Concurso Público do Banco do Nordeste do Brasil – 2003	09
TABELA 02 - Índice Preliminar de Desempenho - Tipo de Escola de Nível Médio .	33
TABELA 03 - Índice Preliminar de Desempenho - Grau de Instrução	34
TABELA 04 - Índice Preliminar de Desempenho - Renda Pessoal Mensal	35
TABELA 05 - Índice Preliminar de Desempenho - Renda Familiar Mensal	36
TABELA 06 - Índice Preliminar de Desempenho - Situação no Mercado de Trabalho.	37
TABELA 07 - Índice Preliminar de Desempenho - Estado Civil	38
TABELA 08 - Índice Preliminar de Desempenho - Faixa Etária	39
TABELA 09 - Índice Preliminar de Desempenho - Local de Origem	40
TABELA 10 - Especificação das Variáveis Explicativas e Sinais Esperados	41
TABELA 11 - Resultado do Modelo <i>Logit</i>	61
TABELA 12 - Poder de Predição da Estimação do Modelo <i>Logit</i>	62
TABELA 13 - Efeitos Marginais sobre a Chance de ser Aprovado no Concurso	67

LISTA DE FIGURAS

FIGURA 01 - Comparativo dos Perfis dos Candidatos quanto ao Tipo de Escola de Ensino Médio – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	12
FIGURA 02 - Comparativo dos Perfis dos Candidatos quanto ao Grau de Instrução – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	13
FIGURA 03 - Comparativo dos Perfis dos Candidatos quanto a Renda Pessoal Mensal – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	14
FIGURA 04 - Comparativo dos Perfis dos Candidatos quanto a Renda Familiar Mensal – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	15
FIGURA 05 - Comparativo dos Perfis dos Candidatos quanto a Situação no Mercado de Trabalho – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	16
FIGURA 06 - Comparativo dos Perfis dos Candidatos quanto ao Estado Civil – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	16
FIGURA 07 - Comparativo dos Perfis dos Candidatos quanto a Faixa Etária – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	17
FIGURA 08 - Comparativo dos Perfis dos Candidatos quanto ao Lugar de Origem – Confronto entre Participação Percentual no Total de Candidatos com a Participação Percentual dentre os Aprovados	18

FIGURA 09 - Desempenho dos Participantes do ENEM/2003 nas Provas Objetiva e de Redação – Por Faixa de Renda da Família	23
FIGURA 10 - Modelo para uma Probabilidade	45

1.0 INTRODUÇÃO

A instituição do concurso público para provimento de cargos do aparelho do Estado Brasileiro tem sido um forte atrator de recursos humanos desde meados do século passado. Assim, uma elite egressa da classe média projetou-se na conformação da burocracia estatal através do concurso público.

A razão para esta atração pode ser explicada por alguns fatores. Um deles é o de que o setor público, em muitos níveis e regiões, vem pagando salários melhores do que os praticados na iniciativa privada. Aliado a isso, existem também benefícios indiretos, tais como planos de saúde e previdência diferenciada.

Outra explicação para esta atração pode ser a idéia de que estar alocado no setor público proporciona um grau de segurança maior quanto à permanência no emprego, dada a legislação que garante certa estabilidade ao funcionário público, o que não se observa no setor privado.

O medo do desemprego também pode ser uma força motivadora deste atrativo. Ou mesmo a vontade das pessoas em seguir uma carreira pública, pois é característico destas funções a existência de planos de cargos e salários diferenciados para alguns concursos.

Todavia, este trabalho toma o desejo comum desta população de candidatos como um fato contemporâneo do cotidiano nacional. Neste contexto, determinadas pessoas buscam uma carreira pública através da aprovação num conjunto de provas, onde apenas algumas são aprovadas.

A questão primordial a ser colocada é: o que faz determinadas pessoas serem aprovadas em detrimento de outras? Será que existem fatores socioeconômicos que aumentam as chances de uma pessoa ter um melhor desempenho? Outras questões subseqüentes a esta são: maior renda implica em maiores chances de passar?; Ter nível de educação acima do exigido no concurso implica em maiores chances de passar?; etc. Assim, o objetivo principal deste trabalho é investigar os determinantes da aprovação em um determinado concurso público.

Para tanto, fez-se um estudo de caso com base em dados e resultados de um concurso público para provimento de cargo administrativo de nível médio para o Banco do Nordeste do Brasil (BNB), realizado no ano de 2003.

Quanto ao caso estudado nesse trabalho, tem-se a particularidade de ter sido um concurso para uma instituição localizada em uma região de menor renda, quando comparada às regiões Sudeste e Sul do Brasil. Isso reflete ainda mais a necessidade de se entender o que explica a aprovação de um candidato. Pois, de forma patente, a região Nordeste se ressentia de ofertas de vagas de trabalho atrativas. Assim, por exemplo, se os candidatos provenientes de famílias de melhor estrutura financeira tiverem maiores chances de serem aprovados, poderá existir um círculo vicioso de concentração de renda.

Na seqüência do trabalho estas questões serão devidamente desenvolvidas. E adiante, além desta introdução, tem-se mais quatro capítulos.

No Capítulo 2, é feito um levantamento da literatura sobre avaliação de desempenho nos casos em que a variável a ser explicada é matematicamente limitada, como é o caso desta pesquisa. Discute-se ainda, neste capítulo, o problema paralelo da super-educação como forma de ilustrar uma possível má alocação de recursos no mercado de trabalho, o que geralmente ocorre em concursos dessa natureza.

No Capítulo 3, é apresentada a base de dados que será utilizada. Nesta base constam os resultados finais dos candidatos que participaram da seleção e o questionário socioeconômico respondido por estes. Ainda nesse capítulo são apresentados elementos que indicam a pertinência da utilização dos dados disponíveis como variáveis explicativas do desempenho dos candidatos no concurso. Já no Capítulo 4 se expõe a metodologia a ser adotada para se buscar a constatação empírica dos determinantes da aprovação do candidato no concurso. Nesta metodologia se discutirá, particularmente, as técnicas econométricas de estimação de modelos de escolha discreta e sua pertinência neste estudo. E por fim, no Capítulo 5, são apresentados os resultados empíricos e feitas as considerações finais.

2.0 REVISÃO DA LITERATURA SOBRE AVALIAÇÃO DE DESEMPENHO

Este capítulo propõe-se a discutir algumas referências bibliográficas, as quais versam sobre métodos que poderiam ser usados para a avaliação do desempenho de candidatos em concursos públicos. O que esses artigos têm em comum, precisamente, é a avaliação de desempenho ou presença utilizando-se de exercícios econométricos com uma variável dependente limitada.

Em *Probit Analysis and Economic Education*, Spector e Mazzeo (1980), buscaram avaliar a efetividade de um novo método de ensino em um curso intermediário de macroeconomia. Para tanto, usaram um modelo de escolha binária *Probit* para avaliar o desempenho dos alunos, onde a variável dependente utilizada foi o conceito final do curso, onde para ocorrência de conceito A a variável dependente assumiu valor unitário e para os conceitos B e C o valor era zero. Como regressores foram usadas, a nota média de admissão do aluno, a pontuação obtida em uma prova que foi aplicada no início do curso para captar os conhecimentos prévios de macroeconomia e uma variável *dummy* que assumia valor unitário para os alunos que passaram pelo novo método de ensino e zero para os que fizeram o curso pelo método vigente. Os resultados apontaram para um impacto positivo no desempenho dos alunos dada à utilização do novo método de ensino.

Considerando que nos Estados Unidos, o curso introdutório para a universidade é o fundamento da maioria dos programas acadêmicos de economia e um pré-requisito para outros programas de estudo, Brasfield, Harrison e McCoy (1993), em *The Impact of High School Economics on College Principles*, usam a estrutura educacional adaptada numa função de produção e técnicas de estimação *Probit*, para avaliar a importância do estudo da economia no *college* para o desempenho dos alunos de turmas introdutórias de economia da *Murray State University*, de *kentucky*. Nesse estudo encontrou-se uma relação positiva entre os alunos que tiveram aulas de economia no colégio e suas notas nas classes introdutórias de Macroeconomia e Microeconomia.

Em *Economic Literacy Among Corporate Employees*, Wood e Doyle (1999) relatam os resultados de uma pesquisa feita por telefone entre 1.001 empregados de grandes corporações norte-americanas. A pesquisa constou de um conjunto de 20 questões, sobre problemas econômicos, e buscou saber se acertar mais questões podia ser explicado por um maior nível de educação, uma maior renda, sexo, etc.

Nesse trabalho, Wood e Doyle modelaram o número de respostas corretas como um processo *Poisson* com média condicional e variância determinada por um vetor de variáveis explicativas (esta técnica é apresentada em detalhes por Cameron & Trivedi, 1998). Nesse estudo, a variável dependente representava o número de questões corretas.

Como variáveis explicativas foram utilizadas *dummies* para gerência e outros cargos ocupados pelos questionados; níveis de educação; faixas de renda; origem (hispânica, afro-americana e asiática); faixas de idade; sexo; identidade política (liberal ou conservador); tipo de graduação; e se o entrevistado, em algum momento, cursou uma disciplina de economia.

Os resultados obtidos foram que um melhor desempenho nas respostas do questionário (acertar mais) está associado a um maior nível de educação, ter feito disciplina de economia na faculdade, ter renda mais alta e ser do sexo masculino. O restante das possíveis variáveis explicativas foi estatisticamente não significante neste estudo.

Em *Personality Type and Student Performance in Upper-Level Economics Courses: The Importance of Race and Gender*, Borg e Stranaham (2002) buscam testar a hipótese de que o tipo de personalidade é uma importante variável explicativa na performance dos cursos avançados de economia. Os autores, baseados em uma teoria do psicanalista Carl Gustav Jung, utilizaram a hipótese de que estudantes de economia que fossem introvertidos se sairiam melhor que os estudantes extrovertidos.

Para o estudo, tomaram 119 estudantes das disciplinas de "princípios de macroeconomia", "economia do trabalho" e "finanças públicas" na Universidade do Norte da Flórida, e mais 617 estudantes das disciplinas de "princípios de microeconomia" e outras na Universidade de Miami.

Quanto à metodologia, Borg e Stranaham construíram um indicador de personalidade chamado *Myers Briggs Type Indicator* (MBTI), baseados no trabalho de Myers (1975), feito a partir de questionários aplicados aos estudantes observados. O MTBI foi estruturado nas

classificações de consciência mental da teoria de *Carl Gustav Jung*, que acreditava que certas personalidades são melhor adaptáveis a certas profissões, ou seja, extrovertidos têm maiores chances de boas notas em cursos de publicidade, por exemplo.

Para testar a hipótese colocada, os autores usaram a estimação de um modelo de escolha binária *Probit*, onde intervalos de notas em provas foram utilizados como variável dependente. As variáveis explicativas utilizadas foram *dummies* para: sexo do aluno observado; raça; tipo de disciplina que cursava (macro, micro, etc.); existência de monitor para a disciplina; e os oito tipos de personalidades *Junguianas* identificadas pelo MTBI.

Depois de vários arranjos para as regressões, os autores apresentam os três melhores resultados. Além das interações das personalidades com as notas, fizeram interações das raças com as personalidades, e identificaram que o tipo de personalidade é uma importante variável explicativa para o desempenho dos estudantes nos curso de economia, confirmando a hipótese levantada.

Uma outra referência bibliográfica sobre esse tema é *Gender and the Study of Economics: The Role of Gender of the Instructor*. Nesse trabalho *Robb e Robb* (1999) procuraram identificar o motivo do declínio do número de matrículas nos cursos de economia nos EUA e no Canadá. Para tanto, eles consideraram que a decisão do estudante em cursar economia envolve dois estágios (conforme a estrutura de ensino local). No primeiro estágio o estudante decide fazer economia no primeiro ano da universidade; e, no segundo estágio, o estudante decide se vai continuar a cursar economia depois dos estudos introdutórios.

A informação relevante, até então percebida pelo meio acadêmico local, é que a taxa de desistência das mulheres é significativamente superior a dos homens. Em estudos anteriores descobriu-se que a decisão pelo curso envolve a facilidade de tirar boas notas, além do interesse prévio por assuntos econômicos. O trabalho de *Robb e Robb* buscou aprofundar e testar a hipótese de que o tipo de professor encontrado nos cursos introdutórios baliza a decisão do estudante em continuar cursando economia.

Quanto aos dados, foram observados 2.552 estudantes (823 mulheres) da primeira classe de microeconomia na *Brock University* de *Ontario* entre 1989 e 1995. Neste período, foram oito os professores que lecionaram, quatro homens e quatro mulheres. Todos os professores

usaram os mesmos livros e as mesmas estruturas de provas. E as diferenças básicas entre os professores foram: sexo; ter doutorado ou não; e ser de dedicação exclusiva ou não.

A estrutura da estimação foi desenvolvida em duas etapas. Na primeira etapa foi feita uma estimação com mínimos quadrados onde a equação tinha como variável dependente a nota do aluno na classe de microeconomia, e como variáveis explicativas um vetor de características dos estudantes; um vetor de características dos professores; *dummies* para alunos que estudaram cálculo, álgebra e outras disciplinas de economia.

Visto que o estudante faz microeconomia e disciplinas de outros cursos no primeiro ano de faculdade (direito, sociologia, fotografia, etc.), concluiu-se que tirar boa nota em microeconomia está relacionado com tirar, concomitantemente, boa nota em todas as outras disciplinas (se o aluno é bom estudante tira boa nota em todas as disciplinas que se matriculou). Quanto a continuar estudando economia isso depende, então, de outros fatores.

Na segunda etapa do trabalho os autores usaram um modelo de escolha binária *Probit* com a equação da regressão representando a variável dependente por estudante segue ($y=1$) ou não ($y=0$) no curso de economia e essa é explicada por notas no curso introdutório; vetor de características dos estudantes e vetor de características dos professores. Concluindo-se que se o primeiro instrutor for do sexo feminino aumenta em muito as chances das alunas (mulheres) seguirem no curso. Entretanto, os demais resultados não foram muito conclusivos.

Em *Gender Economics Courses in Liberal Arts Colleges: Further Results*, Greene (1998) toma a base de dados de Burnett (1997) e aplica uma técnica de dois estágios para observar se chegaria aos mesmos resultados por maneiras distintas. A idéia de fundo é a de simplesmente explicar a presença de cursos de economia em algumas escolas de arte.

Assim, Greene desenvolveu um modelo *Probit* de estimação simultânea. Durante o trabalho discute o problema de especificação nos modelos *Probit*, sugerindo que muitas destas estimações devem ser feitas de maneira simultânea.

Greene aponta que a escola de arte ter ou não um curso de economia pode ter relação com ela ser só para mulheres (escolas só para mulheres tenderiam a não ter cursos de economia). Mas que o fato da escola ser só para mulheres pode ser explicado por muitas das variáveis explicativas do fato da escola ter ou não curso de economia, o que levaria a um problema de má especificação do modelo.

O procedimento usado pelo autor é feito com duas equações simultânea descritas assim:

- I. O Curso é só para Mulheres (1) ou Não (0) = Função (Um escore de reputação acadêmica da universidade, variando entre 0 e 1, pior e melhor, respectivamente; *dummy* se a faculdade é de cunho religioso ou não; e *dummies* para regiões geográficas onde é localizada a faculdade);
- II. Possuir (1) ou não (0) uma disciplina de economia = Função (O mesmo escore de reputação acadêmica da universidade; *dummy* se o curso é só para mulheres; *dummy* se faculdade possui ou não departamento de economia; e *dummies* para regiões geográficas).

Segundo argumenta *Greene* em suas considerações finais, seus resultados foram melhores que os de *Burnett* porque um problema de especificação do modelo teria sido corrigido pela estimação simultânea.

Nota-se nesta breve resenha que todos os autores subentendem que uma teoria pode ser criada a partir de alguns resultados empíricos encontrados. Em geral, eles tomam uma questão particular e testam suas hipóteses através de métodos econométricos de variável dependente limitada. Portanto, estas equações não possuem como fundamento um modelo econômico que indique o seu formato ou comportamento.

Por fim, vale destacar que com a notável exceção do artigo de *Greene* (1998), o qual se preocupa com o viés de simultaneidade, nenhum destes trabalhos enfrentou com algum rigor econométrico o problema de erros de especificação no modelo.

3.0 A BASE DE DADOS DISPONÍVEL E SUA UTILIZAÇÃO PARA DETECTAR OS DETERMINANTES DE DESEMPENHO

3.1. A BASE DE DADOS

Dentro do contexto de dificuldades de obtenção de informações para pesquisas que utilizam micro-dados, vale antes de mais nada ressaltar a relevância da base de dados utilizada nesse estudo. Ao contrário, dos Estados Unidos da América e de países europeus, a instituição do concurso para admissão no serviço público é, no Brasil, não só praticada, mas também condição necessária para o ingresso no serviço público. Dos concursos em que a instituição organizadora toma o cuidado em solicitar aos candidatos o preenchimento de um questionário socioeconômico, gera-se uma base de micro-dados que não é de fácil obtenção e disponibilidade. Uma das temáticas que se ressentem de base de dados ampla é a que trata de estudos sobre determinantes de desempenho. E, para se ter idéia da relevância da base de dados utilizada nesse trabalho, participaram do concurso em questão 232.771 candidatos.

Os dados disponíveis referem-se ao concurso público de seleção externa, do Banco do Nordeste do Brasil, realizado no ano de 2003, para provimento de vagas para uma função¹ que requeria como educação formal o Ensino Médio. Era exigido, também, idade mínima de 18 anos. As vantagens e remuneração oferecidas eram: salário base de R\$ 777,06, pouco mais de 3 salários mínimos da época², direitos previstos na Consolidação das Leis do Trabalho (CLT); e possibilidade de participação em planos assistenciais de saúde e de previdência complementar.

Os candidatos responderam um questionário socioeconômico de múltipla escolha com quesitos referentes à instrução, atributos pessoais, renda, local de origem e situação quanto ao

¹ No edital do concurso é feita uma discriminação sumaria das atividades as serem desenvolvidas, quais sejam: Realizar atividades operacionais, executando as rotinas e processos de serviço de sua unidade, dentro dos padrões de qualidade, produtividade e tempestividade estabelecidos pelo Banco, para atendimento dos clientes internos e externos.

² Salário Mínimo em 2003, na época do concurso era de R\$240,00.

mercado de trabalho. Na Tabela 01 são relacionados os quesitos do questionário socioeconômico passíveis de serem utilizados no presente trabalho, assim como suas opções de resposta.

TABELA 01 – QUESITOS E RESPOSTAS POSSÍVEIS DO QUESTIONÁRIO SOCIOECONÔMICO RESPONDIDO PELOS CANDIDATOS AO CONCURSO PÚBLICO DO BANCO DO NORDESTE DO BRASIL – 2003

QUESITO	RESPOSTAS POSSÍVEIS
Tipo de Escola do Nível Médio	Toda Pública Toda Privada Maior Parte Pública Maior Parte Privada Metade Pública Metade Privada
Grau de Instrução	Médio Superior Incompleto Superior Completo Pós-Graduação
Renda Pessoal Mensal (em Salários Mínimos)	até 1 de 1 a 2 de 2 a 4 de 4 a 6 ou acima de 6
Renda Familiar Mensal (em Salários Mínimos)	até 3 de 3 a 10 de 10 a 20 de 20 a 30 ou acima de 30
Situação no Mercado de Trabalho	Empregado Público Empregado Privado Autônomo Estagiário Desempregado

Continua...

TABELA 01 – QUESITOS E RESPOSTAS POSSÍVEIS DO QUESTIONÁRIO SOCIOECONÔMICO RESPONDIDO PELOS CANDIDATOS AO CONCURSO PÚBLICO DO BANCO DO NORDESTE DO BRASIL – 2003 (CONTINUAÇÃO)

Continuação...

QUESITO	RESPOSTAS POSSÍVEIS
Estado Civil	Solteiro Casado Separado Viúvo outros
Faixa Etária	de 18 a 22 de 23 a 27 de 28 a 32 de 33 a 37 acima de 37
Local de Origem	Região Metropolitana Interior Outra Unidade Federativa Exterior

Fonte: Associação Cearense de Estudos e Pesquisas (ACEP) / Banco do Nordeste do Brasil (BNB)

Da base de dados foram excluídas as informações referentes aos candidatos que deixaram de responder algumas das perguntas, exceto à questão referente a renda pessoal, em virtude de não se ter nesse quesito uma opção “sem rendimentos”.

Excluiu-se, então, informações de 463 candidatos, que representa apenas 0,2% dos candidatos. Nenhum dos excluídos foi aprovado no concurso.

Dado que o número de candidatos que atingiu o perfil mínimo da prova foi inferior ao número de vagas ofertadas, não houve candidato aprovado se valendo da condição de cotas exclusivas a portadores de deficiências físicas.

Isso implica na inexistência de candidatos aprovados em condição diferenciada. Ou seja, todos os aprovados chegaram a essa condição dentro de um mesmo critério de melhor desempenho, observado o atendimento do perfil mínimo exigido em edital.

3.1.1. ANÁLISE DESCRITIVA DA BASE DE DADOS

Visto que a problemática da identificação de determinantes de desempenho é o foco desse trabalho, a exposição das características da base de dados disponível direcionou-se nesse sentido. Nessa subseção tratam-se os dados de cada quesito considerado do questionário socioeconômico (Tabela 01) de forma a se expor o percentual de ocorrência das respostas tanto dentro o total de candidatos que participaram do exame, como dentro do universo dos aprovados.

A intenção é comparar a participação relativa de uma resposta dentro do total de candidatos com a participação relativa da mesma resposta no universo dos aprovados. Ou seja, se, por exemplo, 70% dos candidatos que participaram do exame são solteiros e o percentual de participação desse atributo é maior que os 70% dentre os que foram aprovados, isso indica vestígio quanto à influência positiva do estado civil solteiro sobre o desempenho nos exames. Caso ocorra o contrário, o percentual de solteiros aprovados for inferior a 70%, os indícios seriam de um impacto negativo de ser solteiro nos resultados do exame.

Sendo assim, será exposto, para cada quesito (Tabela 01), o comparativo dos percentuais de participação de cada resposta dentre os candidatos que participaram do concurso, com relação aos percentuais dentro do universo dos aprovados. Cabe explicitar que esse tipo de abordagem trata-se de uma observação preliminar, onde a intenção é balizar o aproveitamento das informações quanto a sua utilização na composição do estudo empírico que se seguirá nos capítulos seguintes.

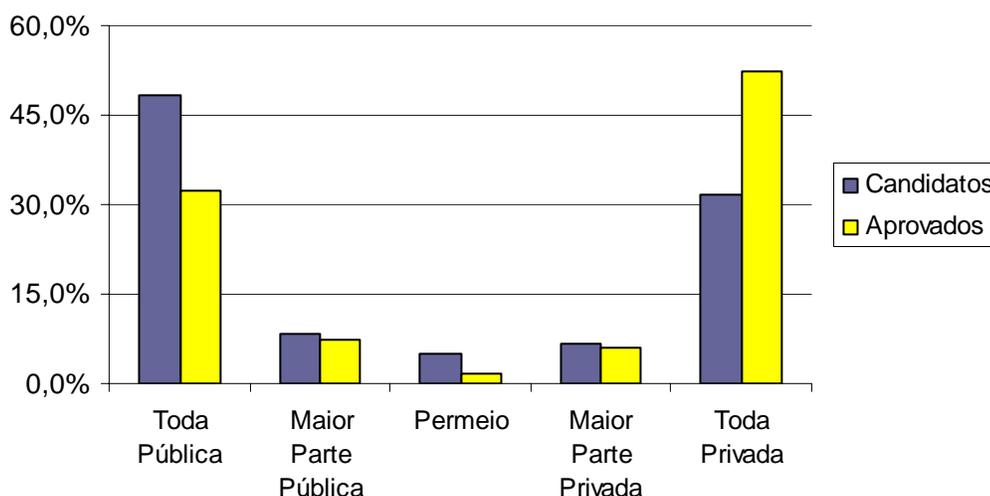
Na Seção 3.3 retoma-se as informações trabalhadas em todo esse capítulo de forma a se obter um diagnóstico consolidado das variáveis disponíveis para a determinação do desempenho dos candidatos no concurso público em questão.

3.1.1.1. TIPO DE ESCOLA DO NÍVEL MÉDIO

Parcela significativa dos candidatos (80%) freqüentou o ensino médio todo em escola pública, ou todo em escola privada. Observa-se na contraposição dos percentuais de participação

(ver Figura 01) que, à medida que se possui uma maior quantidade de tempo de estudo em escolas do ensino privado, a indicação quanto ao desempenho avança no sentido de se ter uma influência positiva. Destaca-se um percentual maior dentre os aprovados face ao percentual dentre os candidatos, somente entre os que fizeram todo o nível médio em escola particular. No outro extremo, os alunos que fizeram todo o ensino médio em escola pública apresentam uma queda considerável na participação entre os aprovados.

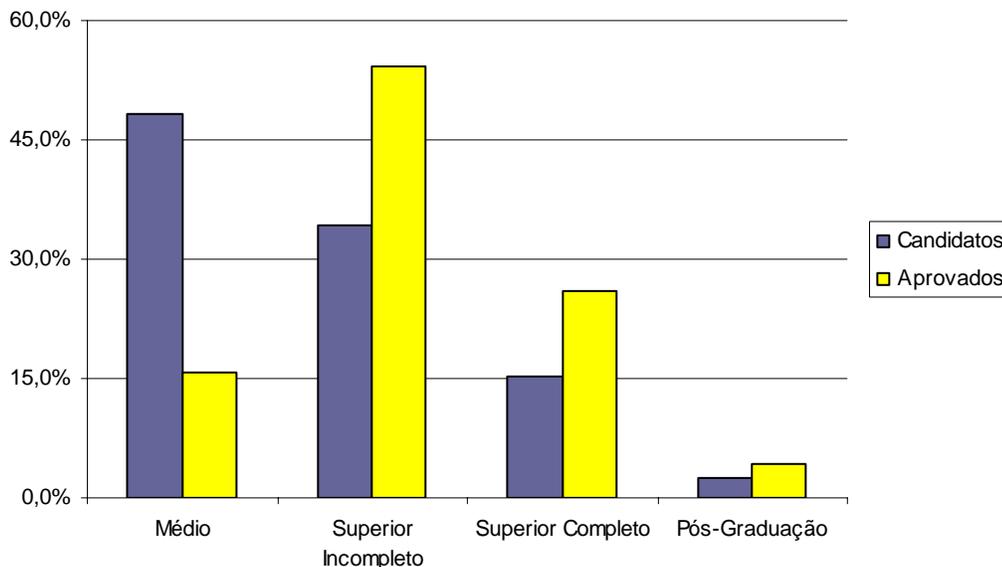
FIGURA 01 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO AO TIPO DE ESCOLA DE ENSINO MÉDIO – CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS



3.1.1.2. GRAU DE INSTRUÇÃO

Parcela significativa dos candidatos (82,5%) possuía ensino médio ou estava por completar estudos em cursos de nível superior. Destaca-se nesse quesito o percentual maior dentre os aprovados face ao percentual dentre os candidatos, dos que tem formação acima da exigida pelo concurso (ver Figura 02). Os que possuem somente o ensino médio representavam quase metade dos candidatos, contudo somente 15,5% dos aprovados possuíam apenas o nível médio.

FIGURA 02 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO AO GRAU DE INSTRUÇÃO – CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS



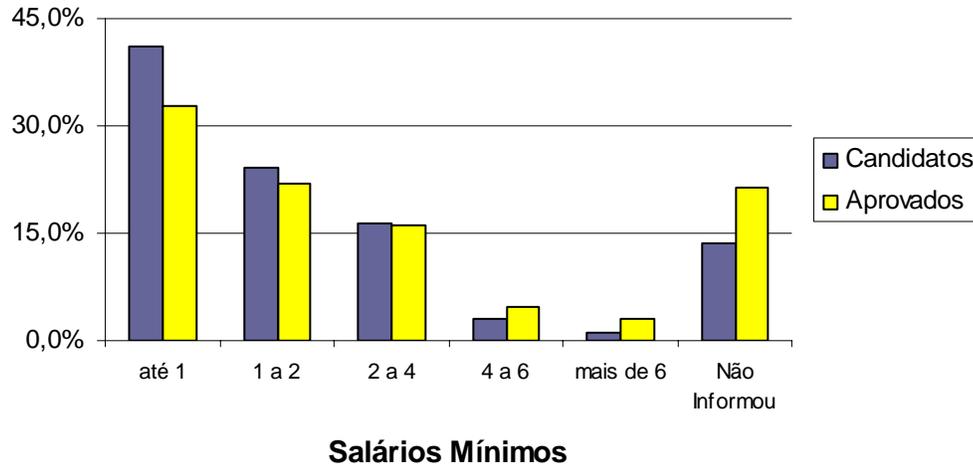
3.1.1.3. RENDA PESSOAL MENSAL

Do total de concorrentes 65,5% eram pessoas com renda pessoal abaixo de 2 salários mínimos. Observa-se nesse quesito que à medida que a faixa de renda pessoal é maior o percentual dentre os aprovados se avoluma face ao percentual dentre os participantes do concurso (ver Figura 03).

Somente a partir da faixa de 4 a 6 salários mínimos de renda pessoal tem-se um percentual maior dentre os aprovados em contraposição ao percentual dentre os candidatos. Tem-se, então, indícios de que uma maior renda pessoal pode fazer parte do perfil do candidato que apresenta boa performance.

No item renda “não informada” não se pode ter certeza de se tratar de uma faixa homogênea. Apesar da questão não dar a opção de resposta “sem renda pessoal”, o que a princípio, poderia sugerir que essa faixa representaria as pessoas sem rendimentos, deve-se considerar comum as pessoas guardarem reservas quanto a informação de sua renda pessoal.

FIGURA 03 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO A RENDA PESSOAL MENSAL– CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS



3.1.1.4. RENDA FAMILIAR MENSAL

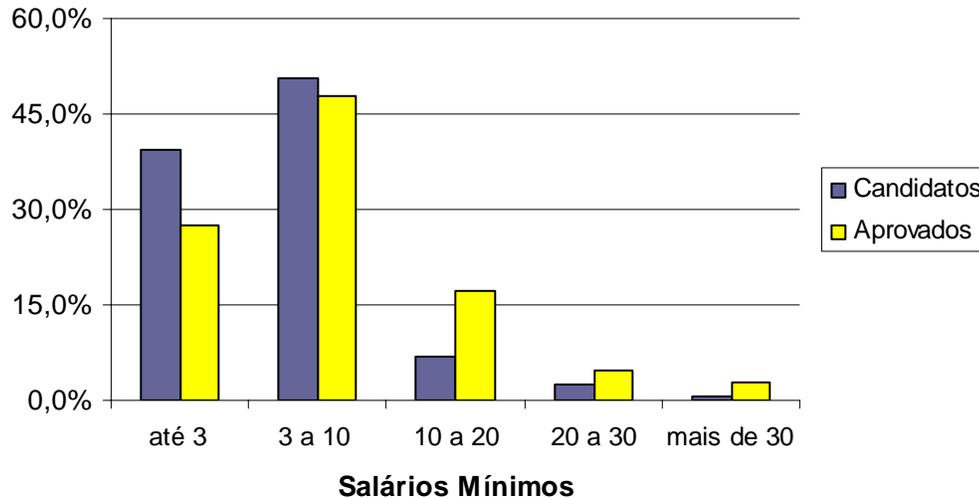
Na Figura 04 são informados o percentual de participação no total de candidatos para cada faixa de renda familiar, e o percentual de participação dentre os aprovados, também para cada faixa de renda familiar.

É de se destacar que a participação de pessoas com renda baixa é bem maior dentre o número total de candidatos que entre os aprovados.

Essa situação vai gradativamente se invertendo à medida que a faixa de renda familiar observada é maior. Tem-se, por exemplo, que enquanto o número de candidatos inscritos com renda familiar até 1 salário mínimo representava quase 40% do total, a participação de candidatos com essa faixa de renda familiar, dentre os aprovado, foi de apenas 27,5%. Ou seja, a participação relativa dos que têm menor renda familiar caiu em cerca de 70%, se comparada a participação dentre aprovados com a dentre o total de participantes.

O contrário ocorre com os que estão nas faixas de maior renda familiar, o que aponta para indícios de que quanto maior a renda familiar maior a possibilidade de aprovação do candidato.

FIGURA 04 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO A RENDA FAMILIAR MENSAL– CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS



3.1.1.5. SITUAÇÃO NO MERCADO DE TRABALHO

Praticamente metade dos candidatos (44%) encontravam-se desempregados na época do concurso, os demais estavam distribuídos de forma quase equitativa entre as outras possibilidades apresentadas no questionário. Quanto ao desempenho, indícios apontam (ver Figura 05) que candidatos desempregados ou na condição de estagiário (que trabalham em tempo parcial) podem ter melhor performance, dado que foram as únicas categorias a apresentar maior presença relativa dentre os aprovados em contraposição a participação que tinham dentro do total de candidatos.

3.1.1.6. ESTADO CIVIL

O estado civil predominante entre os candidatos era o de solteiro (70,5%). Observa-se que a participação percentual dos solteiros dentre os aprovados é a única relativamente maior que a participação dentre os candidatos como um todo, o que pode indicar uma maior possibilidade de aprovação para os que têm esse atributo.

FIGURA 05 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO A SITUAÇÃO NO MERCADO DE TRABALHO – CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS

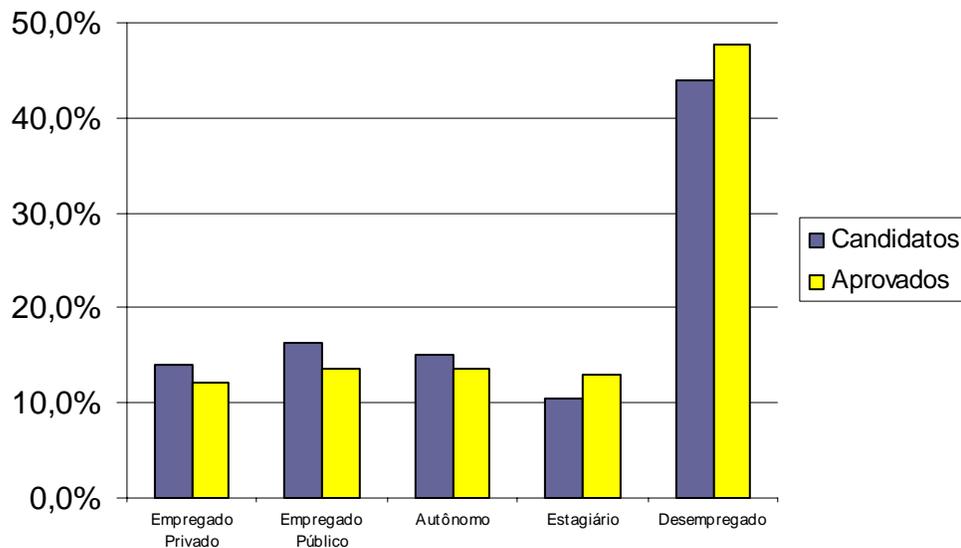
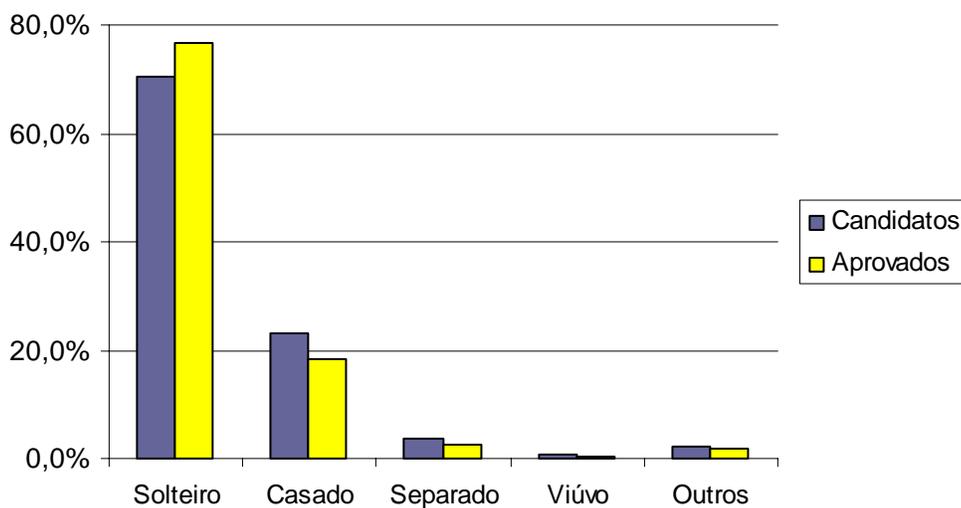


FIGURA 06 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO AO ESTADO CIVIL – CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS



3.1.1.7. FAIXA ETÁRIA

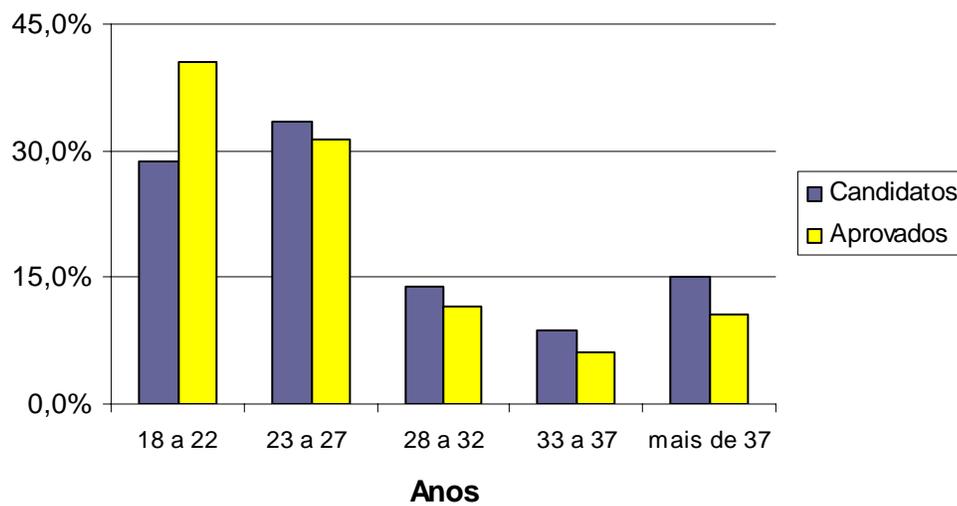
É de se destacar a maior participação de candidatos de faixa etária baixa, onde 71,9% tinham menos de 27 anos de idade.

Também observa-se uma quantidade considerável para a faixa etária mais alta, 15,1% dos candidatos tinham mais de 37 anos de idade.

São exatamente as pessoas que se encontram nos extremos das faixas etárias que apresentam, em geral, maior dificuldades de inserção no mercado de trabalho. Os mais novos por falta de experiência, e os mais velhos por geralmente terem dificuldades de requalificação que os reintegrem de forma competitiva ao mercado de trabalho.

Quanto ao desempenho, a faixa de candidatos de idade entre 18 a 22 anos é a única a apresentar participação percentual dentre os aprovados relativamente maior que a participação dentre os candidatos como um todo, o que pode apontar para um indício de maior possibilidade de aprovação para os que estão nessa faixa de idade.

FIGURA 07 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO A FAIXA ETÁRIA – CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS

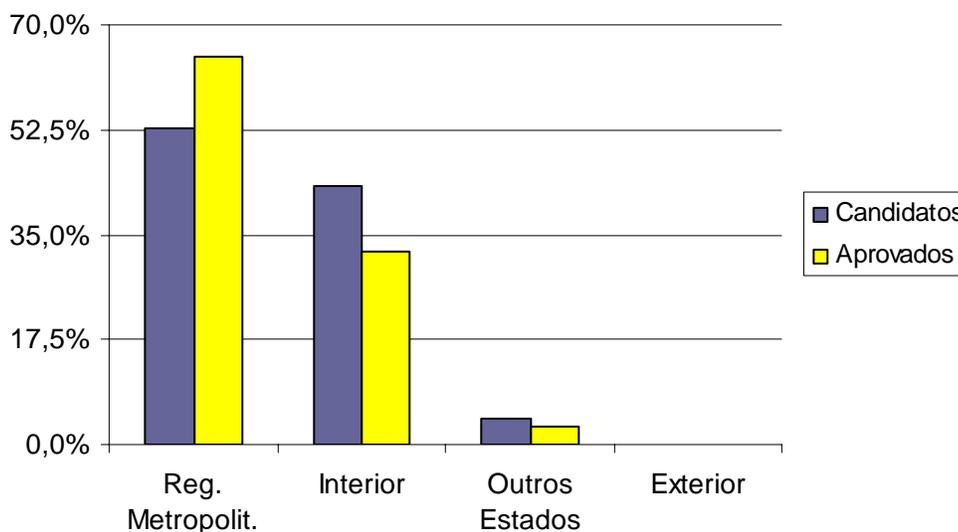


3.1.1.8. LOCAL DE ORIGEM

Os candidatos residentes em regiões metropolitanas representaram pouco mais da metade dos inscritos. Dos demais, 43,2% eram de municípios do interior e 4,2% eram candidatos de outros estados, que, por conseguinte, apresentavam-se dispostos a migrar caso obtivessem a aprovação no concurso.

Os candidatos residentes em regiões metropolitanas apresentaram sinais de melhor performance, dado ser a única categoria nesse quesito a apresentar percentual entre os aprovados maior que dentre o total de candidatos (Ver Figura 08)

FIGURA 08 – COMPARATIVO DOS PERFIS DOS CANDIDATOS QUANTO AO LUGAR DE ORIGEM – CONFRONTO ENTRE PARTICIPAÇÃO PERCENTUAL NO TOTAL DE CANDIDATOS COM A PARTICIPAÇÃO PERCENTUAL DENTRE OS APROVADOS



Nessa seção foi feita uma breve, mas reveladora exposição de informações emanadas pelos dados que serão utilizados para tentar identificar os determinantes de obter aprovação no concurso objeto de nosso estudo.

Na seção 3.2. será analisado com mais detalhes as características do exame que foi aplicado aos candidatos para se agregar elementos que definam mais claramente a relevância dos dados que se tem disponíveis.

Ainda na seção 3.2. faz-se uso de informações do Exame Nacional do Ensino Médio (ENEM), no ano de 2003, para respaldar algumas das constatações indicadas nesse trabalho. A utilização dos dados do ENEM é justificada em virtude da constatação da relevância da qualidade do ensino médio para o desempenho no exame de concurso público em questão.

No final da seção 3.2. é feita uma digressão sobre o caso em que a educação formal do candidato é superior à exigida. Essa digressão foi motivada pela ocorrência de alto índice de aprovação de candidatos com formação superior à requerida pelo concurso (ver item 3.1.1.2).

E na seção 3.3. serão retomadas todas as informações trabalhadas no presente capítulo com o intuito de se obter um diagnóstico consolidado das variáveis disponíveis para a determinação do desempenho dos candidatos no concurso público em questão.

3.2. IDENTIFICANDO DETERMINAÇÕES RELEVANTES QUANTO AO DESEMPENHO DOS CANDIDATOS NO CONCURSO

Para indicar os prováveis determinantes do desempenho dos candidatos no concurso em questão o ponto de partida óbvio é descrever o exame que foi aplicado, suas regras e conteúdo programático explorados, para daí se ter as referências iniciais desses determinantes.

3.2.1. CARACTERÍSTICAS DO EXAME DO CONCURSO DO BANCO DO NORDESTE DO BRASIL

Para a realização da seleção dos candidatos foi aplicado um exame, no intuito de aferir por meio de aplicação de provas escritas, os conhecimentos e habilidades dos candidatos.

O exame era não só classificatório, mas apresentava, também caráter eliminatório; ou seja, não bastava o candidato estar entre os que obtiveram melhores performances dentre os que participaram, era exigido também um perfil mínimo de desempenho a ser cumprido para o candidato ser considerado aprovado.

Nesse concurso foram aplicadas quatro provas objetivas: Língua Portuguesa, Matemática, Conhecimentos Gerais e Conhecimentos Bancários.

Na prova de Língua Portuguesa exigia-se um nível de conhecimento de acordo com o adquirido no ensino médio.

Na prova de Matemática, também se destaca a base a ser adquirida até o ensino médio; contudo, dando-se destaque a alguns tópicos que podem ser desenvolvidos em certos cursos de nível superior, como em alguns das ciências administrativas, por exemplo.

No programa da prova de conhecimento gerais os pontos exigidos eram:

- Tópicos relevantes e atuais de diversas áreas, tais como política, economia, sociedade, educação, tecnologia, energia, relações internacionais, desenvolvimento sustentável, segurança e ecologia e suas vinculações históricas.
- O Nordeste brasileiro: atividades econômicas, contrastes sub-regionais, o polígono das secas e as características das regiões naturais do Nordeste; o Nordeste no contexto nacional.
- O papel do Banco do Nordeste como agente impulsionador do desenvolvimento sustentável da região Nordeste.

Apesar de se tratar de um concurso para nível médio, mais uma vez, observa-se a relevância de conhecimentos que extrapolam o conteúdo programático das escolas de ensino médio. Para a prova de conhecimentos gerais candidatos que tivessem formação superior em áreas como economia, sociologia e geografia, por exemplo, teriam maiores possibilidades de dominarem os temas explorados na prova de conhecimentos gerais.

Na prova de conhecimentos bancários o conteúdo tratava de um tipo de conhecimento bem específico, que para adquiri-lo seria necessário frequentar cursos especializados em concurso público (cursinhos) e/ou se valer de material didático específico.

Conhecimentos básicos adquiridos ao longo da formação educacional até o ensino médio (domínio da língua portuguesa e de matemática, por exemplo) e certas questões de conhecimentos gerais reforçam a relevância da qualidade do ensino médio como indicador de desempenho.

Também é relevante o candidato possuir curso de nível superior. Além de, em geral, aumentar o grau de amadurecimento intelectual do candidato, em particular, observa-se uma gama de cursos que desenvolvem especificamente parte do conteúdo exigido nas provas do concurso em questão.

Frequentar cursos preparatórios específicos (cursinhos) é outro elemento importante que interfere positivamente no desempenho do candidato, principalmente no que se refere ao conteúdo das provas de conhecimentos gerais e bancários, onde nesses cursinhos se costuma tratar de forma direcionada essas esferas de conhecimento, dadas as peculiaridades das provas.

Finalmente, por motivos óbvios, contribui ao bom desempenho dos candidatos quaisquer determinantes que possibilitem um maior tempo de dedicação aos estudos.

Dada a importância destacada ao ensino médio, observar-se-á na próxima seção os resultados do Exame Nacional do Ensino Médio (ENEM) no intuito de obter mais elementos que ajudem a balizar quais fatores influenciam o bom desempenho dos egressos desse nível de ensino.

Já na seção 3.2.3., em função da relevância observada do ensino de nível superior nos resultados apresentados no concurso em questão, refletido no exposto no item 3.1.1.2, é feita um pequeno desvio da temática central desse trabalho para se melhor conhecer o fenômeno de pessoas buscando ocupações que exigem grau de escolaridade inferior ao que elas possuem.

3.2.2. INFORMAÇÕES DO EXAME NACIONAL DO ENSINO MÉDIO - ENEM

Visto que a formação básica, até a conclusão do ensino médio, é um dos principais elementos para o bom desempenho nas provas do concurso ora objeto desse estudo, conforme se discorreu sobre as características do referido exame na seção 3.2.1., vale observar o desempenho dos participantes do ENEM associado às características socioeconômicas, para

daí se consolidar melhor algumas das referências que se obteve na seção 3.1.1., a qual tratou da parte descritiva da base de dados que será utilizada no presente trabalho.

O Exame Nacional do Ensino Médio (ENEM) é aplicado, desde 1998, pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), ligado ao Ministério da Educação. O teste é aplicado anualmente aos alunos concludentes e aos egressos deste nível de ensino, com o objetivo fundamental de avaliar o desempenho do aluno ao término da escolaridade básica e com vistas a aferir o desenvolvimento de competências fundamentais.

O ENEM/2003 teve 1.882.393 inscritos e 1.322.644 participantes que efetivamente fizeram as provas. O exame era composto por duas provas, uma objetiva que abrangia o conteúdo programático básico exigido para o ensino médio e uma prova de redação.

Nos itens a seguir faz-se uma síntese dos principais resultados observados da confrontação do desempenho dos alunos no ENEM/2003 com seus respectivos perfis socioeconômicos.

A) Escola Pública *versus* Escola Privada

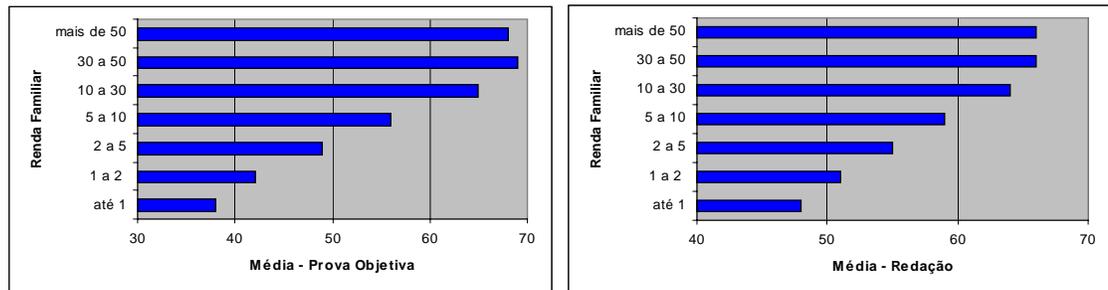
Tanto na prova objetiva como na redação evidencia-se o melhor desempenho dos alunos que cursaram escolas privadas. Comparando os alunos que fizeram todo o ensino médio em escolas públicas, com alunos que estudaram parte em escola pública, parte em escola privada, esses obtiveram média 15% superior nas provas objetivas e 7,5% superior na redação. A diferença de performance fica mais patente quando a comparação é feita entre os alunos de escolas públicas com os alunos que fizeram todo o ensino médio em escola particular. Esses obtiveram média 42% maior na prova objetiva e 21% na redação.

B) Renda Familiar

Quanto à renda familiar, é claro o aumento do desempenho no ENEM/2003 à medida que se tem uma maior renda familiar. Os dados constantes na Figura 09 mostram claramente um melhor desempenho nas notas à medida que se tem uma maior renda familiar. A influência da

renda familiar revela-se tão forte que na prova objetiva chega-se a observar uma diferença próxima de 100% em favor dos que possuem as maiores rendas, frente aos que estão nas faixas de menor renda.

FIGURA 09 – DESEMPENHO DOS PARTICIPANTES DO ENEM/2003 NAS PROVAS OBJETIVA E DE REDAÇÃO – POR FAIXA DE RENDA DA FAMÍLIA



C) Atividades Extracurriculares

Os participantes do ENEM/2003 responderam se freqüentavam atividades extracurriculares durante o ensino médio, dentre eles cursos preparatórios para vestibular (cursinho). Os participantes que freqüentavam cursinho tiveram desempenho em torno de 10% maior que os demais. Cabe observar que a possibilidade de se fazer cursinho reflete um nível de renda familiar maior ou, em certas circunstâncias, de renda pessoal maior.

D) Perspectiva quanto ao Mercado de Trabalho e Continuidade dos Estudos

Foi perguntado aos participantes qual a principal decisão a ser tomada quando da conclusão do ensino médio. Os que afirmaram que pretendiam trabalhar por conta própria ou conseguir um emprego tiveram um desempenho inferior aos que afirmaram que pretendiam continuar os estudos (uma diferença de 20% na prova objetiva e 25% na redação em prol dos que

afirmaram querer continuar os estudos). Isso pode estar refletindo a necessidade de acréscimo na renda pessoal ou familiar por parte dos que não pretendiam dar continuidade aos estudos.

3.2.3. DIGRESSÃO SOBRE O CASO DA EDUCAÇÃO FORMAL DO CANDIDATO SER SUPERIOR AO EXIGIDO NO CONCURSO

Antes de fazer considerações quanto às possíveis variáveis explicativas do desempenho em concursos públicos, vale dedicar um espaço para uma digressão acerca de pessoas que demandam por ocupações que exigem educação formal inferior a que elas possuem. Visto que o concurso exigia do candidato apenas o nível médio de ensino, essa digressão é justificável, uma vez que se observou um grande número de inscritos no concurso com educação superior ao ensino médio (51,8%), e dentre os aprovados o percentual se apresentou mais elevado ainda (84,3%). Ademais, tem-se as considerações na subseção 3.2.1. que apontam para uma possível vantagem dos candidatos com nível de educação formal superior.

Neste contexto, o fenômeno de *Overeducation* é o referencial teórico do aspecto a ser observado. A intenção é adotar uma adequada argumentação, em termos teóricos, para definir com mais clareza o que leva a falta de sincronia entre o sistema educacional e o mercado de trabalho.

Nos últimos anos se produziu uma quantidade substancial de pesquisas constatando evidências empíricas na Europa e na América do Norte do fenômeno do *Overeducation* (Green *et al.*, 1999). Contudo, sua conceituação, apesar de ser bem elementar, a princípio guarda algumas nuances que podem gerar controvérsias quanto à precisão do conceito.

Sicherman (1991) indica que os estudos referentes a *Overeducation* não têm explicado satisfatoriamente esse fenômeno. Em um discurso de orientação geral na *Applied Econometrics Association*, *Joop Hargot* afirma que “... ainda se ressentir de uma relação sólida da literatura de *Overeducation* com uma teoria formal do mercado de trabalho” (Hargot, 1997). Uma dificuldade crucial encontra-se no aspecto conceitual, onde não se tem uma abordagem única para a definição de *Overeducation*. Assim, faz-se necessário expor as nuances que perpassam esse conceito, para que se defina de forma clara o que se adotará como referencial teórico nesse trabalho.

A existência de um hiato entre a formação adquirida pelo agente ofertador de mão-de-obra no mercado de trabalho e a formação necessária ao desempenho da ocupação obtida por esse, constitui-se no fenômeno denominado *Overeducation* (sendo *Undereducation* o complementar desse conceito).

Richard Freeman (1976)³ e *Ronald Dore* (1976)⁴ foram os primeiros economistas a expressarem interesse no problema do hiper-investimento em educação superior. *Freeman*, argumentava que a considerável queda na taxa de retorno para um graduado americano, em 1970, indicava a existência de hiper-investimento em educação superior, resultando em um excesso de oferta de graduados no mercado de trabalho. Ele também argumentava que, embora esse hiper-investimento fosse possível no curto prazo, a queda da taxa de retorno levaria a redução, no longo prazo, da oferta de mão-de-obra graduada. Essa redução não se constatou. Nos Estados Unidos, durante toda a década de 1980 a oferta de graduados cresceu, assim como seus salários relativos (*Murphy e Welch*, 1989). Isso sugere uma forte e crescente demanda por esse tipo de trabalho.

Em países em desenvolvimento, como o Brasil, não poderia ser diferente a condição de necessidade de pessoas com educação superior. Na verdade essa carência tende a apresentar-se mais acentuada. A existência de “super-educados” em países como o Brasil pode estar sinalizando problemas de ineficiência do sistema educacional, no que tange sua sincronia com o sistema produtivo. Isso revelaria alocações ineficientes (inadequadas) de recursos, já tão escassos, em educação. O significado de *Overeducation* tem freqüentemente dependido das suposições feitas pelos pesquisadores (*Green at al*, 1999). As circunstâncias sob as quais imputa-se a ocorrência desse fenômeno indicam a racionalidade do significado a ele conferido.

3.2.3.1. EM QUE CIRCUNSTÂNCIA OCORRE O *OVEREDUCATION*?

É fundamental a identificação das circunstâncias que desencadeiam esse fenômeno. Dependendo da motivação que se entenda como causadora de sua ocorrência, podem-se ter distintas formas de interpretar seus efeitos. Descartada a possibilidade de excesso de mão-de-obra

³ *Apud Green, at al* (1999).

⁴ *idem*.

qualificada⁵, pode-se, de forma esquemática, atribuir a existência do *Overeducation* como consequência da:

- Ocorrência de um jogo "ruim" (*mismatch*) no mercado de trabalho - denominado de **Teoria do *Mismatch***;
- Forte complementaridade entre as várias formas de aquisição de Capital Humano (educação formal, experiência, treinamento em serviço) - **Modelos de Mobilidade da Carreira**;
- Heterogeneidade da força de trabalho, o que leva os menos hábeis a ter que incorporar educação suplementar - **Teoria do Fator Trabalho Heterogêneo**;
- Ineficiência do Sistema Educacional, no sentido de haver uma falta de sincronia desse com o mercado de trabalho ou de falta de qualidade do Sistema - **Ineficiência do Sistema Educacional**.

3.2.3.1.1. A Teoria do *Mismatch*

A discrepância entre a formação do trabalhador e a exigida pelo emprego ocupado ocorreria, dentro da ótica da Teoria do *Mismatch*, em virtude da má qualidade do *matching* no mercado de trabalho. Em decorrência de informação imperfeita ou rigidez no mercado de trabalho, os trabalhadores não achariam as ocupações mais apropriadas para as suas habilidades.

Uma informação incompleta levaria os indivíduos a não encontrarem o trabalho mais apropriado as suas habilidades. De acordo com esta teoria, os custos de procura e informações imperfeitas resultam em um jogo inicialmente "ruim" (chamado *mismatch*).

Todavia, um processo de ajuste gradual, na seqüência do jogo, levaria a obtenção das informações necessárias a se chegar finalmente a uma partida perfeita, com resultados racionais para os agentes envolvidos (*Jovanovic, 1979*)⁶.

⁵ Dado que o fenômeno de *overeducation* é observado em uma gama de países [para citar alguns: Inglaterra (Dolton e Vignoles, 1997); Bélgica (Verhaest e Omey, 2002); Irlanda (McGuinness, 2002); Itália (Di Pietro e Urwin, 2002); Espanha (Alba-Ramírez, 2002)] seria inadequado concluir a existência de um excesso de trabalhadores qualificados difundido no mundo industrializado. Mais inadequado ainda em se tratando de países em desenvolvimento, como o Brasil.

⁶ *Apud* Verhaest e Omey (2002)

Nessa abordagem *Overeducation* seria um fenômeno que retrata ineficiência de alocação do fator trabalho, contudo, o caráter temporário conferido ao fenômeno atenua conseqüências negativas por ele causadas.

3.2.3.1.2. Modelos de Mobilidade da Carreira

Nessa abordagem, *Overeducation* é interpretado como uma conseqüência de um *trade-off* entre educação e outras formas de obtenção de capital humano, como a experiência e o treinamento em serviço, por exemplo (*Sisherman, 1991*).

A Teoria da Mobilidade da Carreira, indica que os trabalhadores podem temporariamente ocupar atividades aquém de sua educação formal, na intenção dessas atividades lhes propiciarem qualificações que não são ofertadas, por natureza, pelo ensino formal. Dada a absorção dessa qualificação “extra-curricular”, essa contribuiria na recolocação desses trabalhadores no mercado de trabalho de forma a lhes propiciar acesso a um emprego de nível mais elevado, condizente com sua formação.

Assim como na Teoria do *Mismatch*, a perspectiva da Teoria da Mobilidade da Carreira aponta o fenômeno do *Overeducation* como temporário. Por outro lado, a ocorrência do fenômeno é tida como racional nessa teoria, diferente do que indica a Teoria do *Mismatch*. Na Teoria da Mobilidade da Carreira o trabalhador em situação de *Overeducation* despenderia tempo, por um curto prazo, em uma ocupação que exige uma formação inferior, com o intuito de obter qualificações inatas à prática do trabalho, com o fim de habilitar-se plenamente a ocupar uma vaga em outra atividade para o qual ele tenha se preparado através da educação formal.

3.2.3.1.3. A Teoria do Fator Trabalho Heterogêneo

Nesta teoria tem-se que indivíduos com as mesmas qualificações podem ter níveis gerais de habilidade diferentes e, conseqüentemente, diferentes níveis de produtividade. Os que possuem menos habilidades, cientes dessa defasagem, e no intuito de ganhar melhores condições para se inserir no mercado de trabalho, acabam por realizar hiper-investimentos em educação.

Com uma educação formal adicional esses buscam serem absorvidos por atividades as quais eles foram capacitados antes de lançar mão do recurso do hiper-investimento em educação formal.

Geralmente a ocorrência desse fenômeno se dá em períodos marcados por quebras estruturais nos níveis de emprego. Em circunstâncias normais, ou seja, sem quebras estruturais, a condição de possuir habilidades abaixo da média, via de regra, desmotiva esses trabalhadores quanto à iniciativa de obter capacitação adicional, dados os baixos retornos esperados.

Contudo, nos períodos em que quebras estruturais elevam os níveis da taxa de desemprego, a necessidade de se colocar no mercado de trabalho, mesmo que com retornos aquém do esperado torna-se menos indesejado que a possibilidade do desemprego.

Assim como para a Teoria do *Mismatch*, a existência simultânea dos fenômenos de *Overeducation* e *Undereducation* testemunha a favor tanto da argumentação da heterogeneidade do fator trabalho quanto para existência de um jogo ruim. Já quanto à duração, a heterogeneidade da força de trabalho aponta para uma situação de *Overeducation* permanente, visto que o cerne da questão gira em torno de características natas dos ofertadores de força de trabalho.

3.2.3.1.4. Ineficiência do Sistema Educacional

A ocorrência de *Overeducation* pode se dar em função de desequilíbrios entre o mercado de trabalho e o sistema educacional, ou da baixa qualidade de elementos desse último. Dentro dessa abordagem, *Rumberger* (1981) afirma que os trabalhadores que apresentam formação superior às atividades por eles exercidas estão sendo sub-utilizados.

A falta de sincronia do sistema educacional com o mercado de trabalho se dá quando as alternativas de educação formal não suprem de forma eficiente às necessidades de capacitação do fator trabalho apontadas pelo mercado. Daí pode-se ter pessoas que precisem, por exemplo, de um curso pós-ensino médio, que não tenha o *status* de nível superior, que os capacitem a ocupações que se apresentem como demandantes de mão-de-obra. A não existência desse curso pós-médio faz com que alguns busquem um curso de nível superior para ter a possibilidade de ocupar uma vaga no mercado de trabalho. Esses estarão obtendo uma formação que poderá ser sub-aproveitada.

Outra circunstância seria a baixa qualidade de parte do sistema educacional. Se por exemplo, uma parte do ensino médio for de baixa qualidade, os oriundos desses cursos que apresentarem dificuldades de se inserir no mercado de trabalho podem buscar curso de ensino superior para, com isso, sinalizar para o mercado sua possibilidade produtiva.

Evidências apontam para uma relativa improdutividade dos trabalhadores que se encontram na situação de *Overeducation*. Segundo *Groot* (1996) os indivíduos em situação de *Overeducation* ganham menos do que aqueles corretamente alocados. Além disso, esses trabalhadores, em geral, passam longos períodos nessa situação, o que sugere que esses trabalhadores recebem menos ofertas de trabalho. A ocorrência de *Overeducation* ocasionado pela ineficiência do sistema educacional, confere a esse fenômeno uma condição de permanente.

Cada argumentação dá conta de formas diferentes de desencadeamento de *Overeducation*. Isso não significa que a procedência de uma teoria anule, necessariamente, a validade das outras. Dependendo de questões conjunturais (como, o nível de atividade da economia) e estruturais (como a qualidade da estrutura do sistema de educacional) uma das argumentações teriam maior possibilidade de explicar com propriedade a ocorrência do fenômeno.

Em situações em que a economia apresenta-se em crescimento e o sistema de educação responde a contento as necessidades de capacitação demandadas pelo setor produtivo, as Teorias do *Mismatch* e da Mobilidade da Carreira seriam as explicações mais adequadas. Em períodos em que ocorre, por exemplo, difusão de um novo paradigma tecnológico, como o que aconteceu no Brasil no início da década de 1990 (Almeida, 1996), a teoria do trabalho heterogêneo ganharia uma maior possibilidade de explicar com melhor propriedade a ocorrências do *Overeducation*. E numa circunstância de haver constatações de baixa integração entre o sistema e produtivo a estruturação do sistema educacional, ou de baixa qualidade desse, se teria a Teoria da Ineficiência do Sistema Educacional como pertinente para explicar o fenômeno.

Apesar das condições estruturais e conjunturais poderem determinar qual teoria seria a mais pertinente, nada impede que o fenômeno de *Overeducation* seja desencadeado simultaneamente por razões explicadas por duas ou mais dessas teorias. Pode-se ter, por exemplo, uma quebra estrutural, oriunda da mudança de paradigma tecnológico, que justifique a explicação do trabalho heterogêneo e, ao mesmo tempo, ocorrer um *matching* ruim, devido a informação imperfeita no mercado de trabalho, e simultaneamente se observar a existência de um sistema

educacional inadequado. Além disso, nenhuma dessas características anulam a validade da complementaridade das diversas formas de se adquirir capital humano.

Das motivações que explicam a geração de *Overeducation*, as que o indicam como um fenômeno de longa duração, apresentam-se como as que mais podem retratar efeitos indesejados. Uma alocação indevida do fator trabalho, caracterizada como permanente, indica um sub-aproveitamento crônico de mão-de-obra qualificada. Duas teorias apontam *Overeducation* como sendo de longa duração: a do fator trabalho heterogêneo e a da ineficiência do sistema educacional.

De certo modo, se pode considerar que a existência de um sistema educacional inadequado pode agravar o grau de desperdício (formação obtida e não utilizada) dos que tem habilidades abaixo da média e buscam um ensino adicional para sinalizar para o mercado suas possibilidades produtivas. Poucas opções de educação formal (baixa segmentação e especialização dos cursos ofertados) podem levar as pessoas de baixa produtividade a terem que buscar uma educação adicional bem acima do que elas realmente necessitariam para poder se inserir melhor no mercado de trabalho.

Então, a ineficiência do sistema educacional, além dos problemas que essa abordagem aponta, ainda apresenta efeitos indiretos sobre os efeitos negativos apontados pela outra abordagem que indica a *Overeducation* como um fenômeno de longa duração.

Uma situação em que o sistema educacional do país não esteja privilegiando uma adequada formação produtiva dos indivíduos, ou esteja dando-lhes educação formal de baixa qualidade indica a má utilização de recursos públicos e privados envolvidos no financiamento da educação.

3.3. DEFINIÇÃO DAS VARIÁVEIS EXPLICATIVAS

Nessa seção são retomadas as informações tratadas no transcórre desse capítulo de forma a definir quais e como se conformaram as variáveis a serem utilizadas para explicar o desempenho dos candidatos no concurso do BNB. Para a síntese da análise de cada quesito tomou-se o seguinte procedimento:

1. É feito o resgate das considerações discutidas no transcorrer da seção 3.2 para se apontar à relevância da variável como fator que influencia a performance do candidato;
2. Observou-se para cada possível resposta a contraposição da participação relativa dentre o total de candidatos com a participação relativa dentre os aprovados. Para melhor visualização dessa informação se fez a razão entre o percentual observado da resposta dentre os aprovados e o percentual observado entre os participantes, o que se denominou de Índice Preliminar de Desempenho (IPD). Caso o IPD seja maior que um (1), indicará que a característica da resposta é um atributo que pode representar uma interferência positiva no desempenho do candidato. No caso do IPD se apresentar inferior a um (1), o atributo a ele relacionado representaram, a priori, uma influência negativa no desempenho;
3. Finalmente foi definida a montagem de variáveis *dummies* a serem utilizadas como regressores, dada a observação do IPD, onde se indica, inclusive, o sinal esperado do um efeito marginal da variável sobre a chance de aprovação no concurso.

3.3.1. RELEVÂNCIA E DEFINIÇÃO DO FORMATO DE CADA VARIÁVEL EXPLICATIVA

3.3.1.1. TIPO DE ESCOLA DO NÍVEL MÉDIO

A importância destacada na subseção 3.2.1. da qualidade do ensino médio no desempenho do candidato no concurso do BNB, justifica a utilização da variável Tipo de Escola em que o candidato cursou o ensino médio.

Observa-se na Tabela 02 que a condição de fazer todo o ensino médio em escola privada é o único perfil que apresenta IPD maior que um. Sendo assim, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato tiver feito todo o ensino médio em escola particular, e zero (0) para os demais casos.

TABELA 02 – ÍNDICE PRELIMINAR DE DESEMPENHO - TIPO DE ESCOLA DE NÍVEL MÉDIO

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
Toda Pública	48,4%	32,4%	0,7
Maior Parte Pública	8,3%	7,3%	0,9
Permeio	5,1%	1,8%	0,3
Maior Parte Privada	6,6%	6,1%	0,9
Toda Privada	31,6%	52,4%	1,7
	100,0%	100,0%	-

Os dados sobre o desempenho dos participantes do ENEM/2003 (ver seção 3.2.1.1.), exame que busca, dentre outras coisas identificar os determinantes da qualidade do ensino de nível médio, reitera a melhor performance dos alunos de escolas privadas.

É patente o melhor nível do ensino médio privado no Brasil, em comparação ao público, isso é reiterado pelos dados apresentados pelo ENEM/2003. Espera-se, então que a condição de egresso do ensino médio privado (*dummy* com valor unitário) apresente um efeito marginal positivo sobre a chance de aprovação no concurso;

3.3.1.2. GRAU DE INSTRUÇÃO

Na subseção 3.2.1. foram destacados vários aspectos ligados às provas do concurso que indicam a relevância do maior grau de instrução quanto ao desempenho dos candidatos.

Na prova de matemática, por exemplo, é dado destaque a alguns tópicos que podem ser desenvolvidos em certos cursos de nível superior, como em alguns das ciências administrativas, por exemplo.

Na prova de conhecimentos gerais são exigidos elementos que extrapolam o conteúdo programático das escolas de ensino médio, onde candidatos com formação superior em áreas

como economia, sociologia e geografia, por exemplo, apresentam maiores possibilidades de dominarem os temas ali tratados.

Ademais, freqüentar um curso de nível superior, em geral, contribui para o aumento do grau de amadurecimento intelectual do candidato, o que contribui na performance, tendo em vista que as provas de exames buscam avaliar habilidades mentais que abrangem compreensão, aplicação, análise, síntese e avaliação, valorizando a capacidade de raciocínio.

Observa-se na Tabela 03 que os candidatos que estão cursando, ou que já concluíram curso superior, apresentam IPD maior que um, e o perfil de se ter ensino médio apresentou IPD bem abaixo de um. Dada a exegese desenvolvida na subseção 3.2.3., acerca de candidatos que possuem nível superior ao exigido pela vaga a qual se pleiteia, serão geradas duas variáveis para representar o grau de instrução.

- Dados os valores do IPD, e seguindo o procedimento adotado, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato tiver apenas o ensino médio, e zero (0) para os demais casos; ou seja, superior incompleto, superior completo e pós-graduação.
- Os candidatos com nível superior incompleto, para todos os efeitos, não têm formalmente grau de instrução superior ao nível médio, mesmo que estejam frequentando um curso de nível superior, o que já dá um certo grau de conhecimento adicional ao adquirido no ensino médio. A condição da carga de informação adicional adquirida pelos universitários já é considerada na variável *dummy* na qual se equiparam todas as alternativas de instrução possíveis nesse concurso (com valor zero) que não sejam simplesmente o ensino médio (que assume valor um). Isolando-se os que já possuem o nível superior dos demais se pretende captar a relevância dos que estão legitimamente na condição de “super-educados”. Quando se fala legitimamente é considerando o candidato que já possui nível superior e está demandando um emprego de nível médio. É relevante, então, observar, isoladamente, o desempenho dos “super-educados”. Sendo assim, a variável explicativa em questão será uma *dummy* que assume valor zero (0) quando o candidato possui ensino médio ou

superior incompleto, e um (1) para os que têm nível superior completo ou sejam pós-graduados, ou seja, os “super-educados”.

TABELA 03 – ÍNDICE PRELIMINAR DE DESEMPENHO - GRAU DE INSTRUÇÃO

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
Médio	48,2%	15,7%	0,3
Superior Incompleto	34,3%	54,2%	1,6
Superior Completo	15,2%	25,9%	1,7
Pós-Graduação	2,4%	4,2%	1,8
	100,0%	100,0%	-

Para a primeira variável de grau de instrução espera-se que a condição do candidato ter apenas o ensino médio acarrete um efeito marginal negativo sobre a chance de aprovação no concurso. Na outra variável, espera-se que a condição de “super-educado”, ou seja, ter nível superior completo ou pós graduação, acarrete um efeito marginal positivo sobre a chance de aprovação no concurso.

3.3.1.3. RENDA PESSOAL MENSAL

A existência de uma prova de conhecimentos bancários no exame de seleção do BNB (subseção 3.2.1.) indica a necessidade de freqüentar cursos especializados em concurso público e/ou aquisição de material didático específico, visto que o conteúdo programático dessa prova não faz parte da grade curricular dos cursos de nível médio, ou mesmo de nível superior. Entende-se, portanto, que a renda pessoal é determinante para se ter acesso a esses cursos e materiais didáticos de qualidade.

Os dados sobre o desempenho dos participantes do ENEM/2003, reiteram essa argumentação quando apontam uma melhor performance de alunos que freqüentam atividades extracurriculares, dentre as quais destaca-se os cursos preparatórios para vestibular - cursinhos. Ora, sendo os cursinhos pré-vestibulares relevantes quanto ao desempenho dos alunos, mesmo

sendo apenas um reforço de matérias já conhecidas, imagina-se que os cursos preparatórios para temas que, em geral, são desconhecidos, como conhecimentos bancários, por exemplo, tenham uma maior relevância no impacto sobre o desempenho do candidato. Então, a influência positiva no desempenho do candidato no concurso do BNB para os que têm acesso a cursos preparatórios, justifica a utilização da variável Renda Pessoal. Observa-se na Tabela 04 que quanto maior a faixa de renda pessoal maior o IPD, sendo que esse índice apresenta valor menor que 1 para as faixas de renda até dois salários mínimos. Sendo assim, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato tiver renda pessoal até 2 salários mínimos, e zero (0) para os que informaram ter renda superior a 2 salários mínimos.

TABELA 04 – ÍNDICE PRELIMINAR DE DESEMPENHO - RENDA PESSOAL MENSAL

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
até 1 SM	41,2%	32,8%	0,8
1 a 2 SM	24,3%	21,8%	0,9
2 a 4 sM	16,5%	16,2%	1,0
4 a 6 SM	3,1%	4,7%	1,5
mais de 6 SM	1,2%	3,1%	2,6
Não Informou	13,7%	21,4%	1,6
	100,0%	100,0%	-

Pelo exposto, espera-se que a renda pessoal mensal menor que 2 salários mínimos (*dummy* com valor unitário) apresente um efeito marginal negativo sobre a chance de aprovação no concurso.

3.3.1.4. Renda Familiar Mensal

Uma família possuir renda mensal alta sinaliza sua possível influência positiva no desempenho do candidato.

A renda familiar elevada representa, inicialmente melhores condições materiais para o processo de aprendizagem, dadas as possibilidades de se pagar por serviços de escolas melhores; por atividades extracurriculares na época dos estudos básicos; de possuir computador; de acesso a canais de televisão por assinatura; de poder adquirir livros, jornais e revistas. Essas condições criam possibilidades que enriquecem não só a formação dada no nível médio como quanto a base sobre conhecimentos gerais, além do melhor desenvolvimento das habilidades cognitivas.

Outro aspecto relevante é que famílias com renda mais alta, em geral, é reflexo de um nível de escolaridade mais alta dos pais. Os pais terem um nível de escolaridade alto redundam em melhores possibilidades de orientação adequada quanto ao encaminhamento dessa educação e nível de informações passados à criança, até a vida profissional dos filhos.

Observa-se na Tabela 05 que nas faixas de renda familiar mais altas o IPD é maior, sendo que esse índice apresenta valor maior que 1 para as faixas de renda acima de 10 salários mínimos. Sendo assim, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato tiver renda familiar acima de 10 salários mínimos, e zero (0) para os que informaram ter renda inferior a 10 salários mínimos.

TABELA 05 – ÍNDICE PRELIMINAR DE DESEMPENHO - RENDA FAMILIAR MENSAL

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
Até 3 SM	39,3%	27,5%	0,7
3 a 10 SM	50,7%	47,7%	0,9
10 a 20 SM	7,0%	17,3%	2,5
20 a 30 SM	2,5%	4,7%	1,9
Mais de 30 SM	0,5%	2,8%	5,5
	100,0%	100,0%	-

Os dados sobre o desempenho dos participantes do ENEM/2003, na seção 3.2.2., reiteram a melhor performance dos alunos que têm renda familiar elevada.

Espera-se que os candidatos que possuem maior renda familiar apresente um efeito marginal positivo sobre a chance de aprovação no concurso dadas as melhores condições materiais e de orientação de se obter uma adequada preparação no transcorrer do ensino básico.

3.3.1.5. SITUAÇÃO NO MERCADO DE TRABALHO

Entende-se que a disponibilidade de tempo para se dedicar aos estudos para obter aprovação no concurso é um elemento importante. Espera-se, então, que pessoas desempregadas ou na condição de estagiário (que trabalha em tempo parcial) tenham nesse tempo disponível um diferencial positivo quanto a possibilidade de ser aprovado.

Observa-se na Tabela 06 que quanto ao mercado de trabalho, estar desempregado ou estagiando são as únicas condições que apresentam IPD maior que um. Sendo assim, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato for desempregado ou estagiário, e zero (0) para os demais casos.

TABELA 06 – ÍNDICE PRELIMINAR DE DESEMPENHO - SITUAÇÃO NO MERCADO DE TRABALHO

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
Empregado Privado	14,1%	12,1%	0,9
Empregado Público	16,4%	13,6%	0,8
Autônomo	15,0%	13,6%	0,9
Estagiário	10,5%	12,9%	1,2
Desempregado	44,0%	47,7%	1,1
	100,0%	100,0%	-

Espera-se que a condição de desempregado ou estagiário apresente um efeito marginal positivo sobre a chance de aprovação no concurso, dada a maior disponibilidade de tempo para preparação, frente aos que estão trabalhando em tempo integral.

3.3.1.6. ESTADO CIVIL

É de se esperar que pessoas solteiras tenham mais tempo disponível para se dedicar aos estudos para o concurso, bem como tenham menos responsabilidades, que por vezes pode tirar o candidato do foco de se dedicar a preparação adequada para o concurso. Observa-se na Tabela 07 que o atributo ser solteiro é o único perfil que apresenta IPD maior que um. Sendo assim, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato for solteiro, e zero (0) para os demais casos.

TABELA 07 – ÍNDICE PRELIMINAR DE DESEMPENHO - ESTADO CIVIL

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
Solteiro	70,5%	76,9%	1,1
Casado	23,0%	18,4%	0,8
Separado	3,6%	2,4%	0,7
Viúvo	0,8%	0,3%	0,4
Outros	2,1%	1,9%	0,9
	100,0%	100,0%	-

Espera-se que ter o estado civil de solteiro apresente um efeito marginal positivo sobre a chance de aprovação no concurso.

3.3.1.7. FAIXA ETÁRIA

É de se esperar que pessoas em faixas de idade mais baixas tenham um melhor ritmo de estudo, visto estarem ainda sob o ritmo de estudo para conclusão do ensino médio e/ou para vestibular, e para os que avançam nos estudos, é nessa faixa de idade (18 a 22 anos) que normalmente se está na universidade, ou se é recém formado. O vigor da idade pode ser também

apontado como uma condição que pode propiciar um maior esforço sem que o candidato se ressinta do desgaste físico e mental.

Observa-se na Tabela 08 que somente na faixa de idade de 18 a 22 anos de idade o IPD apresenta-se maior que um. Sendo assim, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato estiver na faixa de idade de 18 a 22 , e zero (0) para as faixas superiores.

TABELA 08 – ÍNDICE PRELIMINAR DE DESEMPENHO - FAIXA ETÁRIA

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
18 a 22 anos	28,8%	40,5%	1,4
23 a 27 anos	33,4%	31,4%	0,9
28 a 32 anos	14,0%	11,5%	0,8
33 a 37 anos	8,7%	6,1%	0,7
mais de 37 anos	15,1%	10,5%	0,7
	100,0%	100,0%	-

Espera-se que estar na faixa de idade mais baixa (entre 18 e 22 anos) apresente um efeito marginal positivo sobre a chance de aprovação no concurso.

3.3.1.8. LOCAL DE ORIGEM

Centros mais urbanizados apresentam melhor estrutura de oferta de serviços, onde se pode incluir os serviços educacionais, como escolas, cursos especializados (cursinhos, cursos de línguas) além de outras condições como a existência de uma vasta rede de livrarias com uma oferta de uma gama de materiais didáticos que podem não ser encontrados em regiões menos centrais. Além disso, os grandes centros urbanos concentram cursos de nível superior que ofertam mão-de-obra de qualidade para os serviços educacionais ali existentes.

Essas condições propiciam aos residentes de regiões metropolitanas melhores serviços nas escolas de ensino de nível médio, bem como na disponibilidade de cursinhos de qualidade. Isso traz rebatimentos positivos tanto para o desempenho nas provas que têm conteúdos relacionados com o ensino médio como nas provas de conhecimentos específicos.

Observa-se na Tabela 09 que a condição de ser oriundo de regiões metropolitanas é a única que apresenta IPD maior que um. Sendo assim, se utilizará como uma das variáveis explicativas uma *dummy* que assume valor um (1) quando o candidato for de região metropolitana, e zero (0) para os demais casos.

TABELA 09 – ÍNDICE PRELIMINAR DE DESEMPENHO - LOCAL DE ORIGEM

PERFIL	CANDIDATOS (A)	APROVADOS (B)	ÍNDICE PRELIMINAR DE DESEMPENHO IPD =(B)/(A)
Região Metropolitana	52,7%	64,8%	1,2
Interior	43,1%	32,3%	0,7
Outros Estados	4,2%	2,9%	0,7
Exterior	0,0%	0,0%	-
	100,0%	100,0%	-

Espera-se que a condição de ser oriundo de regiões metropolitanas apresente um efeito marginal positivo sobre a chance de aprovação no concurso.

3.3.2. SÍNTESE DA DEFINIÇÃO DAS VARIÁVEIS EXPLICATIVAS

Na Tabela 10 é exposto o conjunto de variáveis *dummies* que comporá o vetor de variáveis independentes, com suas especificações e o sinal esperado de cada uma quanto ao efeito marginal sobre o desempenho no concurso:

TABELA 10 – ESPECIFICAÇÃO DAS VARIÁVEIS EXPLICATIVAS E SINAIS ESPERADOS

VARIÁVEIS		ESPECIFICAÇÃO		SINAL ESPERADO
		1	0	
Grau de Escolaridade	EMED	Ensino Médio	Outros	-
	ESUPO	Superior Completo ou com Pós-graduação	Outros	+
Estado Civil	SOLT	Solteiro	Outros	+
Renda Mensal da Família	YFALT	Maior que 10 Salários Mínimos (SM)	Menor ou igual a 10 SM.	+
Renda Pessoal Mensal	YPBAI	Menor ou igual a 2 SM	Maior que 2 SM.	-
Escola em que Courseou o Ensino Médio	PRIV	Todo em Escola Privada	Outros	+
Tipo de Curso do Ensino Médio	TIPC	Ensino comum	Outros	+
Situação no Mercado de Trabalho	DEST	Estagiário ou Desempregado	Outros	+
Faixa Etária	IDBAI	De 18 a 22 anos	Acima de 22 anos	+
Local de Origem	RMETR	Oriundos de Regiões Metropolitanas	Outras	+

4.0 METODOLOGIA

A variável que ocupa a posição central neste estudo é o resultado do candidato no concurso. Essa variável só pode assumir dois resultados; ou o candidato é **aprovado** ou é **reprovado**. Tem-se, então, uma *dummy* que indica o resultado do candidato no concurso: de maneira que:

Se $Y = 1$ o candidato foi **aprovado** no concurso;

$Y = 0$ o candidato foi **reprovado**.

Sendo então a variável dependente do tipo que dá resposta sim ou não; ou seja, de natureza dicotômica, cabe determinar o modelo econométrico mais adequado para tratamento desse tipo de variável resposta.

4.1. ESPECIFICAÇÃO DO MODELO ECONOMÉTRICO

4.1.1. O MODELO DE REGRESSÃO

Modelos para tratar escolhas binárias caem na categoria de variáveis latentes, os quais podem ser representados da maneira abaixo:

$$Y^* = \beta'X + \varepsilon \quad (1)$$

Onde Y^* representa uma variável não observável e X um conjunto de fatores explicativos de Y^* . No entanto quando Y^* ultrapassa um certo limite ela assume o valor 1, que representa um sucesso. Já quando Y^* fica aquém do limite estabelecido acima ele passa a assumir o valor zero.

Portanto, o modelo inicial pode ser convertido para:

$$Y = \beta'X + \varepsilon, \text{ para } 0 < \beta'X + \varepsilon \quad (2)$$

$$Y = \beta'X + \varepsilon, \text{ para } \beta'X + \varepsilon > 0$$

onde $Y = 1$ ou $Y = 0$

Modelos de variáveis latentes podem então ser convertidos em modelo probabilísticos do tipo:

$$P(Y=1/X) = F(-\beta'X < \varepsilon) \text{ e } P(Y=0/X) = F(-\beta'X \geq \varepsilon) \quad (3)$$

Onde F é a função de probabilidade cumulativa de ε e, por facilidade, assume-se que o limite crítico de Y^* é zero.

A equação de regressão que realmente interessa é:

$$E(Y/X) = 1 \cdot P(Y=1/X) + 0 \cdot P(Y=0/X)$$

$$E(Y/X) = F(-\beta'X < \varepsilon) = F(\varepsilon > -\beta X)$$

$$E(Y/X) = F(-\beta'X) = 1 - F(-\beta'X) = F(\beta'X) \quad (4)$$

O conjunto de parâmetros β reflete o impacto das mudanças de x sobre a probabilidade. Por exemplo, entre os fatores que podem interferir neste problema específico está o efeito marginal do nível de escolaridade sobre a probabilidade de ser aprovado no concurso. O

problema neste ponto é elaborar um modelo adequado para o lado direito da equação. Uma das possibilidades é utilizar um modelo linear.

$$F(x,\beta) = \beta \cdot x \quad (5)$$

Desde que $E[y | x] = \beta \cdot x$, nós podemos construir o modelo de regressão

$$\begin{aligned} y &= E[y | x] + (y - E[y | x]) \\ y &= \beta \cdot x + \varepsilon \end{aligned} \quad (6)$$

Contudo, o modelo de probabilidade linear tem um grande número de imperfeições. Uma das complicações é que ε é heterocedástico, dado que depende de β . Desde que $\beta \cdot x + \varepsilon$ deva ser igual a 0 ou 1, então, ε é igual a $-\beta \cdot x$ ou $1 - \beta \cdot x$, com probabilidades $1-F$ e F , respectivamente. Portanto, pode-se facilmente mostrar que:

$$Var[\varepsilon / x] = \beta \cdot x(1 - \beta \cdot x) \quad (7)$$

É claro, que na ausência de outros problemas, poder-se-ia solucionar a questão utilizando-se um estimador de FGLS (Estimador de Mínimos Quadrados Factíveis). Para o presente propósito, contudo, uma falha mais séria é que sem uma intervenção *ad hoc* nos distúrbios, não se pode estar seguro de que as predições destes modelos irão realmente parecer probabilidades. Não se pode restringir $\beta \cdot x$ ao intervalo $[0-1]$, isso acarretaria probabilidades absurdas e variâncias negativas. Em vista disso, e dada a disponibilidade de *softwares* para estimação com maiores possibilidades de processamento, o modelo linear tem se tornado cada vez menos usado, exceto nos casos em que ele é utilizado para comparação com alguns outros modelos mais apropriados. Requer-se, então, um modelo que possa produzir predições consistentes com a base teórica; isto é,

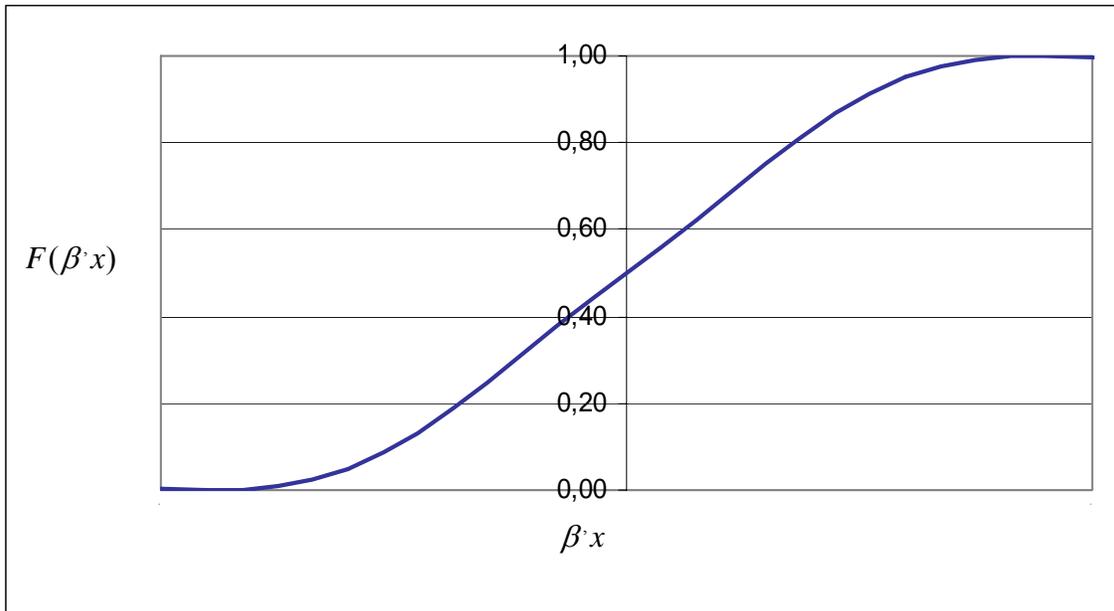
$$\text{Prob (evento } j \text{ ocorrer)} = \text{Prob}(Y=j) = F(\text{efeitos relevantes: parâmetros}) \quad (8)$$

Para um dado vetor de regressores, deve-se esperar, então, que:

$$\begin{aligned} \lim_{\beta \cdot x \rightarrow \infty} \text{Prob}(Y = 1) &= 1 \\ \lim_{\beta \cdot x \rightarrow -\infty} \text{Prob}(Y = 1) &= 0 \end{aligned} \quad (9)$$

Conforme se pode observar na Figura 10, em princípio, qualquer distribuição de probabilidade contínua definida nos reais seria adequada.

FIGURA 10 – MODELO PARA UMA PROBABILIDADE



A distribuição normal tem sido largamente utilizada em muitas análises, o que deu origem ao modelo *Probit*:

$$\begin{aligned} \text{Prob}(y = 1) &= \int_{-\infty}^{\beta \cdot x} \phi(t) dt = \\ &= \Phi(\beta \cdot x) \end{aligned} \quad (10)$$

A função $\Phi(\cdot)$ é uma notação comumente usada para a distribuição normal padrão.

Particularmente, devido a sua conveniência matemática, a distribuição logística também tem sido usada em muitas aplicações. Ou seja,

$$\begin{aligned} \text{Prob}(Y = 1) &= \frac{e^{\beta \cdot x}}{1 + e^{\beta \cdot x}} \\ &= \Lambda(\beta \cdot x) \end{aligned} \quad (11)$$

A notação $\Lambda(\cdot)$ refere-se a função distribuição acumulada logística. Este modelo é chamado de *modelo Logit*. Apesar de outras distribuições poderem ser utilizadas, têm sido usados quase que exclusivamente os modelos *Logit* e *Probit* em aplicações econométricas (Greene, 1993).

Quanto à questão de qual dessas distribuições seria a mais adequada, a distribuição logística é bastante similar a distribuição normal, sendo aquela um pouco mais densas nas caudas (se aproximando de uma t com 7 graus de liberdade).

Portanto, para valores intermediários de $\beta \cdot x$, as duas distribuições tendem a dar probabilidades similares.

A distribuição logística tende a dar probabilidades maiores para $y=0$ quando $\beta \cdot x$ é extremamente pequeno (e probabilidades menores para $y=0$ quando $\beta \cdot x$ é muito grande) em relação a distribuição normal.

É difícil estabelecer generalidades práticas sobre esta base, pois elas requerem o conhecimento de β .

Deve-se, então, esperar diferentes predições destes dois modelos, se nossa amostra contém:

- (a) pouquíssimas respostas $y = 1$, ou pouquíssimas respostas $y = 0$;
- (b) variação muita extensa em uma variável independente, particularmente se também ocorrer.

Existem razões para favorecer uma ou outra distribuição, em alguns casos, por conveniência matemática, mas é difícil justificar a escolha no terreno teórico. Amemya (1981)⁷

⁷ Apud Greene (1993)

discute algumas situações referentes a essa problemática, mas como proposição geral, a questão não é resolvida.

Em muitas aplicações, não parece fazer muita diferença a utilização de um modelo ou outro. Mas, como em certas aplicações estas diferenças são perceptíveis, discute-se adiante uma maneira mais segura de optar entre um destes dois modelos e até mesmo de considerar uma terceira alternativas.

O modelo de probabilidade, então, é uma regressão do tipo:

$$\begin{aligned} E[y/x] &= 0[1 - F(\beta \cdot x)] + 1[F(\beta \cdot x)] \\ &= [F(\beta \cdot x)] \end{aligned} \quad (12)$$

Para qualquer distribuição usada, torna-se importante notar que os parâmetros do modelo, como em qualquer modelo de regressão não-linear, não são necessariamente os efeitos marginais que se costuma obter. Em geral,

$$\begin{aligned} \frac{\partial E[y/x]}{\partial x} &= \left\{ \frac{dF(\beta \cdot x)}{d(\beta \cdot x)} \right\} \beta \\ &= f(\beta \cdot x) \beta \end{aligned} \quad (13)$$

onde $f(\cdot)$ é a função densidade que corresponde à distribuição acumulada $F(\cdot)$.

Para o caso da distribuição normal:

$$\frac{\partial E[y/x]}{\partial x} = \phi(\beta \cdot x) \beta \quad (14)$$

onde $\phi(t)$ é a densidade da normal padrão.

Já no caso da distribuição logística,

$$\begin{aligned}\frac{\partial \Lambda[\beta'x]}{\partial(\beta'x)} &= \frac{e^{\beta'x}}{(1+e^{\beta'x})^2} \\ &= \Lambda(\beta'x)[1-\Lambda(\beta'x)]\end{aligned}\quad (15)$$

a qual é particularmente conveniente. Portanto em um modelo **Logit**,

$$\frac{\partial E[y/x]}{\partial x} = \Lambda(\beta'x)[1-\Lambda(\beta'x)]\beta \quad (16)$$

4.1.2. ESTIMAÇÃO E INFERÊNCIA EM MODELOS DE ESCOLHA BINÁRIA

Com exceção do modelo de probabilidade linear, a estimação do modelo de escolha binária é geralmente baseada no método de máxima verossimilhança, onde cada observação é tratada como uma única loteria de uma distribuição *Beunoulli* (binomial com uma loteria).

Considerando uma função de distribuição conjunta, os parâmetros estão explicitamente incluídos. Então, dada uma forma específica da função de distribuição, e dado que os parâmetros assumem determinado valor busca-se saber qual a probabilidade de se observar a ocorrência simultânea de determinado conjunto de valores em uma amostra.

Contudo, na prática não se sabe o valor dos parâmetros, que é o que se quer estimar. Nesse sentido o objetivo é posto de forma inversa. Dado que se observa um conjunto de informações de uma amostra, e que se admite conhecer a forma da função de distribuição das variáveis aleatórias, das quais esses dados são uma realização particular, pretende-se saber qual o conjunto de valores dos parâmetros dessa distribuição que teria, com maior probabilidade, gerado os dados observados. Isso define a função de verossimilhança.

O modelo, com probabilidade de sucesso $F(\hat{\beta}'x)$, e observações independentes, conduz a probabilidade conjunta, ou função de máxima verossimilhança

$$\text{Pr ob}(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{y_i=0} [1 - F(\beta'x_i)] \prod_{y_i=1} F(\beta'x_i) \quad (17)$$

Reescrevendo-se (17) de uma forma mais conveniente:

$$L = \prod_{i=1}^n [F(\beta' x_i)]^{y_i} [1 - F(\beta' x_i)]^{1-y_i} \quad (18)$$

Essa é a máxima verossimilhança para uma amostra de n observações. Aplicando o logaritmo em (18), obtém-se:

$$\ln L = \sum_{i=1}^n [y_i \ln F(\beta' x_i) + (1 - y_i) \ln(1 - F(\beta' x_i))] \quad (19)$$

Para maximização de (19), requer-se que as condições de primeira ordem sejam iguais a zero, ou seja,

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f_i}{(1 - F_i)} \right] x_i = 0 \quad (20)$$

Na equação (20), utilizou-se subscrito i para indicar que a função tem argumento $\beta' x_i$. A escolha de uma forma particular para F_i leva ao modelo empírico. A menos que se esteja usando um modelo de probabilidade linear, as equações geradas em (20) serão não-lineares, o que requererá uma solução iterativa.

Estas condições de 1ª ordem são equivalentes às de mínimos quadrados ponderados de um modelo não linear do tipo:

$$Y_i = F(X_i' \beta) + V_i \quad (21)$$

onde o peso de cada observação é definido como:

$$w_i = (F(X_i' \beta)(1 - F(X_i' \beta)))^{-1/2} \quad (22)$$

Notando-se que por construção V_i tem média zero e variância dada por:

$$E(V_i^2) = F(X_i' \beta)((1 - F(X_i' \beta))) \quad (23)$$

Logo o modelo é heterocedástico, o que requer um método de estimação ponderado para que se obtenha estimadores eficientes dos parâmetros. (Ruud,2004).

Usando o modelo *Logit*, insere-se (11) e (15) em (20) chegando, após uma pequena manipulação, às condições necessárias:

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n (y_i - \Lambda_i) x_i = 0 \quad (24)$$

Note-se que se x_i contém um termo constante, as condições de primeira ordem implicam que a média das probabilidades previstas deve ser igual à proporção de 1's na amostra. A expressão acima é análoga às equações normais de mínimos quadrados se se considerar o termo $y_i - \Lambda_i$ como um resíduo. Para a distribuição normal a log-verossimilhança é dada por:

$$\ln L = \sum_{y_i=0} \ln[1 - \Phi(\beta \cdot x_i)] + \sum_{y_i=1} \ln \Phi(\beta \cdot x_i) \quad (25)$$

As condições de primeira ordem para maximização de L são:

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta} &= \sum_{y_i=0} \frac{-\phi_i}{1 - \Phi_i} x_i + \sum_{y_i=1} \frac{\phi_i}{\Phi_i} x_i \\ &= \sum_{i=1}^n \left(\frac{q_i \phi(q_i \beta \cdot x_i)}{\Phi(q_i \beta \cdot x_i)} \right) x_i \\ &= \sum_{i=1}^n \lambda_i x_i = 0 \end{aligned} \quad (26)$$

Onde, $q_i = 2y_i - 1$

Para ambos os modelos, o método de Newton é uma maneira simples de computar as estimativas dos parâmetros. A segunda derivada para o modelo *Logit*, com base em (19), é dada por:

$$H = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_i \Lambda_i (1 - \Lambda_i) x_i x_i' \quad (27)$$

Desde que a segunda derivada não envolve a variável aleatória y_i , o método de Newton também é o adequado para modelo *Logit*. Note-se que o Hessiano é sempre negativo definido, então a log-verossimilhança é globalmente côncava.

O método de Newton irá, então, convergir para o máximo da log-verossimilhança em algumas poucas interações, a menos que os dados sejam especialmente mal condicionados.

A computação é ligeiramente mais complicada no modelo **Probit**. Uma simplificação útil é obtida pelo uso da variável $\lambda(y_i, \beta' x_i) = \lambda_i$ que é definida em (26). A segunda derivada pode ser obtida usando o resultado que para qualquer z ,

$$\frac{d\phi(z)}{dz} = -z\phi(z)$$

Então,

$$H = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \sum_{i=1}^n -\lambda_i (\lambda_i + \beta' x_i) x_i x_i' \quad (28)$$

Esta também é negativa definida para todos os valores de β . A prova é menos óbvia do que para o modelo **Logit**. Contudo, é bastante notar que a parte escalar no somatório é $1 - \text{Var}[\varepsilon / \varepsilon \langle \beta' x \rangle]$ quando $y=1$ e $1 - \text{Var}[\varepsilon / \varepsilon \langle \beta' x \rangle]$ quando $y=0$. Em ambos os casos, a variância está entre 0 e 1.

A matriz de covariância assintótica para o estimador de máxima verossimilhança pode ser estimado usando a inversa do negativo do Hessiano avaliada nas estimativas de máxima verossimilhança.

Existem também dois outros estimadores disponíveis. O estimador BHHH de *Breandt, Hall, Hall e Hausman* (1974):

$$B = \sum_i g_i^2 x_i x_i'$$

Onde $g_i = (y_i - \Lambda_i)$ para o modelo **Logit** (24) e $g_i = y_i$ para o modelo **Probit** (26). O terceiro estimador é baseado no valor esperado do Hessiano. O Hessiano para o modelo **Logit** não envolve y_i , então $H = E[H]$. Mas como λ_i é uma função de y_i em (26), isso não é verdade para o modelo **Probit**. *Amemiya* (1981) mostrou que para o modelo **Probit**,

$$E \left[\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \right]_{probit} = \sum_{i=1}^n \lambda_{0i} \lambda_{1i} x_i x_i' \quad (29)$$

Novamente, a parte escalar da expressão é sempre negativa (26). A estimativa da matriz de covariância assintótica para as estimativas de máxima verossimilhança é, portanto, o inverso negativo de qualquer matriz usada para estimar o Hessiano esperado.

As probabilidades previstas $f(\hat{\beta}, x) = \hat{F}$ e os efeitos marginais estimados $f(\hat{\beta}, x) \times \hat{\beta} = \hat{f} \hat{\beta}$ são funções não lineares das estimativas dos parâmetros. Para calcular os erros padrões, pode-se usar aproximações lineares [método delta (*Greene*, 1993, p.678)]. Para as probabilidades previstas:

$$Asy.Var[\hat{F}] = [\partial \hat{F} / \partial \hat{\beta}]' V [\partial \hat{F} / \partial \hat{\beta}]$$

onde:

$$V = Asy.Var[\hat{\beta}]$$

A matriz de covariância assintótica estimada de $\hat{\beta}$ pode ser qualquer uma das três descritas anteriormente. Sendo $z = x' \hat{\beta}$. Então o vetor de derivadas é:

$$[\partial \hat{F} / \partial \hat{\beta}] = [d \hat{F} / dz] [\partial z / \partial \hat{\beta}] = \hat{f} x$$

A combinação dos termos origina,

$$Asy.Var[\hat{F}] = \hat{f}' x' V x$$

A qual depende de um particular vetor x utilizado.

Para efeitos marginais, chamando $\hat{\gamma} = \hat{f} \hat{\beta}$, tem-se:

$$Asy.Var[\hat{\gamma}] = \begin{bmatrix} \frac{\partial \hat{\gamma}}{\partial \hat{\beta}'} \\ \frac{\partial \hat{\gamma}}{\partial \hat{\beta}} \end{bmatrix} V \begin{bmatrix} \frac{\partial \hat{\gamma}}{\partial \hat{\beta}'} \\ \frac{\partial \hat{\gamma}}{\partial \hat{\beta}} \end{bmatrix},$$

A matriz de derivadas é:

$$\hat{f} \begin{pmatrix} \frac{\partial \hat{\beta}}{\partial \hat{\beta}'} \\ \frac{\partial \hat{\beta}}{\partial \hat{\beta}} \end{pmatrix} + \hat{\beta} \begin{pmatrix} \frac{d \hat{f}}{dz} \end{pmatrix} \begin{pmatrix} \frac{\partial z}{\partial \hat{\beta}'} \\ \frac{\partial z}{\partial \hat{\beta}} \end{pmatrix} = \hat{f} I + \begin{pmatrix} \frac{d \hat{f}}{dz} \end{pmatrix} \hat{\beta} x'$$

Para o modelo **Probit**, $df/dz = -z\phi$, então:

$$Asy.Var[\hat{\gamma}] = \phi [I - (\hat{\beta}' x) \hat{\beta} x'] V [I - (\hat{\beta}' x) \hat{\beta} x'] \quad (30)$$

Para o modelo **Logit**, $\hat{f} = \hat{\Lambda}(1 - \hat{\Lambda})$, então:

$$\frac{d \hat{f}}{dz} = (1 - 2\hat{\Lambda}) \left(\frac{d \hat{\Lambda}}{dz} \right) = (1 - 2\hat{\Lambda}) \hat{\Lambda}(1 - \hat{\Lambda})$$

Reunindo os termos, obtém-se:

$$Asy.Var[\hat{\gamma}] = [\hat{\Lambda}(1 - \hat{\Lambda})]^2 [I + (1 - 2\hat{\Lambda}) \hat{\beta} x'] V [I + (1 - 2\hat{\Lambda}) x \hat{\beta}'] \quad (31)$$

Cujo valor dependerá do valor do vetor x utilizado.

Existem problemas em modelos de escolha binária semelhantes aos encontrados em regressões em geral. No entanto, estes problemas podem, em alguns casos, ter conseqüências mais graves com escolha discreta do que em modelos tradicionais.

Particularmente, existem quatro tipos básicos de erros de especificação que podem ocorrer neste tipo de modelo. O primeiro é o erro de especificação sobre a presença ou omissão de variáveis explicativas, no qual se inclui o problema da heterogeneidade não observada. O segundo tipo de erro é sobre a estrutura de variância do erro; a qual, em princípio, pode ser

heterocedástica. O terceiro tipo de erro é sobre a determinação simultânea das variáveis do modelo; ou seja, variáveis do lado direito da equação requerem equações adicionais para explicá-las. Finalmente, o quarto tipo de erro é sobre a especificação funcional do modelo; ou seja, a função de probabilidade escolhida não é correta. Discute-se com maior detalhe abaixo as maneiras de identificar e solucionar estes problemas.

O problema de heterogeneidade não observada pode ser melhor ilustrado por uma regressão de dados em painel do tipo:

$$Y_{it} = \alpha + \beta X_{it} + C_i + \varepsilon_{it} \quad (32)$$

onde C_i é uma variável não observada. Em equações de determinação de salário, por exemplo, X é um vetor de atributos do trabalhador e C_i é um efeito específico da unidade observacional, diferenciando-a das demais. C_i pode representar a habilidade individual de cada trabalhador. No entanto, se o vetor X contém variáveis de escolaridade é muito provável que C_i seja correlacionado com X , levando a estimativas inconsistentes dos parâmetros do modelo (Hsiao, 2003). Em dados em painel esta questão é subjacente à escolha entre tratamento de efeitos fixos e tratamento de efeitos aleatórios.

No modelo a ser especificado mais adiante, onde o sucesso em um concurso é modelado como função de um vetor de atributos, o qual inclui variáveis de grau de instrução, não é possível observar a habilidade específica de cada candidato. Mesmo se fosse possível observá-la seria difícil mensurá-la, embora o Quociente de Inteligência (QI) seja uma boa *proxy* de habilidade.

Em modelos de dados em painel o teste de *Hausman* é a ferramenta mais usual para discriminar se a heterogeneidade é correlacionada com o vetor X . Com dados transversais é impossível trabalhar-se com uma modelagem de efeitos fixos se as diferenças individuais são correlacionadas com as demais variáveis independentes, pois o teste de *Hausman* é baseado em contrastes de MQO e MQG.

No entanto, como heterogeneidade não observada cai na categoria mais geral de erro de especificação, existe uma estratégia para testar a sua existência ou, mais importante, testar se a sua ocorrência é danosa para a estimação. *White* (1980) alerta que o seu teste de heterocedasticidade é também um teste para erro de especificação. Pois, se a matriz:

$$M = \frac{1}{n} \sum_{i=1}^n e_i^2 X_i X_i' \quad (33)$$

não convergir em probabilidade para $\sigma^2 = (X'X)^{-1}$, tanto pode ser por razão de heterocedasticidade como pode ser pelo motivo de que variáveis omitidas do modelo, e embutidas no erro, tornem significativa a correlação com os valores de X .

Como foi discutido anteriormente, estimativas de máxima verossimilhança corrigem o problema de heterocedasticidade ocorrentes em modelos de escolha discreta. Portanto, um teste de *White*, ou seu equivalente como o do multiplicador de *Lagrange*, responderá se a heterogeneidade não observada é danosa à estimação. Na prática o teste consiste em postular uma variância do erro do tipo:

$$V(\varepsilon_i) = \exp(2Z_i'\gamma) \quad (34)$$

onde γ é uma vetor de parâmetros e Z é um subconjunto de X . O resultado do teste é a soma de quadrados explicada da regressão, abaixo,

$$\frac{Y_i - F(X_i'\beta)}{w_i} = \frac{f(X_i'\beta)}{w_i} X_i' b_1 + \frac{f(X_i'\beta)(X_i'\beta)}{w_i} Z_i' b_2 + V_i \quad (35)$$

onde em lugar de β usa-se o estimador de máxima verossimilhança e w_i é dado por

$$w_i = ((1 - F(X_i'\beta))F(X_i'\beta))^{1/2} \quad (36)$$

Mostra-se que este teste é equivalente a um teste de multiplicador de *Lagrange* com distribuição assintótica dada por uma χ^2 com número de graus de liberdade igual ao número de elementos em γ .

Um outro tipo de erro de especificação é a possibilidade de endogeneidade das variáveis do lado direito do modelo. No modelo a ser definido posteriormente renda do candidato, renda da família, nível educacional e mesmo estado civil podem ser simultaneamente determinados. A questão, no entanto, é inteiramente similar a da heterogeneidade negligenciada. Ou seja, o problema fundamental a ser observado é se os erros das equações auxiliares de determinação

destas variáveis são correlacionadas ou não com o erro do modelo originalmente proposto (Sargan, 1975). Na linguagem de séries temporais necessita-se de um teste de exogeneidade, neste caso, um teste de *River-Wuong* (Wooldridge, 2002).

Infelizmente, para este teste precisa-se de instrumentos, os quais inexistem para este problema específico. No questionário base apenas a variável tipo de lazer preferido foi excluída do modelo, e esta nem é apropriada nem suficiente para fazer as vezes de instrumento neste tipo de estimação.

Caso houvesse um número suficiente de instrumentos e fosse constatada a endogeneidade o procedimento correto seria estimar o sistema por máxima verossimilhança completa. Para exemplificar considera-se um modelo de escolha discreta do tipo:

$$Y_1 = I(Z_1\delta_1 + \alpha_1 Y_2 + u_1 > 0) \quad (37)$$

onde I é a função indicador, e Y_2 é uma variável binária endógena descrita por

$$Y_2 = I(Z\delta_2 + V_2 > 0) \quad (38)$$

$Z=[Z_1, Z_2]$ e u_1 e V_2 são correlacionadas. Neste caso a distribuição conjunta de Y_1 e Y_2 , condicionada a Z , deve ser construída e a maximização desta função é equivalente à maximização da função de verossimilhança completa em modelos de equações simultâneas (Wooldridge, 2002). O raciocínio é análogo se existem mais variáveis endógenas na equação que descreve o comportamento de Y_1 .

Na ausência de instrumentos para testar a endogeneidade e, caso esta se confirme, proceder a uma estimação simultânea, não se pode estar plenamente confiante sobre a consistência dos estimadores. Resta, no entanto, uma argumentação razoável sobre a possível ausência de correlação entre os erros de um modelo de determinação simultânea das variáveis a serem tratadas mais adiante.

Existe a possibilidade de que a renda familiar, a renda do indivíduo, o grau de instrução e até mesmo o estado civil, sejam simultaneamente determinados. Contudo, é extremamente improvável que os erros das equações que determinam estas variáveis sejam correlacionados com

o erro da equação determinante do sucesso no concurso público. A razão para isto é que cada variável citada acima é determinada em diferentes mercados, como as rendas, ou por diferentes mecanismos de formação, como grau de instrução e estado civil. Portanto, o que se presume, é que exista correlação entre estas variáveis, mas não a existência de viés de simultaneidade.

O último tipo de erro de especificação a ser discutido é o do formato da escolha funcional. As formas funcionais mais populares para tratar variáveis dependentes binárias são os modelos *Probit* e *Logit*. Embora estes modelos forneçam resultados muito similares, nem por isso deve-se negligenciar um teste sobre a escolha funcional, principalmente pelo motivo da amostra nesse problema apresentar um número de valores nulos de Y excessivamente grande com relação ao número de 1's. Discriminar entre estes dois formatos requer um teste na categoria de modelos não entrelaçados. Silva (2001) mostra que, sob condições adequadas de regularidade, hipóteses alternativas sobre forma funcional em modelos binários podem ser entrelaçada em um modelo dado por:

$$L_{\rho}(Y / X, \rho, \alpha, \beta, \gamma) = \frac{[(1 - \alpha)L_1(Y / X, \beta)^{\rho} + \alpha L_2(Y / X, \gamma)^{\rho}]^{1/\rho}}{\int [(1 - \alpha)L_1(Z / X, \beta)^{\rho} + \alpha L_2(Z / X, \gamma)^{\rho}]^{1/\rho} dZ} \quad (39)$$

onde L denota máxima verossimilhança. Para $\rho = 1$ a equação acima transforma-se em:

$$L(Y / X, \alpha, \beta, \gamma) = (1 - \alpha)L_1(Y / X, \beta) + \alpha L_2(Y / X, \gamma) \quad (40)$$

que é uma combinação linear convexa das formas funcionais alternativas e a decisão sobre a melhor hipótese recai sobre o valor de α .

Esta forma de entrelaçar é equivalente ao teste proposto por *Davidson e Mackinnon* (1993); ou seja,

$$\begin{aligned} H_1 : E(Y_i / \Omega_i) &= F_1(X_{1i} \beta_1) \\ H_2 : E(Y_i / \Omega_i) &= F_2(X_{2i} \beta_2) \end{aligned} \quad (41)$$

onde Ω_i é o conjunto de informação.

A hipótese que entrelaça as duas alternativas é dada por:

$$H_c : E(Y_i / \Omega_i) = (1 - \alpha)F_1(X_{1i}\beta_1) + \alpha F_2(X_{2i}\beta_2) \quad (42)$$

Em termos de uma regressão artificial, supondo que os dois modelos possuem o mesmo conjunto de variáveis explicativas, diferindo apenas pela forma funcional, tem-se que o modelo a ser estimado é

$$w_i^{-1/2}(Y_i - F_{1i}) = w_i^{-1/2}f_{1i}X_{ib} + \alpha(F_{2i} - F_{1i}) + V_i \quad (43)$$

E o resultado do teste novamente recai sobre o valor de α .

Este teste tanto é adequado para discriminar entre os modelos *Probit* e *Logit*, como para contrastar um desses modelos representativos de distribuições simétricas com um modelo de categoria assimétrica.

5.0 RESULTADOS E CONCLUSÕES

5.1. RESULTADOS

Inicialmente escolheu-se um modelo de escolha binária Logit, cujo método de estimação utilizado foi o de Máxima Verossimilhança.

O algoritmo de maximização utilizado precisou de 10 interações para convergir, e o valor da função log-verossimilhança no ponto máximo é apresentado no Tabela 11, bem como o valor da Estatística de Razão de verossimilhança $-2\ln(\lambda)$.

$$\lambda = \frac{L(\hat{\theta}_R)}{L(\hat{\theta})} \xrightarrow{d} \chi_{R_{gl}}^2 \quad (44)$$

Onde $L(\theta)$ é a função de verossimilhança sem restrição e $L(\theta_R)$ a função com restrição; ou seja, a função de verossimilhança de um modelo que inclui apenas o intercepto. O número de restrições (R) dá o número de graus de liberdade. A restrição seria que todos os regressores (nove) fossem iguais a zero.

A idéia por trás do teste da razão de verossimilhança é que se as restrições definidas *a priori* forem válidas, as funções log-verossimilhança com e sem restrição não devem ser diferentes, caso em que λ será igual a 1, ou $\ln(\lambda)$ igual a zero. Caso contrário, esses divergirão.

Em grandes amostras, como é o caso (232.308 observações), λ segue assintoticamente uma distribuição *qui-quadrado*. Ciente da distribuição de $\ln(\lambda)$ e dado um nível de significância arbitrário é possível verificar se a divergência existente é estatisticamente significativa.

O ponto crítico para, uma qui-quadrado com 9 graus de liberdade, e dado um nível de significância de 5%, é de **16,91**. O valor de $-2\ln(\lambda)=$ **453,30** aponta para a rejeição da hipótese nula. O *p-valor* representa a probabilidade de ocorrência de um erro do tipo I. O *p-valor* da

estatística da razão de verossimilhança é inferior a 0,05 (Tabela 11), o que indica que a hipótese nula é rejeitada ao nível de significância de 5%.

Aos resultados do modelo apresentam ainda duas medidas de ajustamento. Uma é o R^2 de *McFadden*, o qual é extremamente baixo. A outra medida é o percentual de valores corretamente preditos (ver Tabela 12), o qual fornece o percentual corretamente predito de 1's, quando a probabilidade é maior do que um certo valor crítico, e de zeros, quando a probabilidade é menor do que um certo valor crítico.

Usualmente este valor crítico é estabelecido como 0.5. No entanto, neste problema específico, como a proporção de zeros é substancialmente maior do que os valores de 1's, usou-se como valor crítico a probabilidade empírica de passar no concurso; ou seja, 0.00266. Assim a proporção de valores corretamente preditos para $Y=1$ é de 61.26% e a proporção de valores corretamente preditos para $Y=0$ é de 74.27%.

O que leva a um valor ponderado de predições corretas de 61.29%. No entanto, é bom se ter em mente que medidas de ajustamento em modelos de escolha binária não são importantes nem do ponto de vista estatístico nem do ponto de vista econômico para avaliar a significância das variáveis explicativas (*Wooldridge*, 2002). O teste da razão de máxima verossimilhança e os testes individuais de significância é tudo que se precisa para avaliar o modelo.

A razão entre o coeficiente estimado e o erro padrão (Tabela 11) representa uma estatística distribuída assintoticamente como uma normal padrão (Estatística z). Esses valores são utilizados para a realização do teste de hipótese sobre a significância estatística de cada um dos parâmetros estimados. Como se pode observar nas estatísticas de teste, somente a variável referente ao estado civil (SOLT) não se mostrou estatisticamente significativa (0,03), ao nível de significância de 5% (ponto crítico **1,96**).

Quanto ao *p-valor*, excetuando-se a variável SOLT, todas as demais covariadas apresentaram *p-valor* inferior a 0,05, o que significa dizer que a hipótese nula (coeficiente ser igual a zero) pode ser rejeitada ao nível de significância de 5%. Assim, a exceção da variável que representa o estado civil (SOLT), todas as demais são estatisticamente significantes.

TABELA 11 – RESULTADO DO MODELO *LOGIT*

COVARIADAS	COEFICIENTE ESTIMADO	ERRO PADRÃO	ESTATÍSTICA Z	P-VALOR
C	- 6,353	0,132	-47,98	0,0000
EMED	- 1,320	0,118	-11,17	0,0000
ESUPO	0,289	0,100	2,88	0,0039
SOLT	0,004	0,109	0,03	0,9696
YFALT	0,545	0,098	5,54	0,0000
YPBAI	-0,182	0,082	-2,21	0,0266
PRIV	0,250	0,088	2,84	0,0045
DEST	0,319	0,087	3,67	0,0002
IDBAI	0,325	0,108	3,00	0,0026
RMETR	0,503	0,088	5,69	0,0000
Média da Variável Dependente	0,002660	Critério de informação de Akaike		0,034995
S.D. da Variável Dependente	0,051509	Critério de Schwarz		0,035441
Erro Padrão da Regressão	0,051451	Critério de Hannan-Quinn		0,035126
Soma do quadrado dos resíduos	614,9497	R ² McFadden		0,052937
Log-verossimilhança				-4.054,8530
Log-verossimilhança restrita				-4.281,5040
Estatística de Razão de verossimilhança [ln(λ)] - 9 graus de liberdade				453,3016
p-valor de λ				0,0000
Obs. com y=0	231.690	Total de Observações		232.308
Obs. com y=1	618			

Nota: Dados gerados através do aplicativo *EViews 4.1*[®]

TABELA 12 – PODER DE PREDIÇÃO DA ESTIMAÇÃO DO MODELO *LOGIT*

	EQUAÇÃO ESTIMADA			PROBABILIDADE CONSTANTE		
	Y=0	Y=1	Total	Y=0	Y=1	Total
P(Y=1)≤C	141928	159	142087	0	0	0
P(Y=1)>C	89762	459	90221	231690	618	232308
Total	231690	618	232308	231690	618	232308
Correto	141928	459	142387	0	618	618
Correto (%)	61.26	74.27	61.29	0.00	100.00	0.27
Incorreto (%)	38.74	25.73	38.71	100.00	0.00	99.73
Acerto Total	61.26	-25.73	61.03			
Acerto Percentual	61.26	NA	61.19			

Nota 1: Resultados gerados através do aplicativo *EViews 4.1*[®]

Nota 2: Corte para sucesso (Y=1), C = 0.00266026

Os coeficientes das variáveis explicativas no modelo *Logit* não expressam os efeitos marginais sobre a probabilidade da variável dependente ser 1.

Assim, são apresentados, inicialmente, os resultados referentes aos sinais dos valores estimados. Para identificar um elemento comum que seja relevante quanto à condição de determinante de desempenho por parte das variáveis explicativas utilizadas, far-se-á observações que indiquem pontos de contato entre essas variáveis.

Com o propósito de organizar a linha de raciocínio a ser desenvolvida, condensar-se-á as considerações em três blocos de variáveis; ou seja, aquelas diretamente ligadas à educação formal; as que se referem à renda; e as que representam determinações mais gerais.

Inicialmente tratar-se-á das determinações chamadas gerais; passando em seguida para as referentes à renda; e, por fim, para as que dizem respeito à educação formal.

DETERMINANTES GERAIS

A) Local de Origem

Os resultado apontaram um efeito marginal positivo sobre a chance de aprovação no concurso para os candidatos originários de regiões metropolitanas. Pode-se afirmar que o elemento diferencial desse determinante é a maior quantidade e melhor qualidade dos serviços ligados à educação e cultura nos grandes centros urbanos, *vis-à-vis*, os municípios do interior. Nesse sentido, o elemento fundamental que está por trás desse determinante é o acesso a um maior volume de serviços educacionais e culturais de melhor qualidade.

B) Faixa Etária

No quesito faixa de idade os resultados apontaram um efeito marginal positivo sobre a chance de aprovação no concurso para os candidatos dentro da faixa mais baixa (18 a 22 anos). O resultado apresentado é considerado procedente com base na argumentação de melhor ritmo de estudo dos candidatos mais jovens, por estarem ainda freqüentando aulas ou terem parado de estudar há pouco tempo. Mais uma vez, o elemento fundamental do determinante de desempenho está ligado a questões educacionais.

C) Situação no Mercado de Trabalho

Um efeito marginal positivo sobre a chance de aprovação dos candidatos que estão desempregados ou são estagiários é tido como esperado, pois essa condição no mercado de trabalho possibilita uma maior disponibilidade de tempo desses para se dedicarem aos estudos. Novamente o estudo se apresenta como elemento central.

Adicionalmente, cabe observar que na pergunta do questionário sócio-econômico que trata da condição no mercado de trabalho não se observa uma alternativa referente a estar desocupado, ou seja, não estar procurando emprego.

Assim, dentre os que se denominaram desempregados, é possível ter uma parcela que tenha optado por se dedicar aos estudos antes de entrar no mercado de trabalho, ou para retornar a esse mercado com melhores condições de colocação ou remuneração.

Então, uma maior quantidade de tempo dedicada à obtenção de educação e cultura, e a boa qualidade desses serviços, representa um elemento constituinte que perpassa os determinantes: Local de Origem, Faixa de Idade e Situação no Mercado de Trabalho.

DETERMINANTES RELACIONADOS À RENDA

A) Renda Pessoal Mensal

Ter renda pessoal baixa apontou para um efeito marginal negativo sobre a chance de aprovação. Um resultado que está em conformidade com a argumentação de serem piores as condições de se preparar para o concurso para os que têm baixos rendimentos, visto que essa situação é limitante do acesso a bens materiais de estudo e cursinhos de qualidade. Destaca-se a relevância do acesso a produtos e serviços que reforcem, ou complementem, o aprendizado já adquirido, ou que iniciem os candidatos em temas desconhecidos, como os exigidos em provas específicas.

B) Renda Familiar Mensal

Possuir uma renda familiar alta redundou num efeito marginal positivo sobre a chance de aprovação. Todo o background existente entre os que são de famílias de maior renda, conforme descrito anteriormente, redundou num perfil de uma pessoa que cursou ensino formal de qualidade, com acesso a materiais didáticos e para-didáticos de qualidade, ambiente propício ao desenvolvimento cognitivo, maior volume de informações em educação e cultura, acesso a atividades extracurriculares, inclusão digital, dentre outros inúmeros elementos materiais e intangíveis.

É nítida a melhor e mais adequada condição de acesso à educação e à cultura quando se fala de renda pessoal ou familiar. Base educacional e cultural é novamente o elemento central.

DETERMINANTES RELACIONADOS À EDUCAÇÃO FORMAL

A) Ser Graduado ou Pós-Graduado (“super-educados”)

Os candidatos “super-educados” apresentam efeito marginal positivo sobre a chance de aprovação. Isso reflete fundamentalmente o baixo nível geral do ensino médio. Os que têm melhores possibilidades vão no sentido de avançar nos estudos para adquirir mais conhecimento e sinalizar ao mercado de trabalho que eles seriam mais produtivos por terem nível superior.

Contudo, não se pode desprezar o fato de que num concurso que exige nível médio, no qual participaram praticamente 80.000 candidatos universitários, mais de 35.000 graduados e aproximadamente 5.600 pós-graduados, 28,5% das vagas ofertadas ficaram ociosas.

Ou seja, de um universo de cerca de 120.600 candidatos que são pelo menos universitários, somente 519 candidatos cumpriram os requisitos mínimos para serem aprovados no concurso, os demais perderam para si mesmos. Isso é uma forte evidência de que o baixo nível da qualidade do ensino tem transbordado para o nível superior.

Um dos componentes no estímulo à disseminação de cursos superiores de baixa qualidade é a incipiente formação de uma quantidade expressiva dos vestibulandos, fruto de um ensino médio de baixa qualidade.

Então, os que freqüentaram cursos de ensino médio de baixa qualidade, para melhor se inserir no mercado de trabalho, buscaram cursar ensino superior em instituições de menor tradição e qualidade, que têm preços mais acessíveis e vestibulares menos exigentes. Contudo, a baixa qualidade desses cursos pode não garantir essa melhor inserção, nem produzir um melhor desempenho em concursos públicos.

B) Ter Cursado Apenas Curso de Nível Médio

Ter cursado apenas o nível médio apresentou efeito marginal negativo sobre a chance de aprovação. Resultado também esperado, considerando a qualidade geral do ensino médio e que a

concorrência com candidatos de grau universitário ou até mesmo com pós-graduação levaria a uma competição desigual principalmente em provas que tratavam de temas que não são nem mesmo discutidos no ensino médio.

C) Ter Cursado o Ensino Médio em Escola Privada

Este item, na verdade, complementa o anterior. Ter cursado o ensino médio em escola privada apresentou efeito marginal positivo na chance de aprovação.

Considerando o sucateamento do ensino público brasileiro de nível médio e a ascensão da escola privada, onde são ofertadas metodologias de ensino, ambiente de estudo e acesso a materiais didáticos complementares incomparáveis com os disponíveis na rede pública, não se constitui em grande surpresa este resultado.

Quanto à magnitude dos coeficientes, como observado na metodologia, por se tratar de uma regressão não linear, os coeficientes apresentados na Tabela 12 não representam os efeitos marginais. O efeito marginal de cada variável independente para a distribuição logística pode ser obtido, conforme a equação 12, através da expressão:

$$\frac{e^{\beta'X}}{1 + e^{\beta'X}} \left(1 - \frac{e^{\beta'X}}{1 + e^{\beta'X}} \right) \beta \quad (45)$$

O produto $\beta'X$ representa o vetor de coeficientes multiplicado por um vetor que contenha valores para as variáveis independentes.

É óbvio que o efeito marginal vai variar de acordo com os valores atribuídos às variáveis contidas em X. Sendo assim, utilizou-se os valores médios das variáveis explicativas para efetuar os cálculos dos efeitos marginais, os quais são apresentados na Tabela 13.

TABELA 13 – EFEITOS MARGINAIS SOBRE A CHANCE DE SER APROVADO NO CONCURSO

VARIÁVEIS EXPLICATIVAS		COEFICIENTE ESTIMADO (B)	VALOR MÉDIO (X)	(B)x(X)	EFEITO MARGINAL
Intercepto	C	-6,3538	-0,0013	0,00829	
Ensino Médio	EMED	-1,3209	0,4815	-0,63599	-3,9%
Nível Superior Completo ou com Pós-graduação	ESUPO	0,2897	0,1536	0,04450	0,8%
Estado Civil – Solteiro	SOLT	0,0042	0,7259	0,00303	0,0%
Renda Familiar superior a 10 Salários Mínimos	YFALT	0,5459	0,1012	0,05525	1,6%
Renda Individual abaixo de 2 Salários Mínimos	YPBAI	-0,1829	0,6382	-0,11672	-0,5%
Toda em escola Privada	PRIV	0,2503	0,3156	0,07901	0,7%
Estagiário ou Desempregado	DEST	0,3199	0,4914	0,15717	0,9%
Oriundos de Regiões Metropolitanas	RMETR	0,5034	0,4377	0,22034	1,5%
Idade de 18 a 22	IDBAI	0,3259	0,6636	0,21624	1,0%
$\beta'X$				0,03113	

O dado relevante quanto à magnitude dos efeitos marginais apresentados na Tabela 13 é que, das variáveis explicativas utilizadas no modelo, a condição do candidato ter o grau de escolaridade adequado, nível médio, para a função a ser exercida foi o que apresentou maior impacto (o qual é negativo) na chance do candidato ser aprovado.

Para avaliar problemas de especificação do modelo dois tipos de teste foram realizados. O primeiro foi o teste de heterocedasticidade, como descrito na metodologia. Vale lembrar que este teste foi usado não como um detector de heterocedasticidade, pois o método de estimação implicitamente aplica uma ponderação de mínimos quadrados não linear.

O objetivo do teste é detectar erro de especificação para o caso de heterogeneidade não observada. Como a suspeita sobre o erro de especificação recai sobre a heterogeneidade não observada, ou mais especificamente sobre a possível ausência de uma medida de

habilidade para cada candidato na equação, optou-se por definir o vetor Z na variância do erro como constituído apenas das variáveis de instrução.

Seguindo a regressão auxiliar, definida na equação (35), e cujos resultados são apresentados no Apêndice A, pode-se verificar que o valor calculado para o teste de *Lagrange* é $LM=0,623415$. Já o valor crítico da distribuição *qui-quadrado*, a um nível de significância de 5% e com dois graus de liberdade é 5,99. O valor-p é de 0,7321. Portanto, aceita-se a hipótese nula de homocedasticidade, o que aqui significa ausência de erro de especificação. Este resultado pode parecer surpreendente à primeira vista; no entanto, é em essência o mesmo encontrado por Griliches e Mason (1972), os quais não encontraram vestígios de correlação entre habilidade e escolaridade.

O segundo teste referente a erro de especificação realizado foi o da forma funcional, também descrito na metodologia. Apesar da argumentação de alguns autores (*Train* (2003), por exemplo) de que os modelos *Probit* e *Logit* são empíricamente indistinguíveis, existem situações em que eles podem divergir, requerendo portanto um teste formal de escolha.

Assim optou-se por um teste de hipótese para discriminar entre os modelos *Logit* e *Probit* seguindo a equação (43). Para tanto o modelo *Probit* foi estimado e seus resultados são apresentados no Apêndice B. Já o resultado do teste é apresentado no Apêndice C. Como pode ser visto, como o valor do coeficiente da variável $Z = F_{2i} - F_{1i}$ é estatisticamente insignificante opta-se pelo modelo *Logit*.

Optar pelo modelo *Logit* tendo como alternativa o modelo *Probit* não leva à conclusão que o modelo *Logit* é o mais apropriado. Realizou-se então o teste tendo como hipótese alternativa uma distribuição *Gompit* como representativa da classe assimétrica.

O modelo de regressão usando a distribuição *Gompit* é definido como

$$P(Y=1/X) = \exp(-\exp(-\beta'X)) \quad (46)$$

Os resultados da estimação deste modelo podem ser visto no Apêndice D. O resultado do teste para discriminar entre o *Logit* e o *Gompit* é mostrado no Apêndice E.

Vê-se novamente que o coeficiente estimado de Z é estatisticamente insignificante ao nível usual de 5%. Portanto, pode-se ter uma razoável confiança de que o modelo escolhido foi bem especificado.

5.2. CONCLUSÕES

No presente estudo de caso de avaliação de desempenho em concurso público foram destacados, na introdução, alguns exemplos de questões primordiais a serem respondidas:

- O que faz determinadas pessoas serem aprovadas em detrimento de outras?
- Será que existem fatores sócio-econômicos que aumentam as chances de uma pessoa ter um melhor desempenho?
- Maior renda implica em maiores chances de passar?
- Ter nível de educação acima do exigido no concurso implica em maiores chances de aprovação?

Após a especificação de um modelo econométrico do tipo *Logit* onde o sucesso no concurso é definido como função de um conjunto de variáveis explicativas constituído por renda familiar e pessoal, grau de instrução, tipo de escola freqüentada, estado civil, região onde habita, situação atual no mercado de trabalho e idade. Constatou-se, após a estimação do modelo, que renda, escolaridade acima do ensino médio, ser oriundo de região metropolitana e estar em baixa faixa de idade, contribuem positivamente para aumentar a probabilidade de passar no concurso. Por outro lado, ser detentor apenas de escolaridade de ensino médio contribui negativamente para aprovação no concurso, e o coeficiente dessa variável representa o maior efeito marginal em termos absolutos, o que dramatiza a situação dos candidatos de mais baixa escolaridade.

Delineia-se assim um perfil de candidato jovem, tipicamente pertencente a uma classe média de renda elevada e que teve acesso à instrução e meios culturais de boa qualidade, como elemento típico de quem obteve sucesso no concurso. Reitera-se então, através dos resultados obtidos, o círculo vicioso de concentração de renda onde candidatos provenientes de famílias de melhor estrutura financeira têm maiores chances de aprovação e acesso a postos de trabalho com maior remuneração média.

Em termos de políticas públicas que contrabalancem estes efeitos perversos nota-se que todos os determinantes do modelo apontam, em última instância, para a necessidade de uma base educacional e cultural de qualidade. O ensino de nível médio parece ser um gargalo no processo de disseminação desse importante condicionante, o qual dentro do escopo desse trabalho, revela-se fundamental não só ao bom desempenho no concurso objeto de estudo, como para quaisquer outros concursos.

Entende-se que o ensino médio público pode ocupar o lugar estratégico quanto à criação de condições mais igualitárias na preparação e no desempenho em concursos. É no âmbito do ensino médio que se pode ter elementos que se não suprem todas as carências dos candidatos menos afortunados, podem atenuar a falta de condição no âmbito familiar, ou a deficiência estrutural de uma região pobre. Portanto, identifica-se o ensino médio público como um *locus* de ação e políticas que atuem para reduzir as desigualdades de condições iniciais de cada candidato a concursos públicos.

Por outro lado, preencher postos de trabalho que requerem apenas uma educação de nível médio com detentores de título universitário, e até mesmo com pós-graduação, aponta para uma má alocação dos recursos públicos conjugada a uma possível deficiência estrutural do mercado de trabalho regional.

BIBLIOGRAFIA

AMEMIYA, T. Qualitative Response Models: A Survey, **Journal of Economic Literature**, v.19, p. 1483-1536. 1981.

ALBA-RAMIREZ **Types of Job Match, Overeducation and Labor Mobility in Spain**. WorkingPaper, University Carlos III of Madrid. 2002.

ALMEIDA, E. S. **Mudanças Institucional e Estrutural na Economia Brasileira do início dos anos 90: uma abordagem evolucionista**. São Paulo: Universidade de São Paulo. 243p. (Dissertação de Mestrado em Economia). 1996.

BERNDT, E. K., *et alli*. Estimation and Inference in Nonlinear Structural Models, **Annals of Economic and Social Measurement**, Oct. 1974, p. 653-65. 1974.

BORG M.O. e STRANAHAN H.A. Personality Type and Student Performance in Upper-Level Economics Courses: The Importance of Race and Gender. **Journal of Economic Education**, Winter. 2002.

BRASFIELD D. W., HARRISON D. E. e McCOY J. P., The Impact of High School Economics on College Principles, **Journal of Economic Education**. v.24, n.2, p. 99-111. 1993.

BRASIL, República Federativa do. **ENEM 2003: Relatório Final**. Brasília: DACC/MEC. 2004.

DAVIDSON, R. e MACKINNON, J.G. **Estimation and Inference in Econometrics**. New York: Oxford University Press. 1993.

DAVIDSON, R. e MACKINNON, J.G. **Econometric Theory and Methods**. New York: Oxford University Press. 2000.

DI PIETRO, G. e URWIN, P. **Education and Skills Mismatch in the Italian Graduate Labour Market**. Westminster Business School, University of Westminster. 2002.

DOLTON, P. J. e A. VIGNOLES **Overeducation Duration: How Long Did Graduates in the 1980s Take to Get a Graduate Job?** Working Paper University of Newcastle, Tyne. 1997.

GREEN, F., McINTOSH, S. e VIGNOLES A. **Overeducation and Skills – Clarifying the Concepts**. Centre for Economic Performance (Published by London School of Economics and Political Science). 1999.

GREENE, W.H. **Econometric Analysis**. New Jersey: Prentice Hall. 1993.

GREENE, W.H. Gender Economics Courses in Liberal Arts Colleges: Further Results, **Journal of Economic Education**. Fall. 1998.

GRILICHES, Z. e MANSON, W. M. Education, Income and Ability, **Journal of Political Economy**, v.80 n.3, part 2 may/june. 1972.

GROOT, W. The Incidence of, and Returns to Overeducation in the UK. **Applied Economics** v.28 n.10, p. 1345-50. 1996.

HARTOG, J. **On Returns to Education: Wandering Along the Hills of ORU Land**, Keynote speech for the LVIIIth Conference of the Applied Econometrics Association, Maastricht. 1997.

HSIAO, C. **Analysis of Panel Data**. Cambridge: Cambridge University Press. 2003.

McGUINNESS, S. **Graduate Overeducation as a Sheepskin Effect: Evidence From Northern Ireland**. Working Paper 70, Northern Ireland Economic Research Centre. 2002.

MURPHY, K.M. and WELCH, F. Wage Premiums for College Graduates, Recent Growth and Possible Explanation, **Educational Researcher**, 18, p.17-26. 1989.

ROBB R.E. e ROBB, A.L. Gender and the Study of Economics: The Role of Gender of the Instructor. **Journal of Economic Education**, (Winter), p. 3-19. 1999.

RUMBERGER, R. W. (**Overeducation in the U.S. Labor Market**. New York: Praeger. 1981.

RUUD, P. A. **An Introduction to Classical Econometric Theory**. New York: Oxford University Press, 2004.

SILVA, A. S. S. A Score test for non-nested Hipoteses with Applications to Discrete Data Model. **Journal of Applied Econometrics**, n.16, p. 577-97. 2001.

SARGAN, D. A. Comment. In: G. A Renton (Ed.). **Modelling the Economy**. 1975.

SICHERMAN, N. Overeducation in Labor Market. **Journal of Labor Economics** v.9, n.02, p.101-122. 1991.

SPECTOR L.C. e MAZZEO M.A. Probit Analysis and Economic Education, **Journal of Economic Education**, p. 37-44. 1980.

TRAIN, K. **Discrete Choice Methods with Simulation**. Cambridge: University Press. Cambridge. 2002.

VERHAEST e OMEY. **Overeducation in the Flemish Youth Labour Market**, WP 2002/141. Department of Social Economics, Ghent University. 2002.

WHITE, H. A. Heteroskedasticity-Consistente Covariance Matrix Estimation and Direct Test for Heteroskedasticity. **Econometrica**. v. 48, n. 4, p. 817-38. 1980.

WOOD, W. C. e DOYLE, J. M. Economic Literacy Among Corporate Employees. **Journal of Economic Education** (Summer) p. 195-205, 2002.

WOOLDRIDGE, J. M. **Econometric Analysis of Cross Section and Panel Data**. Cambridge: MIT Press, 2002.

APÊNDICE A – Regressão Auxiliar para Teste de Heteroscedasticidade

H0: Homoscedasticidade

H1: Heteroscedasticidade

COVARIADAS	COEFICIENTE ESTIMADO	ERRO PADRÃO	ESTATÍSTICA T	P-VALOR
FP	0,0140	0,150	0,093	0,9255
FP*EMED	- 0,7609	1,677	-0,453	0,6500
FP*ESUPO	0,4716	0,919	0,512	0,6082
FP*SOLT	0,0005	0,109	0,004	0,9961
FP*YFALT	-0,0127	0,104	-0,122	0,9029
FP*YPBAI	-0,0009	0,084	-0,011	0,9907
FP*PRIV	-0,0028	0,089	-0,032	0,9744
FP*DEST	-0,0008	0,090	-0,009	0,9923
FP*IDBAI	-0,0071	0,112	-0,063	0,9495
FP*RMETR	-0,0036	0,095	-0,038	0,9697
FP*(-Xβ)*EMED	0,1090	0,242	0,449	0,6531
FP*(-Xβ)*ESUPO	-0,0909	0,177	-0,512	0,6084
R2	0,0000003	Soma do quadrado dos resíduos		234307,7
R2 Ajustado	-0,0000045	Critério de informação de Akaike		2,846552
Média da Variável Dependente	0,000056	Critério de Schwarz		2,847087
S.D. da Variável Dependente	1,004298	Estatística Durbin-Watson		0,292656
Erro Padrão da Regressão	1,004321			
Log-verossimilhança				-330.626,4
Estatística de Razão de verossimilhança $[\ln(\lambda)] - 2$ graus de liberdade				0,62341
p-valor de λ				0,73219

Nota: Resultados gerados através do aplicativo *EViews 4.1*[®]

APÊNDICE B – Resultados do Modelo *Probit*

RESULTADO DO MODELO *PROBIT*

COVARIADAS	COEFICIENTE ESTIMADO	ERRO PADRÃO	ESTATÍSTICA Z	P-VALOR
C	- 2,919	0,043	-67,75	0,0000
EMED	- 0,412	0,036	-11,41	0,0000
ESUPO	0,095	0,034	2,75	0,0058
SOLT	-0,004	0,036	-0,11	0,9093
YFALT	0,190	0,034	5,51	0,0000
YPBAI	-0,06	0,027	-2,30	0,0210
PRIV	0,088	0,029	3,03	0,0024
DEST	0,102	0,028	3,54	0,0004
IDBAI	0,108	0,036	3,00	0,0027
RMETR	0,165	0,028	5,75	0,0000
Média da Variável Dependente	0,002660	Critério de informação de Akaike		0,034998
S.D. da Variável Dependente	0,051509	Critério de Schwarz		0,035444
Erro Padrão da Regressão	0,051452	Critério de Hannan-Quinn		0,035128
Soma do quadrado dos resíduos	614,9534	R ² McFadden		0,052872
Log-verossimilhança				-4.054,7414
Log-verossimilhança restrita				-4.281,5040
Estatística de Razão de verossimilhança [ln(λ)] - 9 graus de liberdade				452,7414
p-valor de λ				0,0000
Obs. com y=0	231.690	Total de observações		232.308
Obs. com y=1	618			

Nota: Dados gerados através do aplicativo *EViews 4.1*[®]

PODER DE PREDIÇÃO DA ESTIMAÇÃO DO MODELO *PROBIT*

	EQUAÇÃO ESTIMADA			PROBABILIDADE CONSTANTE		
	Y=0	Y=1	Total	Y=0	Y=1	Total
P(Y=1)≤C	141379	157	141536	0	0	0
P(Y=1)>C	90311	461	90772	231690	618	232308
Total	231690	618	232308	231690	618	232308
Correto	141379	416	141840	0	618	618
Correto (%)	61.02	74.60	61.06	0.00	100.00	0.27
Incorreto (%)	38.98	25.40	38.94	100.00	0.00	99.73
Acerto Total	61.02	-25.40	60.79			
Acerto Percentual	61.02	NA	60.95			

Nota 1: Dados gerados através do aplicativo *EViews 4.1*[®]

Nota 2: Corte para sucesso (Y=1), C = 0.00266026

APÊNDICE C – Resultados de Teste de Especificação para Forma Funcional – *Logit x Probit*

H0: *Logit*

H1: *Probit*

COVARIADAS	COEFICIENTE ESTIMADO	ERRO PADRÃO	ESTATÍSTICA Z	P-VALOR
C	- 6,3547	0,136	-46,69	0,0000
EMED	- 1,3210	0,118	-11,15	0,0000
ESUPO	0,2901	0,101	2,86	0,0042
SOLT	0,0048	0,111	0,04	0,9654
YFALT	0,5460	0,098	5,54	0,0000
YPBAI	-0,1824	0,08	-2,21	0,0266
PRIV	0,2498	0,089	2,77	0,0055
DEST	0,3202	0,088	3,62	0,0003
IDBAI	0,3259	0,108	3,00	0,0027
RMETR	0,5034	0,088	5,69	0,0000
Z	-0,0403	1,402	-0,03	0,9770
Média da Variável Dependente	0,002660	Critério de informação de Akaike		0,035004
S.D. da Variável Dependente	0,051509	Critério de Schwarz		0,035494
Erro Padrão da Regressão	0,051451	Critério de Hannan-Quinn		0,035147
Soma do quadrado dos resíduos	614,9495	R2 McFadden		0,052937
Log-verossimilhança				4.054,8530
Log-verossimilhança restrita				4.281,5040
Estatística de Razão de verossimilhança $[\ln(\lambda)]$ - 10 graus de liberdade				453,4006
p-valor de λ				0,0000
Obs. com y=0	231.690	Total de observações		232.308
Obs. com y=1	618			

Nota: Resultados gerados através do aplicativo *EViews 4.1*[®]

APÊNDICE D – Resultado Do Modelo *Gompit*

RESULTADO DO MODELO *GOMPIT*

COVARIADAS	COEFICIENTE ESTIMADO	ERRO PADRÃO	ESTATÍSTICA Z	P-VALOR
C	- 1,845	0,022	-83,30	0,0000
EMED	- 0,209	0,018	-11,53	0,0000
ESUPO	0,049	0,018	2,65	0,0080
SOLT	-0,004	0,019	-0,22	0,8228
YFALT	0,104	0,019	5,47	0,0000
YPBAI	-0,03	0,014	-2,37	0,0173
PRIV	0,049	0,015	3,17	0,0015
DEST	0,052	0,015	3,42	0,0006
IDBAI	0,056	0,019	2,99	0,0028
RMETR	0,086	0,015	5,78	0,0000
Média da Variável Dependente	0,002660	Critério de informação de Akaike		0,035003
S.D. da Variável Dependente	0,051509	Critério de Schwarz		0,035449
Erro Padrão da Regressão	0,051452	Critério de Hannan-Quinn		0,035133
Soma do quadrado dos resíduos	614,9591	R ² McFadden		0,052736
Log-verossimilhança				-4.054,714
Log-verossimilhança restrita				-4.281,504
Estatística de Razão de verossimilhança [ln(λ)] - 9 graus de liberdade				451,5786
p-valor de λ				0,0000
Obs. com y=0	231.690	Total de observações		232.308
Obs. com y=1	618			

Nota: Resultados gerados através do aplicativo *EViews 4.1*[®]

PODER DE PREDIÇÃO DA ESTIMAÇÃO DO MODELO *GOMPIT*

	EQUAÇÃO ESTIMADA			PROBABILIDADE CONSTANTE		
	Y=0	Y=1	Total	Y=0	Y=1	Total
P(Y=1)≤C	138271	146	138417	0	0	0
P(Y=1)>C	93419	472	93891	231690	618	232308
Total	231690	618	232308	231690	618	232308
Correto	138271	472	138443	0	618	618
Correto (%)	59.68	76.38	59.72	0.00	100.00	0.27
Incorreto (%)	40.32	23.62	40.28	100.00	0.00	99.73
Acerto Total	59.68	-23.62	59.46			
Acerto Percentual	59.68	NA	59.62			

Nota 1: Resultados gerados através do aplicativo *EViews 4.1*[®]

Nota 2: Corte para sucesso (Y=1), C = 0.00266026

APÊNDICE E – Resultados de Teste de Especificação para Forma Funcional – *Logit x Gompit*

H0: *Logit*

H1: *Gompit*

COVARIADAS	COEFICIENTE ESTIMADO	ERRO PADRÃO	ESTATÍSTICA Z	P-VALOR
C	- 6,3564	0,136	-46,44	0,0000
EMED	- 1,3215	0,118	-11,15	0,0000
ESUPO	0,2908	0,101	2,86	0,0041
SOLT	0,0059	0,111	0,05	0,9575
YFALT	0,5463	0,098	5,54	0,0000
YPBAI	-0,1816	0,083	-2,16	0,0303
PRIV	0,2490	0,089	2,77	0,0055
DEST	0,3211	0,088	3,62	0,0003
IDBAI	0,3261	0,108	3,00	0,0026
RMETR	0,5036	0,088	5,69	0,0000
Z	-0,0614	0,787	-0,07	0,9379
Média da Variável Dependente	0,002660	Critério de informação de Akaike		0,035004
S.D. da Variável Dependente	0,051509	Critério de Schwarz		0,035494
Erro Padrão da Regressão	0,051451	Critério de Hannan-Quinn		0,035147
Soma do quadrado dos resíduos	614,9495	R2 McFadden		0,052938
Log-verossimilhança				4.054,8530
Log-verossimilhança restrita				4.281,5040
Estatística de Razão de verossimilhança $[\ln(\lambda)]$ - 10 graus de liberdade				453,3077
p-valor de λ				0,0000
Obs. com y=0	231.690	Total de observações		232.308
Obs. com y=1	618			